



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΤΜΗΜΑ ΠΟΛΙΤΙΣΜΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΣ

Διδακτορική Διατριβή
του
Νείρου Αντωνίου

ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ
ΑΣΑΦΟΥΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ
ΓΙΑ ΤΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ
ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ
ΣΥΝΑΡΤΗΣΕΩΝ ΑΚΤΙΝΙΚΗΣ
ΒΑΣΗΣ

Τριμελής Συμβουλευτική Επιτροπή

Επιβλέπων:

Τσεκούρας Γεώργιος,
Επίκουρος Καθηγητής

Συνεπιβλέποντες:

Δαρζέντας Ιωάννης, Καθηγητής
Αναγνωστόπουλος Χρήστος,
Επίκουρος Καθηγητής

Δεκέμβριος 2011

Μέλη της Επταμελούς Εξεταστικής Επιτροπής

1. **Τσεκούρας Γεώργιος**, μόνιμος Επίκουρος Καθηγητής, Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας Πανεπιστημίου Αιγαίου (Επιβλέπων)
2. **Αναγνωστόπουλος Χρήστος-Νικόλαος**, Επίκουρος Καθηγητής, Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας Πανεπιστημίου Αιγαίου (μέλος συμβουλευτικής επιτροπής)
3. **Δαρζέντας Ιωάννης**, Καθηγητής, Τμήμα Μηχανικών Σχεδίασης Προϊόντων και Συστημάτων Πανεπιστημίου Αιγαίου (μέλος συμβουλευτικής επιτροπής)
4. **Γαβαλάς Δαμιανός**, μόνιμος Επίκουρος Καθηγητής, Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας Πανεπιστημίου Αιγαίου
5. **Καβακλή Ευαγγελία**, μόνιμη Επίκουρη Καθηγήτρια, Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας Πανεπιστημίου Αιγαίου
6. **Παυλογεωργάτος Γεράσιμος**, Επίκουρος Καθηγητής, Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας Πανεπιστημίου Αιγαίου
7. **Σουλακέλλης Νικόλαος**, Αναπληρωτής Καθηγητής, Τμήμα Γεωγραφίας Πανεπιστημίου Αιγαίου

Αφιερώνεται στην σύζυγό μου Αικατερίνη Νικολαΐδη
στα παιδιά μου Ράνια, Ζωή, Δημήτρη
στους γονείς μου Δημήτρη και Ουρανία
και στον αδερφό μου Παναγιώτη

Ευχαριστίες

Από τη θέση αυτή θα ήθελα πρώτα απ' όλα, να εκφράσω τις ευχαριστίες και την βαθύτατη ευγνωμοσύνη μου στον επιβλέποντα Επίκουρο Καθηγητή κ. Γεώργιο Τσεκούρα για την αμέριστη συμπαράσταση και την εύστοχη καθοδήγησή του κατά την εκπόνηση της παρούσας Διδακτορικής Διατριβής. Το ήθος και η άρτια επιστημονική του κατάρτιση υπήρξαν ουσιαστικά στοιχεία για την ολοκλήρωση της εργασίας. Δεν θα ξεχάσω ποτέ το ζήλο που επέδειξε στο να μου μεταφέρει τις πολύτιμες γνώσεις του και εμπειρίες του. Πάντα δίπλα μου, δεν σταμάτησε ποτέ να μου δίνει κουράγιο και να με καθοδηγεί με το δικό του μοναδικό τρόπο. Τον ευχαριστώ τέλος, που μου έδωσε την ευκαιρία να εργαστώ στο εργαστήριο «Έξυπνων Πολυμέσων και Εικονικής Πραγματικότητας» μέσα στο οποίο πέρασα τέσσερα δημιουργικά χρόνια.

Επίσης θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Δαρζέντα Ιωάννη και τον Επίκουρο Καθηγητή κ. Χρήστο Αναγνωστόπουλο για την πολύτιμη στήριξή τους.

Αξέχαστες θα μου μείνουν οι στιγμές που μοιράστηκα με τους καθηγητές κ. Μαντιμαρούδη Φιλήμονα και κ. Δαμιανό Γαβαλά και τους ευχαριστώ θερμά γιατί παρά το φόρτο εργασίας τους πάντα μου άνοιγαν την πόρτα του γραφείου τους να με ακούσουν και να με καθοδηγήσουν.

Στην ολοκλήρωση της ερευνητικής μου προσπάθειας και της συγγραφικής διαδικασίας σημαντική ήταν η συνδρομή του ΑΠΥΣΔΕ Β. Αιγαίου χάρη στην τριετή εκπαιδευτική άδεια που μου χορήγησε κατά τα σχολικά έτη 2007-2010, αφού χωρίς αυτή θα ήταν πολύ δύσκολη η ολοκληρωτική ενασχόληση με το αντικείμενο της έρευνάς μου.

Τέλος, επιθυμώ να εκφράσω την ευγνωμοσύνη μου στη σύζυγό μου και στα παιδιά μου για την αμέριστη υπομονή τους, την αγάπη τους καθώς και την συμπαράσταση τους καθόλη τη διάρκεια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

Ονοματολογία – Σύμβολα	10
Περίληψη	11
Abstract	13
Λίστα Σχημάτων	15
Λίστα Πινάκων	18
Εισαγωγή.....	21
1. Τεχνητά Νευρωνικά Δίκτυα	28
1.1 Εισαγωγή	28
1.2 Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) - (Artificial Neural Networks – ANNs).....	29
1.3 Τρόπος Λειτουργίας ΤΝΔ.....	30
1.4 Συναρτήσεις Ακτινικής Βάσης (radial basis functions).....	31
1.5 Αρχιτεκτονική Δικτύων	32
1.5.1 Δίκτυα Ενός Στρώματος (Single-Layer Feedforward Networks).....	32
1.5.2 Δίκτυα Πολλών Στρωμάτων (Multilayer Feedforward Networks).....	33
1.5.3 Επαναλαμβανόμενα Δίκτυα (Recurrent Networks).....	35
1.5.4 Δίκτυο Perceptron.....	35
1.5.4.1 Εκπαίδευση Δικτύου Perceptron.....	36
1.5.5 Πολυστρωματικό Perceptron	38
1.5.5.1 Εκπαίδευση Πολυστρωματικού Perceptron.....	39
1.5.6 Δίκτυα Kohonen	40
1.5.7 Νευρωνικά Δίκτυα Συναρτήσεων Ακτινικής Βάσης (RBFNNs)	43
1.5.7.1 Εκπαίδευση (training) Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης.....	46
1.5.7.2 Επιβλεπόμενη Μάθηση (Supervised Learning).....	47
1.5.7.3 Μη – Επιβλεπόμενη Μάθηση (Unsupervised Learning).....	47

2. Αλγόριθμοι Συσταδοποίησης Δεδομένων για την Εκπαίδευση	
Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης.....	49
2.1 Εισαγωγή	49
2.2 Ο Αλγόριθμος των c-Μέσων (ή αλγόριθμος Lloyd).....	49
2.3 Αλγόριθμος Ασαφών c-Μέσων (Fuzzy c-Means)	54
2.3.1 Βασικές Έννοιες Ασαφούς Λογικής.....	54
2.3.2 Ο Αλγόριθμος των Ασαφών c-Μέσων (Fuzzy c-Means)	57
2.4 Αλγόριθμος υπό όρους ασαφής συσταδοποίησης (Conditional Fuzzy Clustering - CFC).....	60
2.4.1 Το πρόβλημα βελτιστοποίησης.....	64
2.5 Ο αλγόριθμος κβάντισης διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ).....	65
2.6 Ο Αλγόριθμος Ασαφούς Κβάντισης Διανύσματος με Εκμάθηση (Fuzzy Learning Vector Quantization –FLVQ).....	66
2.7 Ασαφής Κβάντιση Διανύσματος (Fuzzy Vector Quantization-FVQ)	68
2.8 Ο Βελτιωμένος Αλγόριθμος Ασαφούς Κβάντισης Διανύσματος με Εκμάθηση (IFLVQ).....	70
3. Ανάπτυξη Μεθόδου Εκπαίδευσης Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης με τη χρήση του αλγόριθμου Σταθμισμένων Ασαφών c-Μέσων	76
3.1 Εισαγωγή	76
3.2 Ο προτεινόμενος αλγόριθμος.....	78
3.2.1 Φάση I.....	81
3.2.2 Φάση II.....	83
3.2.3 Φάση III	84
3.3 Πειράματα προσομοίωσης.....	85
3.3.1 Σύστημα Mackey-Glass	85
3.3.2 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης	88
3.4 Συμπεράσματα	89

4. Μέθοδος Βέλτιστης Συσταδοποίησης για την εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης	90
4.1 Εισαγωγή	90
4.2 Ο προτεινόμενος Αλγόριθμος.....	92
4.2.1 Επισκόπηση του αλγορίθμου	92
4.2.2 Ο αλγόριθμος κβάντισης διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ).....	95
4.2.3 Σταθμισμένη (weighted) Ασαφής Συσταδοποίηση (Fuzzy Clustering)	95
4.2.4 Βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering)	97
4.2.5 Εκτίμηση των αρχικών βαρών για το RBF με τη χρησιμοποίηση της μεθόδου των ελαχίστων τετραγώνων (least squares method).....	98
4.2.6 Βελτιστοποίηση του νευρωνικού δικτύου RBF με τη χρησιμοποίηση οπισθοδιάδοσης (back-propagation).....	99
4.3 Πειράματα προσομοίωσης.....	100
4.3.1 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης	100
4.3.2 Προσέγγιση Δισδιάστατης Συνάρτησης	102
4.3.3 Σύστημα Mackey-Glass.....	104
4.4 Συμπεράσματα	106
5. Μέθοδος Ιεραρχικής Ασαφούς Συσταδοποίησης για την Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης.....	107
5.1 Εισαγωγή	107
5.2 Ο προτεινόμενος δείκτης εγκυρότητας.....	108
5.3 Ο προτεινόμενος αλγόριθμος εκπαίδευσης	111
Βήμα 1: Ασαφή Συσταδοποίηση Εισόδου – Εξόδου.....	111
Βήμα 2: Βέλτιστη ασαφής συσταδοποίηση.....	112
Βήμα 3: Δίκτυο εκτίμησης των παραμέτρων.....	114
5.4 Αριθμητικά παραδείγματα	115
5.4.1 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης	115
5.4.2 Προσέγγιση δισδιάστατης συνάρτησης.....	116
5.4.3 Προσέγγιση στατικής δισδιάστατης συνάρτησης.....	118

5.4.4 Η Τρισδιάστατη Συνάρτηση	120
5.4.5 Abalone Δεδομένα	122
5.5 Συμπεράσματα	123
6. Ανάπτυξη Μεθόδου Υβριδικής Ασαφούς Συσταδοποίησης για τον Αποτελεσματικό Σχεδιασμό Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης.....	125
6.1 Εισαγωγή	125
6.2 Ο προτεινόμενος αλγόριθμος εκπαίδευσης	127
6.2.1 Υβριδική ασαφής συσταδοποίηση (hybrid fuzzy clustering)	127
6.2.2 Εκτίμηση παραμέτρων του δικτύου.....	138
6.2.3 Ανάλυση του υπολογιστικού κόστους.....	140
6.3 Μελέτη προσομοίωσης	145
6.4 Συμπεράσματα	167
7. Ανάπτυξη μεθόδου Κοκκοποίησης Πληροφοριών για την Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης.....	169
7.1 Εισαγωγή	169
7.2 Μερικές νέες έννοιες στη ασαφή κοκκοποίηση πληροφοριών.....	171
7.3 Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης (RBFNN) χρησιμοποιώντας Ασαφείς Κόκκους	179
7.3.1 Πρώτο επίπεδο κοκκοποίησης.....	181
7.3.2 Δεύτερο επίπεδο κοκκοποίησης.....	182
7.3.3 Τρίτο επίπεδο κοκκοποίησης.....	188
7.3.4 Εκτίμηση παραμέτρων του δικτύου RBF	190
7.3.5 Περαιτέρω σχόλια στον αλγόριθμο	191
7.4 Μελέτη προσομοίωσης	192
7.5 Συμπεράσματα	200
8. Συζήτηση Αποτελεσμάτων και Μελλοντική Εργασία	201
8.1 Συμπεράσματα	201

8.1.2 Συμπεράσματα κάθε μεθόδου.....	201
8.1.2 Σύγκριση των προτεινόμενων μεθόδων.....	204
8.2 Μελλοντική εργασία (Future work).....	205
ΠΑΡΑΡΤΗΜΑ.....	206
ΒΙΒΛΙΟΓΡΑΦΙΑ	207

Ονοματολογία – Σύμβολα

$\ \cdot \ $	2-νόρμα	u_{jk}	βαθμός συμμετοχής του k -δεδομένου στην j -συστάδα
A	Ασαφές σύνολο	v_i	το κέντρο του i -RBF κόμβου (ή i -συστάδας)
c	αριθμός κρυφών RBF κόμβων (αριθμός συστάδων)	w	διάνυσμα βάρους
e	παράμετρος οπισθοδιάδοσης (back-propagation) εκμάθησης	w_i	βάρος νευρώνα σύνδεσης του i -RBF κόμβου με την έξοδο
g_i	i -οστή γκαουσιανή συνάρτηση πυρήνα	X	σύνολο δεδομένων εισόδου
f_k	περιορισμός πλαισίου	x	είσοδος
$h_{l,l_0,k}$	συνάρτηση που καθορίζει το πιο πρόσφατο φαινόμενο αλληλεπίδρασης μεταξύ των νευρώνων l και l_0 .	x_k	k -οστό διάνυσμα εισόδου
$J_{c\text{-means}}$	αντικειμενική συνάρτηση κόστους που ελαχιστοποιεί ο αλγόριθμος c -Means	y	πραγματική έξοδος
J_m	αντικειμενική συνάρτηση των Ασαφών c -Μέσων (fuzzy c -means)	\tilde{y}	προβλεφθείσα έξοδος
J_r	αντικειμενική συνάρτηση των σταθμισμένων (weighted) Ασαφών c -Μέσων (fuzzy c -means)	y_k	k -οστό διάνυσμα εξόδου
k	αριθμός διανύσματος εισόδου		
m	παράμετρος ασάφειας		Ελληνικά γράμματα
n	αριθμός δειγμάτων εκπαίδευσης,	α	παράμετρος εκμάθησης στον αλγόριθμο LVQ
N	συνολικός αριθμός δεδομένων (εκπαίδευσης και δοκιμής)	α_0	αρχικό ποσοστό εκμάθησης για τον αλγόριθμο LVQ
\aleph	ασαφής πληθικότητα συνόλου	β	ρυθμός εκπαίδευσης (learning rate)
p	διάσταση του διανύσματος εισόδου	θ	ένας μικρός θετικός αριθμός έτσι ώστε $\theta \in (0, 1)$
q	παράμετρος ασάφειας	λ	συντελεστής Lagrange πολλαπλασιαστών
q_j	αριθμός των ασαφών συνόλων	μ_{ij}	συναρτήσεις συμμετοχής του i -δεδομένου στην j -συστάδα
Q	αντικειμενική συνάρτηση στον αλγόριθμο CFC	ξ	προκαθορισμένο όριο με $\xi \in (0,1)$
s	μέτρο διαχωρισμού	π	μέτρο πυκνότητας
S_{PC}	συντελεστής διαμερισμού	ρ_i	βάρος μεγάλης σημασίας της i -συστάδας
S_{PE}	εντροπία διαμερισμού	σ_i	πλάτος (width) της i -συστάδας
S_{XB}	συντελεστής Xie-Beni	τ	αριθμός κοντινότερων γειτόνων
t	αριθμός επανάληψης		

Περίληψη

Στην παρούσα διδακτορική διατριβή λαμβάνει χώρα ερευνητική προσπάθεια στην εκπαίδευση (training) Νευρωνικών Δικτύων και συγκεκριμένα των Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης (Radial Basis Function Neural Networks – RBFNN). Για την εκπαίδευση των Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης χρησιμοποιήθηκαν αλγόριθμοι ασαφούς συσταδοποίησης (fuzzy clustering).

Αναπτύχθηκαν πέντε νέες καινοτόμοι μέθοδοι εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης και οι οποίες έχουν δημοσιευθεί στα πρακτικά διεθνών επιστημονικών συνεδρίων καθώς και σε έγκυρα διεθνή περιοδικά.

Η συνεισφορά της παρούσης διδακτορικής διατριβής είναι ότι διεξήχθη συστηματική έρευνα με σκοπό την διερεύνηση της χρήσης της ασαφούς συσταδοποίησης στην εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Στο πλαίσιο της έρευνας εστίασαμε στην ανάλυση των πλεονεκτημάτων και των μειονεκτημάτων καθώς και της ουσιαστικής επίδρασης της ασαφούς συσταδοποίησης στην διαδικασία εκπαίδευσης τέτοιου τύπου νευρωνικών δικτύων.

Το αποτέλεσμα της εν' λόγω έρευνας συνίσταται στην ανάπτυξη σε πρώτο επίπεδο τριών απλών μεθόδων ασαφούς συσταδοποίησης για την εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Συγκεκριμένα στην πρώτη μέθοδο, χρησιμοποιήθηκε η πολύ γνωστή μέθοδος των Ασαφών c -Μέσων (FCM) για την προεπεξεργασία των δεδομένων και στη συνέχεια η μέθοδος των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) για τον υπολογισμό των παραμέτρων των πυρήνων των κρυφών κόμβων του νευρωνικού δικτύου. Στη δεύτερη μέθοδο, χρησιμοποιήθηκε η βέλτιστη συσταδοποίηση (optimal clustering) για τον καθορισμό του αριθμού των πυρήνων των συναρτήσεων ακτινικής βάσης καθώς και τον παραμέτρων του νευρωνικού δικτύου. Τέλος προτάθηκε μια τρίτη μέθοδος [63] που συνδυάζει ασαφή συσταδοποίηση και τη βέλτιστη ασαφή συσταδοποίηση για τον αποτελεσματικό σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης.

Σε δεύτερο επίπεδο, αναπτύχθηκαν δύο καινοτόμες μέθοδοι εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Συγκεκριμένα στην πρώτη μέθοδο προτείνεται μια νέα καινοτόμος υβριδική μέθοδος συσταδοποίησης, η οποία συνδυάζει με ομοιόμορφο τρόπο την διακριτή (crisp) και την ασαφή (fuzzy) συσταδοποίηση (clustering) για τον υπολογισμό των παραμέτρων των πυρήνων των συναρτήσεων ακτινικής βάσης του νευρωνικού δικτύου. Η δεύτερη προτεινόμενη μέθοδος χρησιμοποιεί κοκκοποίηση πληροφοριών (information granulation) για τον υπολογισμό των κέντρων των πυρήνων των συναρτήσεων ακτινικής βάσης και μία νέα μετρική απόσταση (metric distance) για τον υπολογισμό των πλατών των συναρτήσεων ακτινικής βάσης.

Σε όλες τις προτεινόμενες μεθόδους έγινε πρακτική εφαρμογή σε πραγματικά μοντέλα και σε προσεγγίσεις συναρτήσεων, τα αποτελέσματα των οποίων έδειξαν ότι οι προτεινόμενες μέθοδοι παράγουν αξιόπιστα, ακριβή και συμπαγή νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης. Η σύγκριση με την διεθνή βιβλιογραφία έδειξε ότι τα δίκτυα RBF που δημιουργήθηκαν έχουν καλύτερα αποτελέσματα όσον αφορά το μέσο τετραγωνικό σφάλμα (mean square error – MSE) καθώς και το κανονικοποιημένο μέσο τετραγωνικό σφάλμα (normalized mean square error – NRMSE) τόσο για τα δεδομένα εκπαίδευσης (training data) όσο για τα δεδομένα δοκιμής (test data).

Abstract

In the present doctoral thesis takes place inquiring effort in the training of Neural Networks and concretely the Radial Basis Function Neural Networks – RBFNN. For the training of radial basis function neural networks were used algorithms of fuzzy clustering.

Five new methods for the training of radial basis function neural networks were developed and have been published in international scientific conferences and journals.

The contribution of this doctoral dissertation is that systematic research was conducted with a view to exploring the use of fuzzy clustering in training of radial basis functions neural networks. Within the framework of the research we focused on the analysis of advantages and disadvantages and the effective impact of fuzzy clustering in the process of training of this type of neural networks.

The result of this research is to develop in the first level of three simple methods for clustering fuzzy neural networks for the training of radial basis functions. Concretely in the first method, was used the very known method of fuzzy c -means (FCM) for the preprocessing of the training data and then the method of weighted fuzzy c -means (FCM) for the calculation of the core parameters of the neural network hidden nodes. In the second method, was used the optimal clustering for the determination of the number of cores of radial basis functions as well as the parameters of neural network. Finally, we proposed a new method [63] that combines input-output fuzzy clustering and optimal fuzzy clustering for the efficient design of radial basis function neural networks.

In second level, were developed two innovative methods for the training of radial basis function neural networks. Concretely in first method is proposed a new innovator hybrid method of clustering, which combines in a uniform fashion crisp and fuzzy clustering for the calculation of the neural network core parameters. The second proposed method uses information granulation for the calculation of the core centers and a new metric distance for the calculation of widths of the radial basis functions.

In all the proposed methods became practical application in real models and functions approximation, the results of which showed that the proposed methods produce reliable, precise and compact radial basis function neural networks. The comparison with the international bibliography showed that RBF networks that was created have better results with regard to the mean square error – MSE as well as the normalized mean square error – NRMSE so much for training data as long as for the test data.

Λίστα Σχημάτων

- Σχήμα 1.1** Δίκτυο ενός Στρώματος
- Σχήμα 1.2** Δίκτυο Πολλών Στρωμάτων
- Σχήμα 1.3** Δομή Δικτύου Kohonen
- Σχήμα 1.4** Δομή Επαναλαμβανόμενου δικτύου
- Σχήμα 1.5** Βασική τοπολογία ενός νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης
- Σχήμα 2.1** Επαναληπτική διαδικασία του αλγόριθμου c-μέσων (c-Means)
- Σχήμα 2.2** Ευαισθησία του αλγόριθμου c-Means στην επιλογή αρχικών κέντρων
- Σχήμα 2.3** Ασαφή σύνολα με: (α) τριγωνική, (β) τραπεζοειδή και (γ) γκαουσιανή συνάρτηση συμμετοχής
- Σχήμα 3.1** Το διάγραμμα του προτεινόμενου αλγόριθμου
- Σχήμα 3.2** Η αρχική και προβλεφθείσα έξοδος αντίστοιχα για το παράδειγμα Mackey-Glass. (Δεδομένα εκπαίδευσης)
- Σχήμα 3.3** Η αρχική και προβλεφθείσα έξοδος αντίστοιχα για το συνθετικό μονοδιάστατο παράδειγμα δεδομένων (δεδομένα δοκιμής)
- Σχήμα 4.1** Ο προτεινόμενος αλγόριθμος
- Σχήμα 5.1** Το διάγραμμα ροής του προτεινόμενου αλγόριθμου
- Σχήμα 6.1** Χρήση τριών επαναλήψεων για να περιγράψει τη μετάβαση από ασαφή σε διακριτή κατάσταση ενός διανύσματος δεδομένων εκπαίδευσης που περιβάλλεται από 10 κέντρα συστάδων, όπου η περιοχή καθορίζει το σύνολο κέντρων συστάδων που επηρεάζονται από το διάνυσμα εκπαίδευσης σε κάθε επανάληψη
- Σχήμα 6.2** (α) Ασαφής (σκιασμένη περιοχή) και διακριτοί (λευκή περιοχή) τομείς μιας υβριδικής συστάδας, (β) ένα στιγμιότυπο της διαδικασίας εκπαίδευσης που παρουσιάζει δύο αλληλεπιδρούσες υβριδικές συστάδες, όπου το μαύρο τμήμα αντιστοιχεί στην κοινή ασαφή περιοχή τους

- Σχήμα 6.3** Υβριδικός διαμερισμός με έξι συστάδες, όπου οι ασαφείς (σκιασμένη περιοχή) και διακριτοί (λευκές περιοχές) όροι συνυπάρχουν για: (α) μικρή τιμή θ , και (β) μεγάλη τιμή θ
- Σχήμα 6.4** Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού των επαναλήψεων για: (α) $c = 5$, και (δ) $c = 10$
- Σχήμα 6.5** Ασαφείς (συμπαγείς τετράγωνα) και διακριτοί (συμπαγείς κύκλοι) τομείς του χώρου των εισόδων χρησιμοποιώντας $c = 3$ για: (α) $\theta = 0.1$, (β) $\theta = 0.3$, και (γ) $\theta = 0.5$, όπου τα κέντρα συστάδων συμβολίζονται από τα κενά τρίγωνα
- Σχήμα 6.6** Ασαφείς (συμπαγείς τετράγωνα) και διακριτοί (συμπαγείς κύκλοι) τομείς του χώρου των εισόδων χρησιμοποιώντας $c = 5$ για: (α) $\theta = 0.1$, (β) $\theta = 0.3$, και (γ) $\theta = 0.5$, όπου τα κέντρα συστάδων συμβολίζονται από τα κενά τρίγωνα
- Σχήμα 6.7** Μέσος αριθμός υπολογισμών απόστασης ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 1
- Σχήμα 6.8** Γράφημα διασποράς για $c = 10$ και $\theta = 0.5$ μεταξύ της αρχικής και προβλεφθείσας εξόδου του προτεινόμενου δικτύου
- Σχήμα 6.9** Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού επανάληψης για: (α) $c = 10$, και (β) $c = 20$
- Σχήμα 6.10** Μέσος αριθμός υπολογισμών απόστασης ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 2
- Σχήμα 6.11** RMSE μέσες τιμές που λαμβάνονται από τον προτεινόμενο αλγόριθμο ($\theta = 0.5$), το RBFN 1, το RBFN 2 και το RBFN 3 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής
- Σχήμα 6.12** Γραφήματα διασποράς για $c = 25$ και $\theta = 0.3$ μεταξύ της αρχικής και προβλεφθείσας εξόδου: (α) δεδομένα εκπαίδευσης, (β) δεδομένα δοκιμής
- Σχήμα 6.13** Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού επανάληψης για: (α) $c = 5$, και (β) $c = 10$

- Σχήμα 6.14** Μέσος αριθμός υπολογισμών απόστασης ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 3
- Σχήμα 6.15** Γραφήματα διασποράς για $c = 10$ και $\theta = 0.3$ μεταξύ της αρχικής και προβλεφθείσας εξόδου: (α) δεδομένα εκπαίδευσης, (β) δεδομένα δοκιμής
- Σχήμα 6.17** Δευτερόλεπτα ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα δεδομένων Abalone
- Σχήμα 7.1** Βασική τοπολογία του προτεινόμενου αλγορίθμου
- Σχήμα 7.2** Ασαφής διαμερισμός στο δισδιάστατο χώρο εισόδου
- Σχήμα 7.3** Γεωμετρική ερμηνεία της υπόθεσης του l ασαφές υποχώρου και της σχετικής Ευκλείδειας απόστασης στον δυσδιάστατο χώρο
- Σχήμα 7.4** Προσδιορισμός της ελάχιστης τιμής της παραμέτρου θ στον δισδιάστατο χώρο των εισόδων
- Σχήμα 7.5** Δύο τροποποιήσεις του αλγορίθμου έτσι ώστε να μπορεί να χρησιμοποιηθεί για τον υπολογισμό του κατάλληλου αριθμού κρυφών κόμβων
- Σχήμα 7.6** MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\mathcal{G} = 0.51$), του RBFN 1, και RBFN 2 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής
- Σχήμα 7.7** MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\mathcal{G} = 0.51$), του RBFN 1, και RBFN 2 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής
- Σχήμα 7.8** MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\mathcal{G} = 0.51$), του RBFN 1 και RBFN 2 ως λειτουργία του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής

Λίστα Πινάκων

- Πίνακας 3.1** Σύγκριση αποτελεσμάτων για τα δεδομένα του παραδείγματος Mackey-Glass
- Πίνακας 3.2** Συγκριτικά αποτελέσματα για το παράδειγμα προσέγγισης συνάρτησης
- Πίνακας 4.1** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης μονοδιάστατης συνθετικής συνάρτησης
- Πίνακας 4.2** Αποτελέσματα σύγκρισης για το παράδειγμα προσέγγισης μονοδιάστατης συνθετικής συνάρτησης
- Πίνακας 4.3** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα στατικής συνάρτησης (static function)
- Πίνακας 4.4** Αποτελέσματα σύγκρισης για το παράδειγμα στατικής συνάρτησης (static function)
- Πίνακας 4.5** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα Mackey-Glass
- Πίνακας 4.6** Αποτελέσματα σύγκρισης για το παράδειγμα Mackey-Glass
- Πίνακας 5.1** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης συνθετικής μονοδιάστατης συνάρτησης.
- Πίνακας 5.2** Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές MSE
- Πίνακας 5.3** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης δισδιάστατης συνάρτησης.
- Πίνακας 5.4** Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές NRMSE
- Πίνακας 5.5** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της προσέγγισης στατικής δισδιάστατης συνάρτησης
- Πίνακας 5.6** Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE
- Πίνακας 5.7** Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE
- Πίνακας 5.8** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της συνθετικής τρισδιάστατης συνάρτησης

- Πίνακας 5.9** Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα Abalone
- Πίνακας 5.10** Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE
- Πίνακας 6.1** NRMSE μέσες τιμές και οι αντίστοιχες διαφορές (variances)
- Πίνακας 6.2** Συγκριτική ανάλυση από την άποψη των μέσων τιμών NRMSE
- Πίνακας 6.3** Ένα δίκτυο που παράγεται από τον προτεινόμενο αλγόριθμο για $c = 10$ και $\theta = 0.5$
- Πίνακας 6.4** Μέσες τιμές RMSE και οι αντίστοιχες διαφορές (variances) για τα δεδομένα εκπαίδευσης
- Πίνακας 6.5** Μέσες τιμές RMSE και οι αντίστοιχες διαφορές (variances) για τα δεδομένα δοκιμής
- Πίνακας 6.6** Συγκριτική ανάλυση από την άποψη των μέσων τιμών RMSE
- Πίνακας 6.7** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης.
- Πίνακας 6.8** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής.
- Πίνακας 6.9** Συγκριτική ανάλυση από την άποψη του MSE.
- Πίνακας 6.10** Ένα δίκτυο που παράγεται από τον προτεινόμενο αλγόριθμο για $c = 12$ και $\theta = 0.5$
- Πίνακας 6.11** NRMSE μέσες τιμές και οι αντίστοιχες διαφορές (variances) για τα δεδομένα εκπαίδευσης
- Πίνακας 6.12** NRMSE μέσες τιμές και οι αντίστοιχες διαφορές (variances) για τα δεδομένα δοκιμής
- Πίνακας 6.13** Σύνολο δεδομένων για τη διαδικασία ωρίμανσης αποστάγματος κρασιού.
- Πίνακας 6.14** RMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης και δοκιμής για τα δεδομένα Auto MPG
- Πίνακας 6.15** RMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης και δοκιμής του για τα δεδομένα Abalone
- Πίνακας 6.16** Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Auto MPG

- Πίνακας 6.17** Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Abalone
- Πίνακας 6.18** Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Auto MPG (2 εισόδους)
- Πίνακας 7.1** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης
- Πίνακας 7.2** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής
- Πίνακας 7.3** Συγκριτική ανάλυση από την άποψη του MSE.
- Πίνακας 7.4** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του RMSE για τα δεδομένα εκπαίδευσης
- Πίνακας 7.5** Συγκριτική ανάλυση των τριών δικτύων από την άποψη RMSE για τα δεδομένα δοκιμής
- Πίνακας 7.6** Συγκριτική ανάλυση από την άποψη των μέσων τιμών RMSE
- Πίνακας 7.7** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης.
- Πίνακας 7.8** Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής.
- Πίνακας 8.1** Σύγκριση των προτεινόμενων μεθόδων όσον αφορά τα δεδομένα εκπαίδευσης και τα δεδομένα δοκιμής.

Εισαγωγή

Στην παρούσα εργασία, μελετήθηκαν τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (Radial Basis Function Neural Networks - RBFNN). Τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης είναι μια κατηγορία δικτύων που βασίζεται στις συναρτήσεις ακτινικού τύπου. Χαρακτηριστικό αυτών των συναρτήσεων είναι ότι εξαρτώνται από την ευκλείδεια απόσταση του διανύσματος του κέντρου τους από ένα σημείο x . Τα νευρωνικά δίκτυα ακτινικής βάσης, δομικά, αποτελούνται από το στρώμα εισόδου, το στρώμα εξόδου και ένα μόνο κρυφό στρώμα στο ενδιάμεσο. Χρησιμοποιούν μάθηση με επίβλεψη για να εκπαιδευτούν και αλγόριθμους συσταδοποίησης όπως ο c -Μέσων (c -Means) και ο αλγόριθμος Ασαφών c -Μέσων (Fuzzy c -Means) για τον εντοπισμό των κέντρων των συστάδων τους και την καταχώρηση των δεδομένων που τους δίνονται σε αυτά. Η συγκεκριμένη κατηγορία νευρωνικών δικτύων θεωρείται ότι έχει καλύτερες δυνατότητες παρεμβολής, είναι απλό σε δομή δίκτυο και διαθέτει αλγόριθμους που μαθαίνουν γρήγορα. Γι αυτούς τους λόγους τα τελευταία χρόνια είναι μια από τις πιο δημοφιλείς κατηγορίες νευρωνικών δικτύων.

Ουσιαστικά αναπτύχθηκαν πέντε νέες μέθοδοι εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Σε κάθε μέθοδο που αναπτύχθηκε, έγινε η σύγκριση με άλλες μεθόδους που υπάρχουν στην διεθνή βιβλιογραφία. Η σύγκριση έγινε με σκοπό να εξεταστεί κατά πόσο το νευρωνικό εκφέρει καλύτερα αποτελέσματα στις τιμές του μέσου τετραγωνικού σφάλματος (mean square error – MSE) ή της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (root mean square error - RMSE) τόσο στη διαδικασία εκπαίδευσης του νευρωνικού, όσο και στη δοκιμή του, κατά πόσο ικανοποιητικά δηλαδή πραγματοποιήθηκε η εκπαίδευση. Τα αποτελέσματα των συγκρίσεων υπάρχουν στο τέλος του κάθε κεφαλαίου.

Ο αλγόριθμος που χρησιμοποιήθηκε στην πρώτη μέθοδο [64], είναι επαναληπτικός και υπολογίζει τόσο τον αριθμό των κρυφών κόμβων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBFN) όσο και όλες τις παραμέτρους του

νευρωνικού δικτύου. Αρχικά γίνεται μία προεπεξεργασία των δεδομένων εισόδου χρησιμοποιώντας τον αλγόριθμο ασαφής συσταδοποίησης – Ασαφών c-Μέσων (FCM) στο σύνολο των δεδομένων του. Στο δεύτερο βήμα, τα κέντρα συστάδων που εξήχθησαν από τη FCM θεωρούνται το νέο σύνολο δεδομένων, το οποίο συσταδοποιείται περαιτέρω με τη χρησιμοποίηση του αλγορίθμου σταθμισμένων (weighted) Ασαφών c-Μέσων (FCM). Στο τρίτο βήμα, βασισμένο στο προκύπτον ασαφή διαμερισμό, εκτελούμε μια αρχική εκτίμηση των κέντρων των κρυφών κόμβων. Τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης υπολογίστηκαν με μία νέα μέθοδο που προτάθηκε στην παρούσα διδακτορική διατριβή. Η φάση 2 ασχολείται για την αρχική εκτίμηση παραμέτρων των βαρών σύνδεσης νευρώνων του νευρωνικού δικτύου ακτινικής βάσης, η οποία ολοκληρώνεται με τη μέθοδο των ελαχίστων τετραγώνων. Τέλος, η φάση 3 εκτελεί μια διαδικασία βελτιστοποίησης οπισθοδιάδοσης (back-propagation) για να τελειοποιήσει όλες τις παραμέτρους του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBFN). Οι ανωτέρω φάσεις εφαρμόζονται επαναληπτικά, όπου σε κάθε επανάληψη ο αριθμός κόμβων αυξάνεται από δύο έως ότου βρίσκεται η ακρίβεια προσέγγισης μέσα στα αποδεκτά επίπεδα. Σε αυτό το σημείο, πρέπει να τονιστεί ότι ο αριθμός των κόμβων του RBFN είναι ίσος με τον αριθμό συστάδων που παράγονται από τη μέθοδο σταθμισμένων (weighted) Ασαφών c-Μέσων. Κατά συνέπεια, αρχίζοντας από δύο συστάδες (κόμβοι) λαμβάνουμε τελικά τον αριθμό κόμβων για τον οποίο η απόδοση του δικτύου είναι ακριβής.

Μία δεύτερη μέθοδος [66] προτείνεται για την εκπαίδευση νευρωνικών δικτύων ακτινικής βάσης. Η μέθοδος στοχεύει προς τη μείωση του αριθμού κρυφών κόμβων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης χωρίς απώλεια της απόδοσης. Τρεις καλά γνωστοί δείκτες βέλτιστης ασαφούς συσταδοποίησης έχουν χρησιμοποιηθεί προκειμένου να μειωθεί ο αριθμός κρυφών κόμβων του νευρωνικού δικτύου ακτινικής βάσης. Δύο παραδείγματα προσομοίωσης έχουν χρησιμοποιηθεί για να ελέγξουν την ισχύ αυτής της μεθόδου. Τα αποτελέσματα αποδεικνύουν την αποτελεσματικότητα της προτεινόμενης προσέγγισης.

Η δεύτερη μέθοδος αρχικά συσταδοποιεί τα δεδομένα με τη χρήση του αλγορίθμου ανταγωνιστικής εκμάθησης (competitive learning) που ονομάζεται κβάντιση διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ). Ο LVQ είναι μια

προσέγγιση που τροποποιεί τα διανύσματα κωδικοποίησης σε πραγματικό χρόνο (online) κάθε φορά που εμφανίζεται ένα καινούριο διανυσματικό δεδομένο. Στη συνέχεια υπολογίζει την ευκλείδεια απόσταση ανάμεσα στα κέντρα και τα δεδομένα, και τη συμμετοχή του κάθε δεδομένου στα προκαθορισμένα κέντρα. Έπειτα χρησιμοποιεί τη συνάρτηση συμμετοχής για να υπολογίσει τα ελάχιστα τετράγωνα και τέλος υπολογίζει το μέσο τετραγωνικό σφάλμα για τους νευρώνες εκπαίδευσης και δοκιμής. Η παράμετρος της οποίας οι τιμές μεταβάλλονταν ήταν η παράμετρος ασάφειας.

Μια τρίτη μέθοδος [63], προτάθηκε η οποία συνδυάζει την ασαφή συσταδοποίηση και τη βέλτιστη ασαφή συσταδοποίηση για τον αποτελεσματικό σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Πρώτα, εφαρμόστηκε η μέθοδος Ασαφών *c*-Μέσων (FCM) στο χώρο εισόδου-εξόδου με σκοπό να γίνει η προεπεξεργασία των διαθέσιμων στοιχείων. Οι προκύπτουσες συστάδες προβάλλονται στο χώρο εισόδου. Τα αντίστοιχα κέντρα συστάδων θεωρούνται ως ένα νέο σύνολο δεδομένων, το οποίο στη συνέχεια αναλύεται περισσότερο, χρησιμοποιώντας τη βέλτιστη μέθοδο της ασαφούς συσταδοποίησης από πλευράς της μεθόδου σταθμισμένων (weighted) Ασαφών *c*-Μέσων (FCM). Για να επιτευχθεί αυτό, αναπτύξαμε έναν νέο δείκτη εγκυρότητας, ο οποίος χρησιμοποιείται για να προσδιορίσει τον κατάλληλο αριθμό των κρυφών RBF κόμβων. Ο αλγόριθμος έχει εφαρμοστεί με επιτυχία σε γνωστές σειρές δεδομένων, όπου οι επιδόσεις του έχουν δοκιμαστεί και αξιολογηθεί.

Στην τέταρτη μέθοδο [65], τα δεδομένα συσταδοποιήθηκαν με ένα νέο υβριδικό αλγόριθμο ασαφούς συσταδοποίησης που συνδυάζει με ομοιόμορφο τρόπο την διακριτή (crisp) με την ασαφή (fuzzy) συσταδοποίηση (clustering). Με αυτόν τον αλγόριθμο υπολογίστηκαν τα κέντρα των πυρήνων των συναρτήσεων ακτινικής βάσης. Στη συνέχεια χρησιμοποιήθηκε η μέθοδος που χρησιμοποιήθηκε και στον προηγούμενο αλγόριθμο εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης, για να υπολογιστούν τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης και τέλος χρησιμοποιήθηκε η μέθοδος των ελαχίστων τετραγώνων για τον υπολογισμό των συναπτικών βαρών (synaptic weights) των νευρώνων. Και σε αυτή τη μέθοδο τα αποτελέσματα αποτυπώνονται από το μέσο τετραγωνικό σφάλμα (MSE) των δεδομένων εκπαίδευσης και δοκιμής. Πλήθος προσομοιώσεων πραγματοποιήθηκαν για να συγκριθεί η προτεινόμενη μέθοδος με μεθόδους που υπάρχουν στη διεθνή βιβλιογραφία. Τα

αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος παράγει αξιόπιστα και συμπαγή νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης.

Τέλος, μια πέμπτη μέθοδος, προτάθηκε η οποία χρησιμοποιεί την ασαφή κοκκοποίηση (fuzzy granulation) για την εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Συγκεκριμένα, αρχικά τα δεδομένα εκπαίδευσης συσταδοποιήθηκαν με την μέθοδο του ασαφούς διαμερισμού (fuzzy partition). Τα κέντρα που προέκυψαν θεωρήθηκαν ένα νέο σύνολο δεδομένων που συσταδοποιήθηκε περαιτέρω με την χρήση της μεθόδου των σταθμισμένων (weighted) Ασαφών *c*-Μέσων (FCM). Με τον παραπάνω τρόπο υπολογίστηκαν τα κέντρα των κρυφών κόμβων. Με την βοήθεια μία νέας μετρικής απόστασης που αναπτύχθηκε στην παρούσα διατριβή υπολογίστηκαν τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης και τέλος με τη χρήση των ελαχίστων τετραγώνων υπολογίστηκαν τα συναπτικά βάρη των νευρώνων. Πλήθος πειραμάτων έδειξαν την απόδοση και την υπεροχή της προτεινόμενης μεθόδου για την εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης σε σχέση με άλλες μεθόδους που υπάρχουν στη διεθνή βιβλιογραφία.

Το πειραματικό κομμάτι υλοποιήθηκε στο περιβάλλον του MatLab.

Το περιεχόμενο της παρούσας διδακτορικής διατριβής έχει οργανωθεί ως εξής:

Στο 1^ο κεφάλαιο περιγράφονται ορισμένες έννοιες σχετικές με τα τεχνητά νευρωνικά δίκτυα (artificial neural networks). Πιο συγκεκριμένα, αναφέρονται οι διάφοροι τύποι τεχνητών νευρωνικών δικτύων, οι συναρτήσεις ακτινικής βάσης, τα αντίστοιχα δίκτυα και οι στρατηγικές μάθησης που χρησιμοποιούνται από αυτά.

Στο 2^ο κεφάλαιο λαμβάνει χώρα η βιβλιογραφική ανασκόπηση των βασικών αρχών της διακριτής όσο και της ασαφούς συσταδοποίησης δεδομένων και αναλύονται διάφοροι μέθοδοι εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης με τη χρήση διακριτής και ασαφούς συσταδοποίησης.

Στο 3^ο κεφάλαιο, αναπτύσσεται η πρώτη [64] από τις μεθόδους εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Πιο συγκεκριμένα, σε πρώτη φάση γίνεται προεπεξεργασία των δεδομένων εισόδου με την μέθοδο της ασαφούς συσταδοποίησης Ασαφών *c*-Μέσων (Fuzzy *c*-Means - FCM). Τα κέντρα που προκύπτουν χρησιμοποιούνται σαν τα δεδομένα εισόδου για την μέθοδο συσταδοποίησης σταθμισμένων (weighed) Ασαφών *c*-Μέσων (FCM). Επίσης, γίνεται μία αρχική

εκτίμηση των πλατών (widths) των συναρτήσεων ακτινικής βάσης με μία νέα μέθοδο που αναπτύχθηκε στην παρούσα διδακτορική διατριβή. Στη δεύτερη φάση, γίνεται μια αρχική εκτίμηση των βαρών σύνδεσης μεταξύ των νευρώνων χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων. Τέλος, στη τρίτη φάση χρησιμοποιείται ο αλγόριθμος οπισθοδιάδοσης (back-propagation) για την βελτιστοποίηση των παραμέτρων του νευρωνικού δικτύου. Η μέθοδος χρησιμοποιείται για την αναγνώριση του χαοτικού συστήματος Mackey-Glass καθώς και για την προσέγγιση συνάρτησης.

Στο 4^ο κεφάλαιο παρουσιάζεται μια νέα μέθοδος [66] εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης χρησιμοποιώντας τη μέθοδο της βέλτιστης συσταδοποίησης (optimal clustering). Συγκεκριμένα, σε πρώτο επίπεδο γίνεται μια πρώτη συσταδοποίηση των δεδομένων εισόδου από τη μονάδα προεπεξεργασίας χρησιμοποιώντας τον αλγόριθμο ανταγωνιστικής εκμάθησης (competitive learning) που ονομάζεται κβάντιση διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ) [45]. Ο LVQ είναι μια προσέγγιση που τροποποιεί τα διανύσματα κωδικοποίησης σε πραγματικό χρόνο (on-line) κάθε φορά που εμφανίζεται ένα καινούριο διανυσματικό δεδομένο. Στο δεύτερο επίπεδο γίνεται ο υπολογισμός των κρυφών κόμβων του νευρωνικού δικτύου χρησιμοποιώντας τον αλγόριθμο βέλτιστης συσταδοποίησης. Εδώ πρέπει να αναφερθεί ότι χρησιμοποιούνται τρεις δείκτες βέλτιστης ασαφούς συσταδοποίησης (validity index) και οι οποίοι συγκρίνονται στην ενότητα των πειραματικών εφαρμογών του τετάρτου κεφαλαίου. Επιπλέον σε αυτό το επίπεδο γίνεται αρχική εκτίμηση των πλατών (widths) των συναρτήσεων ακτινικής βάσης (radial basis functions) με μία νέα μέθοδο [58] που προτείνετε στο παρόν κεφάλαιο. Στο επόμενο επίπεδο γίνεται αρχική εκτίμηση των βαρών (weights) σύνδεσης των νευρώνων μεταξύ των κρυφών κόμβων και του κόμβου εξόδου, χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων (least-squares). Τέλος, στο τελευταίο επίπεδο αναλύεται η διαδικασία βελτιστοποίησης οπισθοδιάδοσης (back-propagation) των παραμέτρων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης. Η μέθοδος χρησιμοποιείται για την προσέγγιση συνάρτησης καθώς και την αναγνώριση του χαοτικού συστήματος Mackey-Glass. Και στις δύο περιπτώσεις έλαβε χώρα ένας αριθμός προσομοιώσεων που δείχνουν ότι η μέθοδος χαρακτηρίζεται από ακριβή προσέγγιση των υπό μελέτη συστημάτων με ταυτόχρονη δημιουργία μικρού αριθμού κρυφών κόμβων του νευρωνικού δικτύου.

Στο 5^ο κεφάλαιο προτάθηκε μια τρίτη μέθοδος [63] που συνδυάζει ασαφή συσταδοποίηση και τη βέλτιστη ασαφή συσταδοποίηση για τον αποτελεσματικό σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Πρώτα, εφαρμόστηκε η μέθοδος Ασαφών *c*-Μέσων (FCM) στο χώρο εισόδου-εξόδου με σκοπό να γίνει η προεπεξεργασία των διαθέσιμων δεδομένων. Οι προκύπτουσες συστάδες προβάλλονται στο χώρο εισόδου. Τα αντίστοιχα κέντρα συστάδων θεωρούνται ως ένα νέο σύνολο δεδομένων, το οποίο στη συνέχεια αναλύεται περισσότερο, χρησιμοποιώντας τη βέλτιστη μέθοδο της ασαφούς συσταδοποίησης από πλευράς της μεθόδου σταθμισμένων (weighted) Ασαφών *c*-Μέσων (FCM). Για να επιτευχθεί αυτό, αναπτύξαμε ένα νέο δείκτη εγκυρότητας, ο οποίος χρησιμοποιείται για να προσδιορίσει τον κατάλληλο αριθμό των κρυφών RBF κόμβων. Ο αλγόριθμος έχει εφαρμοστεί με επιτυχία σε γνωστές σειρές δεδομένων, όπου οι επιδόσεις του έχουν δοκιμαστεί και αξιολογηθεί.

Στο 6^ο κεφάλαιο, αναπτύσσεται μία νέα μέθοδος [65] βασισμένη στη συσταδοποίηση για την εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBFNN). Σε αντίθεση με τις υπάρχουσες προσεγγίσεις, αναπτύσσεται μια εξειδικευμένη στρατηγική εκμάθησης που συνδυάζει τις μορφές της ασαφής (fuzzy) και διακριτής (crisp) συσταδοποίησης (clustering). Η διακριτή (crisp) συσταδοποίηση είναι μια πολύ γρήγορη διαδικασία, όμως πολύ ευαίσθητη στην αρχικοποίηση. Αφ' ετέρου, η ασαφής (fuzzy) συσταδοποίηση (clustering) μειώνει την εξάρτηση στην αρχικοποίηση εντούτοις, αποτελεί μια αργή διαδικασία εκπαίδευσης. Η προτεινόμενη στρατηγική στοχεύει να βρει μια ισορροπία μεταξύ αυτών των δύο διαφορετικών αποτελεσμάτων. Οι παραχθείσες συστάδες κατέχουν τις ασαφείς και διακριτές περιοχές και, επομένως, το τελικό αποτέλεσμα είναι ένας υβριδικός διαμερισμός (partition), όπου οι ασαφείς και οι διακριτοί όροι συνυπάρχουν. Οι υβριδικές συστάδες περιλαμβάνονται άμεσα στη διαδικασία εκτίμησης των παραμέτρων του νευρωνικού δικτύου. Συγκεκριμένα, τα κέντρα των συναρτήσεων ακτινικής βάσης συμπίπτουν με τα κέντρα των συστάδων. Προκειμένου να υπολογιστούν αποτελεσματικά τα αντίστοιχα πλάτη (widths), χρησιμοποιήτε η μέθοδος [66] που αναπτύχθηκε στο προηγούμενο κεφάλαιο. Για αυτόν τον λόγο, ο σχεδιασμός του δικτύου γίνεται μια γρήγορη και αποδοτική διαδικασία. Η προτεινόμενη μέθοδος εφαρμόζεται επιτυχώς σε διάφορες πειραματικές περιπτώσεις, όπου τα παραχθέντα δίκτυα αποδεικνύονται ιδιαίτερα ακριβή και συμπαγή στο μέγεθος.

Στο 7^ο κεφάλαιο, προτείνεται ένας τριών-επιπέδων ασαφής αλγόριθμος κοκκοποίησης πληροφοριών (information granulation), ο οποίος περιλαμβάνεται ρητά στη διαδικασία εκπαίδευσης των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Στο πρώτο στάδιο καθορίζεται μια προσανατολισμένη προς το χρήστη κοκκοποίηση σε όλο το πεδίο ορισμού, η οποία απαιτείται για να ικανοποιήσει ορισμένες ιδιότητες. Το δεύτερο και τρίτο στάδιο αποτελούν μια από κάτω προς τα επάνω (bottom-up) αλγοριθμική προσέγγιση βελτιστοποίησης κοκκοποίησης. Η σημαντικότερη καινοτομία αυτής της μεθόδου βρίσκεται στην ανάπτυξη μιας μετρικής απόστασης (metric distance) μεταξύ των ασαφών κόκκων (fuzzy granules) που είναι σε θέση να συνδυάσει τις χαμηλού επιπέδου πληροφορίες που κωδικοποιούνται στο πρώτο στάδιο κοκκοποίησης με τις υψηλού επιπέδου πληροφορίες που κωδικοποιούνται στο τελευταίο στάδιο κοκκοποίησης. Για αυτόν τον λόγο, η εκτίμηση των παραμέτρων του νευρωνικού δικτύου RBF γίνεται μια απλή διαδικασία. Η προτεινόμενη μέθοδος εφαρμόζεται επιτυχώς σε τρεις προσομοιώσεις, όπου τα παραχθέντα δίκτυα αποδείχθηκαν πολύ ακριβή και συμπαγή στο μέγεθος.

Τέλος, στο 8^ο κεφάλαιο γίνεται συζήτηση επί των αποτελεσμάτων και αναφέρονται τα συμπεράσματα της παρούσης διδακτορικής διατριβής καθώς και η μελλοντική εργασία που μπορεί να πραγματοποιηθεί.

Τεχνητά Νευρωνικά Δίκτυα

1.1 Εισαγωγή

Τα Τεχνητά Νευρωνικά Δίκτυα ΤΝΔ (Artificial Neural Networks – ANNs) υπήρξαν το αποτέλεσμα της προσπάθειας του ανθρώπου να μιμηθεί την δομή του ανθρωπίνου νευρικού συστήματος. Συνεπώς η φιλοσοφία των τεχνητών νευρωνικών δικτύων είναι δανεισμένη από την ανθρώπινη φύση και οι ομοιότητες τους με το νευρικό σύστημα του ανθρώπου δεν περιορίζονται μόνο στο όνομα.

Η πρώτη προσπάθεια για την κατανόηση του ανθρωπίνου εγκεφάλου έγινε με την πρωτοποριακή εργασία του Ramon y Cayal το 1911 [32] ο οποίος εισήγαγε την ιδέα των νευρώνων ως δομικά στοιχεία του εγκεφάλου. Σήμερα εκτιμάται ότι υπάρχουν γύρω στα δέκα δισεκατομμύρια νεύρα στον ανθρώπινο εγκεφαλικό φλοιό και εξήντα τρισεκατομμύρια συνδέσεις ή όπως αλλιώς λέγονται συνάψεις, η λειτουργία των οποίων αναλύεται παρακάτω.

Κατά την γέννηση του οργανισμού ο εγκέφαλος έχει ευρεία δομή και την ικανότητα να δημιουργεί τους δικούς του κανόνες τους οποίους εμείς αναφέρουμε συχνά ως εμπειρία. Πράγματι, η εμπειρία είναι μια μακροχρόνια διαδικασία με την πιο σημαντική ανάπτυξη να λαμβάνει χώρα στα δύο πρώτα χρόνια από την γέννηση. Κατά τη διάρκεια του σταδίου αυτού, διαμορφώνονται γύρω στις ένα εκατομμύριο συνάψεις το δευτερόλεπτο.

Οι συνάψεις είναι δομικά και λειτουργικά στοιχεία του εγκεφάλου που μεσολαβούν μεταξύ των αντιδράσεων των νευρώνων. Η πιο κοινή σύναψη είναι η χημική η οποία μετατρέπει τα ηλεκτρικά σήματα που δέχεται από ένα νεύρο σε χημικά για να τα μετατρέψει εκ νέου σε ηλεκτρικά και να τα μεταδώσει σε ένα άλλο νεύρο. Με διαφορετικά λόγια θεωρείται ότι οι συνάψεις εξάπτουν ή διεγείρουν τα νεύρα.

1.2 Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) - (Artificial Neural Networks – ANNs)

Όπως είναι κατανοητό η εγκεφαλική λειτουργία που στηρίζεται στην παραπάνω δομή χαρακτηρίζεται από μεγάλη ικανότητα προσαρμογής στο περιβάλλον. Η πλαστικότητα αυτή του εγκεφάλου επιτυγχάνεται είτε με τη δημιουργία νέων συνδέσεων, είτε με την μετατροπή κάποιων που ήδη υπάρχουν. Όπως η πλαστικότητα αυτή παρουσιάζεται στο φυσικό νευρικό σύστημα έτσι εμφανίζεται και στα τεχνητά νευρικά συστήματα. Σε μια γενική μορφή λοιπόν ένα Τεχνητό Νευρωνικό Δίκτυο (ΤΝΔ) είναι μια προσπάθεια για μοντελοποίηση της εγκεφαλικής λειτουργίας για την επίτευξη κάποιου στόχου. Μια ενδιάμεση διαδικασία απαραίτητη για τη σωστή του λειτουργία είναι η διαδικασία της εκπαίδευσης, όπως ακριβώς στη φύση υπάρχει η εμπειρία. Τέλος σε αντιστοιχία με τις συνάψεις και τα νεύρα έχουμε τις υπολογιστικές μονάδες και τις συνδέσεις μεταξύ αυτών. Μπορούμε να δώσουμε εν συντομία τον παρακάτω ορισμό για τα νευρωνικά δίκτυα.

Ένα ΤΝΔ είναι ένας συμπαγής παράλληλος επεξεργαστής που έχει την ικανότητα να αποκτά γνώση και να την κάνει διαθέσιμη για χρήση. Έχει δε τις ακόλουθες ομοιότητες με τον ανθρώπινο εγκέφαλο.

- Η γνώση αποκτάται από το δίκτυο μέσω μιας διαδικασίας εκμάθησης.
- Τα βάρη των συνδέσεων (γνωστά και ως συναπτικά βάρη) χρησιμοποιούνται για την αποθήκευση της γνώσης.

Με τον όρο παράλληλος επεξεργαστής εννοούμε τον επεξεργαστή αυτόν που έχει την δυνατότητα να κάνει με παράλληλο τρόπο παραπάνω από μια επεξεργασία την φορά. Αυτό το χαρακτηριστικό καθιστά τον εγκέφαλο ισχυρότατο υπολογιστικό εργαλείο λόγω του μεγάλου αριθμού νευρώνων όπως αναφέρθηκε προηγουμένως. Το σημαντικό αυτό χαρακτηριστικό μεταφέρθηκε και στα ΤΝΔ δίνοντας μας την ευκαιρία για ένα νέο υπολογιστικό εργαλείο το οποίο ακόμα απέχει μακριά από τις ικανότητες του προτύπου του. Χαρακτηριστικά αναφέρουμε ότι η ενεργητική ικανότητα του εγκεφάλου πλησιάζει τα 10.000.000.000.000 KJ ανά πράξη ανά δευτερόλεπτο, ενώ η τιμή που αγγίζουν οι τελειότεροι υπολογιστές σήμερα είναι 10.000 KJ ανά πράξη ανά δευτερόλεπτο.

1.3 Τρόπος Λειτουργίας ΤΝΔ

Υπάρχουν πολλά μοντέλα ΤΝΔ και είναι χρήσιμο να οριστούν κάποια κοινά στοιχεία για τα μοντέλα αυτά.

Τα ΤΝΔ λειτουργούν με τους παρακάτω δύο τρόπους.

- **Εκπαίδευση.** Κατά την λειτουργία του δικτύου γίνεται η αποθήκευση των πληροφοριών σε αυτό. Αυτό γίνεται με την συνεχή εναλλαγή κάποιων συντελεστών που ονομάζονται "βάρη των συνδέσεων", με κάποια μεθοδολογία η οποία σκοπεύει στην ελαχιστοποίηση του σφάλματος που προκύπτει από το δίκτυο.
- **Ανάκληση.** Σε αυτή τη φάση το δίκτυο ανακαλεί τις πληροφορίες τις οποίες έχει "μάθει". Δηλαδή χρησιμοποιώντας τις τιμές των βαρών των συνδέσεων τις οποίες απέκτησε στο τέλος της εκπαίδευσης, υπολογίζει στην έξοδο τα στοιχεία που θέλουμε.

Γίνεται σαφές επομένως ότι τα ΤΝΔ δεν μοιάζουν με άλλες συμβατικές υπολογιστικές τεχνικές, όπου η λύση των προβλημάτων έχει από πριν ενσωματωθεί στην κατασκευή τους.

Τα ΤΝΔ οργανώνονται σε διάφορα στρώματα (layers) το κάθε ένα από τα οποία περιλαμβάνει έναν αριθμό κόμβων. Σε κάθε θέση κόμβου τοποθετείται ένας νευρώνας. Σε ένα ΤΝΔ μπορούμε να διακρίνουμε το στρώμα εισόδου, το στρώμα εξόδου και σε ορισμένες υλοποιήσεις ένα ή περισσότερα κρυφά στρώματα. Στον απλούστερο τύπο νευρωνικού δικτύου όλοι οι κόμβοι (νευρώνες) ενός επιπέδου συνδέονται με όλους τους κόμβους του προηγούμενου επιπέδου.

Διακρίνονται δύο είδη νευρώνων:

- Οι νευρώνες εισόδου,
- Οι υπολογιστικοί νευρώνες.

Οι νευρώνες εισόδου δεν επιτελούν κανένα υπολογισμό. Αποτελούν απλά την διαμεσολάβηση ανάμεσα στις εισόδους του συστήματος και στους υπολογιστικούς νευρώνες. Στον αντίποδα, οι υπολογιστικοί νευρώνες πολλαπλασιάζουν τις εισόδους τους και με βάση τα συναπτικά βάρη που τους έχουν ανατεθεί υπολογίζουν τα

αθροίσματα των επιμέρους γινομένων. Το άθροισμα που προκύπτει αποτελεί το όρισμα της συνάρτησης ενεργοποίησης του νευρωνικού δικτύου.

Ένα τυπικός νευρώνας ενός ΤΝΔ περιγράφεται από την παρακάτω εξίσωση:

$$y_k = \phi\left(\sum_{i=0}^N w_{ki} x_{ki}\right) \quad (1.1)$$

όπου y_k : η έξοδος του k νευρώνα,

ϕ : η συνάρτηση ενεργοποίησης (ή ενεργοποίησης του νευρωνικού δικτύου)

x_{ki} : η i -οστή είσοδος του k νευρώνα

w_{ki} : το i -οστό συναπτικό βάρος του k νευρώνα.

Η συνάρτηση ενεργοποίησης του ΤΝΔ μπορεί να είναι:

- Βηματική (step)
- Γραμμική (linear)
- Μη γραμμική (non-linear)
- Στοχαστική (stochastic).

1.4 Συναρτήσεις Ακτινικής Βάσης (radial basis functions)

Μια ειδική κατηγορία συναρτήσεων ενεργοποίησης αποτελούν και οι συναρτήσεις ακτινικής βάσης (radial basis functions). Οι συναρτήσεις αυτές παίρνουν πραγματικές τιμές οι οποίες εξαρτώνται μόνο από την απόσταση του σημείου από την αρχή των αξόνων:

$$\phi(x) = \phi(\|x\|) \quad (1.2)$$

ή εναλλακτικά από ένα σημείο που έχει οριστεί ως κέντρο :

$$\phi(x, c) = \phi(\|x - c\|) \quad (1.3)$$

όπου η νόρμα (norm) $\|\cdot\|$ αναπαριστά συνήθως την Ευκλείδεια απόσταση (Euclidean distance).

Η μεγάλη χρησιμότητα των συναρτήσεων ακτινικής βάσης (radial basis

functions) έγκειται στην δυνατότητα τους να προσεγγίζουν (approximate) συναρτήσεις όταν συνδυαστούν κατάλληλα:

$$y_k = \sum_{i=1}^N w_i \phi(\|\mathbf{x} - c_i\|) \quad (1.4)$$

Στον παραπάνω τύπο η συνάρτηση που θα προσεγγιστεί ($y(\mathbf{x})$) αναπαριστάται ως το άθροισμα N συναρτήσεων ακτινικής βάσης. Θα πρέπει να σημειωθεί ότι κάθε μια από τις N συναρτήσεις ακτινικής βάσης έχει διαφορετικό κέντρο c_i και συντελεστή βάρους w_i .

Οι πιο συχνά χρησιμοποιούμενοι τύποι συναρτήσεων ακτινικής βάσης είναι οι παρακάτω:

- Gaussian ($\beta > 0$) $\phi(r) = e^{-\beta r^2}$
- Multi quadratic ($\beta > 0$) $\phi(r) = \sqrt{r^2 + \beta^2}$
- Polyharmonic spline $\phi(r) = r^k, \quad k = 1, 3, 5, \dots$

Τέλος θα πρέπει να σημειωθεί ότι σύμφωνα με την εξ. (1.4) η συνάρτηση που θα προσεγγιστεί είναι γραμμική ως προς τους συντελεστές βαρών w_i οπότε για τον υπολογισμό τους μπορεί να χρησιμοποιηθεί η μέθοδος των ελάχιστων τετραγώνων (least squares).

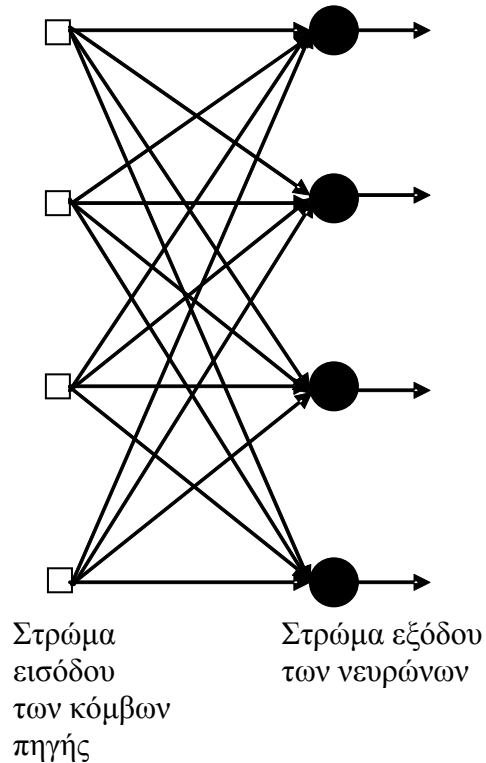
1.5 Αρχιτεκτονική Δικτύων

Ο τρόπος με τον οποίο οι νευρώνες ενός δικτύου είναι δομημένοι σχετίζεται με τον αλγόριθμο μάθησης που χρησιμοποιείται για να εκπαιδευτεί το δίκτυο.

1.5.1 Δίκτυα Ενός Στρώματος (Single-Layer Feedforward Networks)

Σε ένα νευρωνικό δίκτυο που αποτελείται από στρώματα (επίπεδα), οι νευρώνες οργανώνονται στη δομή των στρωμάτων αυτών. Στην πιο απλή μορφή ενός τέτοιου δικτύου έχουμε το στρώμα εισόδου των κόμβων πηγής που προβάλλεται στο στρώμα εξόδου των νευρώνων (computation nodes). Το δίκτυο αυτής της μορφής είναι αυστηρά

εμπροσθοτροφοδοτούμενο (feedforward). Η μορφή του φαίνεται στο παρακάτω σχήμα [32].



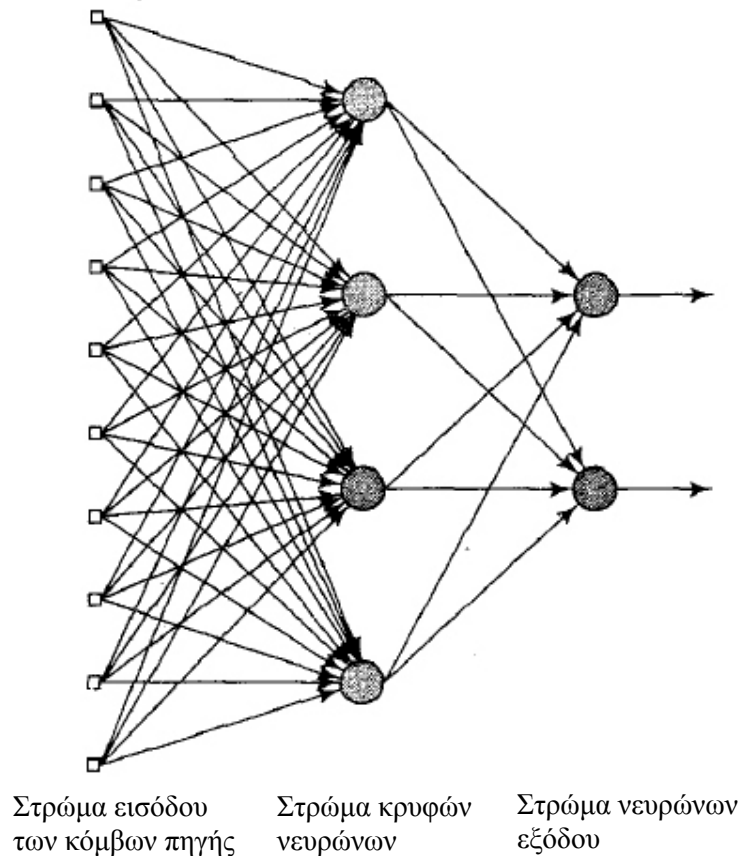
Σχήμα 1.1. Δίκτυο ενός Στρώματος

1.5.2 Δίκτυα Πολλών Στρωμάτων (Multilayer Feedforward Networks)

Μια άλλη μορφή νευρωνικού δικτύου προκύπτει από την παρουσία ενός ή περισσότερων κρυφών στρωμάτων (hidden layers), των οποίων οι νευρώνες που πραγματοποιούν τους υπολογισμούς του δικτύου λέγονται αντίστοιχα κρυφοί νευρώνες ή κρυφές μονάδες (hidden neurons or hidden units). Η λειτουργία των κρυφών νευρώνων είναι να παρεμβάλλεται ανάμεσα στην εξωτερική είσοδο (external input) και την έξοδο του δικτύου κατά κάποιο χρήσιμο τρόπο. Προσθέτοντας ένα ή περισσότερα κρυφά στρώματα, το δίκτυο διευκολύνεται στο να εξάγει στατιστικές υψηλότερης τάξης (higher-order statistics). Κατά μία έννοια το δίκτυο αποκτάει μια παγκόσμια προοπτική παρά την τοπική του συνδεσιμότητα εξαιτίας του επιπλέον συνόλου συναπτικών

συνδέσμων και την επιπλέον διάσταση των νευρωνικών αλληλεπιδράσεων. Η ικανότητα των κρυφών νευρώνων να εξάγει στατιστικές υψηλότερης τάξης είναι ιδιαίτερα χρήσιμη όταν το μέγεθος του στρώματος εισόδου είναι μεγάλο.

Οι κόμβοι εισόδου στο στρώμα εισόδου του δικτύου παρέχουν τα αντίστοιχα στοιχεία για το πρότυπο ενεργοποίησης (activation pattern – input vector), τα οποία συγκροτούν τα σήματα εισόδου που εφαρμόζονται στο δεύτερο στρώμα του νευρώνα (computation nodes), δηλαδή το πρώτο κρυφό στρώμα. Τα σήματα εξόδου του δεύτερου στρώματος χρησιμοποιούνται ως είσοδοι στο τρίτο στρώμα. Αυτή η διαδικασία συνεχίζεται στο υπόλοιπο δίκτυο. Τυπικά οι νευρώνες σε κάθε στρώμα του δικτύου έχουν ως εισόδους τα σήματα εξόδου του προηγούμενου στρώματος μόνο. Το σύνολο των σημάτων εξόδου των νευρώνων του τελικού στρώματος εξόδου του δικτύου αποτελεί τη συνολική απόκριση του δικτύου στο πρότυπο ενεργοποίησης που τροφοδοτήθηκε από τους κόμβους εισόδου στο πρώτο στρώμα.



Σχήμα 1.2. Δίκτυο Πολλών Στρωμάτων

Το παραπάνω σχήμα απεικονίζει τη μορφή ενός δικτύου εμπροσθο-τροφοδοτούμενο (feedforward) με ένα μόνο κρυφό στρώμα. Επιπλέον είναι πλήρως συνδεδεμένο (fully connected) με την έννοια ότι όλοι οι κόμβοι σε κάθε στρώμα του δικτύου ενώνονται σε κάθε ένα από τους νευρώνες στο παρακείμενο προς τα εμπρός επίπεδο. Στην περίπτωση που δεν υπήρχαν ορισμένοι από τους συνδέσμους επικοινωνίας τότε το δίκτυο θα ήταν μερικώς συνδεδεμένο (partially connected) [32].

1.5.3 Επαναλαμβανόμενα Δίκτυα (Recurrent Networks)

Ένα επαναλαμβανόμενο νευρωνικό δίκτυο ξεχωρίζει από ένα εμπροσθοτροφοδοτούμενο (feedforward) δίκτυο στο ότι έχει τουλάχιστον έναν βρόγχο ανάδρασης (feedback loop). Δηλαδή, ένα επαναλαμβανόμενο δίκτυο μπορεί να αποτελείται από ένα μόνο στρώμα νευρώνων με κάθε νευρώνα να τροφοδοτεί το σήμα εξόδου του πίσω στις εισόδους όλων των άλλων νευρώνων, όπως φαίνεται και στο σχήμα 1.2. Στη δομή του δικτύου του σχήματος, δεν υπάρχουν αυτό-τροφοδοτούμενοι βρόγχοι. Αυτό σημαίνει ότι η έξοδος του νευρώνα δεν τροφοδοτεί πίσω την είσοδό του ίδιου νευρώνα [32].

1.5.4 Δίκτυο Perceptron

Το Perceptron είναι μια απλή τοπολογία δικτύου πρόσθιας τροφοδότησης χωρίς κρυφά επίπεδα και αποτελεί ιστορικά μια πρώτη προσέγγιση τεχνητών νευρωνικών δικτύων. Προτάθηκε το 1958 από τον Rosenblatt ως ένας μηχανισμός που μπορεί να εκπαιδευτεί στην κατηγοριοποίηση προτύπων και σε διάφορες παραλλαγές, πιο απλές ή πιο σύνθετες.

Το στοιχειώδες Perceptron (elementary Perceptron) είναι η πιο απλή μορφή αυτού του τύπου δικτύου, καθώς αποτελείται αποκλειστικά από έναν τεχνητό νευρώνα. Εφόσον υπάρχει μόνο ένας νευρώνας, οι μόνες συνδέσεις που υπάρχουν είναι αυτές μεταξύ των εισόδων και του νευρώνα του δικτύου.

Στο μοντέλο Perceptron ο νευρώνας υλοποιεί την παρακάτω συνάρτηση μεταφοράς.

$$v = \sum_{i=1}^n w_i x_i - \theta \tag{1.5}$$

$$y = \varphi(v) \tag{1.6}$$

Η συνάρτηση μεταφοράς απεικονίζει το διάνυσμα εισόδου $\mathbf{X}=[x_1, x_2, \dots, x_n]^T$ στην έξοδο y . Οι παράμετροι w_1, w_2, \dots, w_n είναι τα συναπτικά βάρη. Η παράμετρος θ είναι το κατώφλι ενεργοποίησης. Ο όρος αυτός εξηγείται από το γεγονός ότι η διέγερση v του νευρώνα που προκύπτει από τις εισόδους του δικτύου είναι θετική αν το άθροισμα $\sum_{i=1}^n w_i x_i$ είναι μεγαλύτερο από το θ :

$$v > 0, \quad \text{αν} \quad \sum_{i=1}^n w_i x_i > \theta$$

$$v = 0, \quad \text{αν} \quad \sum_{i=1}^n w_i x_i = \theta$$

$$v < 0, \quad \text{αν} \quad \sum_{i=1}^n w_i x_i < \theta$$

Η μορφή της συνάρτησης μεταφοράς εξαρτάται από την τιμή της διέγερσης v και φυσικά από την επιλογή της συνάρτησης ενεργοποίησης ϕ .

Η συνάρτηση ενεργοποίησης στο μονοστρωματικό Perceptron τροφοδοτείται από τη διέγερση v και δίνει την έξοδο y του νευρώνα. Είναι μη-γραμμική και συνήθως παίρνει την κανονική ή τη διπολική μορφή της βηματικής συνάρτησης ενεργοποίησης.

Ωστόσο, η επιλογή της συνάρτησης ενεργοποίησης είναι δευτερεύουσας σημασίας καθώς η συμπεριφορά του νευρώνα καθορίζεται από το διάνυσμα των συναπτικών βαρών w και το κατώφλι θ [108].

1.5.4.1 Εκπαίδευση Δικτύου Perceptron

Το δίκτυο Perceptron εκπαιδεύεται με επίβλεψη. Μαθαίνει χρησιμοποιώντας έναν επαναληπτικό αλγόριθμο να προσαρμόζει τις παραμέτρους, λαμβάνοντας υπόψη τα επαυξημένα πρότυπα εκπαίδευσης και τους στόχους των προτύπων αυτών.

Ο κανόνας εκπαίδευσης Perceptron λέγεται και κανόνας σταθερής αύξησης. Λειτουργεί με επανάληψη, καθώς τα πρότυπα παρουσιάζονται στο δίκτυο με κυκλική σειρά και όταν τελειώσουν επαναλαμβάνονται από την αρχή. Ένας πλήρης κύκλος χρήσης όλων των προτύπων ονομάζεται εποχή.

Ο κανόνας σταθερής αύξησης, τροποποιεί το επαυξημένο διάνυσμα των συναπτικών βαρών w όταν ο στόχος $d^{(k)}$ για το πρότυπο k διαφέρει από την έξοδο του

δικτύου $y = f(\mathbf{w}(t-1)^T \mathbf{x}^{(k)})$ όπου $\mathbf{w}(t-1)$ είναι το επαυξημένο διάνυσμα των συναπτικών βαρών μετά την επανάληψη $t-1$. Όταν υπάρχει σφάλμα, τότε η διόρθωση των βαρών γίνεται προσθέτοντας ή αφαιρώντας ένα ποσοστό του προτύπου $\mathbf{x}^{(k)}$. Συγκεκριμένα, αν κατά την επανάληψη k , εισάγεται το πρότυπο p τότε ο κανόνας της διόρθωσης είναι: $\mathbf{w}(t) = (\mathbf{w}(t-1) + \beta(d^{(k)} - y)\mathbf{x}^{(k)})$ όπου $\mathbf{w}(t)$ είναι το επαυξημένο διάνυσμα συναπτικών βαρών μετά την επανάληψη t . Και επειδή τα επαυξημένα διανύσματα βαρών και εισόδου ορίζονται ως $\mathbf{w}^T = [w_0, \mathbf{w}^T]$ και $\mathbf{x}^T = [x_0, \mathbf{x}^T]$, τότε ο κανόνας εκπαίδευσης $\mathbf{w}(t) = (\mathbf{w}(t-1) + \beta(d^{(k)} - y)\mathbf{x}^{(k)})$ γράφεται εναλλακτικά ως εξής:

$$\mathbf{w}(t) = (\mathbf{w}(t-1) + \beta(d^{(k)} - y)\mathbf{x}^{(k)})$$

η παράμετρος β ρυθμίζει το μέγεθος της διόρθωσης και καλείται βήμα εκπαίδευσης ή ρυθμός εκπαίδευσης (learning step ή learning rate) και πρέπει να είναι μικρός θετικός αριθμός.

Η εκπαίδευση του \mathbf{w} γίνεται με τέτοιο τρόπο ώστε αν το πρότυπο ταξινομηθεί λανθασμένα, στην επόμενη επανάληψη να ταξινομηθεί σωστά ή να πλησιάζει περισσότερο στο να ταξινομηθεί σωστά.

Θεωρώντας $u_{k, \text{μετά}}^{(k)} = \mathbf{w}(t-1)^T \mathbf{x}^{(k)}$ τη διέγερση του νευρώνα πριν από τη διόρθωση των βαρών κατά την επανάληψη t , τότε μετά τη διόρθωση των βαρών έχουμε:

$$u_{t, \text{μετά}}^{(k)} = \mathbf{w}(t-1)^T \mathbf{x}^{(k)} = \mathbf{w}(t-1)^T \mathbf{x}^{(k)} + \beta(d^{(k)} - y)\mathbf{x}^{(k)T} \mathbf{x}^{(k)} = u_{t, \text{πριν}}^{(k)} + \beta(d^{(k)} - y)\|\mathbf{x}^{(k)}\|^2 \quad (1.7)$$

Υπάρχουν δύο ειδών εσφαλμένες ταξινομήσεις κατά τις οποίες γίνεται εκπαίδευση των βαρών:

1. $d^{(k)} = 1$ και $y = f(u_{t, \text{πριν}}^{(k)}) = 0$, οπότε

$(d^{(k)} - y) = 1 > 0$ και $u_{t, \text{πριν}}^{(k)} \leq 0$. Συνεπώς, $\beta(d^{(k)} - y)\|\mathbf{x}^{(k)}\|^2 > 0$, και σύμφωνα με

την (1.7): $u_{t, \text{μετά}}^{(k)} > u_{t, \text{πριν}}^{(k)}$.

2. $d^{(k)} = 0$ και $y = f(u_{t, \text{πριν}}^{(k)}) = 1$, οπότε

$(d^{(k)} - y) = -1 < 0$ και $u_{t, \text{πριν}}^{(k)} > 0$. Συνεπώς, $\beta(d^{(k)} - y) \|\mathbf{x}^{(k)}\|^2 < 0$, και σύμφωνα με την (1.7): $u_{t, \text{μετά}}^{(k)} < u_{t, \text{πριν}}^{(k)}$.

Και στις δύο περιπτώσεις η διέγερση $u_i^{(k)}$ βελτιώνεται ώστε το y να πλησιάζει το στόχο $d^{(k)}$:

- Αν $d^{(k)} = 1$ τότε η διέγερση $u_i^{(k)}$ αυξάνεται πλησιάζοντας το επιθυμητό $y = 1$, ενώ
- Αν $d^{(k)} = 0$ τότε η διέγερση $u_i^{(k)}$ μειώνεται πλησιάζοντας το επιθυμητό $y = 0$.

Με τη χρήση του αλγόριθμου εκπαίδευσης που περιγράφηκε σε αυτή την ενότητα το μοντέλο Perceptron συγκλίνει σε μια λύση η οποία ταξινομεί σωστά όλα τα πρότυπα αρκεί το πρόβλημα να είναι γραμμικά διαχωρίσιμο [108].

1.5.5 Πολυστρωματικό Perceptron

Η χρήση του μονοστρωματικού Perceptron παρουσιάζει περιορισμένες δυνατότητες καθώς με ένα μόνο νευρώνα μπορούν να αναπαρασταθούν μόνο επίπεδες επιφάνειες. Ο περιορισμός αυτός αντιμετωπίζεται με τη χρήση περισσότερων νευρώνων. Δημιουργείται έτσι ένα δίκτυο το οποίο αποτελείται από δύο στρώματα, το στρώμα εξόδου, το κρυφό στρώμα και το στρώμα εισόδου το οποίο δεν θεωρείται στρώμα νευρώνων καθώς οι κόμβοι της εισόδου δεν λειτουργούν σαν νευρώνες αλλά μεταδίδουν τιμές στο επόμενο στρώμα. Όσο περισσότεροι είναι οι κόμβοι που χρησιμοποιούνται τόσο πιο μεγάλο είναι το σύνολο των περιοχών που μπορούν να διαχωριστούν στην έξοδο. Αυτού του τύπου τα δίκτυα λέγονται Perceptron πολλών στρωμάτων (Multi Layer Perceptron - MLP). Βασικό χαρακτηριστικό τους είναι ότι οι νευρώνες οποιουδήποτε στρώματος l τροφοδοτούν αποκλειστικά τους νευρώνες του επόμενου στρώματος $l+1$ και τροφοδοτούνται αποκλειστικά από τους νευρώνες του προηγούμενου στρώματος $l-1$.

Τα δίκτυα perceptron πολλών στρωμάτων όπου οι νευρώνες χρησιμοποιούν τη βηματική συνάρτηση μπορούν να υλοποιήσουν συναρτήσεις που δεν μπορεί να υλοποιήσει ένα απλό δίκτυο Perceptron. Παρόλα αυτά η χρήση της βηματικής συνάρτησης δεν προτιμάται διότι οι περισσότεροι κανόνες εκπαίδευσης, βασίζονται σε μεθόδους βελτιστοποίησης και χρησιμοποιούν παραγώγους. Η βηματική συνάρτηση δεν

είναι παραγωγίσιμη γι' αυτό χρησιμοποιείται η σιγμοειδής συνάρτηση $\varphi(v) = \frac{1}{1 + e^{-v}}$ για αυτό τον τύπο δικτύου.

Η σιγμοειδής συνάρτηση, είναι παραγωγίσιμη και σχηματικά μοιάζει με τη βηματική. Επίσης στο perceptron πολλών στρωμάτων χρησιμοποιείται και η υπερβολική εφαιπτομένη $\tanh(v) = \frac{e - e^{-v}}{e + e^{-v}}$, η οποία έχει την παρακάτω απλή σχέση με τη σιγμοειδή συνάρτηση $\tanh(v) = 2\varphi(2v) - 1$.

Τα δίκτυα Perceptron πολλών στρωμάτων που ενσωματώνουν τη σιγμοειδή συνάρτηση έχουν πολλές δυνατότητες αναπαράστασης συναρτήσεων. Ουσιαστικά, τα δίκτυα αυτής της μορφής μπορούν να προσεγγίσουν οποιαδήποτε ομαλή συνάρτηση, όσο κοντά είναι επιθυμητό. Εξαιτίας αυτής τους της ιδιότητας, τα δίκτυα αυτά ονομάζονται και καθολικοί προσεγγιστές [108].

1.5.5.1 Εκπαίδευση Πολυστρωματικού Perceptron

Η εκπαίδευση ενός δικτύου πολλών στρωμάτων είναι η διαδικασία ρύθμισης των συναπτικών βαρών του, έτσι ώστε να ικανοποιείται κάποιο κριτήριο καταλληλότητας που είναι και ο στόχος εκπαίδευσης σε οποιοδήποτε νευρωνικό δίκτυο. Χαρακτηριστικό της εκπαίδευσης ενός Πολυστρωματικού (Multilayer) Perceptron (MLP) είναι η ιδιότητα του καθολικού προσεγγιστή, σύμφωνα με την οποία αν υπάρχει το κατάλληλο σε μέγεθος δίκτυο μπορεί να εκπαιδευτεί ώστε να μάθει οποιαδήποτε συνάρτηση με οποιαδήποτε ποιότητα προσέγγισης. Ένας από τους πιο δημοφιλείς αλγόριθμους εκπαίδευσης MLP είναι ο αλγόριθμος οπισθοδιάδοσης (back-propagation).

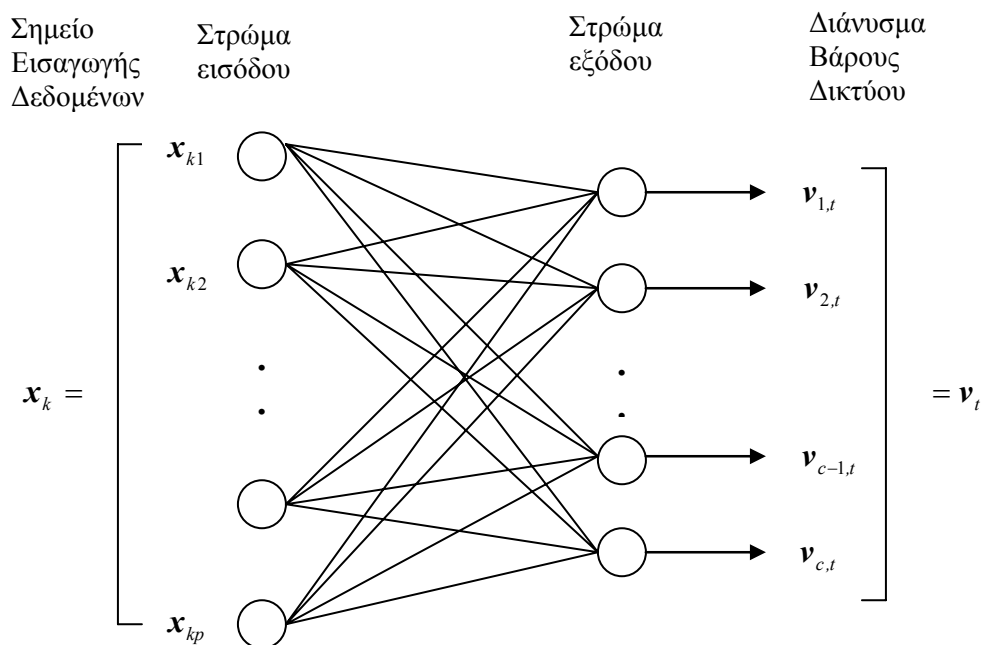
Ο αλγόριθμος οπισθοδιάδοσης (back-propagation) προτάθηκε από τον Paul Werbos τη δεκαετία του 1970. Η ιδιαιτερότητα του αλγορίθμου αυτού συνίσταται στο ότι επιτρέπει σε κάθε στρώμα ενός νευρωνικού δικτύου να ενημερώνει τα προηγούμενα του σε περίπτωση λάθους. Με λίγα λόγια εισάγει την ανάδραση στην διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου. Με αυτό τον τρόπο ένα νευρωνικό δίκτυο δεν εκπαιδεύεται στα “τυφλά” αλλά η εκπαίδευση “κατευθύνεται” προς την περιοχή που παράγει το μικρότερο λάθος.

Βασικό χαρακτηριστικό του αλγόριθμου, είναι η ύπαρξη στόχων και άρα υλοποιεί εκπαίδευση με επίβλεψη. Χρησιμοποιείται στα τεχνητά νευρωνικά δίκτυα χωρίς ανάδραση και απαιτεί η συνάρτηση ενεργοποίησης να είναι διαφορίσιμη. Λειτουργεί σε έξι βήματα:

- Βήμα 1) Δίνεται αρχικά στο νευρωνικό δίκτυο ένα δείγμα από το σύνολο εισόδου για το οποίο η έξοδος είναι γνωστή.
- Βήμα 2) Για κάθε νευρώνα του στρώματος εξόδου, γίνεται σύγκριση της πραγματικής εξόδου με την επιθυμητή.
- Βήμα 3) Υπολογίζεται η επιθυμητή έξοδος για κάθε νευρώνα ξεχωριστά και ένας συντελεστής κλιμάκωσης που επεξεργάζεται την έξοδο του κάθε νευρώνα έτσι ώστε το αποτέλεσμα που θα προκύπτει να είναι η επιθυμητή έξοδος.
- Βήμα 4) Ρυθμίζονται οι συντελεστές βάρους για τη μείωση του τοπικού σφάλματος.
- Βήμα 5) Ανατίθεται στους νευρώνες ένας συντελεστής ευθύνης για το τοπικό σφάλμα. Όσο μεγαλύτερος είναι ο συντελεστής βάρους τόσο μεγαλύτερη είναι και η υπευθυνότητα στους νευρώνες.
- Βήμα 6) Ο αλγόριθμος επιστρέφει στο βήμα 3, έχοντας ως σφάλμα τους συντελεστές ευθύνης της προηγούμενης επανάληψης.

1.5.6 Δίκτυα Kohonen

Η συσταδοποίηση δικτύων του Kohonen [45] είναι μια χωρίς επίβλεψη διαδικασία, η οποία βρίσκει το καλύτερο σύνολο συστάδων με έναν επαναληπτικό τρόπο. Η βασική δομή του δικτύου Kohonen αποτελείται από ένα στρώμα εισόδου και ένα στρώμα εξόδου. Κάθε νευρώνας εξόδου έχει ένα πρότυπο ή διάνυσμα βάρους συνδεδεμένο σε αυτό, και είναι αυτό το διάνυσμα βάρους που προσαρμόζεται κατά τη διάρκεια της εκπαίδευσης. Τα διανύσματα βάρους είναι τα κέντρα των συστάδων. Έχοντας ένα διάνυσμα εισόδου, οι νευρώνες στο επίπεδο εξόδου ανταγωνίζονται μεταξύ τους και ο νικητής, του οποίου το βάρος έχει τη μικρότερη απόσταση από το διάνυσμα εισαγωγής, αναβαθμίζει τα βάρη του και εκείνα από μερικά σύνολα από προκαθορισμένους γείτονες. Η διαδικασία συνεχίζεται μέχρι τα διανύσματα βαρών να σταθεροποιηθούν. Παρακάτω δίνονται τα βήματα λειτουργίας του αλγόριθμου του δικτύου Kohonen.



Σχήμα 1.3. Δομή Δικτύου Kohonen

Αλγόριθμος Δικτύου Kohonen

Βήμα 1). Ορίζεται ο αριθμός των βαρών c και $\varepsilon > 0$ κάποια μικρή σταθερά.

Βήμα 2) Γίνεται τυχαία επιλογή των αρχικών κέντρων $\mathbf{v}_0 = (v_{1,0}, v_{2,0}, \dots, v_{c,0}) \in R^{cp}$ και του ρυθμού μεταβολής $\alpha_{ik,0} \in (1, 0)$.

Βήμα 3). Για $t=1, 2, \dots, t_{\max}$, για $k=1, 2, \dots, n$:

A. Υπολογισμός της Ευκλείδειας απόστασης $d_{ik}^2 = \|\mathbf{x}_k - \mathbf{v}_{i,t}\|^2$, για $i=1, \dots, c$.

B. Ταξινόμηση της $\{d_{ik}^2\}$ με σειρά από το μικρότερο στο μεγαλύτερο:

$$d_{1k}^2 \leq d_{2k}^2 \leq \dots \leq d_{ck}^2.$$

C. Ενημέρωση του νικητή: $\mathbf{v}_{1,t} = \mathbf{v}_{1,t-1} + a_{1k,t}(\mathbf{x}_k - \mathbf{v}_{1,t-1})$.

D. Αν $|N_t| = n_t$ ενημερώνονται οι $(n_t - 1)$ δεσμοί οι οποίοι είναι κοντινότερα

στο \mathbf{x}_k στο B: $\mathbf{v}_{i,t} = \mathbf{v}_{i,t-1} + a_{ik,t}(\mathbf{x}_k - \mathbf{v}_{i,t-1})$, $i = 2, 3, \dots, (n_t - 1)$.

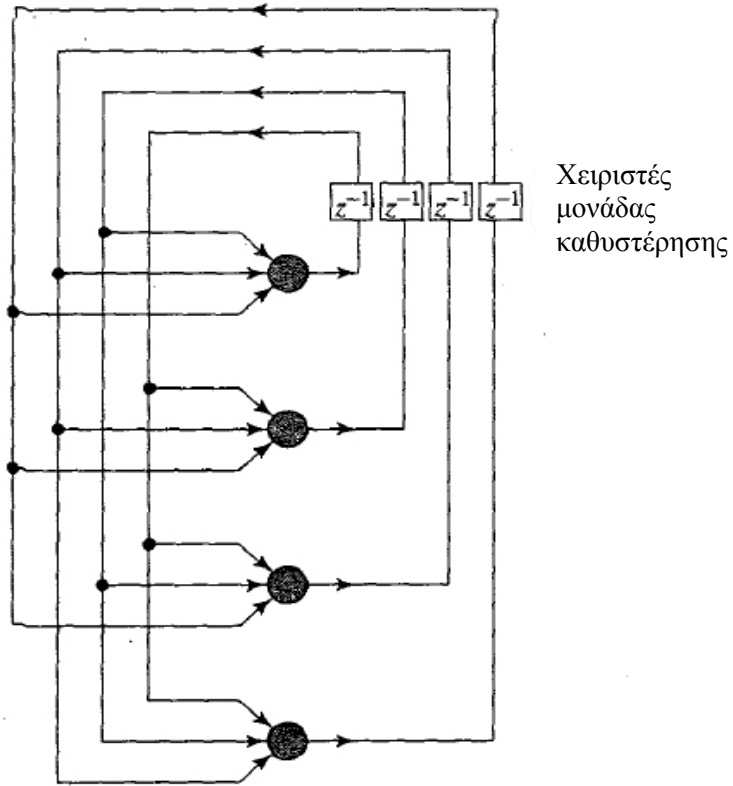
E. Επόμενο k .

Βήμα 4). Υπολογισμός του $E_t = \|\mathbf{v}_t - \mathbf{v}_{t-1}\|^2 = \sum_i \|\mathbf{v}_{i,t} - \mathbf{v}_{i,t-1}\|^2$.

Βήμα 5) Αν $E_t \leq \varepsilon$ τότε ο αλγόριθμος σταματάει, διαφορετικά οι ρυθμοί μεταβολής $\{a_{ik,t}\}$ αλλάζουν και ενημερώνεται η γειτονιά N_t . Επόμενο t .

Τα χαρακτηριστικά του αλγόριθμου Kohonen είναι τα εξής:

- Ο αλγόριθμος Kohonen είναι σειριακός αλγόριθμος – όταν ένα δεδομένο εισόδου x_k τροφοδοτείται στο δίκτυο, κάποια βάρη ενημερώνονται αμέσως.
- Η παράμετρος a έχει συνταχθεί στο i, k και t . Αυτό γίνεται εξαιτίας του ότι οι τιμές του a που χρησιμοποιήθηκαν στην ενημέρωση των βαρών διανέμονται στο N_t , κεντραρισμένες στο νικητή, το $v_{i,t}$, και η σύνταξη του v_i είναι μια συνάρτηση του x_k . Επιπλέον, το a συνήθως μειώνεται με το χρόνο (t) για να εξαναγκάσει το $E_t \leq \varepsilon$ (καθαρά $E_t \rightarrow 0$ όπως $a_{ik,t} \rightarrow 0$). Αυτό είναι μια τεχνητή στρατηγική τερματισμού, στην οποία ο κανόνας ενημέρωσης εγγυάται την εξάρτηση των διαδοχικών επαναλήψεων μετά από αρκετό χρόνο, ακόμη κι αν αυτοί δεν είναι κοντά σε ένα διάλυμα λύσης.
- Επιλέγοντας το μέγεθος της γειτονιάς, η αναλογία εκπαίδευσης και οι στρατηγικές για να λειτουργήσουν αυτοί οι παράμετροι είναι κρίσιμες για τον τερματισμό της συσταδοποίησης του δικτύου Kohonen. Επιπλέον διαφορετικά σύνολα παραμέτρων αποφέρουν διαφορετικά αποτελέσματα. Ο Kohonen έχει δείξει ότι αυτή η διαδικασία συγκλίνει, υπό την έννοια ότι το $\{v_i\} \rightarrow \{v_*\}$ για $\{a\} \rightarrow 0$, όπου το v_* , είναι απλώς ένα σημείο ορίου της επανάληψης αλληλουχίας (και όχι βέλτιστο σημείο).
- Η συσταδοποίηση του δικτύου Kohonen δεν χρησιμοποιεί ένα συγκεκριμένο διαμερισμό των δεδομένων κατά τη διάρκεια της εκπαίδευσης.



Σχήμα 1.4. Δομή Επαναλαμβανόμενου δικτύου

1.5.7 Νευρωνικά Δίκτυα Συναρτήσεων Ακτινικής Βάσης (RBFNNs)

Αυτού του τύπου τα νευρωνικά δίκτυα χρησιμοποιούνται συνήθως σε προσεγγίσεις συναρτήσεων (functions approximation) αλλά και σε πρόβλεψη χρονοσειρών (time-series prediction), σε μοντελοποίηση δεδομένων (data modeling), σε αναγνώριση προτύπων (pattern recognition) και κατηγοριοποίηση (classification).

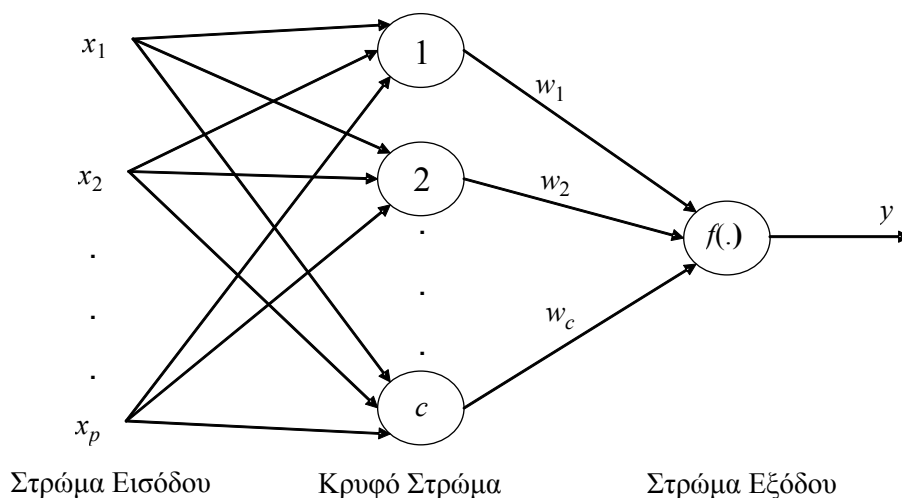
Τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (RBFNNs) είναι νευρωνικά δίκτυα τριών στρωμάτων. Η αιτία δεν είναι ότι απαγορεύεται η χρήση περισσότερων επιπέδων αλλά ότι δεν υπάρχει ικανοποιητικό αλγόριθμος εκπαίδευσης για περισσότερα από τρία επίπεδα. Τα τρία στρώματα αρκούν για να προσεγγιστεί οποιαδήποτε συνεχής συνάρτηση.

- (1) Το **στρώμα εισόδου**. Αποτελείται από νευρώνες εισόδου και η μοναδική τους λειτουργία είναι η διασύνδεση του δικτύου με το περιβάλλον του. Ο αριθμός των νευρώνων εισόδου είναι ίσος με τον αριθμό της διάστασης

του διανύσματος εισόδου.

- (2) Το **κρυφό στρώμα**, με μη γραμμική συνάρτηση ενεργοποίησης που ανήκει στην οικογένεια των συναρτήσεων ακτινικής βάσης. Αποστολή του είναι να εφαρμόζει ένα μη γραμμικό μετασχηματισμό στα δεδομένα της εισόδου και να προωθεί τα αποτελέσματα στο επόμενο επίπεδο. Ο αριθμός των νευρώνων του κρυφού επιπέδου είναι από τα χαρακτηριστικά του νευρωνικού δικτύου για τα οποία θα πρέπει να αποφασίσει ο σχεδιαστής του.
- (3) Το **στρώμα εξόδου** το οποίο είναι γραμμικό και δίνει στο περιβάλλον την έξοδο του νευρωνικού δικτύου ως αποτέλεσμα της διέγερσης του από την είσοδο.

Στο παρακάτω σχήμα παρουσιάζεται ένα τυπικό νευρωνικό δίκτυο συναρτήσεων ακτινικής βάσης (RBF).



Σχήμα 1.5. Βασική τοπολογία ενός νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης

Η βασική τοπολογία του δικτύου RBF περιλαμβάνει στη σειρά ένα κρυφό στρώμα και μια γραμμική μονάδα επεξεργασίας που διαμορφώνουν το στρώμα εξόδου. Το σχήμα 1.5 απεικονίζει αυτήν την τοπολογία για ένα, πολλών εισόδων - μίας εξόδου

(multi-input single-output) δίκτυο, όπου το c αντιπροσωπεύει τον αριθμό κόμβων στο κρυφό στρώμα. Κάθε κρυφός κόμβος αντιστοιχεί σε μια ακτινική συνάρτηση βάσης, ενώ το στρώμα εξόδου υπολογίζει το σταθμισμένο ποσό των αποτελεσμάτων των κόμβων. Μια ακτινική συνάρτηση βάσης αντιπροσωπεύει μια τοπική επίδραση, η σειρά της οποίας καθορίζεται από το κέντρο και το πλάτος (width). Εν τω παρόντι, η ακτινική συνάρτηση βάσης θα αναφερθεί επίσης ως συνάρτηση πυρήνα ή απλά πυρήνας.

Υιοθετώντας την ονοματολογία του σχήματος 1.5, το σύνολο ζευγαριών δεδομένων εισόδου-εξόδου συμβολίζεται ως

$$S = \left\{ (\mathbf{x}_k, y_k) \in R^p \times R \mid f(\mathbf{x}_k) = y_k, 1 \leq k \leq n \right\},$$

όπου το n είναι ο αριθμός δειγμάτων εκπαίδευσης, $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kp}]^T$ είναι το k -οστό διάνυσμα εισόδου και y_k το k -οστό διάνυσμα εξόδου. Επιλέχθηκαν οι γκαουσιανές συναρτήσεις πυρήνων της μορφής,

$$g_i(\mathbf{x}_k) = e^{-\frac{\|\mathbf{x}_k - \mathbf{v}_i\|^2}{\sigma_i^2}} \quad (1.8)$$

όπου $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots, \mathbf{v}_c$ προκύπτουν υπό την μορφή p -διάστατων διανυσμάτων και αναφέρονται ως κέντρα πυρήνων, και $\sigma_1, \sigma_2, \dots, \sigma_i, \dots, \sigma_c$ είναι τα αντίστοιχα πλάτη (widths) των πυρήνων. Η κατ' εκτίμηση έξοδος του δικτύου χρησιμοποιεί την ακόλουθη γραμμική συνάρτηση οπισθοδιάδοσης,

$$f(\mathbf{x}_k) = \tilde{y}_k = \sum_{i=1}^c w_i g_i(\mathbf{x}_k) \quad (1.9)$$

όπου w_i είναι το βάρος σύνδεσης του i -οστού κρυφού κόμβου. Η διαδικασία εκπαίδευσης των δικτύων RBF στοχεύει να λάβει ένα σύνολο τιμών για τις παραμέτρους πυρήνων \mathbf{v}_i , σ_i και τα συναπτικά βάρη w_i . Συνήθως, τα συναπτικά βάρη υπολογίζονται μέσω μιας ανάλυσης βασισμένης σε οπισθοδιάδοση, όπως η μέθοδος των ελαχίστων τετραγώνων, η αποδόμηση μοναδικής αξίας (singular value decomposition) ή οποιαδήποτε παραλλαγή τους.

Αν και πάρα πολλή προσπάθεια έχει τεθεί στον υπολογισμό των κατάλληλων τιμών για τα κέντρα των πυρήνων [1,35,46,51,69,83,84,85,88], υπάρχουν σχετικά λίγες μέθοδοι που αντιμετωπίζουν το ζήτημα του υπολογισμού των πλατών (widths). Σύμφωνα με την αναφορά [14], υπάρχουν μερικές συγκεκριμένες ιδιότητες των πλατών (widths) που έχουν επιπτώσεις στη γενική απόδοση του δικτύου. Παραδείγματος χάριν, εάν τα πλάτη (widths) πάρουν μικρές τιμές τότε το δίκτυο παράγει μια όχι και τόσο καλή αντιπροσώπευση της υποκείμενης δομής των δεδομένων και υπό αυτήν τη έννοια, η απόδοση του δικτύου που προκύπτει είναι μικρή. Αντίθετα, μεγάλες τιμές των πλατών (widths) οδηγούν σε μία απόδοση δικτύου που παράγει επίσης μια όχι και τόσο καλή αντιπροσώπευση της πραγματικής δομής των δεδομένων. Μια άλλη προσέγγιση δηλώνει ότι όλα τα πλάτη (widths) των πυρήνων πρέπει να είναι ίσα μεταξύ τους [14,32]. Αυτή η διαδικασία επιλογής των πλατών (widths) είναι άριστη μόνο στις περιπτώσεις όπου τα δεδομένα διανέμονται ομοιόμορφα [6,86]. Εντούτοις, αυτό είναι μια ιδανική κατάσταση που δεν αντιστοιχεί στα πραγματικά προβλήματα.

Μια άλλη στρατηγική αναγνωρίζει ότι τα πλάτη μπορούν να είναι διαφορετικά για τις διαφορετικές συναρτήσεις βάσης. Παραδείγματος χάριν, οι Moody και Darken [60] υπολογίζουν κάθε πλάτος της συνάρτησης βάσης χρησιμοποιώντας τη μέση απόσταση του αντίστοιχου κέντρου συστάδων στους κοντινότερους τ γείτονές του,

$$\sigma_i = \frac{1}{\tau} \sqrt{\sum_{j=1}^{\tau} d_{ij}^2} \quad (1.10)$$

όπου $d_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|$ με $i \neq j$, και οι χαρακτηριστικές τιμές τ είναι $\tau = 2$ και $\tau = 3$.

Μια ειδική περίπτωση εξ. (1.10) εισήχθη από Saha και Keeler [82], όπου το πλάτος (width) κάθε κόμβου υπολογίστηκε από την απόσταση μεταξύ του κέντρου του πυρήνα και του κοντινότερου γείτονά του, πολλαπλασιασμένος με έναν θετικό παράγοντα.

1.5.7.1 Εκπαίδευση (training) Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

Το βήμα που ακολουθεί την σχεδίαση ενός νευρωνικού δικτύου, είναι αυτό της “Εκπαίδευσης” του. Ως εκπαίδευση (training) ενός νευρωνικού δικτύου ορίζεται η διαδικασία διαμόρφωσης των παραμέτρων του σχεδιασθέντος νευρωνικού δικτύου ώστε

η εφαρμογή ενός συνόλου προτύπων δεδομένων εισόδου να παράγει το επιθυμητό σύνολο δεδομένων εξόδου. Ειδικά στην περίπτωση ενός νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBF) (που θα μας απασχολήσει σε αυτή την διατριβή) οι παράμετροι του νευρωνικού δικτύου οι οποίες θα πρέπει να καθοριστούν είναι τα κέντρα των συναρτήσεων ακτινικής βάσης, τα πλάτη τους (widths) και τα συναπτικά βάρη (synaptic weights) των νευρώνων.

Όπως αναφέρθηκε και παραπάνω η διαδικασία της εκπαίδευσης ενός νευρωνικού δικτύου έχει ως σκοπό την βελτιστοποίηση των τιμών των παραπάνω παραμέτρων με σκοπό την δυνατόν καλύτερη απόκριση του νευρωνικού δικτύου. Για να επιτευχθεί η βελτιστοποίηση αυτή συνήθως χρησιμοποιούνται επαναληπτικές μέθοδοι. Αυτές οι μέθοδοι μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες:

1. Επιβλεπόμενη Μάθηση (Supervised Learning)
2. Μη-επιβλεπόμενη Μάθηση (Unsupervised Learning)

1.5.7.2 Επιβλεπόμενη Μάθηση (Supervised Learning)

Σε αυτού του τύπου την εκπαίδευση ενός νευρωνικού δικτύου υπάρχει ένα σετ δεδομένων (τα δεδομένα εκπαίδευσης – training data set), τα οποία αποτελούνται από δεδομένα εισόδου (για την ακρίβεια διανύσματα εισόδου που μπορούν να είναι οποιασδήποτε διάστασης) με αντίστοιχα δεδομένα (διανύσματα) επιθυμητής εξόδου. Αυτός που επιβλέπει την εκπαίδευση (συνήθως ένας άνθρωπος) τροφοδοτεί το δίκτυο με τα δεδομένα εισόδου και διαμορφώνει τις παραμέτρους του δικτύου ώστε αυτό να παράγει τις επιθυμητές εξόδους.

Η έξοδος του νευρωνικού δικτύου μπορεί να είναι είτε μια συνεχής τιμή μιας συνάρτησης (που αποκαλείται regression) είτε η γενικότερη κατηγορία στην οποία ανήκει το διάνυσμα εισόδου (που αποκαλείται classification).

1.5.7.3 Μη – Επιβλεπόμενη Μάθηση (Unsupervised Learning)

Η Μη – Επιβλεπόμενη Μάθηση [105] είναι μια διαδικασία στην οποία το ζητούμενο είναι να βρεθεί η οργάνωση που συνήθως είναι κρυμμένη των δεδομένων (διανυσμάτων) εισόδου. Την πιο γνωστή εφαρμογή της Μη – Επιβλεπόμενης Μάθησης αποτελεί η

διαδικασία της συσταδοποίησης (clustering) των δεδομένων εισόδου ενός συστήματος.

Σε αντίθεση με την Επιβλεπόμενη Μάθηση, στην Μη – Επιβλεπόμενη δεν υπάρχουν εκ των προτέρων καθορισμένα σύνολα κατηγοριών για την ταξινόμηση των διανυσμάτων της εισόδου αλλά το δίκτυο θα πρέπει να ανακαλύψει τις κατηγορίες αυτές και να ομαδοποιήσει τα ερεθίσματα του (διανύσματα εισόδου) σε αυτές.

Αλγόριθμοι Συσταδοποίησης Δεδομένων για την Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

2.1 Εισαγωγή

Για τον υπολογισμό των κέντρων (centers) των συναρτήσεων ακτινικής βάσης (radial basis functions) χρησιμοποιούνται συνήθως αλγόριθμοι τόσο διακριτής (crisp) όσο και ασαφής (fuzzy) συσταδοποίησης. Παρακάτω περιγράφονται αναλυτικά οι αλγόριθμοι συσταδοποίησης που χρησιμοποιήθηκαν στην παρούσα διδακτορική διατριβή για τον υπολογισμό των παραμέτρων των πυρήνων κατά την εκπαίδευση του νευρωνικού δικτύου.

2.2 Ο Αλγόριθμος των c -Μέσων (ή αλγόριθμος Lloyd)

Ο αλγόριθμος c -Means ονομάζεται και αλγόριθμος Lloyd και ανήκει σε μια μεγάλη κατηγορία αλγορίθμων συσταδοποίησης που είναι γνωστοί ως αλγόριθμοι διαμέρισης (partitioning algorithms). Ουσιαστικά οι αλγόριθμοι αυτοί είναι έτσι φτιαγμένοι, ώστε να διαμερίζουν (διατμηματίζουν) τον χώρο των δεδομένων σε περιοχές και να αντιστοιχεί μια περιοχή σε κάθε συστάδα (cluster). Η μέθοδος θεωρεί πως ο αριθμός των συστάδων που θα προκύψουν είναι γνωστός εκ των προτέρων. Αυτό αποτελεί έναν περιορισμό της μεθόδου, καθώς ο αλγόριθμος θα πρέπει να τρέξει με διαφορετικές επιλογές ως προς το πλήθος των συστάδων ή θα πρέπει με κάποιο τρόπο να έχει αποφασιστεί ο αριθμός των συστάδων. Η μέθοδος δουλεύει επαναληπτικά. Χρησιμοποιεί την έννοια του κέντρου της συστάδας (cluster center) και στη συνέχεια κατατάσσει τις παρατηρήσεις (δηλαδή τα διανυσματικά δεδομένα) ανάλογα με την απόστασή τους από τα κέντρα όλων των

συστάδων. Το κέντρο της συστάδας δεν είναι τίποτα άλλο από τη μέση τιμή (κέντρο βάρους) της αντίστοιχης συστάδας. Στη συνέχεια για κάθε παρατήρηση (διανυσματικό δεδομένο) υπολογίζεται η Ευκλείδεια απόστασή της από τα κέντρα των συστάδων που υπάρχουν και γίνεται η κατάταξη κάθε παρατήρησης στη συστάδα που είναι πιο κοντά (για την ακρίβεια στην συστάδα με κέντρο πιο κοντά στην παρατήρηση). Αφού καταταχθούν όλα τα δεδομένα, τότε υπολογίζονται εκ νέου τα κέντρα, απλώς ως τα διανύσματα των μέσων για τα δεδομένα που ανήκουν στην κάθε συστάδα. Η διαδικασία επαναλαμβάνεται μέχρις ότου δεν υπάρχουν διαφορές ανάμεσα σε δύο διαδοχικές επαναλήψεις. Συνήθως η απόσταση που χρησιμοποιείται για να κατατάξει τα δεδομένα είναι η Ευκλείδεια απόσταση. Για τη χρήση μιας άλλης απόστασης, θα πρέπει να πραγματοποιηθούν ειδικοί μετασχηματισμοί στα δεδομένα πριν τη χρήση της. Όπως αναφέρεται προηγουμένως, ο αλγόριθμος αυτός δουλεύει ικανοποιητικά για μεγάλα σύνολα δεδομένων. Αυτός είναι και ο λόγος που η μέθοδος μερικές φορές καλείται και γρήγορη συσταδοποίηση (quick clustering).

Η αντικειμενική συνάρτηση κόστους που ελαχιστοποιεί ο αλγόριθμος c -Means δίνεται στη συνέχεια:

$$J_{c-Means} = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (2.1)$$

Όπου n είναι ο αριθμός των διανυσματικών δεδομένων, c ο αριθμός των κέντρων των συστάδων, \mathbf{x}_k το k -οστό διανυσματικό δεδομένο, \mathbf{v}_i το κέντρο της i -οστής συστάδας και μ_{ik} είναι η συνάρτηση συμμετοχής του \mathbf{x}_k δεδομένου στην i -οστή συστάδα και δίνεται από την παρακάτω σχέση, που είναι γνωστή ως ο νόμος του κοντινότερου γείτονα,

$$\mu_{ik} = \begin{cases} 1, & \text{αν } \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \min_j \{\|\mathbf{x}_k - \mathbf{v}_j\|^2\} \\ 0, & \text{αλλιώς} \end{cases} \quad (2.2)$$

Η ελαχιστοποίηση της $J_{c-Means}$ επιτυγχάνεται αν υπολογίσουμε τα κέντρα με βάση την παρακάτω εξίσωση,

$$\mathbf{v}_i = \frac{\sum_{k=1}^n \mu_{ik} \mathbf{x}_k}{\sum_{k=1}^n \mu_{ik}} \quad (2.3)$$

Η τελευταία εξίσωση υποδηλώνει ότι ένα κέντρο συστάδας είναι το κέντρο βάρους (μέσος όρος) των διανυσματικών δεδομένων που έχουν βαθμό συμμετοχής ίσον με τη μονάδα ως προς αυτή τη συστάδα. Η επαναληπτική διαδικασία, η οποία ελαχιστοποιεί τη συνάρτηση της εξίσωσης (2.1) δίνεται στη συνέχεια.

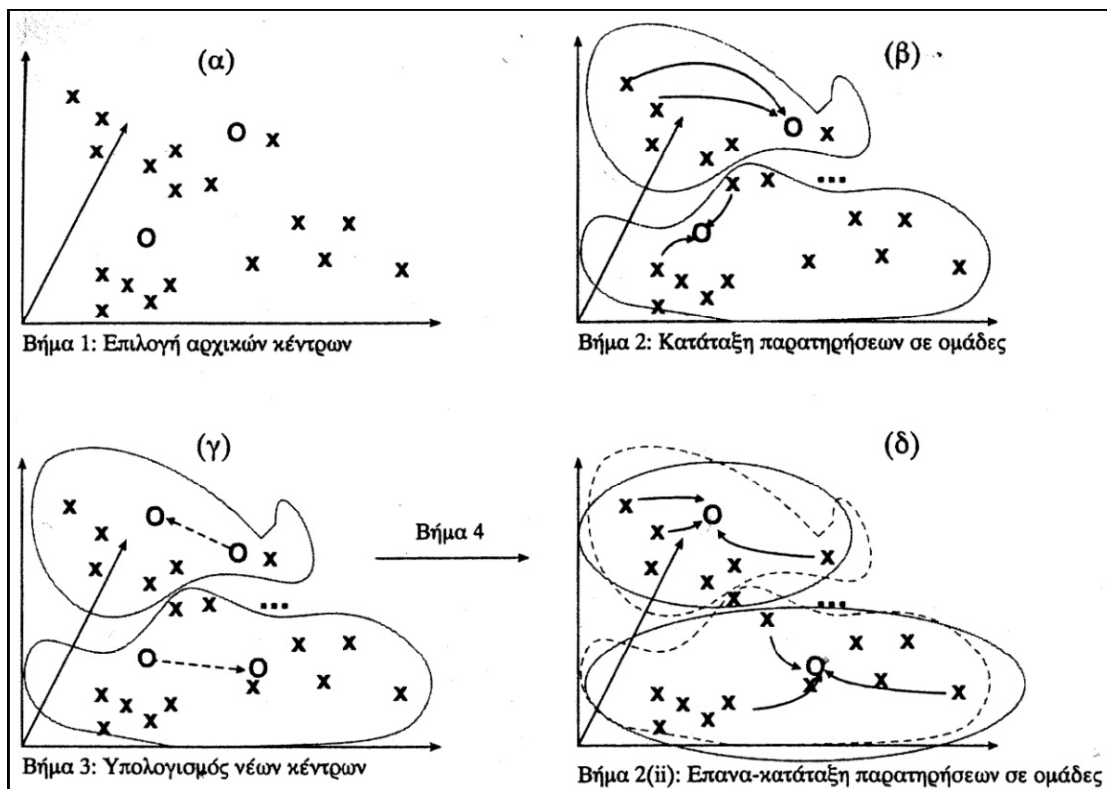
Αλγόριθμος *c*-Means

Βήμα 1) Τυχαία επιλογή των αρχικών τιμών των κέντρων των συστάδων v_i ($1 \leq i \leq c$).

Βήμα 2) Προσδιορισμός των συναρτήσεων συμμετοχής με βάση την εξίσωση (2.2) και κατάταξη των διανυσματικών δεδομένων στις συστάδες.

Βήμα 3) Υπολογισμός των καινούργιων κέντρων των συστάδων (διανύσματα κωδικοποίησης) με βάση την εξίσωση (2.3).

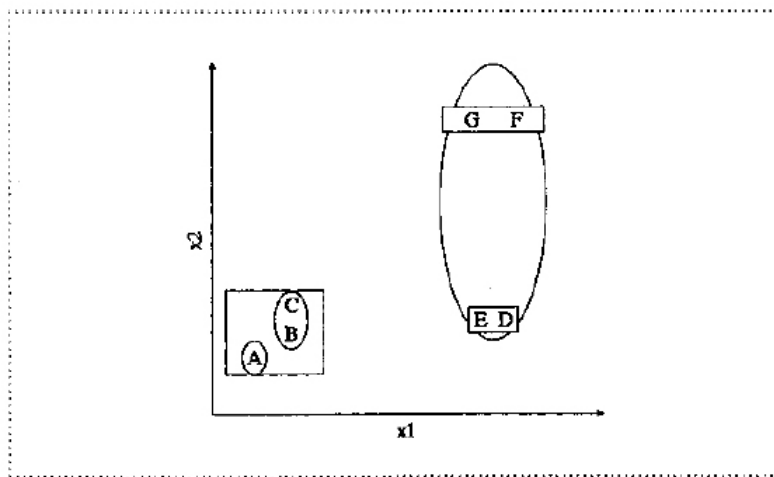
Βήμα 4) Προσδιορισμός των συναρτήσεων συμμετοχής με βάση την εξίσωση (2.2) και ανακατάταξη των διανυσματικών δεδομένων στις συστάδες.



Σχήμα 2.1. Επαναληπτική διαδικασία του αλγόριθμου *c*-Μέσων (*c*-Means).

Το σχήμα 2.1 δείχνει πώς λειτουργεί ο αλγόριθμος των c -Μέσων (c -Means), όπου τα δεδομένα συμβολίζονται με « x » και τα κέντρα με « o ». Για κάθε δεδομένο μετρήθηκε η απόσταση από κάθε κέντρο και κατατάχθηκε στην συστάδα με το πλησιέστερο κέντρο. Σχηματίστηκαν δηλαδή τα δύο νέφη του σχήματος 2.1 (β). Στη συνέχεια, από όλες τις παρατηρήσεις που έχουν καταταχθεί υπολογίστηκαν τα νέα κέντρα των συστάδων. Έτσι, στο σχήμα 2.1 (γ) φαίνεται πως μετατοπίζονται τα κέντρα. Με βάση αυτά τα νέα κέντρα η διαδικασία ξεκινάει από την αρχή κατατάσσοντας πάλι παρατηρήσεις κ.λπ. Τα αρχικά κέντρα μπορούν είτε να οριστούν από το χρήστη είτε υπολογίζονται με κάποιο συγκεκριμένο αλγόριθμο.

Ο παραπάνω αλγόριθμος είναι ιδιαίτερα γρήγορος και στην πράξη σταματάει συνήθως μετά από σχετικά λίγες επαναλήψεις. Αυτό τον κάνει ιδιαίτερα χρήσιμο για τις περιπτώσεις μεγάλων συνόλων δεδομένων. Επίσης, δε χρειάζεται να κρατά στη μνήμη πολλά στοιχεία και επομένως δε χρειάζεται ιδιαίτερα μεγάλη υπολογιστική ισχύ.



Σχήμα 2.2. Ευαισθησία του αλγορίθμου c -Means στην επιλογή αρχικών κέντρων

Το μεγάλο μειονέκτημα του αλγορίθμου είναι ότι εξαρτάται από τις αρχικές τιμές οι οποίες, αν δεν βρεθούν με καλό τρόπο, μπορεί να οδηγήσουν σε κακή συσταδοποίηση. Για να μη συμβεί κάτι τέτοιο, μια λύση είναι να τρέξει ο αλγόριθμος με διάφορες αρχικές τιμές, ώστε να εξασφαλιστεί ότι δεν παγιδεύτηκε σε κάποια μη βέλτιστη λύση. Το παράδειγμα στο σχήμα 2.2 αναφέρεται στο πρόβλημα αυτό. Με βάση το σχήμα αυτό, έστω ότι υπάρχουν 7 παρατηρήσεις και θέλουμε να τις κατατάξουμε σε 3 συστάδες. Αν

χρησιμοποιηθούν ως αρχικές τιμές τα σημεία A, B, C, τότε οι συστάδες που προκύπτουν συμβολίζονται με τις ελλείψεις. Οι συστάδες είναι {A}, {B,C}, {E,D,G,F}. Αν όμως γίνει έναρξη από τα σημεία A, D και F, τότε οι συστάδες που φτιάχνουμε συμβολίζονται με τα τετράγωνα και είναι {A,B,C}, {E,D}, {G,F}). Στην πρώτη περίπτωση οι διαφορές μέσα στις συστάδες είναι πολύ μεγαλύτερες απ' ό,τι στη δεύτερη.

Ένα άλλο πρόβλημα έχει να κάνει με την επιλογή του αριθμού των συστάδων. Μια τακτική θα μπορούσε να είναι η συσταδοποίηση με διαφορετικό κάθε φορά αριθμό συστάδων και στο τέλος την επιλογή της συστάδας που είναι κατά κάποιον τρόπο βέλτιστη. Με βάση, λοιπόν, τα παραπάνω μερικές χρήσιμες στρατηγικές είναι οι ακόλουθες:

- Η επιλογή των αρχικών κέντρων πρέπει να γίνεται έτσι ώστε αυτά να είναι όσο γίνεται πιο μακριά μεταξύ τους.
- Ένας τρόπος για να αποφευχθεί η υλοποίηση μεγάλου αριθμού διαφορετικών συσταδοποιήσεων είναι να μελετηθεί η ήδη διαθέσιμη, ώστε να ενωθούν ή να διαλυθούν οι συστάδες που θα μπορούσαν να βελτιώσουν τη μέση απόσταση των δεδομένων από το κέντρο της συστάδας που ανήκουν. Για παράδειγμα, αν υπάρχει μια συστάδα με μερικά δεδομένα πιο απομακρυσμένα, θα ήταν μια καλή ιδέα να ξανατρέξει ο αλγόριθμος, αυξάνοντας κατά ένα τον αριθμό των συστάδων.

Είναι μάλλον σπάνιο να επιτευχθεί η βέλτιστη λύση με μια μόνο επιλογή αριθμού συστάδων. Συνεπώς, θα πρέπει να δοκιμαστούν διάφορες επιλογές και να χρησιμοποιηθεί και η διαίσθηση του προγραμματιστή, ώστε να υπάρξει η καλύτερη συσταδοποίηση.

Ακόμα, παρατηρείται πως η δυναμική του αλγορίθμου *c*-Μέσων (*c*-Means) είναι ότι με τις πρώτες λίγες επαναλήψεις πλησιάζει πολύ κοντά στην τελική λύση και στις υπόλοιπες επαναλήψεις, οι όποιες διαφορές οφείλονται σε μετακίνηση κάποιων λίγων κέντρων. Επομένως, δεν είναι απαραίτητος ένας μεγάλος αριθμός επαναλήψεων, καθώς η βασική δομή θα σχηματιστεί πολύ γρήγορα.

2.3 Αλγόριθμος Ασαφών c -Μέσων (Fuzzy c -Means)

Πριν γίνει αναφορά στον συγκεκριμένο αλγόριθμο, κρίνεται απαραίτητο να αναλυθούν οι βασικές έννοιες της ασαφούς λογικής.

2.3.1 Βασικές Έννοιες Ασαφούς Λογικής

Η θεωρία της ασαφούς λογικής αναπτύχθηκε από την ανάγκη να εκφραστούν πραγματικά γεγονότα με πιο ρεαλιστικό τρόπο, χωρίς να λαμβάνονται υπόψη διακριτοί περιορισμοί. Το βασικό δομικό στοιχείο της ασαφούς λογικής είναι το ασαφές σύνολο. Η κύρια διαφορά μεταξύ του κλασσικού (crisp) και του ασαφούς (fuzzy) συνόλου είναι ότι η συνάρτηση συμμετοχής του πρώτου παίρνει τις τιμές 0 ή 1, ενώ του δεύτερου παίρνει τιμές που ανήκουν στο κλειστό διάστημα $[0, 1]$. Επομένως, η ασαφής λογική είναι μια επέκταση της κλασσικής λογικής, η οποία περιέχει δύο λογικές τιμές (αληθής και ψευδής).

Η θεμελιώδης έννοια της ασαφούς λογικής (fuzzy logic) είναι το ασαφές σύνολο [44,101] (fuzzy set). Η κύρια δομική διαφορά μεταξύ του ασαφούς (fuzzy) συνόλου και του διακριτού (crisp) συνόλου είναι ο τρόπος συμμετοχής των στοιχείων τους σε αυτά. Οι ορισμοί που ακολουθούν προσδιορίζουν τις δομές των δύο συνόλων.

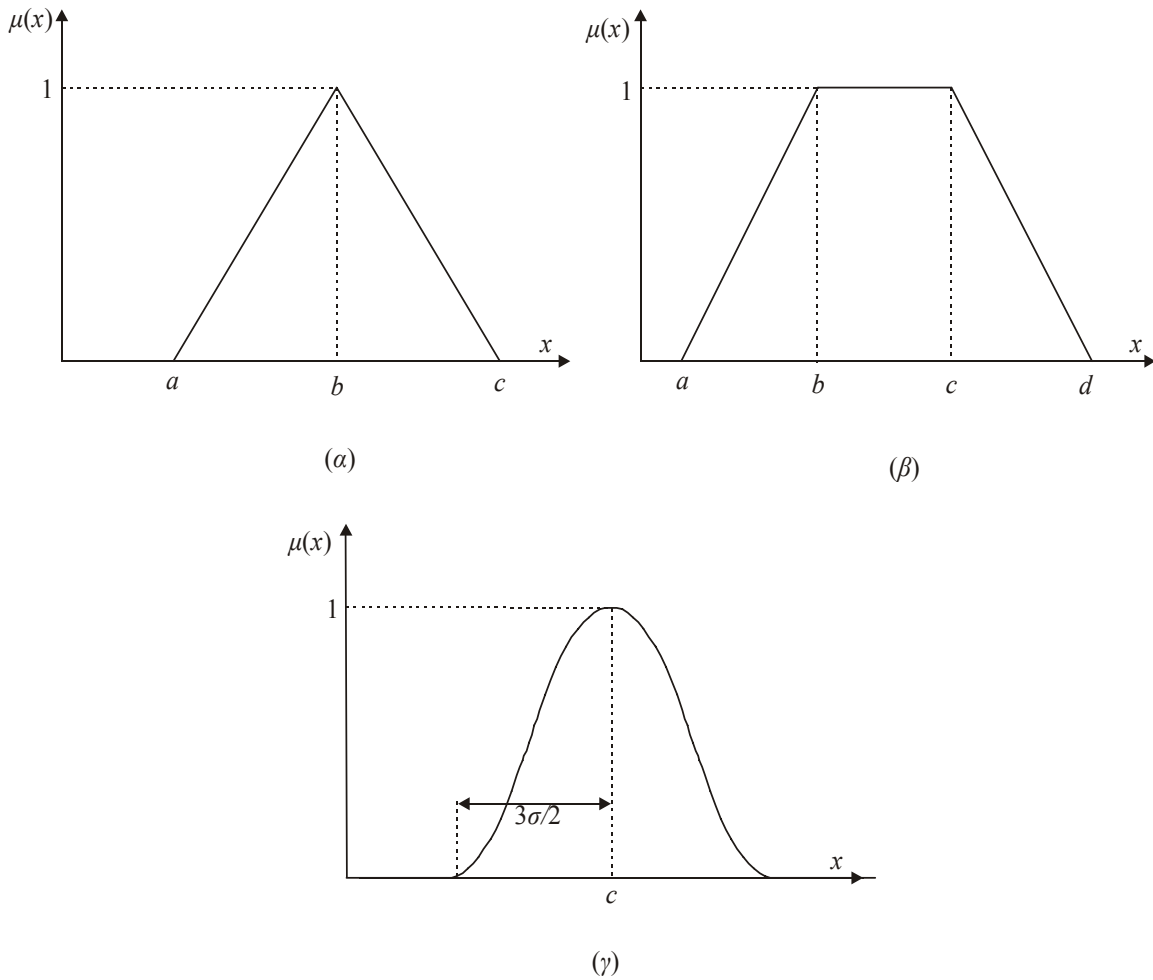
Ορισμός 1: Έστω U το πεδίο ορισμού της μεταβλητής x . Ένα διακριτό σύνολο A ορισμένο στο U χαρακτηρίζεται από την παρακάτω συνάρτηση συμμετοχής των στοιχείων του,

$$\mu_{ik} = \begin{cases} 1, & \text{αν } x \in A \\ 0, & \text{αλλιώς} \end{cases} \quad \text{με } \mu_A : U \rightarrow \{0,1\} \quad (2.4)$$

Ορισμός 2: Έστω U το πεδίο ορισμού της μεταβλητής x . Ένα ασαφές σύνολο A ορισμένο στο U χαρακτηρίζεται από την παρακάτω συνάρτηση συμμετοχής των στοιχείων του,

$$\mu_{ik} = \begin{cases} f(x) \in [0,1], & \text{αν } x \in A \\ 0, & \text{αλλιώς} \end{cases} \quad \text{με } \mu_A : U \rightarrow [0,1] \quad (2.5)$$

Οι πιο διαδεδομένες συναρτήσεις συμμετοχής είναι η τριγωνική, η τραπεζοειδής, και η γκαουσιανή οι οποίες αναλύονται στη συνέχεια:



Σχήμα 2.3. Ασαφή σύνολα με: (α) τριγωνική, (β) τραπεζοειδή και (γ) γκαουσιανή συνάρτηση συμμετοχής

- *Τριγωνική Συνάρτηση Συμμετοχής*

Η τριγωνική συνάρτηση συμμετοχής δίνεται από την παρακάτω σχέση,

$$\mu(x) = \begin{cases} 0 & , \text{αν } x < a \\ \frac{x-a}{b-a} & , \text{αν } a < x < b \\ \frac{x-c}{c-b} & , \text{αν } b < x < c \\ 0 & , \text{αν } x \geq c \end{cases} \quad (2.6)$$

Όπως φαίνεται στο σχήμα 2.3 (α), το κέντρο του τριγώνου είναι το σημείο b στο οποίο αντιστοιχεί ο μέγιστος βαθμός συμμετοχής, και τα σημεία a, c είναι τα ακραία σημεία της βάσης του τριγωνικού συνόλου.

- *Τραπεζοειδής Συνάρτηση Συμμετοχής*

Η τραπεζοειδής συνάρτηση συμμετοχής δίνεται από την παρακάτω σχέση,

$$\mu(x) = \begin{cases} 0 & , \text{αν } x < a \\ \frac{x-a}{b-a} & , \text{αν } a < x < b \\ 1 & , \text{αν } b < x < c \\ -\frac{x-d}{d-b} & , \text{αν } c < x < d \\ 0 & , \text{αν } x \geq d \end{cases} \quad (2.7)$$

Διαγραμματικά η τραπεζοειδής συνάρτηση συμμετοχής φαίνεται στο σχήμα 2.3 (β)

- *Γκαουσιανή Συνάρτηση Συμμετοχής*

Η γκαουσιανή συνάρτηση συμμετοχής δίνεται από την παρακάτω σχέση,

$$\mu(x) = e^{-\frac{(c-x)^2}{\sigma^2}} \quad (2.8)$$

Το σχήμα 2.3 (γ) παρουσιάζει τη δομή της συνάρτησης συμμετοχής.

Ένα από τα βασικά προβλήματα που έχουν να αντιμετωπίσουν οι αλγόριθμοι ανάλυσης συστάδων είναι η μεγάλη εξάρτηση από την επιλογή των αρχικών τιμών τόσο για τον αριθμό των συστάδων όσο και για τα αντίστοιχα κέντρα. Αν δεν επιλεγούν σωστά οι αρχικές τιμές για τις παραπάνω σχεδιαστικές παραμέτρους τότε η εφαρμογή αλγορίθμων όπως ο c -Μέσων (c -Means) θα οδηγήσει σε τοπικό ελάχιστο. Αυτό οφείλεται στο γεγονός ότι υπάρχει περίπτωση πολλά κέντρα συστάδων να μην κινηθούν καθόλου κατά τη διάρκεια εκτέλεσης του αλγορίθμου και συνεπώς στο τέλος πολύ λίγα δεδομένα να αντιστοιχηθούν σε αυτά. Η αιτία είναι ότι οι παραπάνω αλγόριθμοι βασίζονται στη στρατηγική «ο νικητής τα παίρνει όλα» (winner takes all). Η χρήση μεθόδων ασαφούς λογικής μπορεί να βοηθήσει προς αυτή την κατεύθυνση για τους παρακάτω λόγους [43,77,78]:

Η ασαφής λογική παρέχει τεχνικές αξιόπιστης μοντελοποίησης της αβεβαιότητας που υπάρχει σε ένα σύνολο δεδομένων.

- Η ασαφής λογική μπορεί να ελαχιστοποιήσει την εξάρτηση των αλγορίθμων συσταδοποίησης από την αρχικοποίηση των σχεδιαστικών της παραμέτρων γιατί βασίζεται σε πράες υπολογιστικές τεχνικές (soft computing techniques), οι οποίες μειώνουν την επίδραση της στρατηγικής «ο νικητής τα παίρνει όλα (winner takes all)».

Η μείωση της επίδρασης της στρατηγικής «ο νικητής τα παίρνει όλα» οφείλεται κυρίως στο γεγονός ότι, με βάση την ασαφή λογική, τα διανυσματικά δεδομένα δεν ανήκουν μόνο σε μία συστάδα αλλά σε πολλές με διαφορετικούς βαθμούς συμμετοχής. Αυτό πρακτικά σημαίνει ότι τα κέντρα θα κινηθούν, με διαφορετικούς ρυθμούς το καθένα, για να «κερδίσουν» τα δεδομένα [8,12]. Άρα το τελικό αποτέλεσμα θα είναι πιο αξιόπιστο από τις κλασσικές μεθόδους ανάλυσης συστάδων. Επίσης, η χρήση ασαφούς λογικής παρέχει το πλεονέκτημα ότι μπορεί να προσδιορίσει έναν ασαφή διαμερισμό (fuzzy partition) ενός συνόλου δεδομένων σε ασαφείς συστάδες (fuzzy clusters) με αυτόματο τρόπο γιατί στηρίζεται σε αυτοδιδασκόμενους μηχανισμούς. Οι μηχανισμοί αυτοί βασίζονται στην εκπλήρωση κάποιων κριτηρίων ή παραδοχών. Ο διαμερισμός (partition) που τελικά προκύπτει έχει ελεύθερη δομή, πράγμα που καθιστά την μεταχείριση της κάθε μεταβλητής πάρα πολύ εύκολη.

2.3.2 Ο Αλγόριθμος των Ασαφών *c*-Μέσων (Fuzzy *c*-Means)

Η πιο διαδεδομένη μέθοδος ασαφούς ανάλυσης συστάδων είναι ο αλγόριθμος Ασαφών *c*-Μέσων (Fuzzy *c*-Means) [9]. Ο αλγόριθμος αυτός είναι αντίστοιχος με τον κλασικό *c*-Μέσων (*c*-Means) που περιγράφηκε σε προηγούμενη ενότητα. Επιλύει το πρόβλημα της ελαχιστοποίησης των αποστάσεων μέσα σε μια συστάδα και της μεγιστοποίησης των αποστάσεων μεταξύ των συστάδων με βάση το παρακάτω κριτήριο βελτιστοποίησης [8,9,12,34]:

$$J_m = \sum_{k=1}^N \sum_{i=1}^n (u_{ik})^m \| \mathbf{x}_k - \mathbf{v}_i \|^2 \quad (2.9)$$

όπου $u_{ik} \in [0, 1]$ είναι η συνάρτηση συμμετοχής του k -οστού δεδομένου στην i -οστή συστάδα, $m \in [1, \infty)$ είναι η παράμετρος ασάφειας και $\mathbf{v}_i = [v_{i1}, v_{i2}, \dots, v_{ip}]^T$ ($1 \leq i \leq c$) είναι το διάνυσμα που αντιστοιχεί στο κέντρο της i -οστής συστάδας. Το κριτήριο J_m βελτιστοποιείται ως προς τις παραμέτρους u_{ik} και \mathbf{v}_i . Το πρόβλημα είναι η ελαχιστοποίηση της J_m υπό τον ακόλουθο περιορισμό,

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \quad (2.10)$$

Πιο συγκεκριμένα η προσέγγιση έχει ως εξής:

Για σταθερές τιμές των u_{ik} , οι τιμές των \mathbf{v}_i , οι οποίες ελαχιστοποιούν το J_m δίνονται από την εξίσωση [10]:

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_{ik}}{\sum_{k=1}^n (u_{ik})^m} \quad (2.11)$$

Για σταθερές τιμές των \mathbf{v}_i , η ελαχιστοποίηση του J_m ως προς τις u_{ik} είναι ένα πρόβλημα βελτιστοποίησης κατά Lagrange, οι τιμές των u_{ik} που ελαχιστοποιούν την J_m και ικανοποιούν της συνθήκη (2.10) είναι:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}}} \quad (1 \leq k \leq n, 1 \leq i \leq c) \quad (2.12)$$

Με βάση την παραπάνω ανάλυση ο αλγόριθμος Ασαφών c -Μέσων (Fuzzy c -Means) διατυπώνεται ως εξής [9,12]:

Αλγόριθμος Ασαφών c -Μέσων (Fuzzy c -Means)

Έστω ένα σύνολο διανυσματικών δεδομένων $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ με $\mathbf{x}_k \in R^p$ ($1 \leq k \leq n$), τα οποία θέλουμε να ομαδοποιήσουμε σε ασαφείς συστάδες.

- Βήμα 1) Επιλογή του αριθμού c των ασαφών συστάδων, της παραμέτρου m , των αρχικών τιμών για τα διανύσματα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$ και της παραμέτρου c .
- Βήμα 2) Χρήση της εξίσωσης (2.12) για του υπολογισμό των συναρτήσεων συμμετοχής u_{ik} ($1 \leq i \leq c, 1 \leq k \leq n$).
- Βήμα 3) Με βάση την εξίσωση (2.11) γίνεται προσδιορισμός των νέων τιμών για τα κέντρα των ασαφών υποομάδων $\mathbf{v}_1^{new}, \mathbf{v}_2^{new}, \dots, \mathbf{v}_n^{new}$.
- Βήμα 4) Αν $\max_i \{\|\mathbf{v}_i - \mathbf{v}_i^{new}\|^2\} < \varepsilon$ τότε ο αλγόριθμος σταματάει, αλλιώς θέτει $\mathbf{v}_i = \mathbf{v}_i^{new}$ και η ροή του πηγαίνει στο βήμα 2.

Ο παραπάνω αλγόριθμος είναι μια επαναληπτική διαδικασία κατά Picard και όπως έχει αποδειχθεί πάντα συγκλίνει. Έχει όμως το μειονέκτημα ότι για διαφορετικές αρχικές τιμές μπορεί να οδηγήσει σε διαφορετικά τοπικά ελάχιστα. Παρόλα αυτά, είναι ο πιο διαδεδομένος αλγόριθμος στη θεωρία της ασαφούς συσταδοποίησης δεδομένων, διότι συγκλίνει πολύ εύκολα, είναι απλός και οι μέχρι τώρα εφαρμογές έχουν δείξει ότι μπορεί να πετύχει πολύ ικανοποιητικά αποτελέσματα. Πρέπει να σημειωθεί ότι ο αλγόριθμος Ασαφών c -Μέσων (Fuzzy c -Means) έχει ισχύ για $c \geq 2$. Η ασάφεια του παραπάνω αλγόριθμου ελέγχεται από την παράμετρο m που είναι ένας δείκτης της αβεβαιότητας συστάδων. Μέχρι στιγμής, δεν υπάρχουν αξιόπιστες διαδικασίες για να αποφασίσουν την επιλογή της σωστής τιμής για αυτή την παράμετρο. Όταν το m προσεγγίζει τη μονάδα το τελευταίο διαμέρισμα είναι σχεδόν κοφτό λήψης αποφάσεων αποτέλεσμα και τότε οι συστάδες τείνουν να είναι μη ασαφής (τα όρια τους είναι αυστηρά ορισμένα). Από την άλλη πλευρά, αν το m παίρνει μεγάλες τιμές, τότε οι βαθμοί συμμετοχής του κάθε διανύσματος εκπαίδευσης τείνουν να προσεγγίζουν το $\frac{1}{c}$, δηλαδή τα στοιχεία ανήκουν σε όλες τις υπό μελέτη συστάδες με τον ίδιο βαθμό πίστης [106]. Σε αυτή την περίπτωση όλα τα κέντρα συστάδων θα τραβηχτούν προς τα διανυσματικά δεδομένα πολύ αργά.

2.4 Αλγόριθμος υπό όρους ασαφής συσταδοποίησης (Conditional Fuzzy Clustering - CFC)

Ο συγκεκριμένος αλγόριθμος αναπτύχθηκε από τον Pedrycz [72,73] και περιγράφεται αναλυτικά παρακάτω, διότι συγκρίνεται με τις μεθόδους που αναπτύχθηκαν στην παρούσα διδακτορική διατριβή στα κεφάλαια 5, 6 και 7.

Ας θεωρηθεί x_1, x_2, \dots, x_n ένα σύνολο δεδομένων p -διάστασης που ορίζονται στο χώρο \mathbb{R}^p . Ο αλγόριθμος προκύπτει ως μία τροποποίηση του γενικού αλγορίθμου των Ασαφών c -Μέσων (FCM). Η αντικειμενική συνάρτηση κόστους (Q) έχει την εξής τυποποιημένη μορφή:

$$Q = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|x_k - v_i\|^2 \quad (2.13)$$

με τα κέντρα (πρωτότυπα) v , και το u ο πίνακας διαμερισμού (partition matrix) που είναι τα συστατικά που βελτιστοποιούνται.

Ο πίνακας $U = [u_{ik}]$ ονομάζεται πίνακας διαμερισμού (partition matrix), $U \in u$, και δίνεται από τον εξής τύπο:

$$u = \left\{ u_{ik} \in [0,1] \sum_{i=1}^c u_{ik} = 1 \quad \forall k \quad \text{και} \quad 0 < \sum_{k=1}^n u_{ik} < n \quad \forall i \right\}. \quad (2.14)$$

Το πρόβλημα βελτιστοποίησης δίνεται από την εξής φόρμα:

$$\min_{U, u_1, u_2, \dots, u_c} Q \quad \text{υπό την προϋπόθεση} \quad U \in u.$$

Η επαναληπτική διαδικασία οδηγεί σε τοπικό ελάχιστο ή σε ένα σημείο Q το οποίο είναι ευρέως γνωστό και αφορά μια σειρά από ενημερώσεις του πίνακα διαμερισμών (partition matrix) και δίνεται από τον εξής τύπο:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}} \quad i = 1, 2, \dots, c \quad , \quad k = 1, 2, \dots, n \quad (2.15)$$

Τα κέντρα (πρωτότυπα) των συστάδων υπολογίζονται από τον παρακάτω τύπο:

$$v_i = \frac{\sum_{i=1}^n (u_{ik})^m x_{ik}}{\sum_{i=1}^n (u_{ik})^m}, \quad i = 1, 2, \dots, c \quad (2.16)$$

Η πτυχή της υπό συνθήκη μηχανισμού συσταδοποίησης εισάγεται στον αλγόριθμο λαμβάνοντας υπόψη την υπό συνθήκη μεταβλητή (εξόδου), υποθέτοντας ότι οι τιμές f_1, f_2, \dots, f_n στα αντίστοιχα διανύσματα. Πιο συγκεκριμένα, η μεταβλητή f_k περιγράφει το επίπεδο συμμετοχής του διανύσματος x_k στην συστάδα που δημιουργήθηκε (s). Παραδείγματος χάριν, αν υποτεθεί ότι ο γλωσσικός όρος που καθορίζεται στο χώρο εξόδου εκφράζεται ως ασαφές σύνολο $B, B: R \rightarrow [0,1]$, τότε ισχύει $f_k = B(y_k), k = 1, 2, \dots, n$ για έναν βαθμό συμμετοχής του y_k στο B . Με άλλα λόγια, θέτουμε το πρόβλημα συσταδοποίησης όπως,

—ομαδοποιήστε το δεδομένο θεωρώντας ότι το y ανήκει στο B .

Ο τρόπος με τον οποίο το f_k μπορεί να συνδεθεί με ή να διατεθεί μεταξύ των υπολογισμένων τιμών βαθμών συμμετοχής του x_k $u_{1k}, u_{2k}, \dots, u_{ck}$ δεν είναι μοναδικός. Οι περιορισμοί στους βαθμούς συμμετοχής (πίνακας διαμερισμών) τροποποιούνται και ενσωματώνουν τις τιμές του πλαισίου f_k . Το πρόβλημα είναι η ελαχιστοποίηση της Q υπό τον ακόλουθο περιορισμό:

$$\sum_{i=1}^c u_{ik} = f_k \quad (2.17)$$

Ο νέος περιορισμός (f_k) περιλαμβάνει τα ακόλουθα:

- Εάν το k -οστο διανυσματικό δεδομένο του συνόλου δεδομένων θεωρείται χαμηλής σημασίας στο πλαίσιο που επιβάλλεται από το $B, B(y_k) = 0$ αυτό έχει ως αποτέλεσμα τον αποκλεισμό αυτού του δεδομένου από την περαιτέρω συμμετοχή στη διαδικασία συσταδοποίησης, $u_{ik} = 0 \quad \forall i$. Όπως φαίνεται από την εξ. (2.16), οι υπολογισμοί των κέντρων (πρωτοτύπων) που προκύπτουν, δεν

επηρεάζονται από τέτοια στοιχεία. Στη συνέχεια, οι υπολογισμοί του πίνακα διαμερισμού (partition matrix) δεν τους λαμβάνουν υπόψη .

- Αφ' ετέρου, το δεδομένο για το οποίο $B(y_k)=1$ συμβάλλει στη διαδικασία συσταδοποίησης στον υψηλότερο βαθμό. Όλα τα στοιχεία με κάποια ενδιάμεση συμβολή πλαισίου (ενεργοποίηση) λαμβάνονται υπόψη μόνο σε μερική βάση.

Λαμβάνοντας τα παραπάνω υπόψη, τροποποιούνται οι απαιτήσεις που καλύπτονται από τους πίνακες διαμερισμών και ορίζουν την ακόλουθη οικογένεια:

$$u(f) = \left\{ u_{ik} \in [0,1] \sum_{i=1}^c u_{ik} = f_k \quad \forall k \quad \text{και} \quad 0 < \sum_{k=1}^n u_{ik} < n \quad \forall i \right\} \quad (2.18)$$

Η κανονική συνθήκη αντικαθίσταται από την συμμετοχή του υπό συνθήκη περιορισμού. Το πρόβλημα βελτιστοποίησης αναδιατυπώνεται ως εξής:

$$\min_{U, u_1, u_2, \dots, u_c} Q \quad \text{υπό την προϋπόθεση} \quad U \in u(f)$$

Η ελαχιστοποίηση της αντικειμενικής συνάρτησης Q γίνεται επαναληπτικά όπου οι συναρτήσεις συμμετοχής (membership functions) υπολογίζονται από τον εξής τύπο:

$$u_{ik} = \frac{f_k}{\sum_{j=1}^c \left(\frac{\| \mathbf{x}_k - \mathbf{v}_i \|}{\| \mathbf{x}_k - \mathbf{v}_j \|} \right)^{\frac{2}{m-1}}} \quad (2.19)$$

Με το να συμβεί αυτό, οι βαθμοί συμμετοχής του k -διανύσματος αθροίζονται στον περιορισμό πλαισίου f_k . Εδώ πρέπει να σημειωθεί ότι το πλαίσιο αρχικά έχει τεθεί ιδανικά ίσο με 1, $f_k = 1$ για όλα τα k , την αρχική διατύπωση του τυποποιημένου αλγορίθμου Ασαφών c -Μέσων (FCM). Οι πίνακες διαμερισμού (partition matrix) που υπονοούνται από το πλαίσιο F υποδηλώνονται από το $U(F)$, με το επιχείρημα που χρησιμοποιείται εκεί για να δώσει έμφαση στην εξάρτηση επάνω στον περιορισμό πλαισίου.

Όπως αναφέρεται, το πλαίσιο F μπορεί να υποτεθεί ότι έχει διάφορες μορφές. Θα μπορούσε να είναι ένα ασαφές σύνολο ή μια ασαφή σχέση, το οποίο δεν είναι σχετικό με

τη διατύπωση του προβλήματος βελτιστοποίησης - αυτός ο περιορισμός γίνεται αντιληπτός μέσω των βαθμών συμμετοχής του F . Το πρόβλημα βελτιστοποίησης της υπό όρους συσταδοποίησης μπορεί να εκφραστεί ως εξής [72]:

$$\min_{U \in U(F), v_1, v_2, \dots, v_c} \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|x_k - v_i\|^2 \quad (2.20)$$

όπου το $U(F)$ δείχνει μια οικογένεια των πινάκων διαμερισμού (partition matrix) που προκαλούνται από τον όρο F , σημαίνοντας ότι ικανοποιεί τον όρο κανονικοποίησης που εκφράζεται από την εξ. (2.17).

Σχετικά με το θέμα του σχηματισμού των ασαφών συνόλων ή των ασαφών σχέσεων, αξίζει να τονιστεί ότι αυτοί θα μπορούσαν να είναι με δύο διαφορετικές μορφές:

- Συσταδοποίηση προσανατολισμένη σε ειδικό (expert-oriented). Θεωρούμε ότι η γνώση περιοχών αποτελείται από τους γλωσσικούς όρους που διαμορφώνονται από το χρήστη που σκοπεύει να στρέψει την ανάλυση δεδομένων σε ένα ορισμένο πλαίσιο. Τα πλαίσια είναι αυτά τα ενδιαφέροντα σημεία εστίασης. Θα μπορούσαν να ρυθμιστούν δυναμικά, ανάλογα με τα προηγούμενα συμπεράσματα για τη δομή στο σύνολο δεδομένων. Ειδικότερα, κάποιος μπορεί να κινηθεί προς τα πιο συγκεκριμένα (λεπτομερή) πλαίσια εάν η δομή που ανιχνεύεται μέχρι τώρα δεν είναι αρκετά πληροφοριακή.
- Το αποτέλεσμα της βοηθητικής συσταδοποίησης που ολοκληρώνεται χωριστά για τη μεταβλητή πλαισίου (s). Το κύριο πλεονέκτημα αυτής της προσέγγισης είναι ότι αυτά τα πλαίσια είναι βασισμένα στα πειραματικά στοιχεία τα οποία είναι διαθέσιμα ως τμήμα του συνόλου δεδομένων.

Η βασισμένη στο πλαίσιο συσταδοποίηση (context-based clustering) έχει διάφορα χαρακτηριστικά γνωρίσματα που υπονοούνται από το χαρακτήρα των πλαισίων. Τα πιο συγκεκριμένα πλαίσια υπονοούν την εκλεπτυσμένη (refined) συσταδοποίηση (clustering). Λαμβάνοντας υπόψη τα πλαίσια A' και A έτσι ώστε το $A' \subset A$, η βασισμένη στο πλαίσιο συσταδοποίηση (context-based clustering) που καθοδηγείται από το A' τείνει να αποκαλύψει περισσότερες δομικές λεπτομέρειες από τη συσταδοποίηση

που καθοδηγείται από το A . Επιπλέον, παίρνοντας έναν απλό όρο ορίου: $D = X$ αυτό οδηγεί στην ανεξάρτητη από το πλαίσιο (context-free) συσταδοποίηση (clustering). Τα σύνθετα πλαίσια δημιουργούνται με το να διαμορφώσουν τα καρτεσιανά γινόμενα των τμημάτων συμβολής τους, για παράδειγμα, $D = A \times B \times C$. Ο βαθμός συμμετοχής του διανύσματος των εξόδων $z = [z_1 \ z_2 \ z_3]^T$ εκφράζεται ως $D(z)$. Όταν το A , το B , και το C διευκρινίζονται κατά τρόπο ρητό, λαμβάνετε το $D(z) = A(z_1) \cdot B(z_2) \cdot C(z_3)$, με έναν ορισμένο t -κανόνα (t -norm) που χρησιμοποιείται για να διαμορφώσει τον τύπο συνάθροισης των συντεταγμένων.

2.4.1 Το πρόβλημα βελτιστοποίησης

Ο αλγόριθμος βελτιστοποίησης περιλαμβάνει δύο φάσεις: υπολογισμοί του πίνακα διαμερισμού (partition) και των κέντρων (πρωτοτύπων). Το τελευταίο είναι απλό: καταλήγουμε με τον ίδιο τύπο όπως στον τυποποιημένο αλγόριθμο Ασαφών c -Μέσων (FCM). Αυτό δεν αποτελεί έκπληξη. Υπολογίζονται τα κέντρα (πρωτότυπα) βασισμένα στην αντικειμενική συνάρτηση, η οποία δεν εξαρτάται από τους περιορισμούς πλαισίου. Οι υπολογισμοί του πίνακα διαμερισμού (partition matrix) ενδιαφέρονται για τη χρήση των περιορισμών πλαισίου. Μετά από μια τυποποιημένη τεχνική Lagrange πολλαπλασιαστών, διαμορφώνετε ο εξής τύπος,

$$V = \sum_{i=1}^c u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 - \lambda \left(\sum_{i=1}^c u_{ik} - f_k \right) \quad (2.21)$$

ο οποίος ισχύει για κάθε διάνυσμα $k=1,2,\dots, n$. Η ελαχιστοποίηση της εξίσωσης (2.21) γίνεται όσον αφορά τα στοιχεία του πίνακα διαμερισμού. Ο περιορισμός f_k προκαλεί την ακόλουθη έκφραση:

$$u_{ik} = \frac{f_k}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}}} \quad (2.22)$$

Όπως σημειώνεται, οι υπολογισμοί των κέντρων (πρωτοτύπων) είναι οι ίδιοι όπως εκείνοι με το ανεξάρτητο από το πλαίσιο (context-free) Ασαφών c -Μέσων (FCM). Κατά τρόπο ενδιαφέροντα, τα πλαίσια δεν παρουσιάζονται ρητά στον τύπο. Προφανώς

προσκρούουν ακόμα στα κέντρα (πρωτότυπα), και αυτό συμβαίνει μέσω των τιμών του πίνακα διαμερισμού.

2.5 Ο αλγόριθμος κβάντισης διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ)

Ο συγκεκριμένος αλγόριθμος αναπτύχθηκε από τον Kohonen [45] και περιγράφεται αναλυτικά παρακάτω, διότι χρησιμοποιείται στη μέθοδο που αναπτύχθηκε στο κεφάλαιο 4 της παρούσης διδακτορικής διατριβής.

Υποθέτουμε ότι μας δίνεται ένα σύνολο διανυσματικών δεδομένων εισόδου $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$. Επίσης, υποθέτουμε ότι το δίκτυο LVQ αποτελείται από τους νευρώνες n , όπου το συναπτικό βάρος του l νευρώνα, είναι ένα σημείο στο p – διάστατο χώρο: $u_l = [u_{l1}, u_{l2}, \dots, u_{lp}]$ ($1 \leq l \leq n$)

Βήμα 1) Επιλέγονται τυχαίες τιμές για τα αρχικά συναπτικά βάρη w_l ($1 \leq l \leq n$).

Επιλέγεται μια τιμή για το μέγιστο αριθμό επαναλήψεων t_{\max} , και τίθεται $t = 0$.

Βήμα 2) Για $k = 1, 2, \dots, N$

Εξακριβώσε τον καλύτερο νευρώνα (l_0) που ταιριάζει στο τρέχον διάνυσμα δεδομένων εισόδου (\mathbf{x}_k) σύμφωνα με τον ακόλουθο κανόνα κοντινότερων γειτόνων (nearest neighbor),

$$\|\mathbf{x}_k - u_{l_0}(t)\|^2 = \min_l \{\|\mathbf{x}_k - u_l(t)\|^2\} \quad (1 \leq l \leq n) \quad (2.23)$$

Ενημέρωσε το συναπτικό βάρος κάθε νευρώνα ως εξής

$$u_l(t+1) = u_l(t) + a(t)h_{l,l_0,k}(t)(\mathbf{x}_k - u_l(t)) \quad (2.24)$$

Τέλος Επανάληψης

Βήμα 3) Εάν καμία αξιοπρόσεχτη αλλαγή στα συναπτικά βάρη δεν παρατηρείται ή

$t > t_{\max}$ τότε σταμάτα. Αλλιώς θέσε $t = t + 1$ και πήγαινε στο βήμα 2.

Η συνάρτηση $h_{l,l_0,k}(t)$ καθορίζει το πιο πρόσφατο φαινόμενο αλληλεπίδρασης μεταξύ των νευρώνων l και l_0 . Η συνάρτηση $h_{l,l_0,k}(t)$ που χρησιμοποιείται σε αυτό το κεφάλαιο είναι,

$$h_{l,l_0,k}(t) = \begin{cases} 1, & \text{αν } l=l_0 \\ 0, & \text{αν } l \neq l_0 \end{cases} \quad (2.25)$$

Στην εξ. (2.24) η παράμετρος $a(t)$ είναι η παράμετρος ποσοστού εκπαίδευσης. Προκειμένου να βεβαιωθεί η σύγκλιση, αυτή η παράμετρος απαιτείται να μειωθεί μονότονα καθώς η διαδικασία εκπαίδευσης προχωρά

$$a(t) = a_0 \left(1 - \frac{t}{t_{\max}} \right) \quad (2.26)$$

όπου a_0 είναι μια αρχική τιμή για την παράμετρο $a(t)$.

Ο ανωτέρω αλγόριθμος ταξινομεί όλα τα δεδομένα εκπαίδευσης εισόδου σε n συστάδες συγκεντρώνεται, όπου η συστάδα l ορίζεται ως

$$U^l = \{ \mathbf{x}_k \in X : \|\mathbf{x}_k - \mathbf{v}_l\|^2 = \min_j \|\mathbf{x}_k - \mathbf{v}_j\|^2 \} \quad (2.27)$$

και το κεντρικό στοιχείο του είναι το αντίστοιχο συναπτικό βάρος u_l . Είναι σημαντικό να σημειωθεί ότι το σύνολο $U = U^1 \cup U^2 \cup \dots \cup U^n$ καθορίζει ένα ιδανικό χόρισμα στο X .

2.6 Ο Αλγόριθμος Ασαφούς Κβάντισης Διανύσματος με Εκμάθηση (Fuzzy Learning Vector Quantization –FLVQ)

Ο Αλγόριθμος FLVQ χρησιμοποιείται στο κεφάλαιο 6 της παρούσης διδακτορικής διατριβής και γι' αυτό αναλύεται λεπτομερώς παρακάτω.

Ας υποθέσουμε ότι μας δίνεται ένα σύνολο από n διανυσματικά δεδομένα εισόδου $X = \{ \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \} \subset \mathbb{R}^p$. Η αντικειμενική συνάρτηση του FLVQ αλγόριθμου δίνεται από την εξίσωση (2.9).

Τα τελικά διανύσματα κωδικοποίησης και οι συναρτήσεις συμμετοχής που λύνουν αυτό το πρόβλημα δίνονται και εδώ από τις σχέσεις (2.11) και (2.12), αντίστοιχα. Έπειτα, ο αλγόριθμος FLVQ μεταβάλλει την παράμετρο ασάφειας m και περιγράφεται ακολούθως,

Αλγόριθμος FLVQ

Βήμα 1. Επέλεξε τον αριθμό των συστάδων c , τις αρχικές τιμές για τα κέντρα συστάδων $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$, και μια τιμή για την παράμετρο ε .

Βήμα 2. Θέσε τον μέγιστο αριθμό επαναλήψεων t_{\max} , και επέλεξε την αρχική τιμή m_0 και την τελική τιμή m_f για την παράμετρο m .

Βήμα 3. Για $t = 0, 1, 2, \dots, t_{\max}$

a. Υπολόγισε την παράμετρο ασάφειας $m(t) = m_0 - [t(m_0 - m_f)] / t_{\max}$

$$b. \quad \text{Θέσε: } a_{ik}(t) = \left[\sum_{i=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{2/(m(t)-1)} \right]^{-m(t)} = u_{ik}^{m(t)}(t) \quad (2.28)$$

$$c. \quad \text{Θέσε: } \eta_i(t) = \left(\sum_{k=1}^n a_{ik}(t) \right)^{-1}, \quad (1 \leq i \leq c) \quad (2.29)$$

d. Ενημέρωσε τα κέντρα συστάδων σύμφωνα με τον παρακάτω κανόνα

$$\text{εκπαίδευσης: } \mathbf{v}_i(t) = \mathbf{v}_i(t-1) + \eta_i(t) \sum_{k=1}^n a_{ik}(t) (\mathbf{x}_k - \mathbf{v}_i(t-1)) \quad (2.30)$$

e. Αν $E(t) = \sum_{i=1}^c \|\mathbf{v}_i(t) - \mathbf{v}_i(t-1)\| < \varepsilon$ τότε σταμάτα. Αλλιώς συνέχισε.

Με το συνδυασμό των σχέσεων (2.29) και (2.30) επιτυγχάνεται ότι,

$$\mathbf{v}_i(t) = \mathbf{v}_i(t-1) + \frac{\sum_{k=1}^n a_{ik}(t) (\mathbf{x}_k - \mathbf{v}_i(t-1))}{\sum_{k=1}^n a_{ik}(t)} \quad (2.31)$$

Κατόπιν η (2.31) μπορεί να τροποποιηθεί ως,

$$\mathbf{v}_i(t) = \frac{\sum_{k=1}^n a_{ik}(t) \mathbf{x}_k}{\sum_{k=1}^n a_{ik}(t)} \quad (2.32)$$

Η κεντρική ιδέα του παραπάνω αλγόριθμου είναι ότι μπορούμε να επιτύχουμε μια ομαλή μετάβαση από την κατάσταση ασάφειας στην σαφή κατάσταση με τον χειρισμό της παραμέτρου m . Επιπλέον, οι εξισώσεις για την FLVQ προϋποθέτουν ότι οι συναρτήσεις συμμετοχής είναι συνδεδεμένες με τον αλγόριθμο Ασαφών c -Μέσων (FCM) και χρησιμοποιούνται για να καθορίσουν τη δύναμη της έλξης μεταξύ του κάθε κέντρου συστάδας και των διανυσματικών δεδομένων. Ωστόσο υπάρχουν δύο δυσκολίες όταν εφαρμόζουμε τον FLVQ. Η πρώτη είναι η ανάγκη για υψηλό υπολογιστικό κόστος. Η δεύτερη είναι η επιλογή της κατάλληλης αρχικής τιμής για την παράμετρο ασάφειας m . Όπως συζητήθηκε από τους Pal και Bezdek [67], αυτό είναι από τα πιο μεγάλα θέματα στην ανάπτυξη αξιόπιστων αλγόριθμων ασαφούς συσταδοποίησης.

2.7 Ασαφής Κβάντιση Διανύσματος (Fuzzy Vector Quantization-FVQ)

Ο αλγόριθμος FVQ αναφέρεται στο κεφάλαιο 7 της παρούσης διδακτορικής διατριβής. Η κβάντιση διανύσματος πραγματοποιείται με τη συσχέτιση κάθε διανύσματος εκπαίδευσης σε ένα μονό διάνυσμα βιβλίου κωδικοποίησης. Επιπλέον, η εφαρμογή της μεθόδου Ασαφών c -Μέσων (FCM) στην κβάντιση διανύσματος πρέπει να βασίζεται στην εκχώρηση κάθε διανύσματος εκπαίδευσης στο διάνυσμα βιβλίου κωδικοποίησης, το οποίο εμφανίζει τον μέγιστο βαθμό συμμετοχής αναφορικά με αυτό το διανυσματικό δεδομένο. Ωστόσο, μια τέτοια «κοφτή» ερμηνεία της μεθόδου Ασαφών c -Μέσων (FCM) και του FLVQ κατά τη διάρκεια του σχεδιασμού βιβλίου κωδικοποίησης μπορεί να έχει σοβαρές επιπτώσεις στην ποιότητα του τελικού βιβλίου κωδικοποίησης. Αυτή η συμπεριφορά της μεθόδου Ασαφών c -Μέσων (FCM) δικαιολογεί την ανάγκη για τη σταδιακή μετάβαση από τις ασαφείς σε σαφείς αποφάσεις.

Οι Καραγιάννης and Παί [39] έχουν αναπτύξει τρεις προσεγγίσεις ασαφούς κβάντισης διανύσματος, όπου το βασικό θέμα σχεδιασμού είναι η σταδιακή μετάβαση από την ασαφή στην διακριτή (crisp) κατάσταση. Πιο συγκεκριμένα, στα αρχικά στάδια της διαδικασίας σχεδιασμού βιβλίου κωδικοποίησης, κάθε διανυσματικό δεδομένο εκχωρείται στα διανύσματα βιβλίου κωδικοποίησης που ανήκουν σε μια υπερ-σφαίρα κεντραρισμένη σ' εκείνο το διάνυσμα. Το σύνολο των διανυσμάτων που ανήκουν στην

υπερ-σφαίρα κεντραρισμένη στο k -στό διανυσματικό δεδομένο συμβολίζεται ως T_k . Έπειτα, η σταδιακή μετάβαση από την ασαφή στην σαφή κατάσταση επιτυγχάνεται με την βαθμιαία συρρίκνωση των επικαλυπτόμενων υπέρ-σφαιρών κατά τη διάρκεια της διαδικασίας συσταδοποίησης. Επιπλέον, βρέθηκε ότι η ταχύτητα μετάβασης ευθέως επηρεάζει την ποιότητα του προκύπτοντος βιβλίου κωδικοποίησης. Καθώς η διαδικασία σχεδιασμού προχωρά, το σύνολο T_k ενημερώνεται σύμφωνα με την ακόλουθη διαδικασία:

Στην t -οστή επανάληψη το σύνολο $T_k^{(t)}$ περιέχει $\aleph(T_k^{(t)})$ διανύσματα βιβλίου κωδικοποίησης. Με τον προσδιορισμό της ακόλουθης μέσης απόστασης:

$$\tilde{d}_k^{(t)} = \frac{1}{\aleph(T_k^{(t)})} \sum_{\mathbf{v}_i \in T_k^{(t)}} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (2.33)$$

στην $(t+1)$ -οστή επανάληψη το σύνολο T_k επιτυγχάνεται ως εξής:

$$T_k^{(t+1)} = \left\{ \mathbf{v}_i \in T_k^{(t)} : \|\mathbf{x}_k - \mathbf{v}_i\|^2 \leq \tilde{d}_k^{(t)} \right\} \quad (2.34)$$

Ο παραπάνω κανόνας προϋποθέτει ότι οι βαθμοί συμμετοχής του \mathbf{x}_k στα διανύσματα κωδικοποίησης, τα οποία κατά τη διάρκεια της $(t+1)$ στής επανάληψης αφαιρούνται από το σύνολο T_k , θέτονται ίσα με το μηδέν. Επομένως, αρχικά, όλα τα διανύσματα κωδικοποίησης ανήκουν στο T_k . Αυτό το γεγονός εγγυάται την συμμετοχή όλων των διανυσμάτων κωδικοποίησης στην διαδικασία σχεδιασμού βιβλίου κωδικοποίησης. Καθώς αυτή η διαδικασία προχωρά, ο αριθμός στοιχείων $\aleph(T_k^{(t)})$ του συνόλου T_k μειώνεται, μέχρι το T_k να περιλαμβάνει μόνο ένα στοιχείο. Σε αυτή την περίπτωση το k -στό διανυσματικό δεδομένο μεταφέρεται από την ασαφή στη σαφή κατάσταση.

Η παραπάνω διαδικασία προϋποθέτει ότι τα διανύσματα βιβλίου κωδικοποίησης ενημερώνονται χρησιμοποιώντας την εξίσωση (2.11). Επιπλέον, στην t -στή επανάληψη, αν το διανυσματικό δεδομένο \mathbf{x}_k είναι στην ασαφή κατάσταση οι βαθμοί συμμετοχής του υπολογίζονται ως εξής,

$$u_{ik} = \frac{1}{\sum_{v_j \in T_k^{(t)}} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{2/(m-1)}}, \quad \text{με } \mathbf{v}_i \in T_k^{(t)} \quad (2.35)$$

δεδομένου ότι αν το \mathbf{x}_k είναι στην κατάσταση κοψίματος τότε οι βαθμοί συμμετοχής δίνονται από τη συνθήκη του επόμενου κοντινότερου γείτονα,

$$u_{ik} = \begin{cases} 1 & \text{αν } \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \min_{1 \leq i \leq c} \{\|\mathbf{x}_k - \mathbf{v}_i\|^2\} \\ 0 & \text{αλλιώς} \end{cases} \quad (2.36)$$

Αναφορικά με την τελευταία σχέση, η χρήση της ισότητας (2.11) είναι ακόμη έγκυρη, από τη στιγμή που σ' αυτή την περίπτωση διατηρείται το $(u_{ik})^m = u_{ik}$ άσχετα από την τιμή του m .

2.8 Ο Βελτιωμένος Αλγόριθμος Ασαφούς Κβάντισης Διανύσματος με Εκμάθηση (IFLVQ)

Στην παρούσα παράγραφο αναλύεται ο βελτιωμένος αλγόριθμος ασφούς κβάντισης διανύσματος improved FLVQ (IFLVQ) ο οποίος χρησιμοποιείται για σύγκριση στο κεφάλαιο 6 της παρούσης διδακτορικής διατριβής. Ο αλγόριθμος επιτυγχάνει τη μετάβαση από την κατάσταση ασάφειας στις σαφείς συνθήκες με ένα διαφορετικό τρόπο από την FLVQ που αναλύθηκε στην ενότητα 2.6. Πιο συγκεκριμένα, μετατρέπεται η αντικειμενική συνάρτηση του αλγόριθμου Ασαφών c -Μέσων (FCM),

$$J_R = \theta \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 + (1-\theta) \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (2.37)$$

όπου $u_{ik} \in [0,1]$ και $\theta \in [0,1)$.

Όταν $\theta=0$ η παραπάνω αντικειμενική συνάρτηση συμπίπτει με την κλασική FCM. Όταν το θ είναι ίσο με τη μονάδα η αντικειμενική συνάρτηση J_R γίνεται η μέθοδος c -Μέσων (c -Means). Όμως, όπως θα δειχθεί αργότερα, η παράμετρος θ δεν πρέπει να

πάρει την τιμή $\theta=1$. Βασισμένοι σε αυτή την ανάλυση, μπορούμε να συμπεράνουμε ότι κατευθύνοντας την παράμετρο θ από το μηδέν στη μονάδα επιτυγχάνουμε μία FLVQ επανάληψη. Ο στόχος είναι να ελαχιστοποιηθεί η αντικειμενική συνάρτηση στην (2.37) υπό τον περιορισμό που δίνεται από την εξίσωση (2.10). Οι συναρτήσεις συμμετοχής και τα κέντρα συστάδων που επιλύουν αυτό το πρόβλημα ελαχιστοποίησης δίνονται αντίστοιχα παρακάτω,

$$u_{ik} = \frac{2+(c-2)\theta}{2(1-\theta)} \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_j\|}{\|\mathbf{x}_k - \mathbf{v}_i\|} \right)^2} - \frac{\theta}{2(1-\theta)} \quad (2.38)$$

και

$$\mathbf{v}_i = \frac{\sum_{k=1}^n [\theta u_{ik} + (1-\theta)(u_{ik})^2] \mathbf{x}_k}{\sum_{k=1}^n [\theta u_{ik} + (1-\theta)(u_{ik})^2]}, \quad (1 \leq i \leq c) \quad (2.39)$$

Στο επόμενο βήμα, ανασηματίζουμε την αντικειμενική συνάρτηση στην εξίσωση (2.37) με την αντικατάσταση του u_{ik} από την εξίσωση (2.38). Συνεπώς, η J_R γίνεται,

$$J_R = \frac{[2+(c-2)\theta]^2}{4(1-\theta)} \sum_{k=1}^n \left[\sum_{i=1}^c \left(\frac{1}{\|\mathbf{x}_k - \mathbf{v}_i\|} \right)^2 \right]^{-1} - \frac{\theta^2}{4(1-\theta)} \sum_{k=1}^n \sum_{i=1}^c \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (2.40)$$

$$\mathbf{v}_i(t) = \frac{\sum_{k=1}^n [\theta(t)u_{ik} + (1-\theta(t))(u_{ik})^2] \mathbf{x}_k}{\sum_{k=1}^n [\theta(t)u_{ik} + (1-\theta(t))(u_{ik})^2]} \quad (1 \leq i \leq c) \quad (2.41)$$

Συνοψίζοντας, η επαναληπτική εφαρμογή των εξισώσεων (2.38) και (2.41) ορίζει ένα σχήμα κβάντισης διανύσματος, όπου η μετάβαση από την ασαφή στην σαφή κατάσταση επιτυγχάνεται με την σταδιακή αύξηση της παραμέτρου ασάφειας θ από 0 σε 1. Ωστόσο,

η απευθείας εφαρμογή της εν' λόγω επαναληπτικής διαδικασίας προϋποθέτει μεγάλες υπολογιστικές απαιτήσεις. Επιπλέον, από τη στιγμή που η παράμετρος θ δεν θα φτάσει ποτέ τη μονάδα, η πλήρης μετάβαση από την ασαφή στην σαφή κατάσταση δεν θα επιτευχθεί ποτέ. Συνεπώς, στο τέλος της διαδικασίας εκπαίδευσης, υπάρχει ακόμη η ανάγκη να χρησιμοποιηθεί ένα κατώτατο όριο για τους βαθμούς συμμετοχής προκειμένου να μεταφερθούν τα διανυσματικά δεδομένα από την ασαφή στην σαφή κατάσταση.

Οι Tsekouras et al [92] ανέπτυξαν μια στρατηγική για τη μετάβαση από την κατάσταση ασάφειας στην σαφή κατάσταση. Σύμφωνα με αυτή τη διαδικασία, στα αρχικά στάδια της διαδικασίας σχεδιασμού του βιβλίου κωδικοποίησης, κάθε διάνυσμα τιμής δεδομένου αντιστοιχίζεται στα διανύσματα βιβλίων κώδικα (διανύσματα κωδικοποίησης) που ανήκουν σε μια υπερσφαίρα που έχει ως κέντρο το διάνυσμα αυτό. Το σύνολο των διανυσμάτων κωδικοποίησης που ανήκουν σε μια υπερσφαίρα που κεντροθετείται στο k -στο διάνυσμα συμβολίζεται και εδώ ως T_k . Έπειτα η μετάβαση από την κατάσταση ασάφειας στην σαφή κατάσταση ολοκληρώνεται με βαθμιαία συρρίκνωση της επικαλυπτόμενων υπερσφαιρών κατά τη διάρκεια της διαδικασίας συσταδοποίησης. Οι βαθμοί συμμετοχής του x_k στα διανύσματα βιβλίων κώδικα υπολογίζονται από την εξίσωση (2.38). Η μορφή αυτής της εξίσωσης υποδηλώνει ότι ο βαθμός συμμετοχής μπορεί να γίνει αρνητικός αν ισχύει η παρακάτω ανισότητα,

$$\frac{2 + (c - 2)\theta}{2(1 - \theta)} \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_j\|}{\|x_k - v_j\|} \right)^2} \leq \frac{\theta}{2(1 - \theta)} \quad (2.42)$$

Προφανώς, τα διανύσματα βιβλίων κώδικα που αντιστοιχίζουν στο x_k αρνητικούς βαθμούς συμμετοχής, πρέπει να αφαιρεθούν από το σύνολο T_k . Επιπλέον, στην t επανάληψη το σύνολο $T_k^{(t)}$ ορίζεται ως,

$$T_k^{(t)} = \{v_i \in V : u_{ik} > 0\} \quad (2.43)$$

Επομένως, στην πρώτη επανάληψη το σύνολο $T_k^{(0)}$ περιλαμβάνει όλα τα διανύσματα βιβλίου κωδικοποίησης, ενώ στην t -οστή επανάληψη το σύνολο $T_k^{(t)}$ περιέχει $\aleph(T_k^{(t)})$ διανύσματα βιβλίου κωδικοποίησης. Προκειμένου να προσδιοριστούν οι βαθμοί συμμετοχής του \mathbf{x}_k στα διανύσματα συμμετοχής που ανήκουν στο σύνολο $T_k^{(t)}$, τροποποιούμε την εξίσωση (2.38) όπως φαίνεται παρακάτω,

$$u_{ik}(t) = \frac{2 + (\aleph(T_k^{(t)}) - 2)\theta(t)}{2(1 - \theta(t))} \frac{1}{\sum_{\mathbf{v}_j \in T_k^{(t)}} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i(t)\|}{\|\mathbf{x}_k - \mathbf{v}_j(t)\|} \right)^2} - \frac{\theta(t)}{2(1 - \theta(t))}, \quad \mathbf{v}_i(t) \in T_k^{(t)} \quad (2.44)$$

όπου ο αριθμός στοιχείων $\aleph(T_k^{(t)})$ του συνόλου $T_k^{(t)}$ έχει αντικαταστήσει τον αριθμό c στην εξίσωση (2.38). Για ένα διανυσματικό δεδομένο \mathbf{x}_k θέτονται όλοι οι αρνητικοί βαθμοί συμμετοχής ίσοι με το μηδέν, εννοώντας ότι τα αντίστοιχα διανύσματα βιβλίου κωδικοποίησης δεν επηρεάζουν αυτό το διάνυσμα πλέον, και αφαιρούνται από το σύνολο $T_k^{(t)}$. Ωστόσο, θέτοντας όλους τους αρνητικούς βαθμούς συμμετοχής ίσους με μηδέν, το άθροισμα των βαθμών συμμετοχής του \mathbf{x}_k στα διανύσματα βιβλίων κώδικα που παραμένουν στο σύνολο T_k θα είναι μεγαλύτερο από την μονάδα. Επομένως, ο περιορισμός στην εξίσωση (2.10) θα παραβιαστεί, το οποίο συνεπάγεται πως δεν μπορούμε να χρησιμοποιήσουμε την εξίσωση (2.41) για την ενημέρωση των διανυσμάτων βιβλίων κώδικα. Προκειμένου να επιλύσουμε αυτό το πρόβλημα, κανονικοποιούμε τους βαθμούς συμμετοχής σύμφωνα με την ακόλουθη σχέση,

$$w_{ik}(t) = \frac{u_{ik}(t)}{\sum_{i=1}^c u_{ik}(t)}, \quad 1 \leq k \leq n, \quad 1 \leq i \leq c \quad (2.45)$$

Στην (2.45) λάβαμε υπόψη όλα τα διανύσματα βιβλίων κώδικα, γιατί τα διανύσματα βιβλίων κώδικα που δεν ανήκουν πλέον στο σύνολο T_k δεν θα επηρεάσουν το τελικό αποτέλεσμα. Άρα, το άθροισμα αυτών των βαθμών συμμετοχής είναι πάντα ίσο με ένα.

Γι' αυτόν το σκοπό, βασισμένοι στις εξισώσεις (2.41) και (2.45), τα διανύσματα βιβλίων κώδικα μπορούν να ενημερωθούν ως εξής,

$$\mathbf{v}_i(t) = \frac{\sum_{k=1}^n [\theta(t) w_{ik}(t) + (1-\theta(t))(w_{ik}(t))^2] \mathbf{x}_k}{\sum_{k=1}^n [\theta(t) w_{ik}(t) + (1-\theta(t))(w_{ik}(t))^2]} \quad (2.46)$$

Η εξίσωση (2.46) ευθέως υπονοεί ότι σε κάθε επανάληψη, ένα διάνυσμα βιβλίων κώδικα ενημερώνεται λαμβάνοντας υπόψη τα διανυσματικά δεδομένα που εμφανίζουν κανονικοποιημένους βαθμούς συμμετοχής μεγαλύτερους από 0. Βασιζόμενοι σε αυτή την στρατηγική μετάβασης, ο αλγόριθμος ασαφούς κβάντισης διανύσματος δίδεται παρακάτω.

Αλγόριθμος Βελτιωμένης Ασαφούς Κβάντισης Διανύσματος με Εκμάθηση (IFLVQ)

Βήμα 1. Επέλεξε τον αριθμό των διανυσμάτων βιβλίων κώδικα c , τις αντίστοιχες αρχικές τιμές v_1, v_2, \dots, v_c , και μια τιμή για την παράμετρο ε . Επίσης, επέλεξε τον μέγιστο αριθμό επαναλήψεων t_{\max} , το αρχικό $\theta_0 \in [0, 1)$ και το τελικό $\theta_f \in [0, 1)$. Τέλος, θέσε $T_k = V$ και $\aleph(T_k) = c$.

Βήμα 2. Για $t = 0, 1, 2, \dots, t_{\max}$

- a. Υπολόγισε το $\theta(t) = \theta_0 - [t(\theta_0 - \theta_f)] / t_{\max}$
- b. Προσδιόρισε τις συναρτήσεις συμμετοχής χρησιμοποιώντας την εξίσωση (2.44)
- c. Για $k = 1$ έως n ; Για $i = 1$ έως c

Αν $u_{ik} \leq 0$ τότε θέσε $u_{ik} = 0$, αφάιρσε το v_i από την $T_k^{(t)}$ και θέσε

$$\aleph(T_k^{(t)}) = \aleph(T_k^{(t)}) - 1$$

Τέλος_Αν

Τέλος_Επανάληψης; Τέλος_Επανάληψης

- d. Υπολόγισε τις κανονικοποιημένες συναρτήσεις συμμετοχής χρησιμοποιώντας την εξίσωση (2.45)
- e. Ενημέρωσε τα διανύσματα βιβλίων κώδικα χρησιμοποιώντας την (2.46)
- f. Αν $E(t) = \sum_{i=1}^c \|v_i(t) - v_i(t-1)\| < \varepsilon$ τότε σταμάτα; Αλλιώς συνέχισε.

Τέλος_Αν

Τέλος_Επανάληψης

Ο παραπάνω αλγόριθμος χειρίζεται την παράμετρο θ με σκοπό να πραγματοποιήσει την μετάβαση από την ασαφή στη σαφή κατάσταση. Όπως θα δειχθεί στην μελέτη προσομοίωσης, η μέθοδος εμφανίζει μια πολύ εύρωστη συμπεριφορά ως προς την επιλογή της αρχική τιμής της παραμέτρου θ . Επιπλέον, οι αναλυτικές συνθήκες που δίδονται στις παραπάνω εξισώσεις ενισχύουν την μετάβαση αυτή με την αφαίρεση από το T_k των διανυσμάτων βιβλίων κώδικα που δεν επηρεάζουν πλέον το διάνυσμα x_k . Αυτή η στρατηγική μειώνει σημαντικά το υπολογιστικό κόστος, αφού, καθώς η διαδικασία εκπαίδευσης προχωρά ο αριθμός των υπολογισμών Ευκλείδειας απόστασης μειώνεται. Αυτό το γεγονός φαίνεται ξεκάθαρα στο κεφάλαιο 6 της παρούσης Διδακτορικής Διατριβής, όπου οι προσομοιώσεις δείχνουν ότι ο αλγόριθμος είναι γρηγορότερος από τον κλασσικό FLVQ επιτυγχάνοντας ταυτόχρονα καλύτερα ποσοστά συμπίεσης.

Ανάπτυξη Μεθόδου Εκπαίδευσης Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης με τη χρήση του αλγόριθμου Σταθμισμένων Ασαφών c-Μέσων

3.1 Εισαγωγή

Τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (Radial Basis Function Neural Networks - RBFNNs) διαμορφώνουν μια κατηγορία Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks - ANNs), η οποία έχει ορισμένα πλεονεκτήματα πέρα από άλλους τύπους ANNs, συμπεριλαμβανομένων των καλύτερων ικανοτήτων προσέγγισης, απλών δομών δικτύων, και γρηγορότερων αλγορίθμων εκπαίδευσης. Όπως ήταν αναμενόμενο, τα νευρωνικά δίκτυα RBF γίνονται όλο και περισσότερο δημοφιλή σε πολλές επιστημονικές περιοχές [47,50,55,62]. Οι αλγόριθμοι εκπαίδευσης νευρωνικών δικτύων RBF διακρίνονται σε δύο βασικές κατηγορίες: (i) αλγόριθμοι όπου ο αριθμός κρυφών κόμβων προκαθορίζεται και (ii) αλγόριθμοι που περιλαμβάνουν τους μηχανισμούς επιλογής δομών [83].

Η διαδικασία εκπαίδευσης του Γκαουσιανού τύπου (Gaussian type) νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBF) είναι βασισμένη στη βελτιστοποίηση τριών ειδών παραμέτρων: κέντρα (centers), πλάτη (widths) και συναπτικά βάρη (neuro-connecting weights) μεταξύ των νευρώνων. Συνήθως, η εκτίμηση των ανωτέρω παραμέτρων πραγματοποιείται χρησιμοποιώντας δύο επιπέδων στρατηγικές εκπαίδευσης. Στο πρώτο στάδιο, εφαρμόζονται αλγόριθμοι συσταδοποίησης για να υπολογίσουν τις κατάλληλες τιμές των κέντρων και των πλατών (widths) ενώ στο

δεύτερο στάδιο, εποπτευμένες διαδικασίες βελτιστοποίησης, όπως η μέθοδος της καθοδικής κλίσης (gradient descent method) περιλαμβάνεται στη βέλτιστη εκτίμηση των συναπτικών βαρών [2,24,28,38,54].

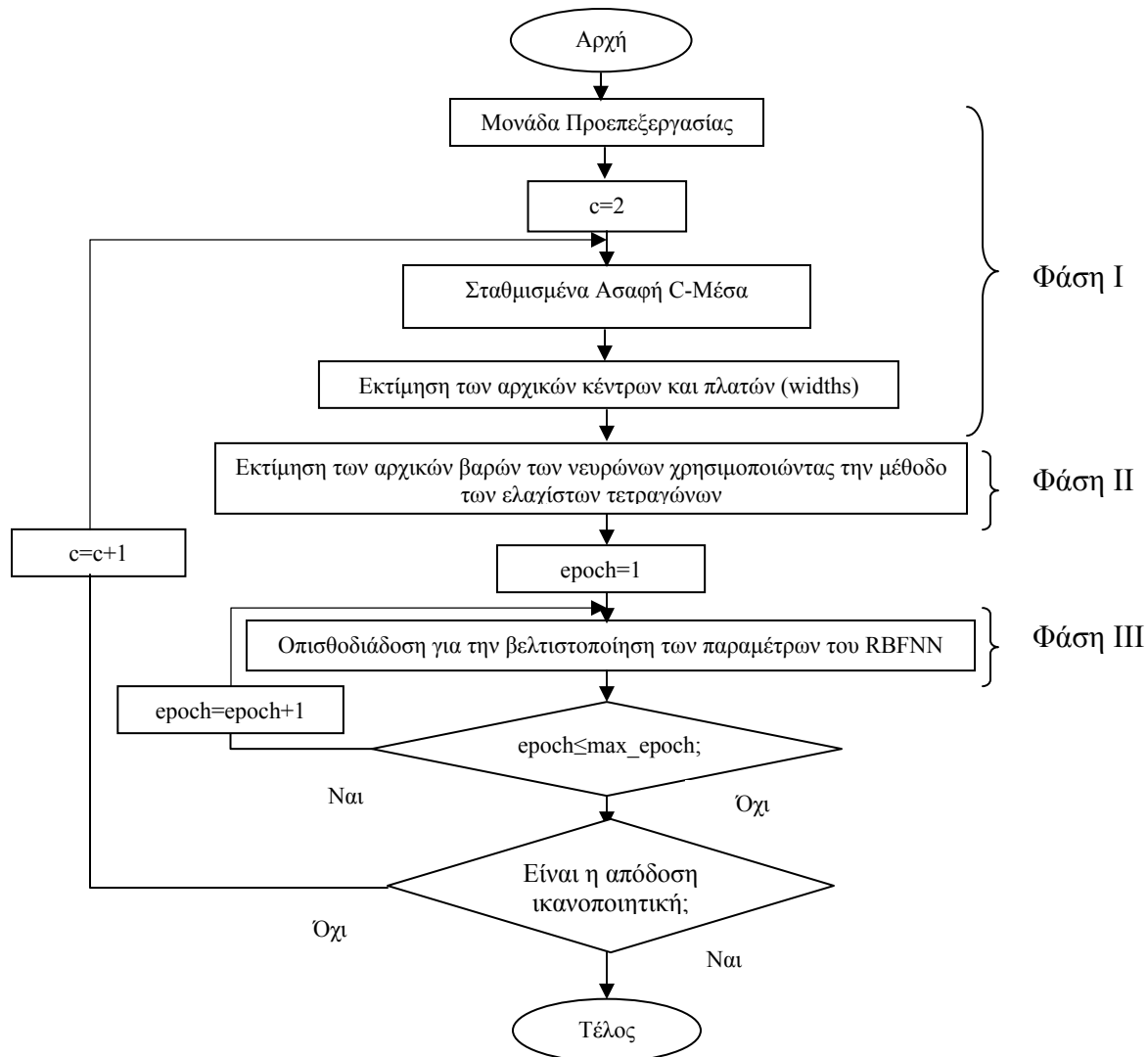
Μέχρι τώρα, η ασαφής συσταδοποίηση (fuzzy clustering) εμφανίζεται να είναι ένα πολύ αποδοτικό εργαλείο για να υπολογίσει τις παραμέτρους ενός γκαουσιανού τύπου (Gaussian type) νευρωνικού δικτύου RBF, δεδομένου ότι είναι σε θέση να ανιχνεύσει τις σημαντικές δομές στο διαθέσιμο σύνολο δεδομένων, που υιοθετεί τις Γκαουσιανές συναρτήσεις συμμετοχής (Gaussian membership functions) αρχικά [32,72,89].

Στο παρόν κεφάλαιο, προτείνεται μια νέα μέθοδος [64] για την εκπαίδευση Γκαουσιανού τύπου (Gaussian type) νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBFNN) με τη χρησιμοποίηση σταθμισμένου (weighted) αλγορίθμου Ασαφών c -Μέσων (FCM). Πιο συγκεκριμένα, ο προτεινόμενος αλγόριθμος χρησιμοποιεί αρχικά τη γνωστή μέθοδο Ασαφών c -Μέσων (FCM) προκειμένου να ομαδοποιηθούν όλα τα δεδομένα εκπαίδευσης σε διάφορες συστάδες. Κατόπιν, σε κάθε κέντρο συστάδων ορίζεται μια τιμή βάρους, η οποία καθορίζεται από τον ασαφή αριθμό δεδομένων του συνόλου της αντίστοιχης συστάδας. Το σύνολο αυτών των κέντρων συστάδων μαζί με τα βάρη τους, συσταδοποιείται περαιτέρω χρησιμοποιώντας τον σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM) [91]. Κατά συνέπεια, αντίθετα από την προσέγγιση που αναπτύσσεται στην αναφορά [52], στον προτεινόμενο αλγόριθμο η συμβολή κάθε κέντρου συστάδων στο τελικό ασαφή διαμερισμό αποφασίζεται από το αντίστοιχο βάρος δεδομένου ότι δύο διαφορετικά κέντρα συστάδων των διαφορετικών μεγεθών δεν πρέπει να ζυγίσουν εξίσου όσον αφορά το τελικό διαμερισμό. Με τη διαμόρφωση των ασαφών συστάδων, που παράγονται από τον αλγόριθμο σταθμισμένων Ασαφών c -Μέσων (weighted FCM), οι αρχικές εκτιμήσεις των κέντρων και τα πλάτη λαμβάνονται. Αυτές οι εκτιμήσεις χρησιμοποιούνται για τον υπολογισμό των βαρών σύνδεσης μεταξύ των κρυφών κόμβων. Τέλος, ολόκληρο το σύνολο των παραμέτρων δικτύου βελτιστοποιείται περαιτέρω από τον αλγόριθμο βελτιστοποίησης οπισθοδιάδοσης (back-propagation).

3.2 Ο προτεινόμενος αλγόριθμος

Ο προτεινόμενος αλγόριθμος χρησιμοποιεί τρεις φάσεις εκπαίδευσης και το διάγραμμα ροής του απεικονίζεται στο σχήμα 3.1. Πριν αναφερθεί η θεωρητική περιγραφή κάθε βήματος, εξετάζεται εν συντομία ολόκληρος ο αλγοριθμικός σχεδιασμός και παρατίθενται ορισμένα ζητήματα σχεδιασμού του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBF).

Η φάση I χρησιμοποιεί τρία βήματα για να εξαγάγει τις αρχικές τιμές για τα κέντρα και τα πλάτη των κρυφών κόμβων δικτύου του νευρωνικού δικτύου RBF. Σε πρώτη φάση, το διαθέσιμο σύνολο δεδομένων προεπεξεργάζεται χρησιμοποιώντας το γνωστό αλγόριθμο Ασαφών c -Μέσων (FCM). Στο δεύτερο βήμα, τα κέντρα συστάδων που εξήχθησαν από τη FCM θεωρούνται το νέο σύνολο δεδομένων, το οποίο συσταδοποιείται περαιτέρω με τη χρησιμοποίηση του σταθμισμένου (weighted) αλγορίθμου Ασαφών c -Μέσων (FCM). Η αιτιολόγηση των ανωτέρω επιλογών δίνεται αργότερα σε αυτό το κεφάλαιο. Στο τρίτο βήμα, βασισμένο στο προκύπτον ασαφή διαμερισμό, εκτελείται μια αρχική εκτίμηση των κέντρων και των πλατών (widths) των κρυφών κόμβων του νευρωνικού δικτύου RBF. Η φάση II ασχολείται με την αρχική εκτίμηση παραμέτρων, των βαρών σύνδεσης νευρώνων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης, η οποία ολοκληρώνεται με τη μέθοδο των ελαχίστων τετραγώνων. Τέλος, η φάση III εκτελεί μια διαδικασία βελτιστοποίησης οπισθοδιάδοσης (back-propagation) για να τελειοποιήσει όλες τις παραμέτρους του νευρωνικού δικτύου RBF. Οι ανωτέρω φάσεις εφαρμόζονται επαναληπτικά, όπου σε κάθε επανάληψη ο αριθμός κόμβων αυξάνεται κατά ένα έως ότου βρεθεί η ακρίβεια προσέγγισης μέσα σε αποδεκτά επίπεδα. Σε αυτό το σημείο, πρέπει να τονιστεί ότι ο αριθμός των κρυφών κόμβων του νευρωνικού δικτύου RBF είναι ίσος με τον αριθμό συστάδων που παράγονται από το σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM). Κατά συνέπεια, αρχίζοντας από δύο συστάδες (κόμβοι) λαμβάνεται τελικά ο αριθμός κόμβων για τον οποίο η απόδοση του δικτύου είναι ακριβής. Αυτή η παρατήρηση άμεσα υπονοεί ότι ο τελικός αριθμός κόμβων θα είναι μικρός αλλά ακόμα αρκετά αξιόπιστος να παραγάγει ένα αποδοτικό νευρωνικό δίκτυο RBF.



Σχήμα 3.1. Το διάγραμμα του προτεινόμενου αλγορίθμου.

Για να κατανοηθεί καλύτερα τη σχέση μεταξύ των βημάτων σχεδιασμού του παραπάνω αλγορίθμου, αναλύονται λεπτομερέστερα τα κύρια κριτήρια που χρησιμοποιήθηκαν στο παρόν κεφάλαιο. Το πρώτο κριτήριο, παραπέμπεται στην επιλογή της χρησιμοποίησης του αλγορίθμου των Ασαφών c -Μέσων (FCM). Ο αλγόριθμος των Ασαφών c -Μέσων (FCM) έχει αναγνωριστεί ως πολύ αποδοτικό εργαλείο για να ανακαλύψει τις δομές που κρύβονται στα στοιχεία του νευρωνικού δικτύου δεδομένου ότι οι παράμετροι των κόμβων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης εξαρτώνται έντονα από αυτές τις δομές. Η χρήση των Ασαφών c -Μέσων (FCM) στην αρχή του αλγορίθμου θα παράσχει ορισμένα οφέλη σχετικά με την

επιλογή μιας "καλής" αφετηρίας για τη διαδικασία εκπαίδευσης του νευρωνικού δικτύου. Ο αλγόριθμος των Ασαφών c -Μέσων χρησιμοποιείται για να προεπεξεργαστούν τα δεδομένα εκπαίδευσης και επομένως, αυτό το βήμα δεν ενδιαφέρεται για το σχεδιασμό ενός αυστηρού αλγορίθμου Ασαφών c -Μέσων (FCM). Ο μόνος περιορισμός εδώ είναι ότι ο αριθμός των συστάδων πρέπει να είναι αρκετά μεγάλος, έτσι ώστε οι περιπτώσεις εκπαίδευσης στο σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM), που ακολουθεί, να είναι επαρκή για να επιτύχουν ένα αξιόπιστο αποτέλεσμα συσταδοποίησης.

Το δεύτερο κριτήριο αναφέρεται στο σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM). Σε παρόμοιες προσεγγίσεις [52] χρησιμοποίησαν τον κλασσικό αλγόριθμο Ασαφών c -Μέσων (FCM) για να διαμορφωθούν τα κέντρα συστάδων που παρήχθησαν από τη μονάδα προεπεξεργασίας. Προφανώς, η εφαρμογή του αλγορίθμου Ασαφών c -Μέσων υποθέτει ότι όλα αυτά τα κέντρα είναι ίσης σημασίας. Εντούτοις, αυτό μπορεί να μην ισχύει, επειδή δύο διαφορετικές αρχικές συστάδες μπορούν να περιέχουν διαφορετικό αριθμό περιπτώσεων εκπαίδευσης. Κατά συνέπεια, με την εφαρμογή του αλγορίθμου Ασαφών c -Μέσων (FCM) διαγράφονται όλες οι πληροφορίες, οι οποίες συσχετίζουν τα αρχικά δεδομένα με τα ανωτέρω κέντρα. Επομένως, η πραγματική δομή δεδομένων δεν μπορεί να ανακαλυφθεί. Για να επιλυθεί αυτό το πρόβλημα ορίζονται τα βάρη στα κέντρα, και διαμορφώνονται χρησιμοποιώντας τη σταθμισμένη έκδοση του αλγορίθμου Ασαφών c -Μέσων (FCM). Κάθε ένα από τα ανωτέρω βάρη ορίζεται ως ο ομαλοποιημένος αριθμός δεδομένων συνόλου της αντίστοιχης αρχικής συστάδας. Κατ' αυτό τον τρόπο, όλες οι απαραίτητες πληροφορίες σχετικές με το αρχικό σύνολο στοιχείων λαμβάνονται υπόψη κατά τη διάρκεια της διαδικασίας σχεδιασμού ολόκληρου του προτύπου.

Το τρίτο κριτήριο αφορά τη χρήση του αλγορίθμου οπισθοδιάδοσης (back-propagation). Ο βασικός αλγόριθμος οπισθοδιάδοσης βρίσκει συχνά το τοπικό ελάχιστο. Είναι επίσης ευαίσθητο στον ορισμό των παραμέτρων. Εντούτοις, πολλοί ερευνητές [38,52,91] χρησιμοποιούν τον αλγόριθμο οπισθοδιάδοσης (back-propagation) εξαιτίας δύο κύριων λόγων: (α) ο αλγόριθμος οπισθοδιάδοσης (back-propagation) έχει μια γρήγορη σύγκλιση, και (β) παρέχοντας τους καλούς αρχικούς όρους στον αλγόριθμο οπισθοδιάδοσης (back-propagation) που είναι σε θέση να επιτύχει πολύ καλά

αποτελέσματα. Ως εκ τούτου, το βασικό σημείο όλων αυτών των προσεγγίσεων, συμπεριλαμβανομένης της προσέγγισής μας εδώ, είναι να παρασχεθούν πολύ καλή αρχικοί όροι στον αλγόριθμο οπισθοδιάδοσης (back-propagation).

3.2.1 Φάση I

Έστω $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ένα σύνολο N διανυσμάτων δεδομένων χαρακτηριστικών γνωρισμάτων με $\mathbf{x}_k \in \mathcal{R}^p$ ($1 \leq k \leq N$), και n ($2 \leq n \leq N$) είναι διάφορες ασαφείς συστάδες. Δεδομένου ότι η συνάρτηση συμμετοχής του k -διανύσματος στην i -συστάδα δίνεται ως $u_{ik} = \{u_i(\mathbf{x}_k), 1 \leq i \leq c, 1 \leq k \leq n\}$ και η παράμετρος ασάφειας με m , ο αλγόριθμος Ασαφών c -Μέσων (FCM) ελαχιστοποιεί την ακόλουθη αντικειμενική συνάρτηση κόστους,

$$J_m = \sum_{k=1}^N \sum_{i=1}^n (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (3.1)$$

κάτω από τον επόμενο περιορισμό,

$$\sum_{i=1}^n u_{ik} = 1 \quad (3.2)$$

Τα κέντρα συστάδων και οι βαθμοί συμμετοχής που λύνουν το ανωτέρω πρόβλημα βελτιστοποίησης δίνονται ως,

$$\mathbf{v}_i = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m}, \quad 1 \leq i \leq n \quad (3.3)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|_A}{\|\mathbf{x}_k - \mathbf{v}_j\|_A} \right)^{\frac{2}{m-1}}} \quad (3.4)$$

Κατόπιν, ο αλγόριθμος Ασαφών c -Μέσων (FCM) εκτελείται επαναληπτικά εφαρμόζοντας τις ανωτέρω δύο εξισώσεις. Στο επόμενο βήμα, τα κέντρα των συστάδων

v_i ($1 \leq i \leq n$) θεωρούνται ως νέο σύνολο δεδομένων. Επειδή το διάνυσμα v_i αναφέρεται σε μια συστάδα του αρχικού συνόλου δεδομένων, η βαθμός σημασίας της i -οστής συστάδας καθορίζεται ως εξής,

$$\rho_i = \frac{\aleph(U^i)}{\sum_i^n \aleph(U^i)} \quad (3.5)$$

όπου $\aleph(U^i)$ είναι η ασαφής πληθικότητα συνόλου (cardinality) της i -οστής συστάδας

$$\aleph(U^i) = \sum_{k=1}^N u_{ik} \quad (1 \leq i \leq n) \quad (3.6)$$

Κατόπιν, η αντικειμενική συνάρτηση του σταθμισμένου (weighted) αλγορίθμου Ασαφών c -Μέσων (FCM) είναι,

$$J_r = \sum_{i=1}^n \sum_{j=1}^c \rho_i (\mu_{ji})^m \|v_i - v_j\|^2 \quad (3.7)$$

όπου το c είναι ο αριθμός των τελικών συστάδων, ο οποίος συμπίπτει με τον αριθμό των κρυφών κόμβων του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBF), v_j ($1 \leq j \leq c$) είναι τα κέντρα των τελικών συστάδων, $m \in (1, \infty)$ είναι ένας παράγοντας για να ρυθμίσει το βαθμό ιδιότητας μέλους σταθμίζοντας την επίδραση και ρ_i είναι το βάρος μεγάλης σημασίας που ορίζεται για το διάνυσμα v_i .

Το πρόβλημα είναι να ελαχιστοποιηθεί η J_r κάτω από τον ακόλουθο περιορισμό,

$$\sum_{j=1}^c \mu_{ji} = 1 \quad \forall i \quad (3.8)$$

Τα τελικά κέντρα συστάδων και οι αντίστοιχες συναρτήσεις συμμετοχής που λύνουν αυτό το περιορισμένο πρόβλημα βελτιστοποίησης δίνονται από τις επόμενες εξισώσεις,

$$\mathbf{v}_j = \frac{\sum_{i=1}^n \rho_i (\mu_{ji})^m \mathbf{v}_i}{\sum_{i=1}^n \rho_i (\mu_{ji})^m} \quad (3.9)$$

$$\mu_{ji} = \frac{1}{\sum_{l=1}^c \left(\frac{\|\mathbf{v}_i - \mathbf{v}_j\|}{\|\mathbf{v}_i - \mathbf{v}_l\|} \right)^{2/(m-1)}} \quad (3.10)$$

Οι εξ. (3.9) και (3.10) αποτελούν μια επαναληπτική διαδικασία βελτιστοποίησης που βελτιστοποιεί την αντικειμενική συνάρτηση εξ. (3.7). Στην περίπτωση όπου όλα τα βάρη είναι ίσα, ο σταθμισμένος αλγόριθμος Ασαφών c -Μέσων είναι ισοδύναμος με τον κλασσικό αλγόριθμο Ασαφών c -Μέσων.

Για αυτόν τον λόγο, ο αριθμός κρυφών κόμβων στο νευρωνικό δίκτυο RBF είναι ίσος με τον αριθμό συστάδων c , ενώ τα κέντρα αυτών των κρυφών κόμβων είναι τα ίδια στα κέντρα συστάδων \mathbf{v}_j ($1 \leq j \leq c$). Αφ' ετέρου, τα πλάτη (widths) των κρυφών κόμβων του νευρωνικού δικτύου RBF υπολογίζονται ως εξής,

$$\sigma^j = \frac{2 * d_{max}^j}{3} \quad (1 \leq j \leq c) \quad (3.11)$$

όπου στην (3.11) το d_{max} δίνεται στην επόμενη εξίσωση,

$$d_{max} = \{ \max_k \{ \|\mathbf{x}_k - \mathbf{v}_j\|^2 \} : \mathbf{x}_k \in C_j \text{ έτσι ώστε } u_{jk} \geq \theta \} \quad (1 \leq j \leq c) \quad (3.12)$$

όπου C_j είναι η j - συστάδα και u_{jk} είναι ο βαθμός συμμετοχής του k - διανύσματος δεδομένων εκπαίδευσης στη j - τελική ασαφή συστάδα, και θ ένας μικρός θετικός αριθμός έτσι ώστε $\theta \in (0,1)$.

3.2.2 Φάση II

Δεδομένου ότι η έξοδος του νευρωνικού δικτύου RBF είναι γραμμική σε βάρη w_j , αυτά τα βάρη μπορούν να υπολογιστούν με μεθόδους ελαχίστων τετραγώνων (least-squares). Για κάθε δεδομένο εισόδου \mathbf{x}_k , οι έξοδοι του νευρώνα υπολογίζονται ως εξής:

$$s_{kj} = g_j(\mathbf{x}_k) \quad (3.13)$$

όπου $g_j(\mathbf{x}_k)$ είναι η Γκαουσιανή συνάρτηση (Gaussian function). Εισάγοντας το διάνυσμα βάρους $\mathbf{w} = [w_1, w_2, \dots, w_c]$ μπορεί να γραφεί η ακόλουθη εξίσωση πινάκων για ολόκληρο το σύνολο δεδομένων:

$$\tilde{\mathbf{y}} = \mathbf{A}\mathbf{w} \quad (3.14)$$

όπου $\mathbf{A} = [s_{kj}]$. Τέλος, με χρήση της μεθόδου των ελαχίστων τετραγώνων (least squares) γίνεται η εκτίμηση των βαρών \mathbf{w} που ελαχιστοποιούν το σφάλμα δικτύου $e = \tilde{\mathbf{y}} - \mathbf{y}$ και ο υπολογισμός δίνεται από την ακόλουθη εξίσωση:

$$\mathbf{w} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \tilde{\mathbf{y}} \quad (3.15)$$

3.2.3 Φάση III

Σε αυτό το βήμα, οι αρχικές εκτιμήσεις των παραμέτρων που υπολογίζονται στα προηγούμενα βήματα τροφοδοτούν τον αλγόριθμο οπισθοδιάδοσης (back-propagation). Όσον αφορά ένα συγκεκριμένο ζευγάρι δεδομένων εισόδου-εξόδου $(\mathbf{x}_k; y_k)$, ο σκοπός είναι να ελαχιστοποιηθεί η ακόλουθη συνάρτηση σφάλματος:

$$e = \frac{1}{2} (\tilde{y}_k - y_k)^2 \quad (1 \leq k \leq N) \quad (3.16)$$

όπου y_k και \tilde{y}_k είναι η πραγματική και προβλεφθείσα έξοδος του νευρωνικού δικτύου, αντίστοιχα.

Η εκπαίδευση σε ένα νευρωνικό δίκτυο συναρτήσεων ακτινικής βάσης (RBF) ολοκληρώνεται με τη χρησιμοποίηση των επόμενων δύο σταδίων:

1. Ρυθμίζοντας τις παραμέτρους του κρυφού στρώματος του νευρωνικού δικτύου RBF, συμπεριλαμβανομένων των κέντρων RBF $\mathbf{v}_j \in R^p$ ($1 \leq j \leq c$) και των πλατών (widths) σ_j ($1 \leq j \leq c$). Το p είναι η διάσταση του διανύσματος εισόδου.
2. Υπολογισμός των βαρών εξόδου w_j ($j = 1, \dots, \gamma$) του δικτύου.

Οι κανόνες εκμάθησης για τα ανωτέρω στάδια είναι:

$$w_j^l(t+1) = w_j^l(t) - e * \frac{\partial e}{\partial w_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (3.17)$$

$$\mathbf{v}_j^l(t+1) = \mathbf{v}_j^l(t) - e * \frac{\partial e}{\partial \mathbf{v}_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (3.18)$$

$$\sigma_j^l(t+1) = \sigma_j^l(t) - e * \frac{\partial e}{\partial \sigma_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (3.19)$$

όπου το e είναι η παράμετρος οπισθοδιάδοσης (back-propagation) εκμάθησης. Τέλος, βασισμένες στις εξισώσεις (3.14) και (3.16), οι μερικές παράγωγοι (3.17), (3.18) και (3.19) μπορούν εύκολα να υπολογιστούν.

3.3 Πειράματα προσομοίωσης

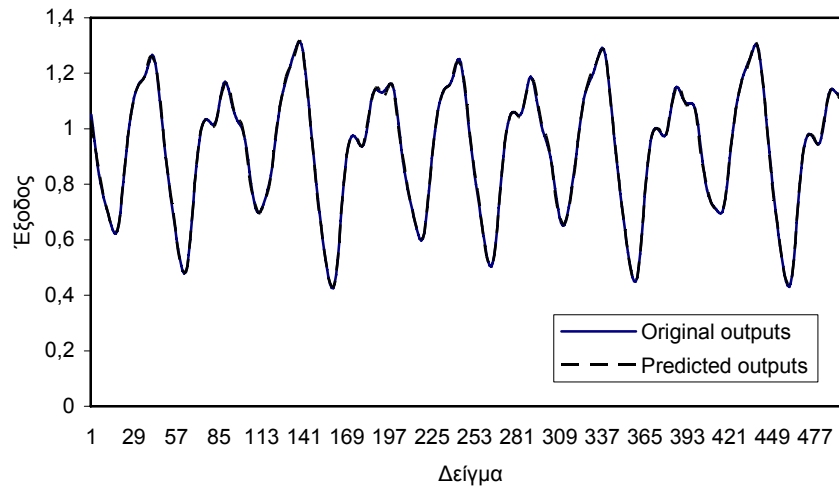
3.3.1 Σύστημα Mackey-Glass

Σε αυτήν την υποενότητα χρησιμοποιούμε τον προτεινόμενο αλγόριθμο προκειμένου να προβλεφθεί η χρονοσειρά Mackey-Glass. Η χρονοσειρά Mackey-Glass παράγεται από την ακόλουθη διαφορική εξίσωση χρονικής καθυστέρησης,

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (3.20)$$

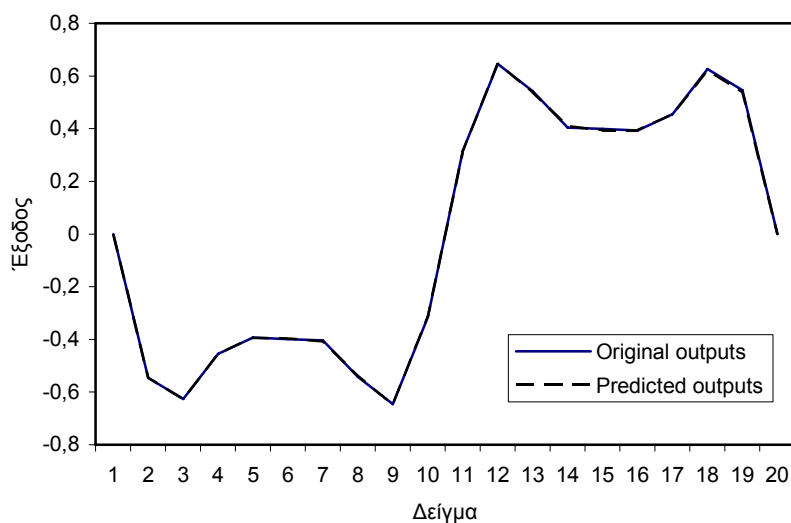
Όταν η παράμετρος τ είναι αρκετά μεγάλη το σύστημα εμφανίζει μια χαοτική συμπεριφορά. Στις προσομοιώσεις τέθηκε το $\tau = 17$ και δημιουργήθηκε ένα δείγμα 1000

σημείων. Τα πρώτα 500 σημεία χρησιμοποιήθηκαν σαν δεδομένα εκπαίδευσης, και τα τελευταία 500 σημεία ως δεδομένα δοκιμής για να επικυρώσουν το νευρωνικό δίκτυο RBF. Ο προτεινόμενος αλγόριθμος εφαρμόστηκε για να δημιουργήσει ένα νευρωνικό δίκτυο RBF με 4 εισόδους: $x(t-18)$, $x(t-12)$, $x(t-6)$ και $x(t)$, ενώ η έξοδος είναι το σημείο $x(t+6)$.



Σχήμα 3.2. Η αρχική και προβλεφθείσα έξοδος αντίστοιχα για το παράδειγμα Mackey-Glass. (Δεδομένα εκπαίδευσης).

Οι παράμετροι για τον αλγόριθμο Ασαφών c -Μέσων και το σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM) επιλέχθηκαν όπως: $m=2$, $n=20$, $\text{μέγιστος_αριθμός_επαναλήψεων}=30000$ και $\text{μικρότερο_ποσό_βελτίωσης} = 10^{-5}$. Ο τελικός αριθμός κρυφών κόμβων ήταν ίσος με $c = 14$. Για τον αλγόριθμο οπισθοδιάδοσης το ποσοστό εκπαίδευσης ήταν ίσο με $\varepsilon = 0.001$, και ο μέγιστος αριθμός επανάληψης (max_epoch) ήταν ίσος με 30000. Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) για τα δεδομένα εκπαίδευσης ήταν ίσο με 0.00517 και 0.00527 για τα δεδομένα δοκιμής. Η έξοδος για τα δεδομένα εκπαίδευσης παρουσιάζεται στο σχήμα 3.2.



Σχήμα 3.3. Η αρχική και προβλεφθείσα έξοδος αντίστοιχα για το συνθετικό μονοδιάστατο παράδειγμα δεδομένων (δεδομένα δοκιμής).

Πίνακας 3.1. Σύγκριση αποτελεσμάτων για τα δεδομένα του παραδείγματος Mackey-Glass

Πρότυπο	Αριθμός κρυφών κόμβων	RMSE
Αυτοανάδρομο πρότυπο (Autoregressive model) [19]	---	0.19
Συσχετισμός NN καταρρακτών (Cascade correlation NN) [19]	---	0.06
Οπισθοδιάδοση (back-propagation) NN [19]	---	0.02
Πολυώνυμο έκτου-βαθμού (Sixth-order polynomial) [19]	---	0.04
Γραμμική μέθοδος πρόβλεψης (Linear prediction method) [19]	---	0.55
Cho , Wang [21]	(Δεδομένα εκπαίδευσης) 23 (Δεδομένα δοκιμής) 23	0.0096 0.0114
Kim D., Kim C. [41]	---	0.049
Chen et al. [19]	(Δεδομένα εκπαίδευσης) --- (Δεδομένα δοκιμής) ---	0.0069 0.0071
Chen et al.[20]	---	0.0076
Jang et al. [37]	16	0.0070
Chen et al. [18]	(Δεδομένα εκπαίδευσης) 21 (Δεδομένα δοκιμής) 21	0.0107 0.0128
Προτεινόμενο δίκτυο RBF	(Δεδομένα εκπαίδευσης) 14 (Δεδομένα δοκιμής) 14	0.00517 0.00527

Τέλος, ο πίνακας 3.1 συγκρίνει διαφορετικά υπολογιστικά παραδείγματα που βρίσκονται στη διεθνή βιβλιογραφία με την προτεινόμενη διαδικασία εκπαίδευσης του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης. Σύμφωνα με αυτόν τον πίνακα, ο προτεινόμενος αλγόριθμος έδωσε το καλύτερο αποτέλεσμα με το μικρότερο αριθμό κρυφών κόμβων.

3.3.2 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης

Εξετάστηκε η ακόλουθη συνάρτηση μίας μεταβλητής,

$$y = 0.6 \eta(\pi x) + 0.3 \eta(3\pi x) + 0.1 \eta(5\pi x) \quad (3.21)$$

όπου η είσοδος x ορίζεται στο διάστημα $[-1, 1]$.

Οι παράμετροι για τον αλγόριθμο Ασαφών c -Μέσων (FCM) και το σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM) επιλέχθηκαν όπως: $m = 2$, $n = 20$, $\text{μέγιστος_αριθμός_επαναλήψεων} = 30000$ και $\text{μικρότερο_ποσό_βελτίωσης} = 10^{-5}$. Η εφαρμογή του αλγορίθμου έδωσε $c = 9$ κρυφούς κόμβους.

Για τον αλγόριθμο βελτιστοποίησης οπισθοδιάδοσης (back-propagation) το ποσοστό εκπαίδευσης ήταν ίσο με $\varepsilon = 0.015$, και ο μέγιστος αριθμός επανάληψης (max_epoch) για την εκπαίδευση του νευρωνικού δικτύου RBF ήταν ίσο με 30000. Η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE) για τα δεδομένα εκπαίδευσης ήταν ίσο με 0.0036 και για τα δεδομένα δοκιμής 0.0041. Η έξοδος για τα δεδομένα εκπαίδευσης παρουσιάζεται στο σχήμα 3.3. Από τα πειραματικά αποτελέσματα σαφώς προκύπτει ότι η προτεινόμενη αρχιτεκτονική νευρωνικού δικτύου RBF πετυχαίνει πολύ καλύτερα αποτελέσματα από τα υπάρχοντα δίκτυα.

Πίνακας 3.2. Συγκριτικά αποτελέσματα για το παράδειγμα προσέγγισης συνάρτησης

Πρότυπο	Αριθμός κρυφών κόμβων	RMSE	MSE
Κλασικό RBF (δεδομένα δοκιμής)	28	0.019	0.1378
Προτεινόμενο RBF (δεδομένα εκπαίδευσης)	14	0.0036	0.060
Προτεινόμενο RBF (δεδομένα δοκιμής)	14	0.0041	0.064

Τέλος, ο πίνακας 3.2 συγκρίνει τις αποδόσεις που λαμβάνονται σε αυτό το παράδειγμα με την απόδοση του κλασσικού RBFNN. Αυτός ο πίνακας δείχνει ότι η απόδοση του κλασσικού RBFNN είναι κατώτερη όταν συγκρίνεται με την προτεινόμενη μεθοδολογία.

3.4 Συμπεράσματα

Σε αυτό το κεφάλαιο, αναπτύχθηκε και αξιολογήθηκε μια νέα μέθοδος βασισμένη στην ασαφή συσταδοποίηση όσον αφορά τη διαδικασία εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Ο αλγόριθμος χρησιμοποιεί τρεις φάσεις εκπαίδευσης για να υπολογίσει τις κατάλληλες τιμές για τις παραμέτρους δικτύου. Στην πρώτη φάση εκτελεί μια δύο σταδίων ασαφή συσταδοποίηση (fuzzy clustering) κατά τέτοιο τρόπο ώστε η υποκείμενη δομή δεδομένων μπορεί να ανακαλυφθεί αποτελεσματικά. Στη δεύτερη φάση, τα βάρη σύνδεσης νευρώνων υπολογίζονται και τελικά, στην τρίτη φάση, όλες οι παράμετροι του δικτύου βελτιστοποιούνται με την εφαρμογή του αλγορίθμου οπισθοδιάδοσης. Ο προτεινόμενος αλγόριθμος προσφέρει δύο σημαντικά πλεονεκτήματα:

1. Η διαδικασία εκπαίδευσης ελέγχεται από τον αριθμό κόμβων. Δηλαδή σε κάθε επανάληψη ο αριθμός κρυφών κόμβων αυξάνεται μέχρι η απόδοση του νευρωνικού δικτύου να είναι αρκετά ακριβής. Αυτό το γεγονός άμεσα υπονοεί ότι η διαδικασία θα σταματήσει μόλις ανιχνεύεται ο μικρότερος αξιόπιστος αριθμός κόμβων διατηρώντας συγχρόνως το μικρότερο αριθμό επαναλήψεων.
2. Η εκτίμηση των παραμέτρων των κρυφών κόμβων δεν εξαρτάται από οποιαδήποτε αρχική τυχαία επιλογή των τιμών τους αλλά μάλλον είναι απλό αποτέλεσμα της διαδικασίας συσταδοποίησης.

Διάφορες προσομοιώσεις εκτελέστηκαν, οι οποίες έδειξαν ότι η μεθοδολογία μπορεί να είναι πολύ επιτυχής όταν εφαρμόζεται στη μοντελοποίηση συστημάτων.

Μέθοδος Βέλτιστης Συσταδοποίησης για την εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

4.1 Εισαγωγή

Η διαδικασία εκπαίδευσης των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF) περιλαμβάνει δύο κύρια ζητήματα: (α) προσδιορισμός του κατάλληλου αριθμού κρυφών κόμβων, και (β) η βέλτιστη εκτίμηση των παραμέτρων πυρήνων των συναρτήσεων ακτινικής βάσης.

Τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης δεν επιτυγχάνουν κανένα λάθος εάν έχουν τόσους κρυφούς κόμβους όσα τα δεδομένα εκπαίδευσης. Εντούτοις, αυτό είναι μη πρακτικό για τις εφαρμογές με έναν μεγάλο αριθμό δεδομένων, δεδομένου ότι το μέγεθος του δικτύου γίνεται μεγάλο [5,15]. Αφ' ετέρου, ένα μικρό δίκτυο RBF μπορεί να έχει την καλύτερη δυνατότητα γενίκευσης, αλλά τα λάθη γίνονται μεγαλύτερα [104]. Επομένως, μια ισορροπία μεταξύ του μεγέθους του δικτύου και των ικανοτήτων διαμόρφωσης πρέπει να ληφθεί υπόψη κατά τη διάρκεια της διαδικασίας εκπαίδευσης του νευρωνικού δικτύου. Σύμφωνα με τους Katayama et al. [40] αυτή η ισορροπία μπορεί να επιτευχθεί χρησιμοποιώντας μια μέγιστη κατ' απόλυτη τιμή βασισμένη στο σφάλμα για την επιλογή κρυφών κόμβων του νευρωνικού δικτύου. Εντούτοις, αυτή η διαμόρφωση δεν μπορεί να περιλάβει όλες τις πληροφορίες που παρέχονται από τα δεδομένα εκπαίδευσης στην τοπολογία δικτύων. Οι Murata et al. [61] πρότειναν μια επαναληπτική μέθοδο που αρχίζει με έναν μεγάλο αριθμό κρυφών κόμβων, ο οποίος μειώνεται βαθμιαία στην πορεία της διαδικασίας εκπαίδευσης. Στην αναφορά [13] αναπτύχθηκε μια εξελικτική μέθοδος για να καθοριστεί η δομή του νευρωνικού δικτύου.

Η εκτίμηση των παραμέτρων των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης αφορά τη βελτιστοποίηση τριών ειδών παραμέτρων: κέντρα και πλάτη (widths) των πυρήνων καθώς επίσης και τα βάρη (weights) σύνδεσης μεταξύ των νευρώνων. Συνήθως, η εκτίμηση των ανωτέρω παραμέτρων πραγματοποιείται χρησιμοποιώντας τις δύο-επιπέδων στρατηγικές εκπαίδευσης [86]. Στο πρώτο στάδιο, η ανάλυση συστάδων (cluster analysis) εφαρμόζεται για να υπολογίσει τις κατάλληλες τιμές των κέντρων (centers) και των πλατών (widths), ενώ στο δεύτερο στάδιο εφαρμόζονται εποπτευμένες διαδικασίες βελτιστοποίησης, όπως η μέθοδος της καθοδικής κλίσης (gradient-descent method), για τον βέλτιστο υπολογισμό των συναπτικών βαρών [28,38,54,87,103]. Όσον αφορά την εκτίμηση των παραμέτρων των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBFNN), η ασαφής συσταδοποίηση (fuzzy clustering) είναι ένα πολύ σημαντικό εργαλείο, δεδομένου ότι είναι σε θέση να ανιχνεύσει τις σημαντικές δομές στο σύνολο δεδομένων, που υιοθετεί τις συναρτήσεις ιδιότητας μέλους πυρήνων αρχικά, όπως οι γκαουσιανές συναρτήσεις [21,30,72,75,89].

Και τα δύο από τα ανωτέρω ζητήματα μπορούν να επιλυθούν με τη χρησιμοποίηση των κατάλληλων τεχνικών συσταδοποίησης, όπως η βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering). Η βέλτιστη ασαφής συσταδοποίηση λαμβάνει υπόψη ορισμένα κριτήρια ισχύος συστάδων προκειμένου να ανιχνευθεί ένας καλός διαμοιρασμός συστάδων [67,99]. Αυτός ο διαμερισμός συσχετίζεται άμεσα με την πραγματική υποκείμενη δομή δεδομένων, δεδομένου ότι το πρώτο είναι ένα αποτέλεσμα του πιο πρόσφατου. Επομένως, η εφαρμογή της βέλτιστης ασαφής συσταδοποίησης (optimal fuzzy clustering) επιτρέπει τον προσδιορισμό του αριθμού συστάδων καθώς επίσης και μεμονωμένων χαρακτηριστικών τους, όπως τα κέντρα και τα αντίστοιχα πλάτη (widths). Για αυτόν τον λόγο, η ανάθεση είναι προφανής, δεδομένου ότι ο αριθμός συστάδων αντιστοιχεί στον αριθμό κρυφών κόμβων και τα μεμονωμένα χαρακτηριστικά συστάδων στις παραμέτρους πυρήνων των συναρτήσεων ακτινικής βάσης.

Σε αυτό το κεφάλαιο, προτείνεται μια νέα συστηματική προσέγγιση [66] στα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης. Η βασική ιδέα είναι να χρησιμοποιηθεί στη σειρά η ιεραρχική ασαφής συσταδοποίηση και η βέλτιστη ασαφής συσταδοποίηση για να ανιχνεύσει την πραγματική δομή των δεδομένων. Η τελική τοπολογία συστάδων δείχνει άμεσα και τον αριθμό κόμβων και των αντίστοιχων παραμέτρων πυρήνων των

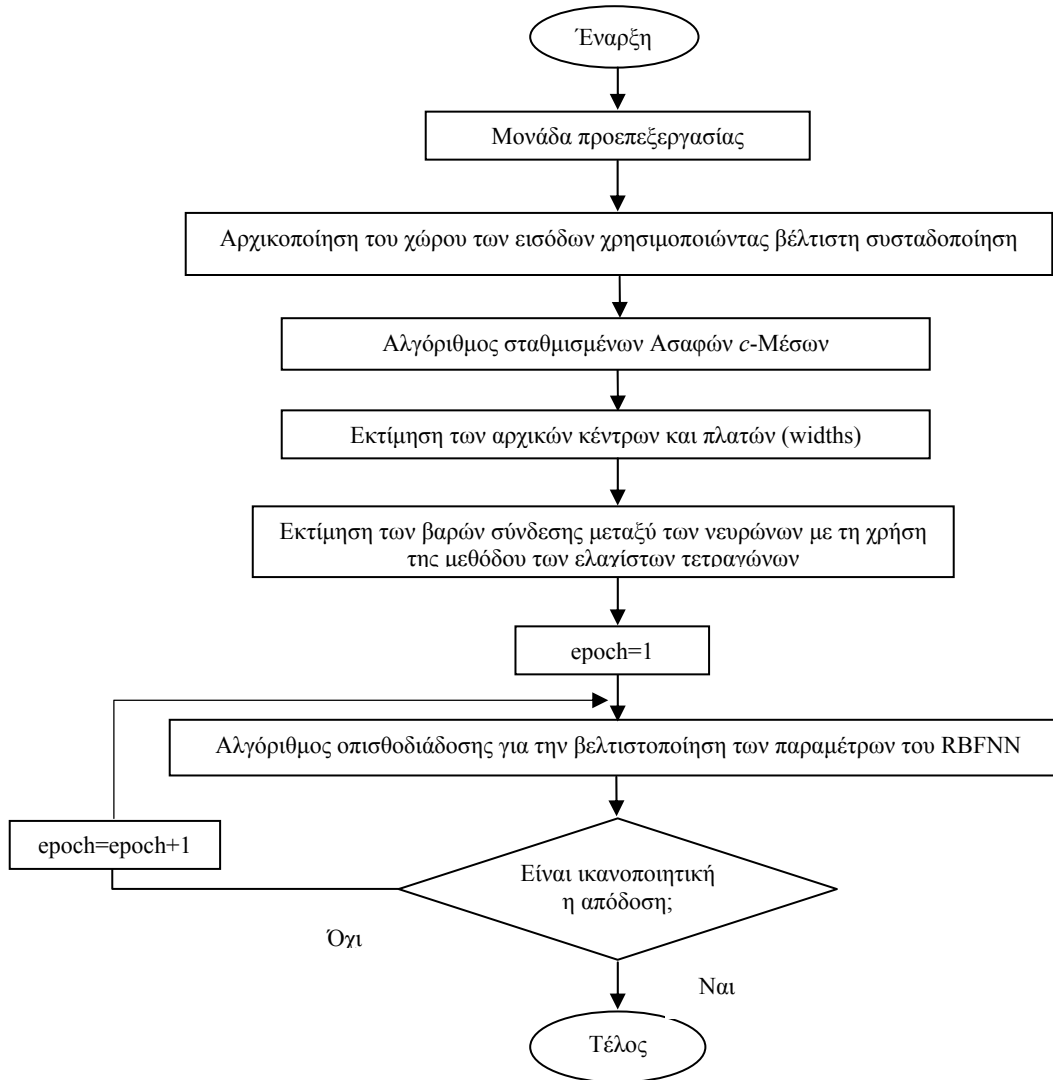
συναρτήσεων ακτινικής βάσης. Ένα σημαντικό χαρακτηριστικό γνώρισμα της προτεινόμενης μεθόδου είναι ότι, αντίθετα προς άλλες προσεγγίσεις [28,38,54,86,87,103], χρησιμοποιούνται τρία στάδια για να βελτιστοποιηθούν οι παράμετροι του νευρωνικού δικτύου. Το πρώτο στάδιο είναι η προηγουμένως αναφερθείσα βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering), η οποία παρέχει τον αριθμό κόμβων και μια αρχική εκτίμηση των παραμέτρων των κρυφών κόμβων. Δεύτερον, χρησιμοποιείται η μέθοδος ελαχίστων τετραγώνων (least squares) για να ληφθεί μια αρχική εκτίμηση των συναπτικών βαρών των νευρώνων (neuroconnecting synaptic weights) και τρίτον, όλες οι ανωτέρω τιμές παραμέτρου χρησιμοποιούνται ως αρχικοί όροι με τη μέθοδο βελτιστοποίησης οπισθοδιάδοσης (back-propagation) για να τελειοποιήσουν ολόκληρο το σύνολο παραμέτρων. Το αποτέλεσμα είναι μια αποδοτική πλατφόρμα εκπαίδευσης, η οποία κατορθώνει να διατηρήσει την ισορροπία μεταξύ του αριθμού κόμβων και της ακρίβειας διαμόρφωσης. Για να εκτελεστεί η βέλτιστη ασαφής συσταδοποίηση, χρησιμοποιήθηκαν τρεις διαφορετικοί δείκτες ασαφούς συσταδοποίησης. Τα αποτελέσματα προσομοίωσης έδειξαν ότι η μέθοδος είναι αποτελεσματική και αποδοτική στην επιλογή του δείκτη βέλτιστης ασαφούς συσταδοποίησης.

4.2 Ο προτεινόμενος Αλγόριθμος

4.2.1 Επισκόπηση του αλγορίθμου

Μια από τις ευρύτετα χρησιμοποιημένες ασαφής προσεγγίσεις συσταδοποίησης είναι η μέθοδος των Ασαφών c -Μέσων (FCM) [67,99]. Εντούτοις, η εφαρμογή της μεθόδου Ασαφών c -Μέσων (FCM) εξαρτάται έντονα από την αρχικοποίηση των κέντρων των συστάδων. Για να υπερνικηθεί αυτό το πρόβλημα χρειάζονται κάποιες πληροφορίες για τις πιθανές δομές που μπορούν να υπάρξουν στο σύνολο δεδομένων εκπαίδευσης. Ένας εφικτός τρόπος να λυθεί αυτό το πρόβλημα είναι να χρησιμοποιηθεί μια μονάδα συσταδοποίησης για να προεπεξεργαστεί τα διαθέσιμα δεδομένα. Αυτή η μονάδα πρέπει να είναι σε θέση να ανιχνεύσει διάφορες κατάλληλα προσδιορισμένες συστάδες που μπορούν να χρησιμοποιηθούν ως αρχικοί όροι στον αλγόριθμο των Ασαφών c -Μέσων (FCM). Στην προτεινόμενη μέθοδο, για να προεπεξεργαστούν τα

δεδομένα εκπαίδευσης χρησιμοποιείται ο γνωστός αλγόριθμος κβάντισης διανύσματος με εκμάθηση (Learning Vector Quantization - LVQ) [45]. Η χρησιμοποίηση του LVQ για να προεπεξεργαστεί τα δεδομένα αποτελεί το πρώτο βήμα του αλγορίθμου.



Σχήμα 4.1. Ο προτεινόμενος αλγόριθμος

Κατόπιν, στα κέντρα των συστάδων που λαμβάνονται στο προηγούμενο βήμα ορίζονται οι τιμές βάρους, οι οποίες καθορίζονται από την πληθικότητα συνόλου (cardinality) των αντίστοιχων συστάδων. Τα κέντρα συστάδων μαζί με τα βάρη τους θεωρούνται ως νέο σύνολο δεδομένων, το οποίο συσταδοποιείται περαιτέρω με τη χρησιμοποίηση των σταθμισμένου (weighted) αλγορίθμου Ασαφών c -Μέσων (FCM) [91]. Χρησιμοποιούμε αυτήν την τροποποίηση της μεθόδου Ασαφών c -Μέσων (FCM)

προκειμένου να μειωθεί η εξάρτηση στην αρχικοποίηση καθώς επίσης και το υπολογιστικό κόστος, δεδομένου ότι ο αριθμός κέντρων συστάδων που προέρχονται από τη μονάδα προεπεξεργασίας είναι πολύ μικρότερος από τον αριθμό των αρχικών δεδομένων εκπαίδευσης

Στο τρίτο βήμα, εφαρμόζεται ένας δείκτης βέλτιστου αριθμού συστάδων, ο οποίος καθορίζει τον κατάλληλο αριθμό συστάδων και των αντίστοιχων χαρακτηριστικών όπως τα κέντρα τους και τα πλάτη τους. Στο τέταρτο βήμα, η αρχική τοπολογία του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης λαμβάνεται, δεδομένου ότι ο αριθμός συστάδων συμπίπτει με τον αριθμό κρυφών κόμβων, ενώ τα χαρακτηριστικά συστάδων με τις παραμέτρους του δικτύου. Τέλος, στο πέμπτο βήμα, χρησιμοποιείται ο αλγόριθμος οπισθοδιάδοσης (back-propagation) για να βελτιστοποιηθεί το σύνολο παραμέτρων. Τα ανωτέρω βήματα απεικονίζονται στο σχήμα 4.1.

Πριν περιγραφεί θεωρητικά το κάθε βήμα σχεδιασμού, αναλύονται λεπτομερέστερα τα κύρια κριτήρια που χρησιμοποιούνται στην ανωτέρω διαδικασία. Το πρώτο κριτήριο αναφέρεται στη χρήση κβάντισης διανύσματος με εκμάθηση (Learning Vector Quantization-LVQ). Το LVQ είναι αρκετά ευαίσθητο όσον αφορά τις παραμέτρους σχεδιασμού του. Ο αντίκτυπος αυτού του γεγονότος είναι ότι, ανάλογα με τις επιλογές των παραμέτρων, μερικές συστάδες μπορούν να περιέχουν έναν μικρό αριθμό δεδομένων ή στη χειρότερη περίπτωση κανένα δεδομένο, λόγω της ανταγωνιστικής φύσης της διαδικασίας εκπαίδευσης (ο νικητής τα παίρνει όλα - winner takes all). Εντούτοις, η χρήση της μεθόδου των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) στο επόμενο βήμα, ακυρώνει τη συμβολή αυτών των συστάδων στον τελικό ασαφή διαμερισμό, δεδομένου ότι τα αντίστοιχα βάρη θα είναι κοντά στο μηδέν ή ίσα με μηδέν, αντίστοιχα. Επομένως, αυτό το βήμα δεν ενδιαφέρεται για το σχεδιασμό ενός αυστηρού αλγορίθμου κβάντισης διανύσματος με εκμάθηση (LVQ). Ο μόνος περιορισμός εδώ είναι, ότι ο αριθμός συστάδων στο LVQ είναι αρκετά μεγάλος, έτσι ώστε οι περιπτώσεις εκπαίδευσης με την μέθοδο των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) να είναι επαρκείς για να επιτύχουν ένα αξιόπιστο αποτέλεσμα συσταδοποίησης. Κατά συνέπεια, λαμβάνοντας αυτήν την παρατήρηση υπόψη, μπορούμε να ωφεληθούμε από τα πλεονεκτήματα που παρέχονται από τον αλγόριθμο

κβάντισης διανύσματος με εκμάθηση (LVQ) όπως η απλότητα της και η ταχύτητά της. Το δεύτερο κριτήριο αφορά τη χρήση του αλγορίθμου οπισθοδιάδοσης (back-propagation). Ο βασικός αλγόριθμος οπισθοδιάδοσης (back-propagation) βρίσκει συχνά το τοπικό ελάχιστο. Είναι επίσης ευαίσθητο στις παραμέτρους σχεδιασμού του. Εντούτοις, χρησιμοποιούμε διαδικασία οπισθοδιάδοσης (back-propagation) λόγω δύο κυρίως λόγων: (α) η οπισθοδιάδοση (back-propagation) παρουσιάζει γρήγορη σύγκλιση, και (β) παρέχοντας καλούς αρχικούς όρους στον αλγόριθμο οπισθοδιάδοσης (back-propagation) τότε αυτός είναι σε θέση να επιτύχει πολύ καλά αποτελέσματα.

4.2.2 Ο αλγόριθμος κβάντισης διανύσματος με εκμάθηση (*Learning Vector Quantization-LVQ*)

Ο αλγόριθμος αναλύεται λεπτομερώς στην παράγραφο 2.5 του κεφαλαίου 2 της παρούσης διδακτορικής διατριβής.

4.2.3 Σταθμισμένη (*weighted*) Ασαφής Συσταδοποίηση (*Fuzzy Clustering*)

Στο επόμενο βήμα, τα κέντρα συστάδων v_i ($1 \leq i \leq n$) θεωρούνται ως ένα νέο σύνολο δεδομένων. Επειδή το διάνυσμα v_i αναφέρεται σε μια συστάδα του αρχικού συνόλου δεδομένων, το βάρος σημασίας του καθορίζεται ως εξής,

$$\rho_i = \frac{\aleph(U^i)}{\sum_{i=1}^n \aleph(U^i)} \quad , \quad (1 \leq i \leq n) \quad (4.1)$$

όπου $\aleph(U^i)$ είναι η ασαφής πληθικότητα συνόλου (cardinality) της i -οστής συστάδας

$$\aleph(U^i) = \sum_{k=1}^N u_{ik} \quad (1 \leq i \leq n) \quad (4.2)$$

Κατόπιν, η αντικειμενική συνάρτηση κόστους της μεθόδου σταθμισμένων (*weighted*) Ασαφών c -Μέσων (FCM) [91] είναι,

$$J_\rho = \sum_{i=1}^n \sum_{j=1}^c \rho_i (\mu_{ji})^m \|v_i - v_j\|^2 \quad (4.3)$$

όπου το c είναι ο αριθμός των τελικών συστάδων, ο οποίος συμπίπτει με τον αριθμό κρυφών RBF κόμβων, \mathbf{v}_j ($1 \leq j \leq c$) είναι τα κέντρα των τελικών συστάδων, $m \in (1, \infty)$ είναι ένας παράγοντας για να ρυθμίσει το βαθμό συμμετοχής (membership degree) σταθμίζοντας την επίδραση, και ρ_i είναι το βάρος μεγάλης σημασίας του διανύσματος \mathbf{v}_i .

Το πρόβλημα βελτιστοποίησης είναι να ελαχιστοποιηθεί η αντικειμενική συνάρτηση J_ρ κάτω από τον ακόλουθο περιορισμό:

$$\sum_{j=1}^c \mu_{ji} = 1 \quad \forall i \quad (4.4)$$

Τα τελικά κέντρα και οι αντίστοιχες συναρτήσεις συμμετοχής (membership functions) που λύνουν αυτό το περιορισμένο πρόβλημα βελτιστοποίησης δίνονται από τις επόμενες εξισώσεις [91]:

$$\mathbf{v}_j = \frac{\sum_{i=1}^n \rho_i (\mu_{ji})^m \mathbf{v}_i}{\sum_{i=1}^n \rho_i (\mu_{ji})^m}, \quad (1 \leq j \leq c) \quad (4.5)$$

και

$$\mu_{ji} = \frac{1}{\sum_{l=1}^c \left(\frac{\|\mathbf{v}_i - \mathbf{v}_j\|}{\|\mathbf{v}_i - \mathbf{v}_l\|} \right)^{2/(m-1)}} \quad (4.6)$$

Οι εξ. (4.5) και (4.6) αποτελούν μια επαναληπτική διαδικασία βελτιστοποίησης που ελαχιστοποιεί την αντικειμενική συνάρτηση κόστους (4.3). Στην περίπτωση όπου όλα τα βάρη είναι ίσα, ο σταθμισμένος (weighted) αλγόριθμος Ασαφών c -Μέσων (FCM) είναι ισοδύναμος με το κλασσικό αλγόριθμο ασαφούς συσταδοποίησης (FCM).

Για αυτόν τον λόγο, ο αριθμός κρυφών κόμβων στο νευρωνικό δίκτυο συναρτήσεων ακτινικής βάσης είναι ίσος με τον αριθμό συστάδων c , ενώ τα κέντρα αυτών των κόμβων είναι τα ίδια με τα κέντρα των συστάδων \mathbf{v}_j ($1 \leq j \leq c$). Αφ' ετέρου, τα πλάτη (widths) των κόμβων του νευρωνικού δικτύου (RBFNN) υπολογίζονται ως εξής,

$$\sigma^j = \frac{2 * d_{\max}^j}{3} \quad (1 \leq j \leq c) \quad (4.7)$$

Στην εξ. (4.7) το d_{\max} δίνεται από την επόμενη εξίσωση,

$$d_{\max} = \{ \max_k \{ \| \mathbf{x}_k - \mathbf{v}_j \|^2 \} : \mathbf{x}_k \in C_j \text{ έτσι ώστε } u_{jk} \geq \theta \} \quad (1 \leq j \leq c) \quad (4.8)$$

όπου C_j είναι η j -οστή συστάδα και u_{jk} είναι ο βαθμός συμμετοχής του k -διανύσματος δεδομένων εκπαίδευσης στη j -τελική ασαφή συστάδα, και θ ένας μικρός θετικός αριθμός έτσι ώστε $\theta \in (0, 1)$.

4.2.4 Βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering)

Η βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering) αφορά την εκτίμηση του αριθμού συστάδων που δίνει ένας διαμερισμός στον οποίο τα δεδομένα που ανήκουν στην ίδια συστάδα είναι όσο το δυνατόν πιο όμοια μεταξύ τους.

Δεν υπάρχει κανένας δείκτης βέλτιστου ασαφούς διαμερισμού συστάδων που να έχει ισχύ για όλες τις εφαρμογές (σύνολα δεδομένων). Προτείνουμε μερικούς από τους πιο αξιόπιστους ταυτόχρονα.

Παρά το γεγονός ότι και ο σταθμισμένος (weighted) και ο κλασσικός αλγόριθμος ασαφούς συσταδοποίησης (FCM) είναι σε θέση να ανιχνεύσουν τις ομοιότητες μέσα σε ένα σύνολο δεδομένων, δεν μπορούν να εκτελέσουν τη βέλτιστη ασαφή συσταδοποίηση (optimal fuzzy clustering) επειδή απαιτούν μια εκ των προτέρων γνώση του αριθμού συστάδων. Ένας κατάλληλος τρόπος να δημιουργηθεί η βέλτιστη ασαφής συσταδοποίηση και να παραχθούν οι καλύτερες δυνατές λύσεις συσταδοποίησης είναι να χρησιμοποιηθεί ένας αξιόπιστος δείκτης για να συνοδεύσει την αντικειμενική συνάρτηση κόστους (4.3).

(α) Ο δείκτης Xie-Beni, για το σταθμισμένο (weighted) αλγόριθμο Ασαφών c -Μέσων (FCM) υπολογίζονται από την ακόλουθη εξίσωση [99]:

$$S_{XB} = \frac{\sum_{k=1}^n \sum_{i=1}^c \rho_k (u_{ik})^p \| a^k - u_i \|^2}{n \min_{i \neq j} \{ \| u_i - u_j \|^2 \}} \quad (4.9)$$

Η ανωτέρω εξίσωση συνδυάζει το ακόλουθο μέτρο πυκνότητας:

$$\pi = \frac{\sum_{k=1}^n \sum_{i=1}^c \rho_k (u_{ik})^p \|a^k - u_i\|^2}{n} \quad (4.10)$$

και το μέτρο διαχωρισμού του τελικού ασαφούς διαχωρισμού, το οποίο δίνεται όπως

$$s = \min_{i \neq j} \left\{ \|u_i - u_j\|^2 \right\} \quad (4.11)$$

Ο βέλτιστος αριθμός συστάδων που δίνει ο δείκτης Xie-Beni είναι αυτός που αντιστοιχεί στη χαμηλότερη τιμή του S_{XB} .

(β) ο συντελεστής διαμερισμού (Partition Coefficient - PC) [67] που καθορίζεται από:

$$S_{PC} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \quad (4.12)$$

Ο βέλτιστος αριθμός συστάδων που δίνει ο συντελεστής διαμερισμού (PC) είναι αυτός που αντιστοιχεί στη χαμηλότερη τιμή του S_{PC} .

(γ) η εντροπία διαμερισμού (Partition Entropy - PE) [67] καθορίζεται από:

$$S_{PE} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log u_{ij} \quad (4.13)$$

Ο βέλτιστος αριθμός συστάδων που δίνει ο δείκτης εντροπίας διαμερισμού είναι αυτός που αντιστοιχεί στη μέγιστη τιμή S_{PE} .

4.2.5 Εκτίμηση των αρχικών βαρών για το RBF με τη χρησιμοποίηση της μεθόδου των ελαχίστων τετραγώνων (least squares method)

Δεδομένου ότι η έξοδος του νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης (RBF) είναι γραμμική σε βάρη w_j , αυτά τα βάρη μπορούν να υπολογιστούν με τη μέθοδο των ελαχίστων τετραγώνων (least squares method). Για κάθε σημείο δεδομένων εισόδου x_k , τα αποτελέσματα του νευρώνα υπολογίζονται όπως:

$$s_{kj} = g_j(\mathbf{x}_k) \quad (4.14)$$

και τοποθετούνται στον πίνακα (μήτρα) $A=[s_{kj}]$, όπου $g_j(\mathbf{x}_k)$ είναι η Γκαουσιανή συνάρτηση (gaussian function). Εισάγοντας το διάνυσμα βαρών $\mathbf{w}=[w_1, w_2, \dots, w_c]$, μπορεί να γραφεί η ακόλουθη εξίσωση πινάκων για ολόκληρο το σύνολο δεδομένων:

$$\tilde{\mathbf{y}} = A\mathbf{w} \quad (4.15)$$

Η εκτίμηση των βαρών \mathbf{w} που ελαχιστοποιούν το σφάλμα δικτύου $e = \tilde{\mathbf{y}} - \mathbf{y}$ γίνεται με την μέθοδο ελαχίστων τετραγώνων και τα βάρη υπολογίζονται από τον παρακάτω τύπο:

$$\mathbf{w} = [A^T A]^{-1} A^T \tilde{\mathbf{y}} \quad (4.16)$$

4.2.6 Βελτιστοποίηση του νευρωνικού δικτύου RBF με τη χρησιμοποίηση οπισθοδιάδοσης (back-propagation)

Σε αυτό το βήμα, οι αρχικές εκτιμήσεις των παραμέτρων v_i , w_i και σ_i ($1 \leq i \leq c$), που υπολογίστηκαν στο προηγούμενο βήμα τροφοδοτούνται στον αλγόριθμο οπισθοδιάδοσης (back-propagation). Ο στόχος είναι να ελαχιστοποιηθεί το ακόλουθο μέσο τετραγωνικό σφάλμα,

$$e = \frac{1}{2}(\tilde{y}_k - y_k)^2 \quad (4.17)$$

Οι κανόνες εκμάθησης για τις ανωτέρω παραμέτρους είναι,

$$w_j^l(t+1) = w_j^l(t) - e^* \frac{\partial e}{\partial w_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (4.18)$$

$$v_j^l(t+1) = v_j^l(t) - e^* \frac{\partial e}{\partial v_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (4.19)$$

$$\sigma_j^l(t+1) = \sigma_j^l(t) - e^* \frac{\partial e}{\partial \sigma_j^l} \Big|_t \quad (1 \leq l \leq p) \quad (4.20)$$

όπου το e είναι το ποσοστό εκμάθησης του αλγορίθμου οπισθοδιάδοσης (back-propagation). Τελικά, οι μερικές παράγωγοι στις εξ. (4.18), (4.19) και (4.20), βασισμένοι στις εξ. (4.15) και (4.17), μπορούν πολύ εύκολα να υπολογιστούν.

4.3 Πειράματα προσομοίωσης

4.3.1 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης

Η συνάρτηση είναι η:

$$y = 0.6 \eta\mu(\pi x) + 0.3 \eta\mu(3\pi x) + 0.1 \eta\mu(5\pi x) \quad \text{με } x \in [-1, 1].$$

Χρησιμοποιώντας την παραπάνω εξίσωση δημιουργήθηκαν διακόσια δεδομένα εισόδου-εξόδου όπως και στις αναφορές [70,75]. Το αρχικό ποσοστό εκμάθησης για το LVQ επιλέχτηκε όπως: $a_0 = 0.5$. Επιπλέον, ο αρχικός αριθμός συστάδων τέθηκε $n = 20$ σε όλες τις προσομοιώσεις και η παράμετρος ασάφειας ήταν: $m = 2$. Ο πίνακας 4.1 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες βέλτιστης ασαφούς συσταδοποίησης. Σύμφωνα με αυτόν τον πίνακα ο βέλτιστος αριθμός συστάδων για το Χει- Beni, το PC και το PE ήταν 8, 5 και 6, αντίστοιχα.

Πίνακας 4.1. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης μονοδιάστατης συνθετικής συνάρτησης.

Αριθμός Συστάδων	XB	PC	PE
2	3.4386	0.9091	0.1531
3	4.2730	0.9273	0.1439
4	6.4105	0.9072	0.1905
5	5.3075	0.8783	0.2467
6	6.4605	0.8833	0.2512
7	4.1647	0.9073	0.2106
8	2.3468	0.9085	0.2049
9	3.7923	0.9162	0.1902
10	4.9628	0.9005	0.2100
11	4.2014	0.9046	0.1986
12	4.5745	0.9197	0.1680

Το μέσο τετραγωνικό σφάλμα (MSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 4.2. Αυτός ο πίνακας συγκρίνει τις αποδόσεις των δικτύων που παράγονται από την προτεινόμενη μέθοδο, με τις αποδόσεις άλλων μεθόδων που υπάρχουν στην διεθνή βιβλιογραφία. Συγκεκριμένα, το *αρχικό δίκτυο* αναφέρεται στη δομή νευρωνικού δικτύου συναρτήσεων ακτινικής βάσης προτού να εφαρμοστεί ο αλγόριθμος οπισθοδιάδοσης (back-propagation), και το *τελικό δίκτυο* αντιστοιχεί στο νευρωνικό δίκτυο συναρτήσεων ακτινικής βάσης μετά από την εφαρμογή του αλγορίθμου βελτιστοποίησης οπισθοδιάδοσης (back-propagation). Ως εκ τούτου, μπορεί να εξαχθεί το συμπέρασμα ότι η επίδραση της οπισθοδιάδοσης απεικονίζεται στο τελικό πρότυπο.

Πίνακας 4.2. Συγκριτική ανάλυση από την άποψη του MSE.

Μοντέλο	Αρ. κόμβων	MSE στα δεδομένα εκπαίδευσης	MSE στα δεδομένα δοκιμής
Pedrycz et al. [75]	≤ 33	0.0029 ± 0.001	0.0052 ± 0.004
	≤ 33	0.0033 ± 0.001	0.0035 ± 0.002
Park et al. [70]	6	$0.0045 \pm 6.2E-4$	0.0281 ± 0.014
Προτεινόμενο RBF	(αρχικό δίκτυο) PE	5	0.023 ± 0.008
	(τελικό δίκτυο) PE	5	0.0067 ± 0.001
	(αρχικό δίκτυο) PE	6	0.019 ± 0.009
	(τελικό δίκτυο) PC	6	0.0059 ± 0.001
	(αρχικό δίκτυο) XB	8	0.011 ± 0.007
	(αρχικό δίκτυο) XB	8	0.0058 ± 0.001

Από αυτόν τον πίνακα, μπορούμε εύκολα να ελέγξουμε ότι η προτεινόμενη μέθοδος παράγει τα πολύ ακριβή πρότυπα που χρησιμοποιούν έναν μικρό αριθμό κρυφών κόμβων. Επιπλέον, οι τρεις δείκτες παράγουν τις ισοδύναμες αποδόσεις. Η μέθοδος του Pedrycz et al. [75], έδωσε το καλύτερο μέσο τετραγωνικό σφάλμα (MSE) σε σχέση με το προτεινόμενο νευρωνικό δίκτυο. Όμως αυτό έγινε με μεγαλύτερο αριθμό κρυφών RBF κόμβων. Η μέθοδος του Park et al. [70] έδωσε καλύτερο μέσο τετραγωνικό σφάλμα (MSE) σε σχέση με το προτεινόμενο νευρωνικό δίκτυο όσον αφορά τα δεδομένα εκπαίδευσης. Όσον αφορά όμως τα δεδομένα δοκιμής το προτεινόμενο

νευρωνικό δίκτυο έδωσε πολύ καλύτερο μέσο τετραγωνικό σφάλμα σε σχέση με τη μέθοδο των Park et al. [70] κάτι που είναι ιδιαίτερα σημαντικό. Τέλος, όταν συγκρίνουμε τα μέσα τετραγωνικά σφάλματα (MSE) που λαμβάνονται από τα αρχικά και τελικά δίκτυα RBF μπορούμε σαφώς να δούμε την επίδραση του αλγορίθμου οπισθοδιάδοσης (back-propagation).

4.3.2 Προσέγγιση Δισδιάστατης Συνάρτησης

Σε αυτήν την υποενότητα μελετώνται οι ικανότητες προσέγγισης συνάρτησης του προτεινόμενου αλγορίθμου. Η στατική συνάρτηση (static function) που θα προσεγγιστεί περιγράφεται από την ακόλουθη εξίσωση,

$$y = (1 + x_1^{-2} + x_2^{-1.5}), \quad 1 \leq x_1, x_2 \leq 5 \quad (4.21)$$

Χρησιμοποιώντας την εξίσωση (4.21) δημιουργήθηκαν πενήντα δεδομένα εισόδου-εξόδου. Το αρχικό ποσοστό εκμάθησης για το LVQ επιλέχτηκε όπως: $a_0 = 0.5$. Επιπλέον, ο αρχικός αριθμός συστάδων τέθηκε $n = 20$ σε όλες τις προσομοιώσεις και η παράμετρος ασάφειας ήταν: $m = 2$. Ο πίνακας 4.3 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες βέλτιστης ασαφούς συσταδοποίησης. Σύμφωνα με αυτόν τον πίνακα ο βέλτιστος αριθμός συστάδων για το Xei- Beni, το PC και το PE ήταν 10, 10 και 14, αντίστοιχα.

Το μέσο τετραγωνικό σφάλμα (MSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 4.4. Αυτός ο πίνακας συγκρίνει τις αποδόσεις των δικτύων που παράγονται από την προτεινόμενη μέθοδο, με τις αποδόσεις άλλων μεθόδων που υπάρχουν στην διεθνή βιβλιογραφία. Συγκεκριμένα, το *αρχικό δίκτυο* αναφέρεται στη δομή νευρωνικού δικτύου RBF προτού να εφαρμοστεί ο αλγόριθμος οπισθοδιάδοσης (back-propagation), και το *τελικό δίκτυο* αντιστοιχεί στο νευρωνικό δίκτυο RBF μετά από την εφαρμογή του αλγορίθμου βελτιστοποίησης οπισθοδιάδοσης (back-propagation). Ως εκ τούτου, μπορεί να εξαχθεί το συμπέρασμα ότι η επίδραση της οπισθοδιάδοσης απεικονίζεται στο τελικό πρότυπο.

Πίνακας 4.3. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα στατικής συνάρτησης (static function)

Αριθμός συστάδων	XB	PC	PE
2	1.3627	0.0382	0.0346
3	3.9715	0.0339	0.0547
4	1.5814	0.0356	0.0664
5	1.9883	0.0330	0.0762
6	1.1419	0.0335	0.0827
7	2.8228	0.0325	0.0883
8	1.1287	0.0338	0.0939
9	1.4649	0.0356	0.0974
10	1.001	0.0316	0.1010
11	1.1549	0.0358	0.1020
12	1.2171	0.0422	0.1103
13	1.4123	0.0361	0.1079
14	1.6619	0.0444	0.1111
15	1.7589	0.0471	0.1086

Πίνακας 4.4. Αποτελέσματα σύγκρισης για το παράδειγμα στατικής συνάρτησης (static function)

Πρότυπο	Αριθμός κρυφών κόμβων	MSE	
Κλασσικό RBF	26	0.0307	
Tsekouras et al [91]	8	0.0042	
Lee et al. [48]	8	0.0148	
Kim et al. [42]	3	0.0090	
Το προτεινόμενο δίκτυο RBF	(αρχικό δίκτυο) XB	10	0.0911
	(τελικό δίκτυο) XB	10	0.0035
	(αρχικό δίκτυο) PC	10	0.0912
	(τελικό δίκτυο) PC	10	0.0028
	(αρχικό δίκτυο) PE	14	0.0830
	(τελικό δίκτυο) PE	14	0.0007

Από αυτόν τον πίνακα, μπορούμε εύκολα να ελέγξουμε ότι η προτεινόμενη μέθοδος παράγει τα πολύ ακριβή πρότυπα που χρησιμοποιούν έναν μικρό αριθμό κρυφών κόμβων. Επιπλέον, οι τρεις δείκτες παράγουν ισοδύναμες αποδόσεις. Το κλασσικό νευρωνικό δίκτυο RBF με 26 κρυφούς κόμβους, έδωσε το χειρότερο μέσο τετραγωνικό σφάλμα (MSE) σε σχέση με το προτεινόμενο νευρωνικό δίκτυο. Τέλος, όταν συγκρίνουμε τα μέσα τετραγωνικά σφάλματα (MSE) που λαμβάνονται από τα αρχικά και τελικά δίκτυα RBF μπορούμε σαφώς να δούμε την επίδραση του αλγορίθμου

οπισθοδιάδοσης (back-propagation).

4.3.3 Σύστημα Mackey-Glass

Σε αυτήν την υποενότητα, χρησιμοποιείται η προτεινόμενη μέθοδος προκειμένου να προβλεφθεί η χρονοσειρά Mackey-Glass. Η χρονοσειρά Mackey-Glass παράγεται από την ακόλουθη διαφορική εξίσωση χρονικής καθυστέρησης

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \quad (4.22)$$

Όταν η παράμετρος τ είναι αρκετά μεγάλη, το σύστημα εμφανίζει μια χαοτική συμπεριφορά. Στις προσομοιώσεις τέθηκε το $\tau = 17$ και δημιουργήθηκε ένα δείγμα 1000 σημείων. Τα πρώτα 500 σημεία χρησιμοποιήθηκαν σαν δεδομένα εκπαίδευσης, και τελευταία 500 σημεία ως δεδομένα δοκιμής για να επικυρώσουν το δίκτυο RBF. Ο προτεινόμενος αλγόριθμος εφαρμόστηκε για να δημιουργήσει το νευρωνικό δίκτυο RBF με 4 εισόδους: $x(t-18)$, $x(t-12)$, $x(t-6)$ και $x(t)$, ενώ η έξοδος ήταν το σημείο $x(t+6)$. Για τον αλγόριθμο LVQ επιλέχτηκε $a_0 = 0.5$. Επιπλέον, ο αρχικός αριθμός συστάδων επιλέχτηκε $n = 20$. Σε όλες τις προσομοιώσεις η παράμετρος ασάφειας ήταν: $m = 2$.

Πίνακας 4.5. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα Mackey-Glass

Αριθμός συστάδων	XB	PC	PE
2	2.2264	0.0346	0.0289
3	1.6717	0.0350	0.0373
4	1.2985	0.0322	0.0456
5	0.8807	0.0302	0.0553
6	1.0443	0.0260	0.0661
7	1.0314	0.0248	0.0710
8	1.1591	0.0386	0.0667
9	1.0492	0.0243	0.0829
10	1.2452	0.0220	0.0901
11	1.4007	0.0220	0.0913
12	1.2411	0.0163	0.1078
13	1.2759	0.0402	0.0749
14	1.4977	0.0470	0.0738
15	1.5244	0.0415	0.0755

Πίνακας 4.6. Αποτελέσματα σύγκρισης για το παράδειγμα Mackey-Glass

Πρότυπο	Αριθμός κρυφών κόμβων	RMSE	
Κλασσικό RBF	50	0.0033	
Cho, Wang [21](δεδομένα δοκιμής)	23	0.0114	
Kim D., Kim C. [41]	5	0.0491	
Jang, Sun, Mizutani [37]	16	0.0072	
Το προτεινόμενο δίκτυο RBF	(αρχικό δίκτυο) XB	5	0.1271
	(τελικό δίκτυο - δεδομένα εκπαίδευσης) XB	5	0.0252
	(τελικό δίκτυο - δεδομένα δοκιμής) XB	5	0.0271
	(αρχικό δίκτυο) PC	12	0.1154
	(τελικό δίκτυο - δεδομένα εκπαίδευσης) PC	12	0.0160
	(τελικό δίκτυο - δεδομένα δοκιμής) PC	12	0.0180
	(αρχικό δίκτυο) PE	12	0.1521
	(τελικό δίκτυο - δεδομένα εκπαίδευσης) PE	12	0.0126
	(τελικό δίκτυο - δεδομένα δοκιμής) PE	12	0.0152

Ο πίνακας 4.5 παρουσιάζει το βέλτιστο αριθμό συστάδων (optimal fuzzy clusters) που λαμβάνονται από τους τρεις δείκτες. Συγκεκριμένα, ο Xei-Beni, δείκτης έδωσε 5 ασαφείς συστάδες σαν βέλτιστο αριθμό συστάδων. Οι δείκτες PC και PE έδωσαν 12 ασαφείς συστάδες. Η τετραγωνική ρίζα των μέσων τετραγωνικών σφαλμάτων (RMSE) απεικονίζεται στον πίνακα 4.6, όπου τα συγκριτικά αποτελέσματα δίνονται επίσης. Από αυτόν τον πίνακα παρατηρείται, ότι το καλύτερο αποτέλεσμα αναφέρεται από το κλασσικό νευρωνικό δίκτυο RBF, το οποίο εντούτοις χρησιμοποιεί 50 κρυφούς κόμβους. Αφ' ετέρου, οι Cho και Wang [21] λαμβάνουν ένα πολύ καλό RMSE με τη χρησιμοποίηση 23 κρυφών κόμβων. Η προτεινόμενη μέθοδος χρησιμοποίησε 5 (Xei-Beni) και 12 (PC και PE) κρυφούς κόμβους για να λάβει RMSE που είναι ελαφρώς μικρότερα από τις ανωτέρω δύο περιπτώσεις. Τέλος, άλλη μια φορά μπορεί να φανεί η επίδραση του αλγορίθμου οπισθοδιάδοσης (back-propagation) επάνω στην ακρίβεια του νευρωνικού δικτύου με τη σύγκριση των αποδόσεων των αρχικών και τελικών δικτύων.

4.4 Συμπεράσματα

Σε αυτό το κεφάλαιο, έγινε ενσωμάτωση της ιεραρχικής και βέλτιστης συσταδοποίησης για τον σχεδιασμό αποδοτικών νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBFNN). Η βασική ιδέα είναι να σχετίζεται άμεσα η τοπολογία του δικτύου με την υποκείμενη δομή δεδομένων, και να χρησιμοποιηθεί η βέλτιστη ασαφής συσταδοποίηση (optimal fuzzy clustering) για να ανιχνεύσει αυτήν την δομή. Κατόπιν η ανάθεση είναι προφανής, δεδομένου ότι τα κέντρα συστάδων και οι ασαφείς διαφορές συστάδων αντιστοιχούν στα κέντρα (centers) και τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης, αντίστοιχα. Τα συναπτικά βάρη (synaptic weights) υπολογίστηκαν από τον αλγόριθμο ελαχίστων τετραγώνων (least squares), ενώ ολόκληρο το σύνολο παραμέτρων (κέντρα, πλάτη, συναπτικά βάρη) βελτιστοποιήθηκε με τη βοήθεια του αλγορίθμου οπισθοδιάδοσης (back-propagation). Τέλος, οι προσομοιώσεις έδειξαν ότι η μέθοδος είναι σε θέση να λάβει την ακριβή απόδοση που χρησιμοποιεί έναν μικρό αριθμό κόμβων.

Μέθοδος Ιεραρχικής Ασαφούς Συσταδοποίησης για την Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

5.1 Εισαγωγή

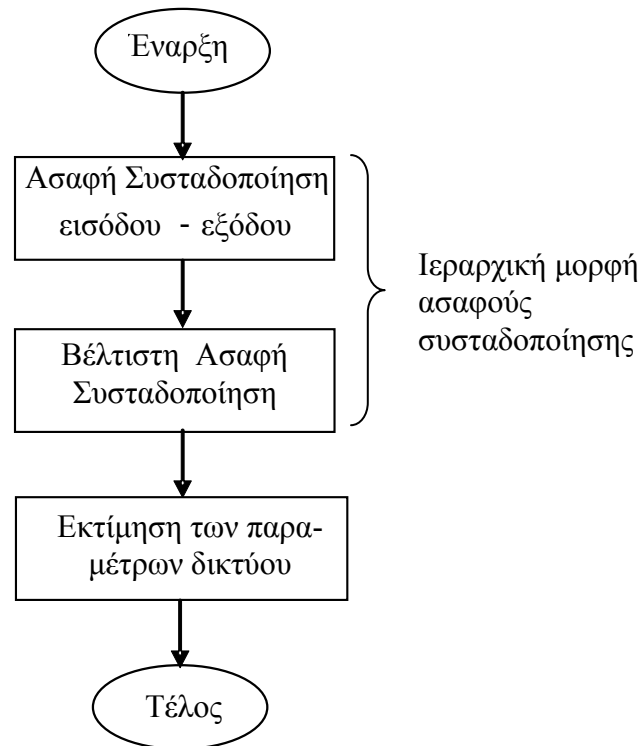
Ο προτεινόμενος αλγόριθμος χρησιμοποιεί τρεις βαθμίδες εκπαίδευσης και το διάγραμμα ροής του απεικονίζεται στο σχήμα 5.1. Το Βήμα 1 εφαρμόζει τον αλγόριθμο Ασαφών c -Μέσων (βλ. παράγραφο 2.3) στο χώρο των χαρακτηριστικών γνωρισμάτων εισόδου-εξόδου (input-output feature space), προκειμένου να προ-επεξεργαστεί τα δεδομένα. Στο Βήμα 2, τα παραγόμενα κέντρα των συστάδων προβάλλονται στο χώρο εισόδου (input space) και θεωρούνται ως το νέο σύνολο δεδομένων, το οποίο συσταδοποιείται περαιτέρω (further clustered) μέσω της βέλτιστης ασαφούς συσταδοποίησης, αποκτώντας έτσι τον αριθμό των κρυφών κόμβων (hidden nodes). Η βέλτιστη ασαφής συσταδοποίηση βασίζεται στην εφαρμογή των σταθμισμένων Ασαφών c -Μέσων (weighted fuzzy c -means). Τέλος, στο Βήμα 3 προχωρούμε στην εκτίμηση των παραμέτρων του δικτύου.

Επίσης, για να κατανοήσουμε καλύτερα τη σχέση μεταξύ των βημάτων σχεδιασμού, αναλύουμε σε περισσότερες λεπτομέρειες τη χρήση της μεθόδου σταθμισμένων Ασαφών c -Μέσων. Ακόμα μπορούμε να χρησιμοποιήσουμε την κλασική μέθοδο Ασαφών c -Μέσων για να επεξεργαστούμε τα κέντρα των συστάδων (cluster centers) που παράγονται από την μονάδα προ-επεξεργασίας (pre-processing unit).

Προφανώς, η εφαρμογή της μεθόδου Ασαφών c -Μέσων προϋποθέτει ότι όλα αυτά τα κέντρα έχουν την ίδια σπουδαιότητα. Ωστόσο, αυτό μπορεί να μην είναι αλήθεια, γιατί δύο διαφορετικές αρχικές συστάδες μπορεί να περιέχουν διαφορετικό

αριθμό των διανυσμάτων εκπαίδευσης. Συνεπώς, με την εφαρμογή της μεθόδου Ασαφών c -Μέσων διαγράφουμε όλες τις πληροφορίες που αφορούν το αρχικό σύνολο δεδομένων με τα ανωτέρω κέντρα.

Ως εκ τούτου, η πραγματική δομή των δεδομένων ενδέχεται να μην εντοπιστεί. Για να επιλύσουμε αυτό το πρόβλημα αναθέτουμε βάρη στα κέντρα, και τα επεξεργαζόμαστε με χρήση της σταθμισμένης (weighted) εκδοχής της μεθόδου Ασαφών c -Μέσων [91].



Σχήμα 5.1. Το διάγραμμα ροής του προτεινόμενου αλγόριθμου.

5.2 Ο προτεινόμενος δείκτης εγκυρότητας

Ας υποθέσουμε ότι έχουμε ένα σύνολο από διανύσματα δεδομένων εκπαίδευσης $\{\mathbf{x}_k\}_{k=1}^N$ που πρέπει να καταταμηθούν σε c συστάδες των οποίων τα κέντρα συμβολίζονται ως $\{\mathbf{v}_i\}_{i=1}^c$. Επιπλέον, δηλώνουμε ως $u_i(\mathbf{x}_k)$ το βαθμό συμμετοχής του δεδομένου εκπαίδευσης \mathbf{x}_k στο κέντρο \mathbf{v}_i της i -συστάδας. Ο δείκτης εγκυρότητας βασίζεται στην

πυκνότητα και σε κριτήρια διαχωρισμού. Η διακύμανση (σ_i) και η ασαφής πληθικότητα συνόλου (n_i) της i -συστάδας δίδονται ως [7, 99],

$$\sigma_i = \sum_{k=1}^N (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2, \quad 1 \leq i \leq c \quad (5.1)$$

$$n_i = \sum_{k=1}^N u_{ik}, \quad 1 \leq i \leq c \quad (5.2)$$

Το μέτρο πυκνότητας (π_i) της i -οστής συστάδας, ορίζεται ως ο λόγος μεταβολής της διακύμανσης (σ_i) και της ασαφής πληθικότητας συνόλου (n_i) και δίνεται από την παρακάτω εξίσωση [7],

$$\pi_i = \frac{\sigma_i}{n_i}, \quad 1 \leq i \leq c \quad (5.3)$$

Το μέτρο καθολικής πυκνότητας (π) του c -ασαφούς διαμερισμού δίνεται ως το άθροισμα των π_i όλων των συστάδων [102],

$$\pi = \sum_{i=1}^c \pi_i \quad (5.4)$$

Προκειμένου να καθορισθεί ο ασαφής διαχωρισμός των c -ασαφών διαμερισμών, κάποια έννοιες εισάγονται. Ας θεωρήσουμε ότι έχουμε δημιουργήσει έναν προσανατολισμένο προς τον χρήστη ασαφή διαμερισμό σε κάθε πεδίο ορισμού X_j ($1 \leq j \leq p$) με τον ισομερή διαμελισμό σε q_j συμμετρικά τριγωνικά ασαφή σύνολα $A_j^1, A_j^2, \dots, A_j^{q_j}$. Επιπλέον, ας υποθέσουμε ότι έχουμε εφαρμόσει ανεξάρτητη ασαφή συσταδοποίηση στα διαθέσιμα δεδομένα και η οποία παρήγαγε μια σειρά από c συστάδες $P = \{P_1, P_2, \dots, P_c\}$. Έτσι έχουμε καταλήξει σε ένα υβριδικό σύστημα, το οποίο φαίνεται να είναι ευέλικτο με την έννοια ότι έχουμε ταυτόχρονη πρόσβαση σε χαμηλού επιπέδου πληροφορίες που κωδικοποιούνται από τον προσανατολισμένο προς τον χρήστη ασαφή διαμερισμό και σε υψηλού επιπέδου πληροφορίες που κωδικοποιούνται από την ασαφή συσταδοποίηση. Απώτερος στόχος μας είναι να συνδυάσουμε αυτά τα δύο είδη πληροφοριών, προκειμένου να καθοριστούν οι ασαφής διαχωρισμοί των ασαφών c -διαμερισμών. Ας δούμε το πρόβλημα από τη σκοπιά του τυχαίου διανύσματος των δεδομένων

$\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kp}]^T$. Ο βαθμός συμμετοχής της j -οστής συνιστώσας στο ασαφές σύνολο A_j^l συμβολίζεται ως $\mu_j^l(x_{kj})$ με $1 \leq l \leq q_j$. Στη συνέχεια, η μερική συνεισφορά του ασαφούς συνόλου A_j^l στην ασαφή συστάδα P_i ορίζεται ως,

$$\Gamma_{i,j,l,k} = \begin{cases} u_i(\mathbf{x}_k), & \text{if } \mu_j^l(x_{kj}) = \max_{1 \leq r \leq q_j} \{\mu_j^r(x_{kj})\} \\ 0, & \text{otherwise} \end{cases} \quad (5.5)$$

Σχετικά, η συνολική συνεισφορά των ασαφούς συνόλου A_j^l στην ασαφή συστάδα P_i ορίζεται σε σχέση με το σύνολο των δεδομένων $\{\mathbf{x}_k\}_{k=1}^N$,

$$Z_{i,j,l} = \sum_{k=1}^N \Gamma_{i,j,l,k} \quad (5.6)$$

Στη συνέχεια, η ανομοιότητα ανάμεσα σε δύο συστάδες P_i και P_t σε σχέση με το ασαφές σύνολο A_j^l ορίζεται ως,

$$\Delta Z_{i,t,j,l} = |Z_{i,j,l} - Z_{t,j,l}| \quad (5.7)$$

Η ανομοιότητα ανάμεσα σε δύο συστάδες P_i και P_t σε σχέση με το πεδίο ορισμού X_j είναι,

$$D_j(P_i, P_t) = \sum_{l=1}^{q_j} \Delta Z_{i,t,j,l} \quad (5.8)$$

Τέλος, η ανομοιότητα μεταξύ των συστάδων P_i και P_t σε σχέση με ολόκληρο το χώρο των χαρακτηριστικών γνωρισμάτων $X = X_1 \times X_2 \times \dots \times X_p$ δίνεται όπως,

$$D(P_i, P_t) = \sum_{j=1}^p D_j(P_i, P_t) \quad (5.9)$$

Λαμβάνοντας υπόψη τις εξισώσεις (5.7), (5.8) και (5.9),

$$D(P_i, P_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} |Z_{i,j,l} - Z_{t,j,l}| \quad (5.10)$$

Για την i -ασαφή συστάδα ορίζουμε την ελάχιστη ανομοιοτήτά της ως,

$$D_{\min}(P_i) = \min_{1 \leq t \leq c, t \neq i} \{D(P_i, P_t)\} \quad \forall i \in \{1, 2, \dots, c\} \quad (5.11)$$

Στη συνέχεια, ο διαχωρισμός της ασαφούς κατάτμησης ορίζεται σύμφωνα με το ακόλουθο τύπο,

$$s = \min_{1 \leq i \leq c} \{D_{\min}(P_i)\} \quad (5.12)$$

Για το σκοπό αυτό, συνδυάζοντας την (5.4) και (5.12), ο δείκτης εγκυρότητας συστάδων ορίζεται ως,

$$VI = \frac{\pi}{s} = \frac{\sum_{i=1}^c \frac{\sum_{k=1}^N (u_i(\mathbf{x}_k))^m \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n_i}}{\min_{1 \leq i \leq c} \{D_{\min}(P_i)\}} \quad (5.13)$$

5.3 Ο προτεινόμενος αλγόριθμος εκπαίδευσης

Βήμα 1: Ασαφή Συσταδοποίηση Εισόδου – Εξόδου

Έστω $\{\mathbf{z}_k = [\mathbf{x}_k, y_k]^T\}_{k=1}^N$ είναι ένα σύνολο από N μη επισημασμένα χαρακτηριστικά διανυσμάτων δεδομένων εισόδου – εξόδου (unlabeled feature input-output data vectors), με $\mathbf{x}_k \in \mathcal{R}^p$ και $y_k \in \mathcal{R}$. Έστω n , ένας αριθμός από ασαφής συστάδες. Δεδομένου ότι η συνάρτηση μέλους (membership function) του k -οστού διανύσματος στην l -οστή συστάδα συμβολίζεται με $u_l(\mathbf{z}_k)$ και η παράμετρος ασάφειας ως m , η μέθοδος Ασαφών c -Μέσων (fuzzy c -means) ελαχιστοποιεί την ακόλουθη αντικειμενική συνάρτηση (objective function) :

$$J_m = \sum_{k=1}^N \sum_{l=1}^n (u_l(\mathbf{z}_k))^m \|\mathbf{z}_k - \boldsymbol{\alpha}_l\|^2 \quad (5.14)$$

βάσει του επόμενου περιορισμού,

$$\sum_{l=1}^n u_l(\mathbf{z}_k) = 1 \quad \forall k \quad (5.15)$$

Τα κέντρα των συστάδων και οι βαθμοί συμμετοχής (membership degrees) που επιλύουν το παραπάνω πρόβλημα βελτιστοποίησης δίνονται ως,

$$\mathbf{a}_l = \frac{\sum_{k=1}^N (u_l(\mathbf{z}_k))^m \mathbf{z}_k}{\sum_{k=1}^N (u_l(\mathbf{z}_k))^m} \quad (5.16)$$

$$u_l(\mathbf{z}_k) = \frac{1}{\sum_{j=1}^n \left(\frac{\|\mathbf{z}_k - \mathbf{a}_l\|}{\|\mathbf{z}_k - \mathbf{a}_j\|} \right)^{2/(m-1)}} \quad (5.17)$$

Στη συνέχεια, η μέθοδος Ασαφών c -Μέσων αποτελείται από επαναληπτικές χρήσεις της εφαρμογής των δύο παραπάνω εξισώσεων.

Βήμα 2: Βέλτιστη ασαφής συσταδοποίηση

Σε αυτό το βήμα, τα κέντρα των συστάδων (cluster centers) \mathbf{v}_l ($1 \leq l \leq n$) θεωρούνται ως το νέο σύνολο δεδομένων. Το διάνυσμα \mathbf{v}_l είναι η προβολή του διανύσματος \mathbf{a}_l για το χώρο των εισόδων (input space),

$$\mathbf{v}_l = \underset{X}{proj}(\mathbf{a}_l) \quad (5.18)$$

Επειδή το διάνυσμα \mathbf{v}_l αναφέρεται σε ένα σύμπλεγμα από το αρχικό σύνολο δεδομένων, το βάρος της σημασίας (significance) του ορίζεται ως εξής,

$$r_i = \mathfrak{N}(U^l) / \sum_{j=1}^n \mathfrak{N}(U^j) \quad (5.19)$$

όπου $\mathfrak{N}(U^l)$ είναι η ασαφής πληθικότητα συνόλου της l -οστής συστάδας

$$\mathfrak{N}(U^l) = \sum_{k=1}^N u_l(z_k) \quad (1 \leq l \leq n) \quad (5.20)$$

Τότε, η αντικειμενική συνάρτηση (objective function) της μεθόδου σταθμισμένων Ασαφών c -Μέσων είναι [88],

$$J_r = \sum_{l=1}^n \sum_{i=1}^c r_l (h_i(\mathbf{v}_l))^q \|\mathbf{v}_l - \mathbf{v}_i\|^2 \quad (5.21)$$

όπου c είναι ο αριθμός των τελικών συστάδων, ο οποίος συμπίπτει με τον αριθμό των κρυφών κόμβων, \mathbf{v}_i ($1 \leq i \leq c$) είναι τα κέντρα συστάδων της τελικής κατάτμησης (final partition), $h_i(\mathbf{v}_l)$ είναι ο βαθμός συμμετοχής του διανύσματος \mathbf{v}_l στην i -οστή συστάδα, $q \in (1, \infty)$ είναι ένας παράγοντας για να ρυθμίσουμε το βαθμό συμμετοχής της επίδρασης στάθμισης, και r_l είναι το βάρος της σημασίας (significance) που έχει εκχωρηθεί στο \mathbf{v}_l .

Το πρόβλημα είναι να ελαχιστοποιηθεί το J_r υπό τον ακόλουθο περιορισμό,

$$\sum_{i=1}^c h_i(\mathbf{v}_l) = 1 \quad \forall l \quad (5.22)$$

Τα τελικά κέντρα και οι αντίστοιχες συναρτήσεις μέλους που λύνουν αυτό το πρόβλημα περιορισμού βελτιστοποίησης δίνονται από τις επόμενες εξισώσεις,

$$\mathbf{v}_i = \frac{\sum_{l=1}^n r_l (h_i(\mathbf{v}_l))^q \mathbf{v}_l}{\sum_{l=1}^n r_l (h_i(\mathbf{v}_l))^q} \quad (5.23)$$

$$h_i(\mathbf{v}_l) = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{v}_l - \mathbf{v}_i\|}{\|\mathbf{v}_l - \mathbf{v}_j\|} \right)^{2/(q-1)}} \quad (5.24)$$

Η (5.23) και η (5.24) αποτελούν μια επαναληπτική διαδικασία βελτιστοποίησης. Για να εκτελέσουμε τη βέλτιστη ασαφή συσταδοποίηση εφαρμόζουμε το δείκτη εγκυρότητας της εξίσωσης (5.13),

$$VI = \frac{\pi}{s} = \frac{\sum_{i=1}^c \frac{\sum_{k=1}^N (u_i(\mathbf{x}_k))^m \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n_i}}{\min_{1 \leq i \leq c} \{D_{min}(P_i)\}} \quad (5.25)$$

σε σχέση με τα κέντρα συστάδων \mathbf{v}_i ($1 \leq i \leq c$) και το διάνυσμα εκπαίδευσης \mathbf{x}_k ($1 \leq k \leq N$), όπου οι βαθμοί συμμετοχής υπολογίζονται ως,

$$u_i(\mathbf{x}_k) = 1 / \sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}} \quad (5.26)$$

Βήμα 3: Δίκτυο εκτίμησης των παραμέτρων

Ο αριθμός των κόμβων του δικτύου RBF είναι ίσος με τον αριθμό των συστάδων c , τα οποία παρήχθησαν στο προηγούμενο βήμα, ενώ τα κέντρα των κόμβων συμπίπτουν με τα κέντρα συστάδων \mathbf{v}_i ($1 \leq i \leq c$). Για να υπολογίσουμε τα πλάτη, χρησιμοποιούμε τη μέθοδο, η οποία αναπτύχθηκε από τον Νείρο και Τσεκούρα στο [66],

$$\sigma_i = \frac{2 * d_{max}^i}{3} \quad (1 \leq i \leq c) \quad (5.27)$$

όπου d_{max}^i δίνεται στην επόμενη εξίσωση,

$$d_{max}^i = \{ \max_{1 \leq k \leq N} \{ \|\mathbf{x}_k - \mathbf{v}_i\|^2 \} : \mathbf{x}_k \in C_i \text{ ώστε } u_i(\mathbf{x}_k) \geq \theta \} \quad (5.28)$$

όπου C_i είναι η i -οστή συστάδα και $u_i(\mathbf{x}_k)$ είναι ο βαθμός συμμετοχής του k -οστού δεδομένου εκπαίδευσης (training data vector) στην i -οστή τελική ασαφή συστάδα και θ ένας μικρός θετικός αριθμός, έτσι ώστε $\theta \ll 1$.

Τελικά, τα βάρη σύνδεσης των νευρώνων υπολογίζονται χρησιμοποιώντας την προσέγγιση των ελαχίστων τετραγώνων.

5.4 Αριθμητικά παραδείγματα

Στην ενότητα αυτή αξιολογούμε την προτεινόμενη μέθοδο και συγκρίνουμε τις επιδόσεις της με δύο διαφορετικά δίκτυα. Το πρώτο, που ονομάζεται RBFN 1, είναι ένα πρότυπο δίκτυο RBF, το οποίο χρησιμοποιεί τον αλγόριθμο *c*-Μέσων (*c*-Means) (βλέπε παράγραφο 2.2) για τον προσδιορισμό των κέντρων των πυρήνων στην εξ. (1.8), ενώ τα αντίστοιχα πλάτη ορίζονται ίσα με 1. Το δεύτερο δίκτυο, που ονομάζεται RBFN 2 έχει εκπαιδευτεί από τον αλγόριθμο Ασαφών *c*-Μέσων (βλέπε παράγραφο 2.3) και οι ακτινικές συναρτήσεις βάσεων συμπίπτουν με τις συναρτήσεις συμμετοχής των αντιστοίχων ασαφών συστάδων. Όλες οι προσομοιώσεις πραγματοποιήθηκαν με τη χρήση 10 διαφορετικών αρχικοποιήσεων (δηλαδή 10 «τρεξίματα» για κάθε πειραματική περίπτωση) για την μέθοδο Ασαφών *c*-Μέσων, η παράμετρος ασάφειας ορίστηκε ίση με 2. Για την εκτέλεση των πειραμάτων για τον προτεινόμενο αλγόριθμο, χρησιμοποιήσαμε το Xie-Beni (XB) δείκτη [99], το δείκτη SV που αναπτύχθηκε στο [93], και τον προτεινόμενο αλγόριθμο. Οι πειραματικές περιπτώσεις είναι τέσσερα γνωστά σύνολα δεδομένων, και η πειραματική ανάλυση δίνεται μέσα στις επόμενες υποενότητες.

5.4.1 Προσέγγιση μονοδιάστατης συνθετικής συνάρτησης

Η συνάρτηση είναι η ακόλουθη:

$$y = 0.6 \eta\mu(\pi x) + 0.3 \eta\mu(3\pi x) + 0.1 \eta\mu(5\pi x) \quad \text{με } x \in [-1, 1].$$

Όπως και στις αναφορές [70,75], δημιουργήθηκαν 200 ζευγάρια δεδομένων εισόδου-εξόδου. Το καλύτερο αποτέλεσμα για τα δεδομένα δοκιμής έρχεται με MSE=0.0034, το οποίο αναφέρεται, όταν $c=12$, ενώ για τα δεδομένα δοκιμής το καλύτερο αποτέλεσμα είναι MSE=0.0036 πάλι για $c=12$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας ήταν $m=2$ και $q=2$ και ο αρχικός αριθμός των συστάδων για την μέθοδο Ασαφών *c*-Μέσων επιλέχθηκε ως $n = 30$.

Ο πίνακας 5.1 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες εγκυρότητας. Σύμφωνα με τον πίνακα αυτό, ο βέλτιστος αριθμός των συστάδων για τον XB, SV και τον προτεινόμενο δείκτη ήταν 4, 11 και 12 αντίστοιχα. Το προκύπτον μέσο τετραγωνικό σφάλμα (MSE), για τους τρεις δείκτες απεικονίζεται στον

πίνακα 5.2. Επιπλέον, αυτός ο πίνακας συγκρίνει τις επιδόσεις των δικτύων που δημιουργούνται με την προτεινόμενη μέθοδο με τις επιδόσεις των άλλων μεθόδων που υπάρχουν στη βιβλιογραφία.

Πίνακας 5.1. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης μονοδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB Δείκτης [99]	SV Δείκτης [93]	Προτεινόμενος Δείκτης
2	0.0015	0.0729	0.00014
3	0.0035	0.0753	0.00010
4	0.0014	0.0237	6.9757e-05
5	0.0127	0.0326	3.8138e-05
6	0.0120	0.0208	3.0910e-05
7	0.0184	0.0257	2.7002e-05
8	0.0182	0.0227	1.6585e-05
9	0.0179	0.0191	2.0676e-05
10	0.0196	0.0178	1.8778e-05
11	0.0197	0.0161	1.3651e-05
12	0.0248	0.0175	1.0298e-05

Πίνακας 5.2. Συγκριτική ανάλυση από την άποψη του MSE.

Μοντέλο	Αρ. κρυφών κόμβων	MSE στα δεδομένα εκπαίδευσης	MSE στα δεδομένα δοκιμής
Pedrycz et al. [75]	≤ 33	0.0029 ± 0.001	0.0052 ± 0.004
	≤ 33	0.0033 ± 0.001	0.0035 ± 0.002
Park et al. [70]	6	$0.0045 \pm 6.2E-4$	0.0281 ± 0.014
Το μοντέλο μας	XB δείκτης	4	0.0538 ± 0.008
	SV δείκτης	11	0.0041 ± 0.002
	Προτεινόμενος δείκτης	12	0.0034 ± 0.001

5.4.2 Προσέγγιση δισδιάστατης συνάρτησης

Χρησιμοποιούμε μια δισδιάστατη συνάρτηση για να μελετήσουμε την απόδοση του προτεινόμενου δικτύου καθώς επίσης και μια σειρά από τις βασικές ιδιότητες του. Εξετάσαμε την ακόλουθη συνάρτηση,

$$y = f(x_1, x_2) = \frac{(x_1 - 2)(2x_1 + 1)}{1 + x_1^2} \frac{(x_2 - 2)(2x_2 + 1)}{1 + x_2^2} \quad \text{με } x_1, x_2 \in [-5, 5]$$

Η συνάρτηση αυτή προσεγγίζεται από ένα σύνολο 441 δειγμάτων, που εξήχθησαν από ένα πλέγμα 21 x 21 σημείων ισομερώς κατανεμημένων στο χώρο των εισόδων. Τα τρία δίκτυα συγκρίθηκαν όσον αφορά την ομαλοποιημένη τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (NRMSE):

$$NRMSE = \sqrt{\frac{\sum_{k=1}^n (y_k - \tilde{y}_k)^2}{\sum_{k=1}^n (y_k - \bar{y})^2}}$$

όπου \bar{y} είναι η μέση τιμή των δεδομένων εξόδου.

Πίνακας 5.3. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης δισδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB Δείκτης [99]	SV Δείκτης [93]	Προτεινόμενος δείκτης
2	0.00019	0.18897	0.01406
3	0.00010	0.05753	0.00796
4	0.00011	0.03432	0.00685
5	0.00014	0.02886	0.00500
6	0.00019	0.02308	0.00483
7	0.00016	0.02391	0.00469
8	0.00017	0.02193	0.00391
9	0.00015	0.03022	0.00269
10	0.00018	0.03144	0.00295

Το καλύτερο αποτέλεσμα έρχεται με NRMSE=0.703, το οποίο αναφέρεται όταν $c=9$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας ήταν $m=2$ και $q=2$ και ο αρχικός αριθμός των συστάδων για την μέθοδο Ασαφών c -Μέσων επιλέχθηκε ως $n = 30$.

Ο πίνακας 5.3 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες εγκυρότητας. Σύμφωνα με τον πίνακα αυτό, ο βέλτιστος αριθμός των συστάδων για τον XB, SV και τον προτεινόμενο δείκτη ήταν 3, 8 και 9 αντίστοιχα. Η προκύπτουσα ομαλοποιημένη τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (NRMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 5.4. Επιπλέον, αυτός ο πίνακας συγκρίνει

τις επιδόσεις των δικτύων που δημιουργούνται με την προτεινόμενη μέθοδο με τις επιδόσεις άλλων μεθόδων που υπάρχουν στη βιβλιογραφία.

Πίνακας 5.4. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές NRMSE.

Μοντέλο		Αριθμός κρυφών κόμβων	NRMSE
CFC [29]		6	0.838 ± 0.073
		9	0.812 ± 0.100
		12	0.823 ± 0.056
CFA [29]		6	0.812 ± 0.044
		8	0.713 ± 0.016
		9	0.708 ± 0.041
		10	0.705 ± 0.035
Το μοντέλο μας	XB δείκτης	3	$0.914 \pm 1.2E-6$
	SV δείκτης	8	$0.754 \pm 1.0E-4$
	Προτεινόμενος δείκτης	9	$0.703 \pm 1.2E-5$

5.4.3 Προσέγγιση στατικής δισδιάστατης συνάρτησης

Η συνάρτηση αυτή περιγράφεται από την ακόλουθη εξίσωση,

$$y = (1 + x_1^{-2} + x_2^{-1.5}), \quad 1 \leq x_1, x_2 \leq 5 \quad (5.29)$$

Χρησιμοποιώντας την (5.29) δημιουργήσαμε 50 δεδομένα εισόδου - εξόδου.

Ο αρχικός αριθμός των συστάδων για τον αλγόριθμο Ασαφών *c*-Μέσων επιλέχθηκε ως $n = 30$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας για τον αλγόριθμο Ασαφών *c*-Μέσων και για τη μέθοδο σταθμισμένων Ασαφών *c*-Μέσων ήταν $m=2$ και $q=2$, αντιστοίχως.

Ο πίνακας 5.5 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες εγκυρότητας. Σύμφωνα με τον πίνακα αυτό, ο βέλτιστος αριθμός των συστάδων για τον XB, SV και τον προτεινόμενο δείκτη ήταν 4, 13 και 10 αντίστοιχα. Η προκύπτουσα τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 5.6. Επιπλέον, αυτός ο πίνακας συγκρίνει τις

επιδόσεις των δικτύων που δημιουργούνται από την προτεινόμενη μέθοδο με τις επιδόσεις των άλλων μεθόδων που υπάρχουν στη βιβλιογραφία.

Πίνακας 5.5. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της προσέγγισης στατικής δισδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB δείκτης [99]	SV δείκτης [93]	Προτεινόμενος δείκτης
2	0.003200	1.120063	0.015848
3	0.003326	0.496843	0.010449
4	0.002890	0.228777	0.009558
5	0.005181	0.199086	0.006293
6	0.004700	0.154682	0.003777
7	0.005953	0.162363	0.004768
8	0.005709	0.155022	0.005269
9	0.004986	0.139234	0.005151
10	0.005379	0.132253	0.000949
11	0.008261	0.121901	0.001790
12	0.007476	0.127239	0.001057
13	0.006010	0.103375	0.002307
14	0.004726	0.113193	0.004527
15	0.009989	0.123951	0.001278

Πίνακας 5.6. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE.

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE
Lee et al. [48]		8	0.0148
Kim et al. [42]		3	0.0090
RBFN 1		4	0.7201
		10	0.4799
		13	0.3564
RBFN 2		4	0.6387
		10	0.4647
		13	0.3150
Το μοντέλο μας	XB δείκτης	4	0.0738 ± 1.3E-8
	SV δείκτης	13	0.0318 ± 2.6E-5
	Προτεινόμενος δείκτης	10	0.0306 ± 1.0E-4

5.4.4 Η Τρισδιάστατη Συνάρτηση

Σε αυτό το παράδειγμα, θεωρούμε την ακόλουθη μη-γραμμική τρισδιάστατη συνάρτηση,

$$y = (1 + x_1^{-2} + x_2^{-1.5} + x_3^{-1.0})^2 \text{ με } x_1, x_2, x_3 \in [1, 5].$$

Δημιουργήσαμε 200 ζεύγη δεδομένων εισόδου-εξόδου και τυχαία τα διαιρέσαμε σε δεδομένα εκπαίδευσης εμπεριέχοντας το 60% των δεδομένων και σε δεδομένα δοκιμών, τα οποία εμπεριέχουν το υπόλοιπο 40% από το σύνολο των δεδομένων. Ο αρχικός αριθμός των συστάδων για την μέθοδο Ασαφών *c*-Μέσων επιλέχθηκε ως $n = 30$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας ήταν $m=2$ και $q=2$, αντίστοιχα.

Πίνακας 5.7. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της συνθετικής τρισδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB δείκτης [99]	SV δείκτης [93]	Προτεινόμενος δείκτης
2	0.00208647	0.94510584	0.009190
3	0.00157383	0.34411418	0.007174
4	0.00137195	0.21514754	0.005514
5	0.00119763	0.14542452	0.004883
6	0.00165735	0.2224394	0.004007
7	0.00147823	0.16656289	0.002826
8	0.00151256	0.14253819	0.001415
9	0.00161277	0.15653192	0.001575
10	0.00144154	0.1322249	0.001474
11	0.00116679	0.08661824	0.000458
12	0.0033599	0.12995062	0.000594

Ο πίνακας 5.7 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες ισχύος. Σύμφωνα με αυτόν τον πίνακα ο βέλτιστος αριθμός των συστάδων για όλους τους δείκτες ήταν 11. Η προκύπτουσα τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 5.8. Επιπλέον, ο πίνακας αυτός συγκρίνει τις επιδόσεις των προτεινόμενων δικτύων με τις επιδόσεις άλλων μεθόδων που υπάρχουν στη βιβλιογραφία καθώς και με τα δίκτυα RBFN 1 και RBFN 2.

Πίνακας 5.8. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE.

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE
Park et. al. [71]	(δεδομένα εκπαίδευσης)	15	0.1490
		24	0.0973
		30	0.1881
	(δεδομένα δοκιμής)	15	0.3684
		24	0.2544
		30	0.4586
RBFN 1	(δεδομένα εκπαίδευσης)	15	1.605 ± 0.1558
		24	1.260 ± 0.0845
		30	1.143 ± 0.0956
	(δεδομένα δοκιμής)	15	1.853 ± 0.1657
		24	1.576 ± 0.0607
		30	1.517 ± 0.3241
RBFN 2	(δεδομένα εκπαίδευσης)	15	0.855 ± 0.0125
		24	0.782 ± 0.0251
		30	0.733 ± 0.0342
	(δεδομένα δοκιμής)	15	0.902 ± 0.0179
		24	0.879 ± 0.0253
		30	0.853 ± 0.0285
Το μοντέλο μας	XB δείκτης (δεδομένα εκπαίδευσης)	11	0.2051 ± 0.0265
		11	0.3135 ± 0.0257
	SV δείκτης (δεδομένα εκπαίδευσης)	11	0.2051 ± 0.0234
		11	0.3135 ± 0.0416
	Προτεινόμενος δείκτης (δεδομένα εκπαίδευσης)	11	0.2051 ± 0.0591
		11	0.3135 ± 0.0845

5.4.5 Abalone Δεδομένα

Το σύνολο δεδομένων περιλαμβάνει 4177 ζεύγη δεδομένων εισόδου-εξόδου (με 8 εισόδους), τα οποία πάρθηκαν από το UCI-Machine Repository. Τα πειράματα πραγματοποιήθηκαν για την 60–40% διάσπαση των δεδομένων σε δεδομένα εκπαίδευσης (training) και δοκιμής (testing), αντίστοιχα. Ο αρχικός αριθμός των συστάδων για την μέθοδο Ασαφών c -Μέσων επιλέχθηκε ως $n = 40$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας ήταν $m=2$ και $q=2$.

Πίνακας 5.9. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα Abalone.

Αριθμός Συστάδων	XB δείκτης [99]	SV δείκτης [93]	Προτεινόμενος δείκτης
2	0.0001023	0.0069726	0.0000096
3	0.0005577	0.0072756	0.0000056
4	0.0002686	0.0058264	0.0000039
5	0.0003105	0.0029522	0.0000029
6	0.0002555	0.0043748	0.0000045
7	0.0002595	0.0029929	0.0000016
8	0.0007015	0.0072709	0.0000055
9	0.0003903	0.0051361	0.0000011
10	0.0000872	0.0044911	0.0000019
11	0.0006241	0.0036482	0.0000010
12	0.0006724	0.0047256	0.0000014
13	0.0033833	0.0042894	0.0000010
14	0.000735	0.0023515	0.0000002
15	0.004946	0.0045681	0.0000004
16	0.0006241	0.0017903	0.0000005
17	0.0023086	0.0019589	0.0000011
18	0.0025219	0.0017524	0.0000010
19	0.0026223	0.0061961	0.0000009
20	0.0019247	0.0015968	0.0000004

Ο πίνακας 5.9 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες βέλτιστης συσταδοποίησης. Σύμφωνα με αυτόν τον πίνακα ο βέλτιστος αριθμός των συστάδων για τον XB [99], SV [93] και τον προτεινόμενο δείκτη ήταν 10, 20 και 14 αντίστοιχα. Η προκύπτουσα τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 5.10, όπου μπορούμε εύκολα

να ελέγξουμε ότι ο προτεινόμενος αλγόριθμος επιτυγχάνει το καλύτερο αποτέλεσμα με το μικρότερο αριθμό των κρυφών κόμβων.

Πίνακας 5.10. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE .

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE	
Pedrycz et al. [75]	(δεδομένα εκπαίδευσης)	m=2	≤ 33	1.9613
		m=2.5	≤ 18	2.0366
		m=4	≤ 33	2.0904
	(δεδομένα δοκιμής)	m=2	≤ 33	2.1910
		m=2.5	≤ 18	2.2119
		m=4	≤ 33	2.1549
RBFN 1	(δεδομένα εκπαίδευσης)	10		2.506 ± 0.437
		14		2.448 ± 0.243
		20		2.206 ± 0.175
	(δεδομένα δοκιμής)	10		2.739 ± 0.790
		14		2.623 ± 0.541
		20		2.346 ± 0.208
RBFN 2	(δεδομένα εκπαίδευσης)	10		2.578 ± 0.126
		14		2.504 ± 0.147
		20		2.317 ± 0.023
	(δεδομένα δοκιμής)	10		2.807 ± 0.111
		14		2.704 ± 0.096
		20		2.414 ± 0.038
Το μοντέλο μας	XB δείκτης (δεδομένα εκπαίδευσης)	10		2.6901 ± 0.080
		10		2.9338 ± 0.096
	SV δείκτης (δεδομένα εκπαίδευσης)	20		2.0825 ± 0.046
		20		2.2114 ± 0.120
	Προτεινόμενος δείκτης (δεδομένα εκπαίδευσης)	14		2.0226 ± 0.043
		14		2.1511 ± 0.023

5.5 Συμπεράσματα

Στο παρόν κεφάλαιο διερευνήσαμε πώς η ασαφή συσταδοποίηση εισόδου-εξόδου μπορεί να συνεργαστεί με τη βέλτιστη ασαφή συσταδοποίηση προκειμένου να εκπαιδεύσουν αποτελεσματικά νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (RBF).

Αρχικά, εφαρμόσαμε την ασαφή συσταδοποίηση στο χώρο εισόδων-εξόδου, και τα προκύπτοντα κέντρα συστάδων προβλήθηκαν στο χώρο των εισόδων. Για κάθε κέντρο συστάδας, αποδώσαμε ένα συντελεστή στάθμισης, ο οποίος σχετίζεται με το μέγεθος (δηλαδή πληθικότητας) της αντίστοιχης συστάδας. Εμείς λοιπόν, αντιμετωπίσαμε τα κέντρα συστάδων μαζί με τα αντίστοιχα βάρη ως νέο σύνολο δεδομένων και πραγματοποιήσαμε τη βέλτιστη ασαφή συσταδοποίηση με την εφαρμογή της μεθόδου σταθμισμένων (weighted) Ασαφών c -Μέσων.

Για τη διεξαγωγή της βέλτιστης συσταδοποίησης έχουμε αναπτύξει ένα νέο δείκτη εγκυρότητας. Για το σκοπό αυτό, το μέγεθος του δικτύου και οι παράμετροι σχεδιασμού του, είναι άμεσα αποτελέσματα της πιο πάνω ιεραρχικής ανάλυσης συστάδων.

Η μέθοδος δοκιμάστηκε σε μια σειρά από πειραματικές μελέτες, όπου η αποτελεσματικότητά της επαληθεύτηκε επιτυχώς.

Ανάπτυξη Μεθόδου Υβριδικής Ασαφούς Συσταδοποίησης για τον Αποτελεσματικό Σχεδιασμό Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

6.1 Εισαγωγή

Νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (RBFNN) έχουν εφαρμοστεί και αξιολογηθεί κάτω από διάφορες παραλλαγές σε πολλούς επιστημονικούς τομείς [23,25,46,51,56,68,80]. Η εκπαίδευση (training) των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBFNN) υλοποιείται μέσω της εκτίμησης τριών ειδών παραμέτρων και συγκεκριμένα των κέντρων και των πλατών (widths) των συναρτήσεων ακτινικής βάσης καθώς και των βαρών σύνδεσης των νευρώνων (interconnecting weights). Ενώ ο υπολογισμός των βαρών σύνδεσης είναι μια τυποποιημένη διαδικασία, η εκτίμηση των παραμέτρων συναρτήσεων βάσης είναι ένα κρίσιμο και προκλητικό πρόβλημα [14,32]. Μια προσέγγιση ικανή να εξετάσει αποτελεσματικά αυτό το πρόβλημα είναι η εφαρμογή της ασαφούς ανάλυσης συστάδων (fuzzy cluster analysis) [5,30,53,64,66,72,81]. Η ασαφής συσταδοποίηση παράγει ένα διαμερισμό ενός συνόλου δεδομένων σε διάφορες συστάδες. Αυτός ο διαμερισμός κατέχει τις ακόλουθες δύο ιδιότητες [5,67]: (α) η ομοιογένεια μέσα στις συστάδες, δηλ. δεδομένα που ανήκουν στην ίδια συστάδα πρέπει να εμφανίζουν υψηλή ομοιότητα, και (β) η ετερογένεια μεταξύ των συστάδων, δηλ. δεδομένα που ανήκουν σε διαφορετικές συστάδες πρέπει να είναι όσο το δυνατόν πιο ανόμοια. Η έννοια "της ομοιότητας" διευκρινίζεται σύμφωνα με κάποιο μέτρο απόστασης που καθορίζει τις μετρικές ιδιότητες του χώρου των χαρακτηριστικών γνωρισμάτων (feature space). Δεδομένου ότι, στις περισσότερες από τις περιπτώσεις τα

δεδομένα είναι διανύσματα με πραγματικές τιμές, η Ευκλείδεια (Euclidean) απόσταση (distance) ή οποιεσδήποτε από την παραλλαγή της (δηλ. η απόσταση Minkowski) μπορεί να χρησιμοποιηθεί για να υπολογίσει την ανομοιότητα (και επομένως η ομοιότητα) μεταξύ των διανυσματικών δεδομένων.

Κατά την εφαρμογή της ασαφούς συσταδοποίησης (fuzzy clustering) στο σχεδιασμό νευρωνικών δικτύων RBF μπορούν να διακριθούν δύο φάσεις [66,72]. Η πρώτη φάση εξετάζει την άμεση συσταδοποίηση των αριθμητικών δεδομένων, και η δεύτερη ενσωματώνει τις παραχθείσες ασαφείς συστάδες στο μηχανισμό εκπαίδευσης. Οι ασαφείς συστάδες συμβάλλουν κατά τρόπο απλό στον υπολογισμό των κέντρων των συναρτήσεων ακτινικής βάσης. Εντούτοις, είναι εξίσου σημαντική και η εκτίμηση των αντίστοιχων πλατών (widths) [6,14]. Πολλοί ερευνητές λύνουν αυτό το πρόβλημα με την έκφραση των συναρτήσεων βάσης από την άποψη των ασαφών συναρτήσεων συμμετοχής, όπως καθορίζονται μέσω της ανάλυσης συστάδων (cluster analysis) [71,72,75,81]. Το κύριο μειονέκτημα αυτής της τεχνικής είναι ότι λαμβάνει υπόψη έναν ενιαίο τύπο συναρτήσεων βάσης, αν και αυτό μπορεί να μην είναι πάντα το ιδανικότερο [31]. Αφ' ετέρου, οι κλασικές προσεγγίσεις περιλαμβάνουν μόνο τις θέσεις των κέντρων των συστάδων [6,14,32,60,82] και, επομένως, η υποκείμενη δομή δεδομένων δεν εξετάζεται καθόλου.

Με την προτεινόμενη μέθοδο, παρέχεται μια διαφορετική προσέγγιση όσον αφορά αυτό το πρόβλημα. Συγκεκριμένα, προτείνουμε μια προσέγγιση που πραγματοποιεί την εκτίμηση των κέντρων και των πλατών (widths) της ακτινικής συνάρτησης βάσης ως άμεσο αποτέλεσμα της διαδικασίας συσταδοποίησης. Για να παραχθούν οι συστάδες έχει αναπτυχθεί ένα νέο υβριδικό σχήμα που συνδυάζει την ασαφή (fuzzy) και διακριτή (crisp) συσταδοποίηση (clustering) σε μια ομοιόμορφη μορφή. Με την εξισορρόπηση αυτών των δύο διαφορετικών αποτελεσμάτων, βρίσκουμε ένα γρήγορο και ακριβές αλγοριθμικό πλαίσιο, το οποίο διατυπώνει μια σύμπραξη μεταξύ των ασαφών συνόλων, των διακριτών συνόλων και του νευρωνικού υπολογισμού. Δεδομένου ότι κάθε συστάδα αντιστοιχεί σε μια ορισμένη συνάρτηση ακτινικής βάσης, χρησιμοποιούμε τις πληροφορίες που παρέχονται από τον υβριδικό διαμερισμό για να προσδιοριστούν αποτελεσματικά τα κέντρα των συναρτήσεων βάσης καθώς και τα αντίστοιχα πλάτη (widths).

6.2 Ο προτεινόμενος αλγόριθμος εκπαίδευσης

6.2.1 Υβριδική ασαφής συσταδοποίηση (hybrid fuzzy clustering)

Ο αλγόριθμος των c -Μέσων (c -Means) [36] είναι πολύ ευαίσθητος στην αρχικοποίηση αλλά είναι μια γρήγορη διαδικασία, ενώ η μέθοδος των Ασαφών c -Μέσων (FCM) [67] είναι σε θέση να μειώσουν την εξάρτηση στην αρχικοποίηση αλλά παραμένει μια αργή διαδικασία. Σε μια πρόσφατη δημοσίευση [92], έχουμε αναπτύξει έναν διανυσματικό ασαφή αλγόριθμο κβαντοποίησης εκπαίδευσης για τους στόχους συμπίεσης εικόνας, οι οποίοι συνδύασαν τη μέθοδο των c -Μέσων (c -Means) και τη μέθοδο των Ασαφών c -Μέσων (FCM). Η βασική ιδέα αυτής της μεθόδου προέρχεται από εκείνο τον αλγόριθμο εκπαίδευσης και χρησιμοποιεί την ακόλουθη αντικειμενική συνάρτηση,

$$J_H = \theta \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 + (1-\theta) \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (6.1)$$

όπου το c είναι ο αριθμός συστάδων, $\theta \in [0,1)$, και $u_{ik} \in [0,1]$ είναι ο βαθμός συμμετοχής του k -διανύσματος εκπαίδευσης στη i -συστάδα. Είναι εύκολα αντιληπτό ότι, όταν $\theta = 0$, μετασχηματίζεται η αντικειμενική συνάρτηση στη αντικειμενική συνάρτηση των Ασαφών c -Μέσων (FCM) με $m = 2$, και όταν $\theta = 1$ γίνεται η αντικειμενική συνάρτηση των c -Μέσων (c -Means). Επομένως, η συνάρτηση J_H παρουσιάζει μια υβριδική δομή που επιτρέπει τη μετατροπή από ασαφείς σε διακριτούς όρους ανάλογα με την τιμή της παραμέτρου θ . Το πρόβλημα είναι να ελαχιστοποιηθεί η συνάρτηση J_H κάτω από τον ακόλουθο περιορισμό,

$$\sum_{i=1}^c u_{ik} = 1, \quad \forall k \quad (6.2)$$

Οι βαθμοί συμμετοχής (membership functions) και τα κέντρα συστάδων που λύνουν αυτό το περιορισμένο πρόβλημα βελτιστοποίησης υπολογίζονται από τις επόμενες εξισώσεις,

$$u_{ik} = \frac{2 + (c-2)\theta}{2(1-\theta)} \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_j\|}{\|\mathbf{x}_k - \mathbf{v}_i\|} \right)^2} - \frac{\theta}{2(1-\theta)} \quad (6.3)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n [\theta u_{ik} + (1-\theta)(u_{ik})^2] \mathbf{x}_k}{\sum_{k=1}^n [\theta u_{ik} + (1-\theta)(u_{ik})^2]} \quad (6.4)$$

Η αναλυτική απόδειξη των εξ. (6.3) και (6.4) δίνεται παρακάτω.

Πρώτα εξετάζεται η εξ. (6.3). Λαμβάνοντας υπόψη τον περιορισμό (6.2), χρησιμοποιώντας τους lagrange πολλαπλασιαστές, η αντικειμενική συνάρτηση (6.1) τροποποιείται ως εξής,

$$F(u_{ik}, \lambda_k) = \theta \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 + (1-\theta) \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 - \sum_{k=1}^n \lambda_k \left(\sum_{i=1}^c u_{ik} - 1 \right) \quad (6.5)$$

Η μερική παράγωγος της $F(u_{ik}, \lambda_k)$ ως προς το u_{ik} είναι,

$$\frac{\partial F(u_{ik}, \lambda_k)}{\partial u_{ik}} = \theta \|\mathbf{x}_k - \mathbf{v}_i\|^2 + 2(1-\theta)u_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 - \lambda_k \quad (6.6)$$

Θέτοντας την ανωτέρω ποσότητα ίση με μηδέν και λύνοντας ως προς u_{ik} παίρνουμε,

$$u_{ik} = \frac{\lambda_k}{2(1-\theta)} \frac{1}{\|\mathbf{x}_k - \mathbf{v}_i\|^2} - \frac{\theta}{2(1-\theta)} \quad (6.7)$$

Αντικαθιστώντας την (6.7) στην εξ. (6.2) και λύνοντας ως προς $\lambda_k / (2(1-\theta))$,

$$\frac{\lambda_k}{2(1-\theta)} = \frac{2 + (c-2)\theta}{2(1-\theta)} \frac{1}{\sum_{j=1}^c \left(\frac{1}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^2} \quad (6.8)$$

Τέλος, αντικαθιστώντας το τελευταίο αποτέλεσμα (6.7) παράγεται εύκολα η εξ. (6.3).

Η μερική παράγωγος της J_H εξ. (6.1) ως προς τα κέντρα \mathbf{v}_i είναι,

$$\frac{\partial J_H}{\partial \mathbf{v}_i} = \theta \sum_{k=1}^n u_{ik} (-2)(\mathbf{x}_k - \mathbf{v}_i) + (1 - \theta) \sum_{k=1}^n (u_{ik})^2 (-2)(\mathbf{x}_k - \mathbf{v}_i) \quad (6.9)$$

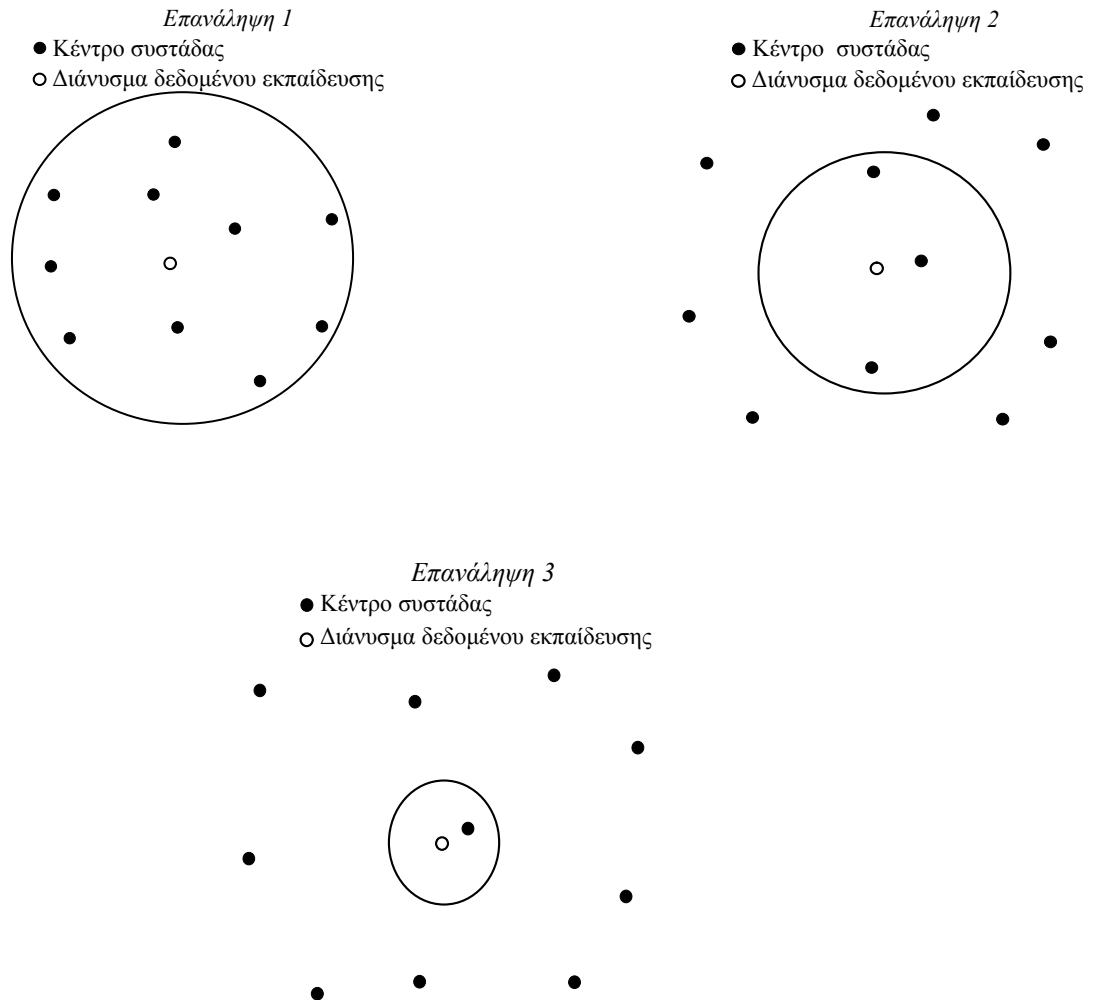
Θέτοντας την (6.9) ίση με το μηδέν και λύνοντας ως προς \mathbf{v}_i προκύπτει η εξ. (6.4). Εξετάζοντας την εξ. (6.3), είναι δυνατό για μερικούς βαθμούς συμμετοχής να ληφθεί μηδέν ή αρνητικές τιμές. Για να αναλυθεί περαιτέρω αυτή η κατάσταση, θέτοντας $u_{ik} \leq 0$ και λύνοντας ως προς την απόσταση $\|\mathbf{x}_k - \mathbf{v}_i\|^2$ έχουμε,

$$\|\mathbf{x}_k - \mathbf{v}_i\|^2 \geq \frac{2 + (c - 2)\theta}{\theta} \frac{1}{\sum_{j=1}^c \left(\frac{1}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^2} \quad (6.10)$$

Κατά συνέπεια, όταν ο όρος (6.10) είναι αληθής, ο βαθμός συμμετοχής του διανύσματος εκπαίδευσης \mathbf{x}_k στο κέντρο \mathbf{v}_i είναι αρνητικός ή ίσος με μηδέν. Δεδομένου ότι οι αρνητικοί βαθμοί συμμετοχής δεν έχουν καμία φυσική έννοια, είναι ασφαλές να υποτεθεί ότι το κέντρο \mathbf{v}_i δεν επηρεάζεται από το διάνυσμα \mathbf{x}_k . Αυτό το σημαντικό αποτέλεσμα επιβάλλει μερικούς περιορισμούς στη διαδικασία εκπαίδευσης, οι οποίοι συσχετίζονται με την εφαρμογή των εξ. (6.3) και (6.4).

Μελετάται αρχικά η υλοποίηση της εξ. (6.3). Για να γίνει αντιληπτό πώς αυτή η εξίσωση μπορεί να εφαρμοστεί σωστά, απεικονίζεται το πρόβλημα από την άποψη του k -διανύσματος εισόδου \mathbf{x}_k . Συγκεκριμένα, ορίζεται το σύνολο T_k ως σύνολο των κέντρων συστάδων που επηρεάζονται από το διάνυσμα \mathbf{x}_k . Αρχικά, το σύνολο T_k περιλαμβάνει όλα τα κέντρα συστάδων και η πληθικότητα συνόλου (cardinality) είναι: $\aleph(T_k^{(0)}) = c$. Καθώς η επαναληπτική διαδικασία προχωρά, μερικοί από τους βαθμούς συμμετοχής γίνονται αρνητικοί ή ίσοι με μηδέν. Θέτοντας όλους αυτούς τους βαθμούς ίσους με μηδέν, και τα αντίστοιχα κέντρα συστάδων αφαιρούνται από το σύνολο T_k . Επομένως, στη ν -επανάληψη, το διάνυσμα \mathbf{x}_k ορίζεται από $\aleph(T_k^{(\nu)}) < c$ συστάδες. Ο επόμενος ορισμός περιγράφει τις έννοιες της ασαφούς (fuzzy) και διακριτής (crisp) κατάστασης (mode).

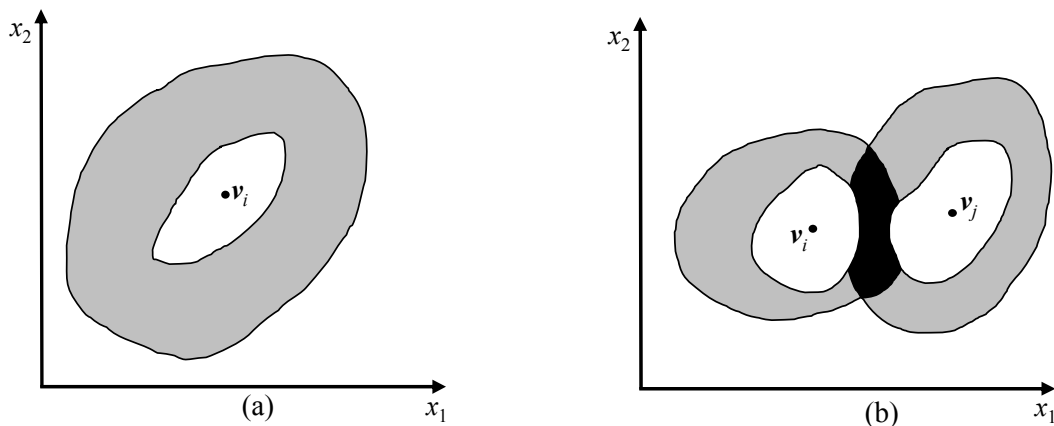
Ορισμός 1. Ένα διάνυσμα εκπαίδευσης x_k είναι στην ασαφή κατάσταση (*fuzzy mode*) εάν $\aleph(T_k) > 1$ και στην διακριτή κατάσταση (*crisp mode*) εάν $\aleph(T_k) = 1$.



Σχήμα 6.1. Χρήση τριών επαναλήψεων για να περιγράψει τη μετάβαση από ασαφή σε διακριτή κατάσταση ενός διανύσματος δεδομένων εκπαίδευσης που περιβάλλεται από 10 κέντρα συστάδων, όπου η περιοχή καθορίζει το σύνολο κέντρων συστάδων που επηρεάζονται από το διάνυσμα εκπαίδευσης σε κάθε επανάληψη.

Δεδομένου ότι ο αριθμός επανάληψης αυξάνεται, οι πληθικότητες (*cardinalities*) των συνόλων T_k ($1 \leq k \leq n$) μειώνονται, ενώ ένα σημαντικό ποσό διανυσμάτων

εκπαίδευσης μεταφέρεται στην διακριτή κατάσταση (crisp mode). Μια γραφική εξήγηση αυτής της στρατηγικής μετάβασης δίνεται στο σχήμα 6.1. Για λόγους απλότητας χρησιμοποιήθηκαν μόνο τρεις επαναλήψεις. Στην πρώτη επανάληψη το διάνυσμα εκπαίδευσης ορίζεται σε όλα τα κέντρα συστάδων. Καθώς κινούμαστε προς την επόμενη επανάληψη τα κέντρα ενημερώνουν τις θέσεις τους που παράγουν μια διαφορετική τοπολογία σύμφωνα με την οποία, το διάνυσμα εκπαίδευσης έχει επιπτώσεις σε μόνο τρία κέντρα. Στην τελευταία επανάληψη η διαδικασία εκπαίδευσης τελειώνει και το διάνυσμα εκπαίδευσης ορίζεται μόνο σε μια συστάδα.



Σχήμα 6.2. (α) Ασαφής (σκιασμένη περιοχή) και διακριτοί (λευκή περιοχή) τομείς μιας υβριδικής συστάδας, (β) ένα στιγμιότυπο της διαδικασίας εκπαίδευσης που παρουσιάζει δύο αλληλεπιδρούσες υβριδικές συστάδες, όπου το μαύρο τμήμα αντιστοιχεί στην κοινή ασαφή περιοχή τους.

Εντούτοις, πρέπει έντονα να υπογραμμιστεί ότι δεν μεταφέρονται όλα τα διανύσματα εκπαίδευσης στην διακριτή κατάσταση. Αυτό είναι μια σημαντική απαίτηση της αντικειμενικής συνάρτησης (6.1). Κατά συνέπεια, κατά τη διάρκεια και μετά από την διαδικασία εκπαίδευσης, πολλά διανύσματα εκπαίδευσης παραμένουν στην ασαφή κατάσταση (fuzzy mode) ακόμα.. Για να απεικονιστεί καλύτερα αυτή η κατάσταση μελετάται το πρόβλημα από την άποψη μιας συστάδας. Μια ασαφής συστάδα συνθέτει ένα υβριδικό αντικείμενο, το οποίο αποτελείται από μια διακριτή και από μία ασαφή περιοχή. Το σχήμα 6.2(a) απεικονίζει τη δομή μιας υβριδικής συστάδας στο δισδιάστατο χώρο. Το σχήμα 6.2(b) παρουσιάζει ένα στιγμιότυπο της διαδικασίας εκπαίδευσης, όπου

δύο υβριδικές συστάδες αλληλεπιδρούν από την άποψη της κοινής ασαφής περιοχής τους. Τα διανύσματα εκπαίδευσης που βρίσκονται σε αυτήν την περιοχή επηρεάζουν και τα δύο κέντρα.

Με βάση την ανωτέρω ανάλυση και να λαμβάνοντας υπόψη την ανίσωση εξ. (6.10), εξάγεται το συμπέρασμα ότι τα σύνολα T_k ($1 \leq k \leq n$) και οι βαθμοί συμμετοχής πρέπει να ενημερωθούν ως εξής,

$$T_k^{(v)} = \left\{ \mathbf{v}_i \in T_k^{(v-1)} : \|\mathbf{x}_k - \mathbf{v}_i\|^2 < \frac{2 + (\aleph(T_k^{(v-1)}) - 2)\theta}{\theta} \frac{1}{\sum_{\mathbf{v}_j \in T_k^{(v-1)}} \left(\frac{1}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^2} \right\} \quad (6.11)$$

$$u_{ik} = \frac{2 + (\aleph(T_k^{(v)}) - 2)\theta}{2(1 - \theta)} \frac{1}{\sum_{\mathbf{v}_j \in T_k^{(v)}} \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^2} - \frac{\theta}{2(1 - \theta)} \quad (6.12)$$

όπου $\mathbf{v}_i \in T_k^{(v)}$. Σημειώνεται ότι ο αριθμός c στις εξ. (6.3) και (6.10) έχει αντικατασταθεί από τις ασαφείς πληθικότητες συνόλου (cardinalities) $\aleph(T_k^{(v-1)})$ και $\aleph(T_k^{(v)})$ στις εξ. (6.11) και (6.12). Οι εξ. (6.11) και (6.12) δείχνουν άμεσα ότι η υπολογιστική πολυπλοκότητα μειώνεται επειδή ο αριθμός υπολογισμών απόστασης μειώνεται επίσης. Μια αναλυτική παρουσίαση αυτού του ζητήματος δίνεται αργότερα σε αυτό το κεφάλαιο.

Μόλις ένα διάνυσμα εκπαίδευσης τίθεται στην διακριτή κατάσταση (crisp mode), έχει επιπτώσεις μόνο σε ένα κέντρο συστάδων (βλ. το σχήμα 6.1 (γ)). Σε αυτήν την περίπτωση, οι βαθμοί συμμετοχής υπολογίζονται με τη χρησιμοποίηση του ακόλουθου όρου κοντινότερου γείτονα,

$$u_{ik} = \begin{cases} 1, & \text{αν } \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \min_{1 \leq j \leq c} \{\|\mathbf{x}_k - \mathbf{v}_j\|^2\} \\ 0, & \text{αλλιώς} \end{cases} \quad (6.13)$$

Αυτός ο όρος υπονοεί ότι είναι δυνατό για ένα διάνυσμα εκπαίδευσης να επανεκχωρηθεί σε μια άλλη συστάδα. Η πιθανότητα για αυτήν την κατάσταση τείνει σε μηδέν εάν το διάνυσμα εκπαίδευσης ανήκει στον διακριτό τομέα μιας υβριδικής συστάδας και βρίσκεται κλειστός στο αντίστοιχο κέντρο. Εντούτοις, αυτό είναι πολύ πιθανό να συμβεί εάν το διάνυσμα εκπαίδευσης είναι κοντά στο όριο του ασαφούς τομέα της συστάδας. Εντούτοις, η χρησιμοποίηση του όρου u_{ik} εξ. (6.13) είναι πολύ σημαντικός επειδή διαφορετικά θα περιοριζόταν ο ανταγωνισμός μεταξύ των συστάδων μόνο για τις συγκεκριμένες περιοχές (δηλ. ασαφείς περιοχές) στο χώρο χαρακτηριστικών γνωρισμάτων (feature space), κάτι που θα οδηγούσε στα ανεπιθύμητα τοπικά ελάχιστα.

Εξετάζεται τώρα πως η εξ. (6.4) πρέπει να τροποποιηθεί. Στο προηγούμενο βήμα όλοι οι αρνητικοί βαθμοί συμμετοχής (membership degrees) τέθηκαν ίσοι με μηδέν. Αυτοί οι βαθμοί αντιστοιχούν στα κέντρα συστάδων που δεν ικανοποιούν τον όρο στην εξ. (6.11). Εντούτοις, μια τέτοια διευθέτηση παραβιάζει τον περιορισμό της εξ. (6.2) επειδή το άθροισμα των υπόλοιπων βαθμών συμμετοχής θα είναι μεγαλύτερο από την μονάδα. Επομένως, η ακόλουθη κανονικοποίηση (normalization) πραγματοποιείται,

$$\hat{u}_{ik} = \frac{u_{ik}}{\sum_{j=1}^c u_{jk}} \quad (1 \leq i, j \leq c) \quad (6.14)$$

και η εξ. (6.4) τροποποιείται ανάλογα ως,

$$\mathbf{v}_i = \frac{\sum_{k=1}^n [\theta \hat{u}_{ik} + (1 - \theta)(\hat{u}_{ik})^2] \mathbf{x}_k}{\sum_{k=1}^n [\theta \hat{u}_{ik} + (1 - \theta)(\hat{u}_{ik})^2]}, \quad (1 \leq i \leq c) \quad (6.15)$$

Στην εξ. (6.14) χρησιμοποιήθηκαν όλοι οι βαθμοί συμμετοχής, δεδομένου ότι εκείνοι που τέθηκαν ίσοι με μηδέν δεν θα επηρεάσουν το τελικό αποτέλεσμα. Ως εκ τούτου, κατά τη διάρκεια της διαδικασίας εκπαίδευσης, στο διάνυσμα \mathbf{x}_k ορίζεται ως μηδέν ή ως θετικοί βαθμοί συμμετοχής όσον αφορά όλες τις συστάδες, το άθροισμα των οποίων είναι η μονάδα, εκπληρώνοντας κατ' αυτό τον τρόπο το βασικό περιορισμό της εξ. (6.2). Παρακάτω περιγράφεται ο προτεινόμενος αλγόριθμος συσταδοποίησης.

Ο προτεινόμενος υβριδικός αλγόριθμος συσταδοποίησης

Επέλεξε τις τιμές για το c και θ .

Τυχαία αρχικοποίησε τα κέντρα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$.

Θέσε $v=0 \quad \forall k : \aleph(T_k^{(0)})=c$, και $T_k^{(0)} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c\}$.

Βήμα 1) Θέσε $v=v+1$.

Βήμα 2) Χρησιμοποίησε την εξ. (6.11) για να ενημερωθούν τα σύνολα $T_k^{(v)}$ και οι ασαφείς πληθικότητες συνόλου (cardinalities) $\aleph(T_k^{(v)})$ ($1 \leq k \leq n$).

Βήμα 3) Εάν $\aleph(T_k^{(v)})=1$ υπολόγισε τους βαθμούς συμμετοχής u_{ik} ($1 \leq k \leq n; 1 \leq i \leq c$) χρησιμοποιώντας την εξ. (6.13) αλλιώς χρησιμοποίησε την εξ. (6.12).

Βήμα 4) Εάν $u_{ik} < 0$ ($1 \leq k \leq n; 1 \leq i \leq c$) θέσε $u_{ik} = 0$.

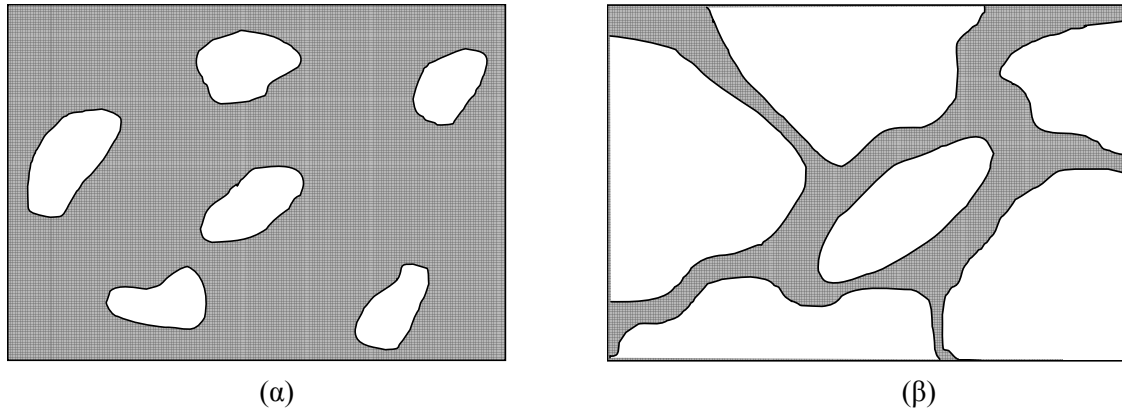
Βήμα 5) Χρησιμοποίησε την εξ. (6.14) για να υπολογίσεις τους ομαλοποιημένους βαθμούς συμμετοχής.

Βήμα 6) Χρησιμοποίησε την εξ. (6.15) για να ενημερωθούν τα κέντρα συστάδων.

Βήμα 7) Εάν δεν υπάρχει καμία αξιοπρόσεχτη αλλαγή για τα κέντρα των συστάδων σταμάτα, αλλιώς πήγαινε στο βήμα 1.

Ο ανωτέρω αλγόριθμος παράγει ένα σύνολο c συστάδων $\{C_1, C_2, \dots, C_c\}$ οι αντιπρόσωποι των οποίων είναι τα διανύσματα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$. Όπως αναφέρθηκε νωρίτερα, λόγω της φύσης της διαδικασίας εκπαίδευσης, ο προκύπτον διαμερισμός δεν είναι ένας καθαρός ασαφής ή διακριτός διαμερισμός αλλά μια υβριδική δομή που συνδυάζει αυτές τις δύο περιπτώσεις. Προκύπτει ότι το ποσοστό μετάβασης από την ασαφή (fuzzy) σε διακριτή (crisp) κατάσταση (mode) ελέγχεται ιδιαίτερα από την παράμετρο θ . Μια μικρή τιμή στην παράμετρο θ υπονοεί ότι η πιο κυρίαρχη επίδραση παρέχεται από τη μέθοδο των Ασαφών c -Μέσων (FCM) και επομένως, το ποσοστό μετάβασης είναι πάρα πολύ μικρό. Οι μεγάλες τιμές για την παράμετρο θ υπονοούν ότι η αρχιτεκτονική του δικτύου κυριαρχείται από την διακριτή κατάσταση (crisp mode). Σε αυτήν την περίπτωση, η επίδραση μετάβασης είναι πάρα πολύ επιθετική και θα

αναγκάσει πολλά διανύσματα εκπαίδευσης να μεταφερθούν στην διακριτή κατάσταση. Το σχήμα 6.3 απεικονίζει αυτά τα δύο διαφορετικά αποτελέσματα, όπου μπορούμε να διακρίνουμε τον αντίκτυπο της παραμέτρου θ .



Σχήμα 6.3. Υβριδικός διαμερισμός με έξι συστάδες, όπου οι ασαφείς (σκιασμένη περιοχή) και διακριτοί (λευκές περιοχές) όροι συνυπάρχουν για: (α) μικρή τιμή θ , και (β) μεγάλη τιμή θ .

Παρατήρηση 1: Ο παραπάνω αλγόριθμος δημιουργεί ένα σύνολο c υβριδικών συστάδων $\{C_1, C_2, \dots, C_c\}$ των οποίων οι εκπρόσωποι είναι τα κέντρα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$. Η δομή των ομάδων τυπικά περιγράφεται ως: $C_i = \Xi_i \cup \{\Phi_i / \hat{u}_{ik} : k = 1, 2, \dots, n\}$ για $1 \leq i \leq c$. Όπως αναφέρθηκε νωρίτερα, λόγω της φύσης της διαδικασίας εκπαίδευσης, η προκύπτουσα διαμέριση δεν είναι μια καθαρή ασαφής ή διακριτή διχοτόμηση, αλλά μια υβριδική δομή που συνδυάζει αυτές τις δύο περιπτώσεις. Αποδεικνύεται ότι το ποσοστό μετάβασης από ασαφή σε διακριτή κατάσταση είναι ιδιαίτερα ελεγχόμενη από την παράμετρο θ . Μια μικρή τιμή του θ σημαίνει ότι η πιο κυρίαρχη επίδραση παρέχεται από τη μέθοδο Ασαφών c -Μέσων και ως εκ τούτου, ο ρυθμός μετάβασης είναι πολύ μικρός. Μεγάλες τιμές θ υπονοεί ότι ο σχεδιασμός του δικτύου κυριαρχείται από διακριτές συνθήκες. Στην περίπτωση αυτή, το αποτέλεσμα μετάβασης είναι πολύ επιθετικό και θα αναγκάσει πολλά διανύσματα εκπαίδευσης να μεταφερθούν στην διακριτή (crisp) κατάσταση. Το σχήμα 6.3 απεικονίζει αυτά τα δύο διαφορετικά αποτελέσματα, όπου μπορούμε να διακρίνουμε την επίδραση της παραμέτρου θ .

Παρατήρηση 2: Στη συνέχεια, αναλύουμε τις βασικές διαφορές μεταξύ του προτεινόμενου αλγορίθμου και του ασαφούς διανυσματικού αλγορίθμου κβαντοποίησης εκπαίδευσης (FLVQ) (βλ. παράγραφο 2.8) που αναπτύσσεται στην αναφορά [92]. Η πρώτη και κρισιμότερη διαφορά αφορά τη στρατηγική βελτιστοποίησης. Ενώ ο προτεινόμενος αλγόριθμος χρησιμοποιεί την καθαρή εναλλασσόμενη (alternate) βελτιστοποίηση για να ελαχιστοποιήσει την αντικειμενική συνάρτηση στην εξ. (6.1), το FLVQ υιοθετεί μια προσέγγιση καθοδικής κλίσης (gradient descent) για να ελαχιστοποιήσει μια αναδιατυπωμένη έκδοση αυτής της συνάρτησης. Η εφαρμογή της καθοδικής κλίσης (gradient descent) πραγματοποιείται μέσω του χειρισμού της παραμέτρου θ από μικρές (δηλ. ασαφείς όροι) στις μεγάλες τιμές (δηλ. διακριτοί όροι) και χρησιμοποιείται κυρίως για να ενημερώσει τις θέσεις των κέντρων συστάδων. Επομένως, ο μηχανισμός εκπαίδευσης πραγματοποιείται όπως στην διανυσματική κβαντοποίηση εκπαίδευσης και είναι εντελώς διαφορετικός όταν συγκρίνεται με την προτεινόμενη μέθοδο εκπαίδευσης. Η δεύτερη διαφορά έγκειται στο γεγονός ότι το FLVQ σχεδιάστηκε ειδικά για τη συμπίεση εικόνας. Υπάρχει μια σημαντική απαίτηση που πρέπει να ικανοποιηθεί κατά τον εξέταση των στόχων συμπίεσης εικόνας. Συγκεκριμένα, κάθε διάνυσμα εκπαίδευσης πρέπει να ανήκει μόνο σε μια συστάδα στον τελικό διαμερισμό, έτσι ώστε οι διαβιβασθείσες πληροφορίες να είναι ελάχιστες. Για να εξετάσει αυτό το πρόβλημα, το FLVQ εκτελεί όλα τα κατάλληλα βήματα και παράγει ένα διακριτό διαμερισμό. Η τελευταία διαφορά αφορά τον ανταγωνισμό μεταξύ των κέντρων συστάδων. Το FLVQ δεν χρησιμοποιεί τον κοντινότερο κοντινό όρο (6.13) για να επιτρέψει τον ανταγωνισμό μεταξύ των συστάδων σε όλο το χώρο χαρακτηριστικών γνωρισμάτων (feature space). Αντ' αυτού, επιλέγει να περιορίσει τον ανταγωνισμό στις ασαφείς περιοχές, μόνο. Προκειμένου να ελαχιστοποιηθεί ο κίνδυνος τοπικών ελάχιστων, αυξάνεται αργά η τιμή της παραμέτρου θ έτσι ώστε να μπορεί να εκτελέσει πολλές επαναλήψεις στην ασαφή κατάσταση. Αυτή η στρατηγική επιτρέπει σε όλα τα κέντρα συστάδων για να κινηθούν και για να ανταγωνιστούν μεταξύ τους, αλλά καθιστά το FLVQ πιο αργό από τον προτεινόμενο αλγόριθμο όταν χρησιμοποιεί ο τελευταίος τις αρκετά μεγάλες τιμές της παραμέτρου θ . Αυτή η παρατήρηση επαληθεύεται επίσης από τα πειράματά μας.

Παρατήρηση 3: Η ιδέα της χρησιμοποίησης αρνητικών βαθμών συμμετοχής για να δείξει ότι ένα διάνυσμα εκπαίδευσης (training data vector) δεν θα πρέπει να επηρεάσει ένα μακρινό κέντρο συστάδας δεν είναι κάτι καινούργιο. Στο [18], οι Miyamoto και Umayahara εφαρμόζουν τετραγωνική κανονικοποίηση (quadratic regularization – QR) σχετικά με την αντικειμενική συνάρτηση του αλγορίθμου c -Μέσων (c -Means), ώστε να ασαφοποιήσει (fuzzify) τους βαθμούς συμμετοχής, ενώ στο [12], οι Honda και Ichihashi έχουν επεκτείνει την παραπάνω μέθοδο. Οι αντίστοιχες συναρτήσεις κόστους είναι,

$$J_{QR} = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2 + \frac{1}{2} \lambda^{-1} \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2 \quad (6.16)$$

και

$$J_{QR}^{Extended} = \sum_{k=1}^n \sum_{i=1}^c \frac{u_{ik} + \lambda (u_{ik})^2}{1 + \lambda} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (6.17)$$

όπου $\lambda > 0$. Η ελαχιστοποίηση της (6.16) και (6.17) υπό τον περιορισμό της εξίσωσης (6.5) έχουν ως αποτέλεσμα ένα σενάριο σύμφωνα με το οποίο, οι βαθμοί συμμετοχής των κέντρων των συστάδων ως προς το διάνυσμα εισόδου \mathbf{x}_k μπορεί να λάβουν αρνητικές τιμές. Και οι δύο μέθοδοι έχουν λύσει αυτό το πρόβλημα μέσω του ορισμού του ακόλουθου συνόλου δεικτών $I_k = \{i : u_{ik} > 0, k = 1, 2, \dots, n\}$ και νέου υπολογισμού των βαθμών συμμετοχής μόνο για τα κέντρα διασποράς, οι δείκτες των οποίων ανήκουν σε αυτό το σύνολο. Είναι πολύ ενδιαφέρον να εντοπιστούν κάποιες βασικές διαφορές μεταξύ της μεθόδου μας, με τους δύο παραπάνω αλγόριθμους. Πρώτον, αν τα θεωρητικά ευρήματα φαίνεται να έχουν κάποια σύνδεση, το φιλοσοφικό υπόβαθρο είναι εντελώς διαφορετικό. Συγκεκριμένα, οι παραπάνω αλγόριθμοι προσπαθούν να ασαφοποιήσουν (fuzzify) τους βαθμούς συμμετοχής που λαμβάνονται από τον αλγόριθμο των c -Μέσων στην πρώτη θέση, ενώ ο προτεινόμενος αλγόριθμος επιχειρεί να συνδυάσει γραμμικά τον αλγόριθμο των c -Μέσων (c -Means) και Ασαφών c -Μέσων (FCM), σύμφωνα με ένα μοναδικό πλαίσιο. Δεύτερον, η διαδικασία βελτιστοποίησης είναι επίσης διαφορετική,

διότι ο προτεινόμενος αλγόριθμος, σταδιακά υπολογίζει και βελτιώνει τα κέντρα των συστάδων τα οποία δεν επηρεάζονται από ένα συγκεκριμένο διάνυσμα εκπαίδευσης (training vector), ενώ οι δύο παραπάνω μέθοδοι αναζητούν το βέλτιστο σετ δεικτών που ελαχιστοποιούν τις συναρτήσεις κόστους (6.16) και (6.17) αντίστοιχα. Τρίτον, οι δύο αλγόριθμοι δεν μπορούν να ελέγξουν τη μετάβαση από ασαφή σε διακριτή κατάσταση, ενώ ο αλγόριθμος μας το κάνει. Ως εκ τούτου, ο προκύπτων διαμερισμός διαθέτει διαφορετικές δομές. Σημειώστε ότι η προτεινόμενη μέθοδος είναι σε θέση να αποκτήσει ακόμη και ένα κοντά-σε-διακριτές ασαφή διαμερισμό κάτι που δεν είναι δυνατόν από τους άλλους δύο αλγόριθμους, δεδομένου ότι ο βέλτιστος δείκτης συνόλων για διάφορους διανύσματα εκπαίδευσης δεν μειώνουν απαραίτητα το μέγεθός τους με τον αριθμό των επαναλήψεων. Τέλος, τα κέντρα των συστάδων υπολογίζονται με αρκετά διαφορετικές φόρμουλες.

6.2.2 Εκτίμηση παραμέτρων του δικτύου

Σε αυτή την ενότητα εξετάζεται η εκτίμηση των παραμέτρων του δικτύου. Ο αριθμός των κρυφών κόμβων είναι ίσος με το c , και τα κέντρα για τις συναρτήσεις ακτινικής βάσης είναι τα διανύσματα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$.

Για να υπολογισμός των πλατών (widths) του πυρήνα της i -οστής συνάρτησης βάσης, χρησιμοποιούμε την προσέγγιση που αναπτύχθηκε από Νείρο και Τσεκούρα στο [66]. Συγκεκριμένα, ορίζουμε το εξής υποσύνολο της i -οστής υβριδικής συστάδας

$$G_i = \{\mathbf{x}_k \in C_i : u_{ik} \geq \xi \text{ με } \xi \in (0,1)\} \quad (6.18)$$

όπου ξ είναι ένα προκαθορισμένο όριο. Για το υποσύνολο G_i θα υπολογιστεί η μέγιστη απόσταση που αναφέρεται στη συνέχεια,

$$d_{\max}^i = \max_{\mathbf{x}_k \in G_i} \{\|\mathbf{x}_k - \mathbf{v}_i\|^2\}$$

Στη συνέχεια, το πλάτος της i -οστής ακτινικής συνάρτησης βάσης είναι,

$$\sigma_i = \frac{2d_{\max}^i}{3} \quad (1 \leq i \leq c) \quad (6.19)$$

Η τιμή του ξ θα πρέπει να είναι αρκετά μικρή ώστε να μπορεί να αποκαλύψει αποτελεσματικά την πραγματική κατανομή των δεδομένων μέσα σε μία συγκεκριμένη συστάδα. Μια αξιόπιστη επιλογή είναι $\xi = 0.001$. Ο λόγος της επιλογής της εξ. (6.19) για τον καθορισμό του πλάτους των έγκειται στη σχέση τους με το βαθμό επικάλυψης κατάτμησης. Συγκεκριμένα, υπάρχουν δύο ζητήματα που πρέπει να λαμβάνονται υπόψη κατά την απόφαση της αλληλεπικάλυψης βαθμών μεταξύ συναρτήσεων βάσης [14]. Η πρώτη ξεκινά από την ανάγκη να διατηρηθεί ένα ελάχιστο ποσό κάλυψης του χώρου χαρακτηριστικών γνωρισμάτων. Δηλαδή, δεν μπορούμε να χρησιμοποιήσουμε πολύ μικρές τιμές πλάτους, επειδή ο προκύπτων βαθμός επικάλυψης θα είναι πολύ μικρός. Σύμφωνα με το δεύτερο ζήτημα, ο βαθμός επικάλυψης δεν πρέπει να είναι πάρα πολύ μεγάλος, διότι το τελικό μοντέλο θα εμφανίζει υπερεκτιμημένη συμπεριφορά, που οδηγεί σε όχι και τόσο καλές δυνατότητες προσέγγισης. Όπως αναφέρεται στην [6], το σενάριο αυτό φαίνεται να λειτουργεί καλά σε περιπτώσεις όπου τα δεδομένα κατανέμονται ομοιόμορφα.

Για να υπολογιστούν τα συναπτικά βάρη (synaptic weights) ορίζονται ο $n \times c$ πίνακας $\mathbf{H} = [g_i(\mathbf{x}_k)]$ όπου $g_i(\mathbf{x}_k)$ δίνεται στην εξ. (1.8), ο $n \times 1$ πίνακας εξόδου $\mathbf{Y} = [y_k]$, και ο $c \times 1$ πίνακας $\mathbf{W} = [w_i]$ όπου w_i είναι το συναπτικό βάρος του i -κόμβου. Ο στόχος είναι να ελαχιστοποιηθεί η ακόλουθη συνάρτηση σφάλματος,

$$E(\mathbf{W}) = \|\tilde{\mathbf{Y}} - \mathbf{Y}\|^2$$

όπου $\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{W}$ είναι ο πίνακας των αποτελεσμάτων του δικτύου. Για να εκτελέσουμε αυτόν τον στόχο, υιοθετούμε τη γνωστή μέθοδο των ελαχίστων τετραγώνων (least squares method),

$$\mathbf{W} = [\mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{Y} \quad (6.20)$$

Τέλος, υπογραμμίζουμε έντονα ότι η εφαρμογή μιας προσέγγισης καθόδου κλίσης (Gradient Descend) για την περαιτέρω βελτίωση των παραμέτρων του δικτύου είναι ακόμα δυνατή, αλλά ο τελικός αλγόριθμος εκπαίδευσης θα γίνει πολύ πιο αργός. Επομένως, δεν χρησιμοποιούμε τέτοιο είδος μηχανισμού εκπαίδευσης.

6.2.3 Ανάλυση του υπολογιστικού κόστους

Υπάρχουν μερικές ενδιαφέρουσες και σημαντικές συνέπειες που αφορούν τις υπολογιστικές προσπάθειες του προτεινόμενου αλγορίθμου. Για να αναλυθεί βαθύτερα το πρόβλημα παρέχεται μια λεπτομερής μαθηματική ανάλυση από την άποψη του αριθμού υπολογισμών απόστασης που εκτελούνται ανά επανάληψη. Κατ' αρχάς, εξετάζουμε τις αντίστοιχες υπολογιστικές δαπάνες των c -Μέσων (c -Means) και του αλγορίθμου των Ασαφών c -Μέσων (FCM). Είναι ευρέως γνωστό ότι ο αριθμός υπολογισμών απόστασης που εκτελούνται από τον αλγόριθμο των c -Μέσων (c -Means) σε μια επανάληψη είναι nc , το οποίο σημαίνει ότι η υπολογιστική πολυπλοκότητα είναι $O(nc)$. Συνεπώς, ο αριθμός υπολογισμών απόστασης που εκτελούνται από τον αλγόριθμο των Ασαφών c -Μέσων (FCM) είναι $(1+c)cn$ [72]. Εφόσον ισχύει ότι $(1+c)cn \leq 2nc^2$, η υπολογιστική πολυπλοκότητα των Ασαφών c -Μέσων (FCM) ανά επανάληψη είναι $O(nc^2)$.

Σε αυτό που ακολουθεί, εστιάζουμε στις υπολογιστικές προσπάθειες του προτεινόμενου αλγορίθμου. Λαμβάνοντας υπόψη την αντικειμενική συνάρτηση εξ. (6.1), είναι αδύνατο όλα τα διανύσματα να παραμείνουν στην ασαφή (fuzzy) κατάσταση. Αν δηλωθεί ότι το σύνολο των διανυσμάτων εκπαίδευσης που παραμένουν στην ασαφή (fuzzy) κατάσταση κατά τη διάρκεια της ν -στης επανάληψης είναι:

$\mathfrak{F}^{(\nu)} = \{ \mathbf{x}_{\delta(1)}, \mathbf{x}_{\delta(2)}, \dots, \mathbf{x}_{\delta(\ell(\nu))} \}$ όπου $\ell(\nu) < n$ και $\delta(k)$ ($k = 1, 2, \dots, \ell(\nu)$) είναι μια αντιμεταβολή των συντεταγμένων (indexed permutation). Με βάση την εξ. (6.12), ο αριθμός υπολογισμών απόστασης που εκτελούνται από τα στοιχεία $\mathfrak{F}^{(\nu)}$ είναι,

$$S_{FM}^{(v)} = [1 + \aleph(T_{\delta(1)}^{(v)})] \aleph(T_{\delta(1)}^{(v)}) + [1 + \aleph(T_{\delta(2)}^{(v)})] \aleph(T_{\delta(2)}^{(v)}) + \dots + [1 + \aleph(T_{\delta(\ell_{(v)})}^{(v)})] \aleph(T_{\delta(\ell_{(v)})}^{(v)})$$

$$= \sum_{k=1}^{\ell_{(v)}} [1 + \aleph(T_{\delta(k)}^{(v)})] \aleph(T_{\delta(k)}^{(v)})$$

Για τα διανύσματα που έχουν μεταφερθεί στην διακριτή (crisp) κατάσταση, ο αριθμός υπολογισμών απόστασης είναι (βλ. εξ. (6.13)),

$$S_{CM}^{(v)} = (n - \ell_{(v)})c$$

Κατά συνέπεια, ο γενικός αριθμός υπολογισμών απόστασης που εκτελούνται από τον αλγόριθμο είναι,

$$S^{(v)} = S_{CM}^{(v)} + S_{FM}^{(v)} = (n - \ell_{(v)})c + \sum_{k=1}^{\ell_{(v)}} [1 + \aleph(T_{\delta(k)}^{(v)})] \aleph(T_{\delta(k)}^{(v)}) \quad (6.21)$$

Εάν επιλέξουμε $\aleph_{\max}^{(v)} = \max_{1 \leq k \leq \ell_{(v)}} \{\aleph(T_{\delta(k)}^{(v)})\}$ τότε το $S^{(v)}$ είναι άνω φραγμένο,

$$S^{(v)} \leq (n - \ell_{(v)})c + \sum_{k=1}^{\ell_{(v)}} (1 + \aleph_{\max}^{(v)}) \aleph_{\max}^{(v)} = (n - \ell_{(v)})c + (1 + \aleph_{\max}^{(v)}) \aleph_{\max}^{(v)} \ell_{(v)}$$

Επίσης, θέτοντας $\aleph_{\min}^{(v)} = \min_{1 \leq k \leq \ell_{(v)}} \{\aleph(T_{\delta(k)}^{(v)})\}$ τότε το $S^{(v)}$ είναι κάτω φραγμένο,

$$(n - \ell_{(v)})c + \sum_{k=1}^{\ell_{(v)}} (1 + \aleph_{\min}^{(v)}) \aleph_{\min}^{(v)} = (n - \ell_{(v)})c + (1 + \aleph_{\min}^{(v)}) \aleph_{\min}^{(v)} \ell_{(v)} \leq S^{(v)}$$

Ας θέσουμε τα χαμηλότερα και ανώτερα όρια ως,

$$LB^{(v)} = (n - \ell_{(v)})c + (1 + \aleph_{\min}^{(v)}) \aleph_{\min}^{(v)} \ell_{(v)} \quad (6.22)$$

$$UB^{(v)} = (n - \ell_{(v)})c + (1 + \aleph_{\max}^{(v)}) \aleph_{\max}^{(v)} \ell_{(v)} \quad (6.23)$$

Οι επόμενες προτάσεις υπολογίζουν τη θεωρητική συμπεριφορά $LB^{(v)}$ και $UB^{(v)}$.

Θεώρημα 1. Για την v -οστη επανάληψη, η ακόλουθη ανισότητα είναι πάντα αληθής,

$$UB^{(v)} < (1 + c)cn \quad (6.24)$$

Απόδειξη

Η ποσότητα $UB^{(v)}$ στην εξ. (6.23) ξαναγράφεται ως,

$$UB^{(v)} = nc + (\aleph_{\max}^{(v)} - c)\ell_{(v)} + (\aleph_{\max}^{(v)})^2 \ell_{(v)}.$$

Αφού $\aleph_{\max}^{(v)} < c$, ο όρος $(\aleph_{\max}^{(v)} - c)$ είναι αρνητικός και η παραπάνω σχέση έχει ως αποτέλεσμα,

$$UB^{(v)} < nc + (\aleph_{\max}^{(v)})^2 \ell_{(v)} < nc + c^2 \ell_{(v)} \quad (6.25)$$

Αλλά $\ell_{(v)} < n$ και επομένως,

$$nc + c^2 \ell_{(v)} < nc + nc^2 = (1+c)nc \quad (6.26)$$

Οι σχέσεις (6.25) και (6.26) έχουν ως αποτέλεσμα την εξ. (6.24). Αυτό ολοκληρώνει την απόδειξη του θεωρήματος 1.

■

Από το θεώρημα 1 αποδεικνύεται ότι η ποσότητα $UB^{(v)}$ είναι μικρότερη από τον αριθμό υπολογισμών απόστασης που εκτελούνται από τον αλγόριθμο των Ασαφών c -Μέσων (FCM) σε κάθε επανάληψη. Επειδή ισχύει ότι $S^{(v)} \leq UB^{(v)}$, η πραγματική υπολογιστική πολυπλοκότητα ανά επανάληψη του αλγορίθμου είναι βεβαίως μικρότερη από $O(nc^2)$.

Θεώρημα 2. Για την v -οστη επανάληψη, εάν ο επόμενος όρος είναι αληθινός,

$$\aleph_{\max}^{(v)} \leq \frac{\sqrt{1+4c} - 1}{2} \quad (6.27)$$

τότε,

$$UB^{(v)} \leq nc \quad (6.28)$$

Απόδειξη

Η εξ. (6.23) μπορεί να τροποποιηθεί ως,

$$UB^{(v)} = nc + \ell_{(v)} \left((\aleph_{\max}^{(v)})^2 + \aleph_{\max}^{(v)} - c \right)$$

Ως εκ τούτου,

$$UB^{(v)} = nc + \ell_{(v)} \left(\aleph_{\max}^{(v)} + \frac{1 + \sqrt{1+4c}}{2} \right) \left(\aleph_{\max}^{(v)} - \frac{\sqrt{1+4c} - 1}{2} \right) \quad (6.29)$$

όπου $-\frac{1+\sqrt{1+4c}}{2}$ και $\frac{\sqrt{1+4c}-1}{2}$ είναι οι ρίζες της εξίσωσης $(\aleph_{\max}^{(v)})^2 + \aleph_{\max}^{(v)} - c = 0$.

Ο όρος (6.27) δίνει $\aleph_{\max}^{(v)} - \frac{\sqrt{1+4c}-1}{2} < 0$ και αφού οι όροι $\ell_{(v)}$ και

$\left(\aleph_{\max}^{(v)} + \frac{1+\sqrt{1+4c}}{2}\right)$ είναι θετικοί, η σχέση (6.29) έχει ως αποτέλεσμα την ανισότητα

(6.28). Αυτό ολοκληρώνει την απόδειξη του θεωρήματος 2.

■

Το αποτέλεσμα της πρότασης 2 είναι πολύ ενδιαφέρον επειδή υπονοεί ότι η πολυπλοκότητα του αλγορίθμου μπορεί να είναι μικρότερη από $O(nc)$. Δηλαδή υπό ορισμένους όρους, ο αλγόριθμος είναι γρηγορότερος ακόμη και από τον αλγόριθμο των c -Μέσων (c -Means). Για παράδειγμα όταν τεθεί ο αριθμός των κρυφών κόμβων $c = 6$. Κατόπιν, όταν $\aleph_{\max}^{(v)} \leq 2$ ο αλγόριθμος είναι γρηγορότερος και από τον αλγόριθμο των c -Μέσων (c -Means) και των Ασαφών c -Μέσων (FCM). Το προηγούμενο παράδειγμα είναι ιδιαίτερα πιθανό να συμβεί. Εντούτοις, το ανωτέρω αποτέλεσμα ισχύει για μια επανάληψη. Εάν συμβαίνει έπειτα αυτό δεν θα είναι στα αρχικά στάδια της διαδικασίας εκπαίδευσης, όπου ο αριθμός υπολογισμών απόστασης είναι πολύ μεγαλύτερος από τον αλγόριθμο των c -Μέσων (c -Means). Επομένως, ο μέσος αριθμός υπολογισμών απόστασης ανά επανάληψη θα είναι μεγαλύτερος από τον αλγόριθμο των c -Μέσων (c -Means). Αυτή η παρατήρηση, επίσης, υποστηρίζεται από τις πειραματικές προσομοιώσεις στο επόμενο τμήμα.

Θεώρημα 3. Για την v -οστη επανάληψη, εάν ο επόμενος όρος είναι αληθής,

$$\aleph_{\min}^{(v)} > \frac{\sqrt{1+4c}-1}{2} \quad (6.30)$$

$$\text{τότε, } nc \leq LB^{(v)} \quad (6.31)$$

Απόδειξη

Από την εξ. (6.22) παίρνουμε,

$$LB^{(v)} = nc + \ell_{(v)} \left(\left(\mathfrak{S}_{\min}^{(v)} \right)^2 + \mathfrak{S}_{\min}^{(v)} - c \right) = nc + \ell_{(v)} \left(\mathfrak{S}_{\min}^{(v)} + \frac{1 + \sqrt{1 + 4c}}{2} \right) \left(\mathfrak{S}_{\min}^{(v)} - \frac{\sqrt{1 + 4c} - 1}{2} \right) \quad (6.32)$$

όπου $-\frac{1 + \sqrt{1 + 4c}}{2}$ και $\frac{\sqrt{1 + 4c} - 1}{2}$ είναι οι ρίζες της εξίσωσης

$$\left(\mathfrak{S}_{\min}^{(v)} \right)^2 + \mathfrak{S}_{\min}^{(v)} - c = 0.$$

Χρησιμοποιώντας τον όρο (6.30) προκύπτει ότι:

$$\ell_{(v)} \left(\mathfrak{S}_{\min}^{(v)} + \frac{1 + \sqrt{1 + 4c}}{2} \right) \left(\mathfrak{S}_{\min}^{(v)} - \frac{\sqrt{1 + 4c} - 1}{2} \right) > 0.$$

Παρεμβάλλοντας το τελευταίο αποτέλεσμα (6.32) παίρνουμε την ανισότητα (6.31). Αυτό ολοκληρώνει την απόδειξη του θεωρήματος 3.

■

Λήμμα 1. Για την v -οστη επανάληψη, εάν ο επόμενος όρος είναι αληθής,

$$\mathfrak{S}_{\min}^{(v)} \leq \frac{\sqrt{1 + 4c} - 1}{2} \quad (6.33)$$

τότε,

$$LB^{(v)} \leq nc \quad (6.34)$$

Απόδειξη

Εάν η ανισότητα εξ. (6.33) είναι αληθής, τότε

$$\ell_{(v)} \left(\mathfrak{S}_{\min}^{(v)} + \frac{1 + \sqrt{1 + 4c}}{2} \right) \left(\mathfrak{S}_{\min}^{(v)} - \frac{\sqrt{1 + 4c} - 1}{2} \right) \leq 0$$

και έτσι χρησιμοποιώντας την (6.32) προκύπτει η ανισότητα (6.34). Αυτό ολοκληρώνει την απόδειξη του λήμματος 1.

■

Σύμφωνα με τα θεωρήματα 1 και 3 η υπολογιστική πολυπλοκότητα του προτεινόμενου αλγορίθμου σε μια ενιαία επανάληψη βρίσκεται ενδιάμεσα $O(nc)$ και $O(nc^2)$. Το λήμμα 1 δείχνει ότι είναι δυνατό (αλλά μη βέβαιο) να συμβαίνει

$S^{(v)} \leq nc$. Παρατηρήστε ότι όταν το αποτέλεσμα του θεωρήματος 2 είναι αληθινό, το αποτέλεσμα του λήμματος 1 είναι αμέσως ικανοποιημένο.

Στην πράξη, εάν επιλέξουμε μια αρκετά μεγάλη τιμή για την παράμετρο θ (κατά προτίμηση στο διάστημα $[0.3, 0.7]$) τότε αναμένουμε ότι οι πραγματικές πληθικότητες συνόλου (cardinalities) να είναι $N(T_{\delta(k)}^{(v)}) \ll c \forall k \in \{1, 2, \dots, \ell_{(v)}\}$, και ο αριθμός διανυσμάτων που μεταφέρονται στην διακριτή (crisp) κατάσταση να είναι αρκετά μεγάλος ώστε να μειώσει περαιτέρω την τιμή $S^{(v)}$, η οποία αποδεικνύεται πολύ μικρότερη από $(1+c)cn$.

6.3 Μελέτη προσομοίωσης

Σε αυτή την ενότητα αξιολογούμε την προτεινόμενη μέθοδο και συγκρίνουμε την απόδοσή της με τρία διαφορετικά δίκτυα. Το πρώτο δίκτυο, αποκαλούμενο ως RBFN 1, είναι ένα τυποποιημένο δίκτυο ακτινικής βάσης (RBF), το οποίο υιοθετεί τον αλγόριθμο των c -Μέσων (c -Means) (βλέπε παράγραφο 2.2) για να καθορίσει τα κέντρα των πυρήνων στην εξ. (1.8), ενώ τα αντίστοιχα πλάτη (widths) τίθενται ίσα με 1. Το δεύτερο δίκτυο, αποκαλούμενο RBFN 2, εκπαιδεύεται από τον αλγόριθμο των Ασαφών c -Μέσων (βλέπε παράγραφο 2.3) και οι συναρτήσεις ακτινικής βάσης συμπίπτουν με τις συναρτήσεις συμμετοχής των αντίστοιχων ασαφών συστάδων. Τέλος, το τρίτο δίκτυο, αποκαλούμενο RBFN 3, χρησιμοποιεί την εξ. (1.8) για να υπολογίσει τις συναρτήσεις βάσης. Τα κέντρα υπολογίζονται από τον ασαφή διανυσματικό αλγόριθμο κβαντοποίησης εκμάθησης (FLVQ) (βλ. παράγραφο 2.8) που αναπτύσσεται στην αναφορά [92] και χρησιμοποιεί την εξ. (1.10) για να καθορίσει τα αντίστοιχα πλάτη (widths).

Όλες οι προσομοιώσεις διευθετήθηκαν χρησιμοποιώντας $\tau = 2$ για την εξ. (1.10) και 25 διαφορετικές αρχικοποιήσεις (δηλ. 25 «τρεξίματα» για κάθε πειραματική περίπτωση) και τέλος για τον αλγόριθμο των Ασαφών c -Μέσων (FCM) η παράμετρος ασάφειας m τέθηκε ίση με 2. Σε όλα τα δίκτυα οι βέλτιστες τιμές των συναπτικών βαρών υπολογίστηκαν με την τυποποιημένη μέθοδο των ελαχίστων τετραγώνων (least squares method).

Για να αξιολογηθεί ο αντίκτυπος της παραμέτρου θ , τα πειράματα εκτελέστηκαν για μια σειρά τιμών της παραμέτρου της θ ίσων με 0.1, 0.3, 0.5, και 0.7. Αυτό το σύνολο των τιμών αντιστοιχεί σε όλες τις αντιπροσωπευτικές περιπτώσεις όσον αφορά στη δομή του υβριδικού διαμερισμού. Αναμένεται ότι οι τιμές 0.1 και 0.3 θα έχουν ως αποτέλεσμα να ληφθούν μικρά ποσοστά μετάβασης από την ασαφή (fuzzy) σε διακριτή (crisp) κατάσταση. Αφ' ετέρου, οι τιμές 0.5 και 0.7 είναι πιθανό να αναγκάσουν έναν σημαντικό αριθμό διανυσμάτων εκπαίδευσης για να μεταφερθούν στην διακριτή (crisp) κατάσταση, που δημιουργεί τις μεγαλύτερες διακριτές περιοχές.

Επιπλέον, γίνεται σύγκριση της προτεινόμενης μεθόδου με τις μεθόδους CFA [29], CFC [72], SFC [89], του δικτύου των Park κ.α.[71], και του δικτύου των Pedrycz κ.α. [75]. Ο αλγόριθμος CFC αναλύεται στην παράγραφο 2.4 του 2^{ου} κεφαλαίου της παρούσης διδακτορικής διατριβής. Η βασική διαφορά μεταξύ της προτεινόμενης μεθόδου και των παραπάνω αλγορίθμων είναι ότι στην περίπτωση μας απασχολούν μόνο συσταδοποίηση των εισόδων ενώ οι υπόλοιπες μέθοδοι εφαρμόζουν συσταδοποίηση εισόδου-εξόδου με διαφορετικούς τρόπους. Σε γενικές γραμμές, αυτές οι μέθοδοι είναι περίπλοκες και έχουν καλύτερες επιδόσεις από τα πρότυπα αυτά. Ωστόσο, όπως θα αποδειχθεί στη συνέχεια, ο αλγόριθμος μας αποδείχθηκε ότι είναι πολύ ανταγωνιστικός με αυτούς και σε πολλές περιπτώσεις τους ξεπερνά σε επιδόσεις, πράγμα που σημαίνει ότι είναι σε θέση να ανακαλύψει αποτελεσματικές δομές για τα δεδομένα εκπαίδευσης.

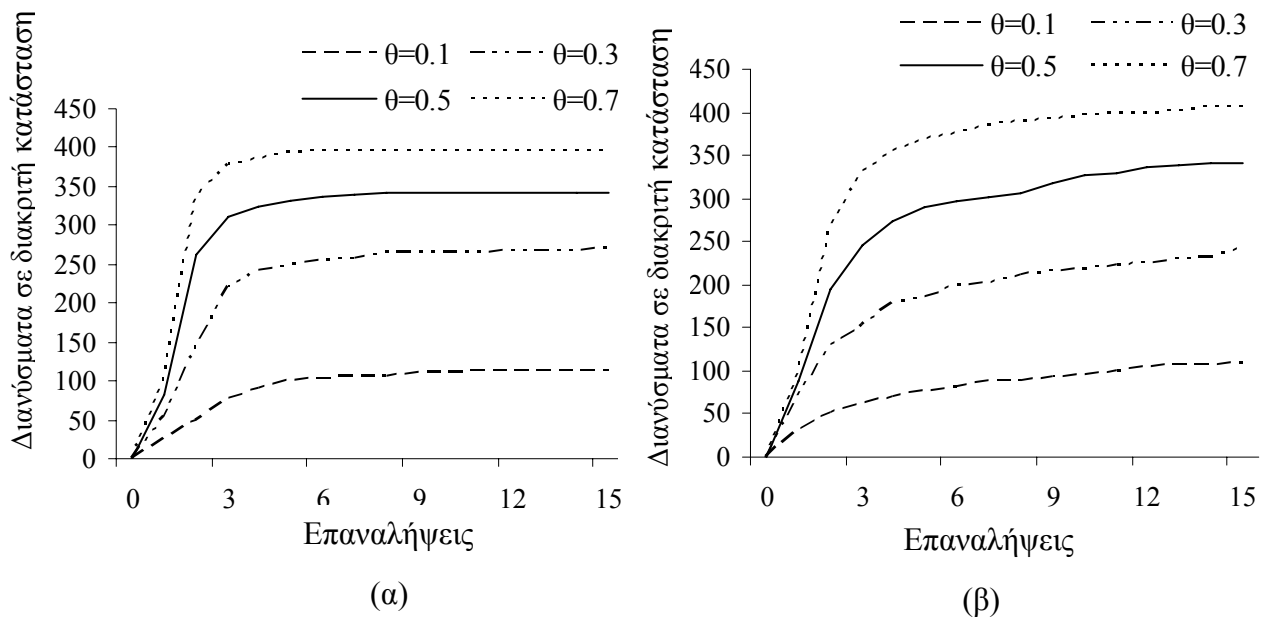
Παράδειγμα 1: Δισδιάστατη συνάρτηση

Χρησιμοποιήθηκε μια δισδιάστατη συνάρτηση για να μελετηθεί η απόδοση του προτεινόμενου δικτύου και διάφορων βασικών ιδιοτήτων του. Συγκεκριμένα, αξιοποιήθηκε ο δισδιάστατος χώρος εισόδου για να απεικονιστεί η μορφή των υβριδικών διαμερισμών. Συγκεκριμένα εξετάστηκε η ακόλουθη συνάρτηση,

$$y = f(x_1, x_2) = \frac{(x_1 - 2)(2x_1 + 1)}{1 + x_1^2} \frac{(x_2 - 2)(2x_2 + 1)}{1 + x_2^2} \quad \text{με } x_1, x_2 \in [-5, 5]$$

Η συνάρτηση προσεγγίστηκε από ένα σύνολο 441 δειγμάτων που εξήχθησαν από ένα πλέγμα 21 x 21 σημείων που ισοκατανεμήθηκαν στο χώρο των εισόδων. Το σχήμα 6.4 παρουσιάζει τον αριθμό διανυσμάτων εκπαίδευσης που μεταφέρονται στην διακριτή

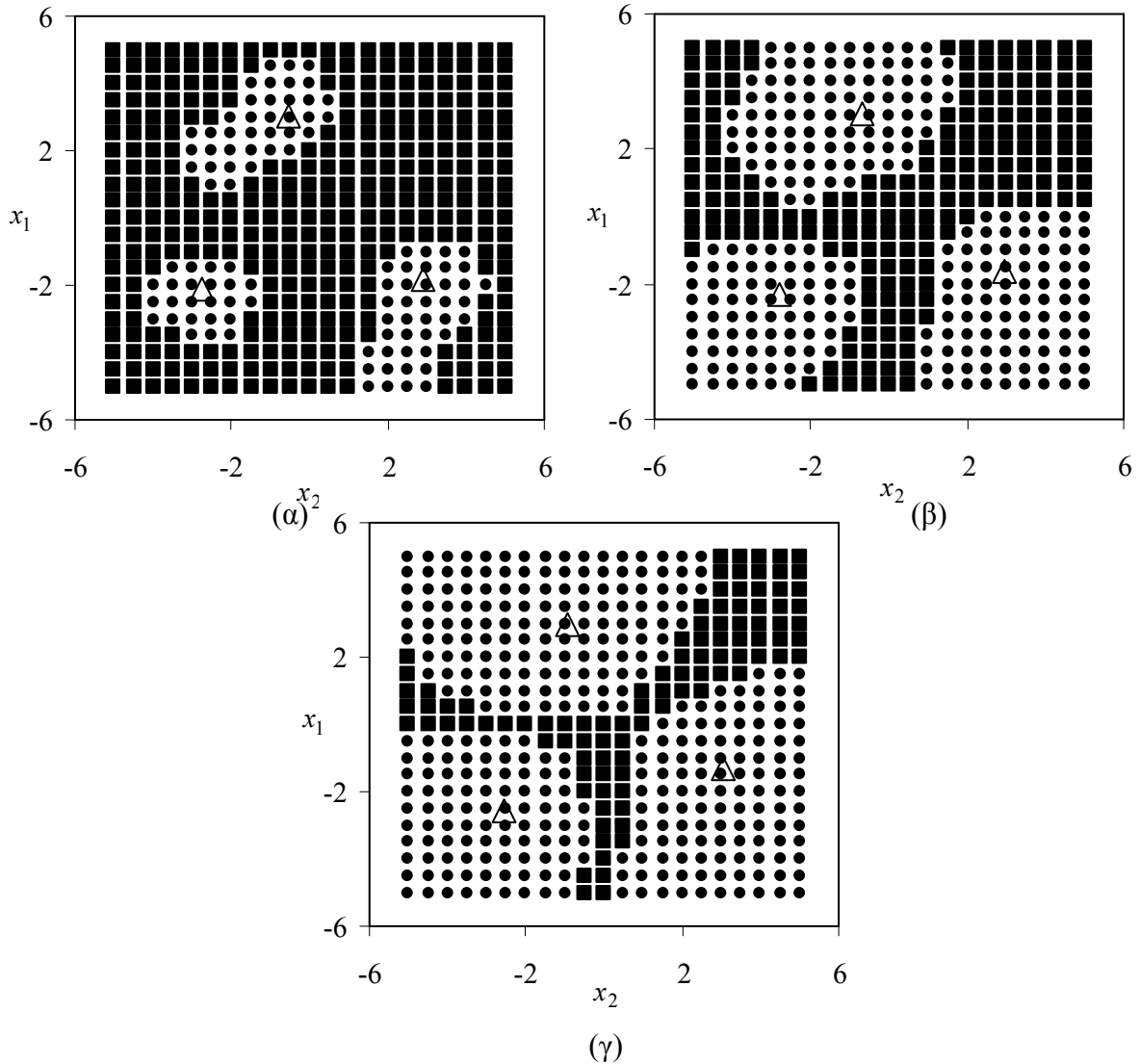
κατάσταση (crisp mode) ως συνάρτηση του αριθμού των επαναλήψεων, για τις διάφορες τιμές θ και c . Αυτός ο αριθμός δικαιολογεί έντονα τα συμπεράσματά μας για την υπολογιστική πολυπλοκότητα του αλγορίθμου. Μπορεί να παρατηρηθεί ότι για $\theta = 0.7$, ένα μεγάλο ποσό διανυσμάτων έχει μεταφερθεί στην διακριτή κατάσταση (crisp mode), ενώ για $\theta = 0.1$ σχεδόν 70% από αυτά παραμένει στην ασαφή κατάσταση. Κατά αρκετά ενδιαφέροντα τρόπο, δεν υπάρχει καμία σημαντική επιρροή του αριθμού κρυφών κόμβων επάνω στον τελικό αριθμό διανυσμάτων που μεταφέρονται στην διακριτή κατάσταση (crisp mode), αν και φαίνεται ότι η τιμή του c παρέχει έναν αντίκτυπο στη δυναμική αλλαγή του ποσοστού μετάβασης. Αυτή η παρατήρηση αναφέρεται επίσης στα ακόλουθα παραδείγματα.



Σχήμα 6.4. Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού των επαναλήψεων για: (α) $c = 5$ και (β) $c = 10$.

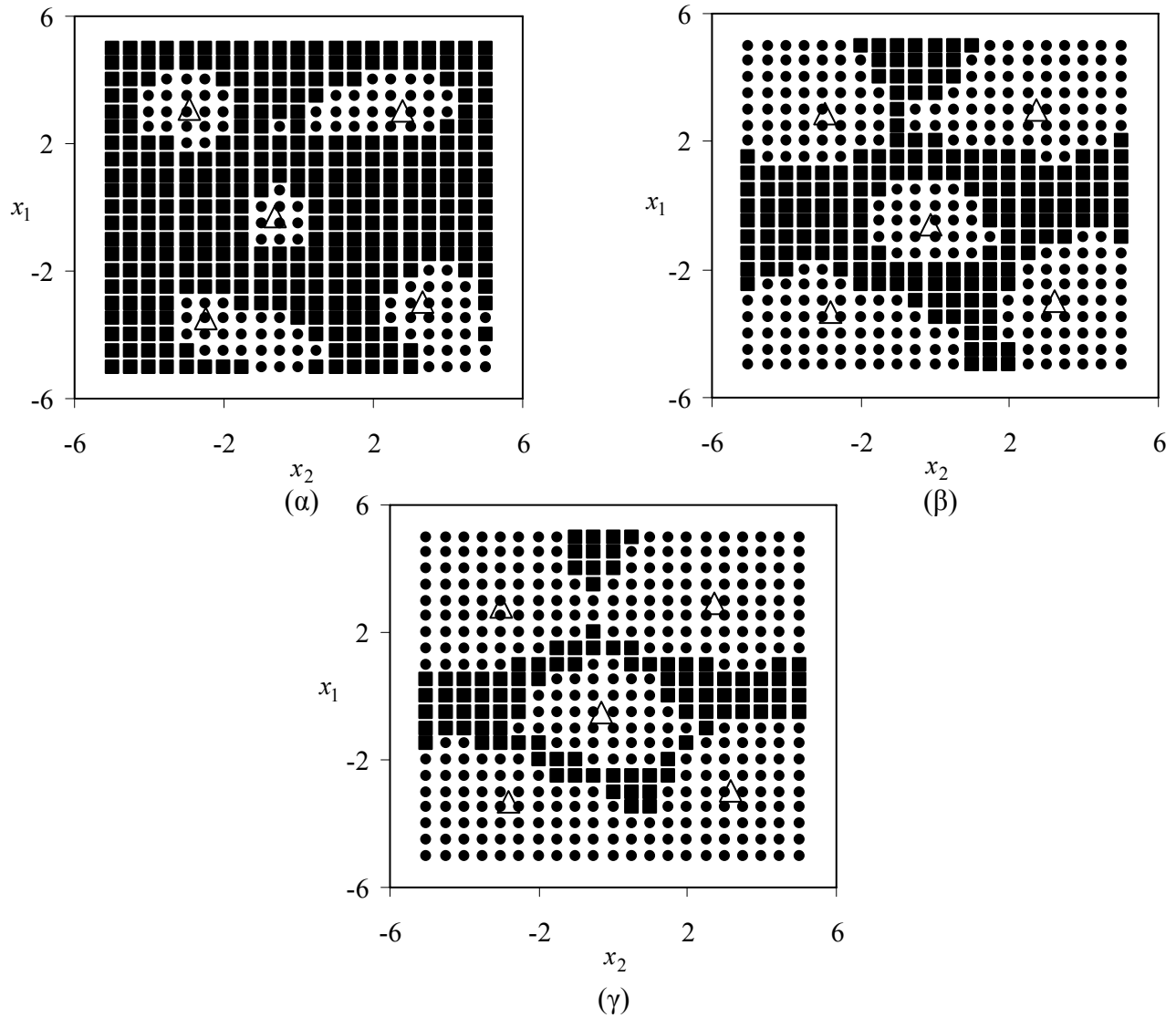
Μπορεί να υιοθετηθεί η ακόλουθη απλή εξήγηση για να δικαιολογηθεί αυτό το γεγονός. Όταν η τιμή του c είναι μικρή, τα κέντρα συστάδων αναμένεται να τοποθετούνται πιο αραιά στο χώρο των χαρακτηριστικών γνωρισμάτων (feature space). Δηλαδή όταν ένα συγκεκριμένο κέντρο συστάδων είναι απόμακρο από τα άλλα κέντρα, τότε αυτό θα αναγκάσει όλα τα διανύσματα που είναι πιο στενά με το παραπάνω κέντρο να μεταφερθούν στην διακριτή κατάσταση (crisp mode). Αφ' ετέρου, όταν ο αριθμός συστάδων είναι μεγάλος, η δυνατότητα ενός διανύσματος εκπαίδευσης για να βρεθεί

αρκετά κοντά σε ένα κέντρο είναι μεγαλύτερη, και επομένως πολλά διανύσματα εκπαίδευσης τίθενται στην διακριτή κατάσταση (crisp mode) επίσης. Κατά συνέπεια, και στις δύο περιπτώσεις αναμένονται παρόμοια ποσοστά μετάβασης.



Σχήμα 6.5. Ασαφείς (συμπαγείς τετράγωνα) και διακριτοί (συμπαγείς κύκλοι) τομείς του χώρου των εισόδων χρησιμοποιώντας $c = 3$ για: (α) $\theta = 0.1$, (β) $\theta = 0.3$, και (γ) $\theta = 0.5$, όπου τα κέντρα συστάδων συμβολίζονται από τα κενά τρίγωνα.

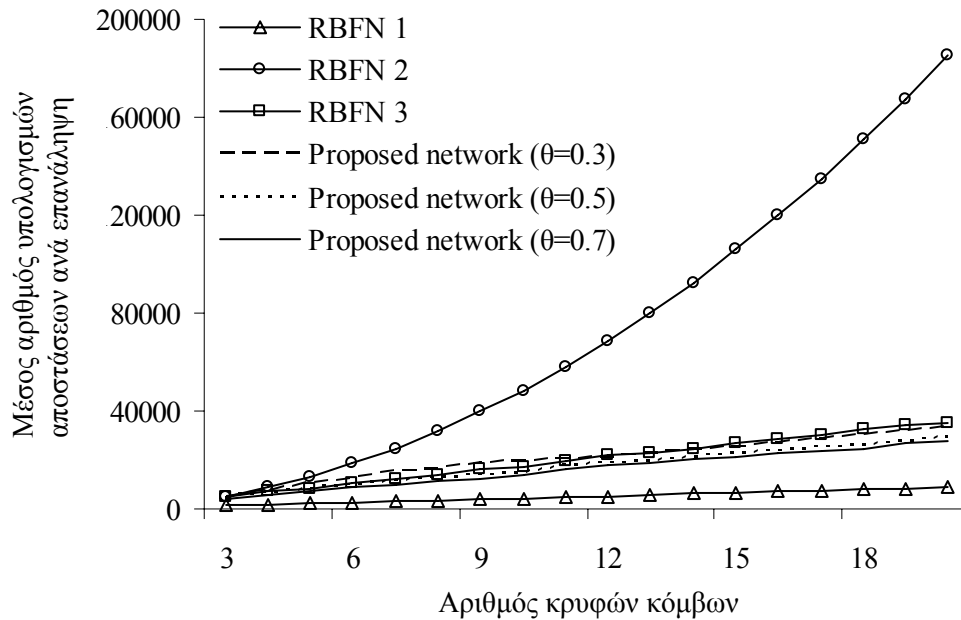
Το σχήμα 6.5 απεικονίζει πώς οι ασαφείς και διακριτές περιοχές δημιουργούνται και συνυπάρχουν και πώς διανέμονται στο χώρο των εισόδων για τρεις κρυφούς κόμβους. Τα κέντρα των πυρήνων αντιστοιχούν στα κενά τρίγωνα, η ασαφής περιοχή στα συμπαγείς τετράγωνα, και η διακριτή περιοχή στους συμπαγείς κύκλους.



Σχήμα 6.6. Ασαφείς (συμπαγείς τετράγωνα) και διακριτοί (συμπαγείς κύκλοι) τομείς του χώρου των εισόδων χρησιμοποιώντας $c = 5$ για: (α) $\theta = 0.1$, (β) $\theta = 0.3$, και (γ) $\theta = 0.5$, όπου τα κέντρα συστάδων συμβολίζονται από τα κενά τρίγωνα.

Η μετάβαση από το σχήμα 6.5 (α) προς το σχήμα 6.5 (γ) δείχνει πως καθώς η τιμή της παραμέτρου θ αυξάνεται από 0.1 σε 0.5, αναπροσαρμόζεται αντίστοιχα η τοπολογία των ασαφών (fuzzy) και διακριτών (crisp) περιοχών. Η ισορροπία μεταξύ αυτών των δύο περιοχών καθορίζεται από την κυρίαρχη επίδραση της παραμέτρου θ για ένα συγκεκριμένο τρέξιμο. Μικρές τιμές του θ (δηλ. σχήμα 6.5 (α)) υπονοούν ότι οι ασαφείς (fuzzy) όροι κυριαρχούν στη διαδικασία εκπαίδευσης, ενώ για μεγάλη τιμή του θ (δηλ.

σχήμα 6.5 (γ)) οι διακριτοί (crisp) όροι κυριαρχούν στην αρχιτεκτονική του δικτύου. Επιπλέον, καθώς αυξάνεται η παράμετρος θ από μικρές σε μεγάλες τιμές, οι ασαφείς (fuzzy) περιοχές αντικαθίσταται ομαλά από διακριτές (crisp) περιοχές. Παρόμοια αποτελέσματα για πέντε κρυφούς κόμβους παρουσιάζονται στο σχήμα 6.6.



Σχήμα 6.7. Μέσος αριθμός υπολογισμών αποστάσεων ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 1.

Με βάση τα σχήματα 6.4-6.6, αναμένεται ότι για $\theta = 0.3$ το προτεινόμενο νευρωνικό δίκτυο ακολουθεί μια πιο αργή διαδικασία από $\theta = 0.5$, δεδομένου ότι περισσότερα διανύσματα εκπαίδευσης παραμένουν στην ασαφή κατάσταση (fuzzy mode). Κατά συνέπεια η εφαρμογή του αλγορίθμου για $\theta = 0.7$ είναι πιθανό να είναι γρηγορότερη από $\theta = 0.5$. Πράγματι, αυτά τα σημαντικά αποτελέσματα παρουσιάζονται στο σχήμα 6.7, το οποίο επιδεικνύει τις υπολογιστικές απαιτήσεις των τεσσάρων δικτύων από την άποψη του μέσου αριθμού υπολογισμών απόστασης ανά επανάληψη. Για να υπολογιστεί αυτός ο αριθμός για την προτεινόμενη μέθοδο, χρησιμοποιήθηκε η εξ. (6.20). Όλες οι προσομοιώσεις σε αυτόν τον αριθμό διευθετήθηκαν χρησιμοποιώντας τους ίδιους αρχικούς όρους για μια συγκεκριμένη τιμή του c . Υπενθυμίζεται ότι το RBFN 1 σχεδιάζεται χρησιμοποιώντας τη μέθοδο των c -Μέσων (c -Means), το RBFN 2 τη μέθοδο των Ασαφών c -Μέσων (FCM), και το RBFN 3 χρησιμοποιεί τη μέθοδο FLVQ

που αναπτύσσεται στην αναφορά [92]. Το RBFN 2 παρουσιάζει μια εκθετική αύξηση του υπολογιστικού κόστους. Αφ' ετέρου το RBFN 3 και το προτεινόμενο δίκτυο διατηρούν τις υπολογιστικές αποδόσεις κοντά στο RBFN 1. Μπορεί επίσης να αναγνωριστεί ότι η απόδοση RBFN 3 είναι ισοδύναμη με το προτεινόμενο νευρωνικό δίκτυο όταν $\theta = 0.3$. Όπως αναφέρεται προηγουμένως, το FLVQ μένει για πολύ χρόνο στην ασαφή κατάσταση (fuzzy mode) και έτσι, ο μέσος αριθμός υπολογισμών απόστασης αυξάνεται, αν και δεν χρησιμοποιεί τον κοντινότερο γειτονικό όρο στην εξ. (6.14).

Πίνακας 6.1. NRMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances)

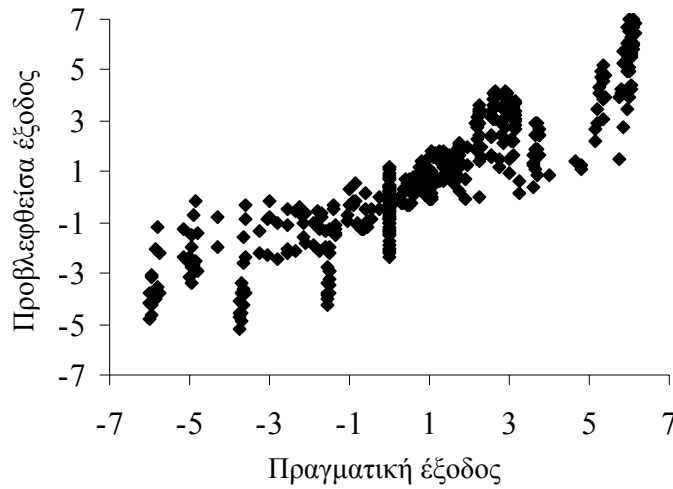
c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
5	0.924 ± 0.004	0.875 ± 0.017	0.970 ± 0.013	0.864 ± 0.015	0.873 ± 0.010	0.835 ± 0.013
10	0.821 ± 0.013	0.654 ± 0.019	0.606 ± 0.012	0.565 ± 0.010	0.564 ± 0.026	0.549 ± 0.016
15	0.725 ± 0.021	0.582 ± 0.009	0.555 ± 0.018	0.483 ± 0.004	0.472 ± 0.017	0.497 ± 0.018
20	0.613 ± 0.016	0.535 ± 0.011	0.505 ± 0.010	0.417 ± 0.005	0.398 ± 0.014	0.408 ± 0.017

Τα τέσσερα δίκτυα συγκρίθηκαν από την άποψη της ομαλοποιημένης τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (NRMSE):

$$NRMSE = \sqrt{\frac{\sum_{k=1}^n (y_k - \tilde{y}_k)^2}{\sum_{k=1}^n (y_k - \bar{y})^2}}$$

όπου \bar{y} είναι η μέση τιμή των δεδομένων εξόδου. Ο πίνακας 6.1 συγκρίνει τις αποδόσεις που λαμβάνονται από τα δίκτυα για τις διάφορες τιμές του c . Σαφώς το προτεινόμενο δίκτυο επιτυγχάνει την καλύτερη απόδοση σε όλες τις περιπτώσεις. Το καλύτερο αποτέλεσμα έρχεται με NRMSE=0.398, το οποίο αναφέρεται όταν $c = 20$ και $\theta = 0.5$. Γενικά, φαίνεται ότι ο αλγόριθμος παρουσιάζει μια ανεξάρτητη συμπεριφορά όσον αφορά την παράμετρο θ . Αυτή η παρατήρηση παρουσιάζεται στα ακόλουθα παραδείγματα.

Το σχήμα 6.8 επεξηγεί τις ικανότητες προσέγγισης και γενίκευσης του δικτύου. Η προσομοίωση για αυτόν τον αριθμό διευθετήθηκε για $c = 10$ και $\theta = 0.5$, και το αντίστοιχο NRMSE υπολογίστηκε ως 0.5472.



Σχήμα 6.8. Γράφημα διασποράς για $c = 10$ και $\theta = 0.5$ μεταξύ της αρχικής και προβλεφθείσας εξόδου του προτεινόμενου δικτύου.

Πίνακας 6.2. Συγκριτική ανάλυση από την άποψη των μέσων τιμών NRMSE

c	CFC [72]	SFC [89]	CFA [29]	Προτεινόμενη Μέθοδος ($\theta = 0.3$)	Προτεινόμενη Μέθοδος ($\theta = 0.5$)	Προτεινόμενη Μέθοδος ($\theta = 0.7$)
6	0.838	0.613	0.812	0.788	0.784	0.791
8	---	0.491	0.713	0.702	0.689	0.669
9	0.812	0.277	0.708	0.575	0.576	0.566
10	---	0.185	0.705	0.565	0.564	0.549

Ο πίνακας 6.2 συγκρίνει τις αποδόσεις του δικτύου για $\theta = 0.3$, $\theta = 0.5$ και $\theta = 0.7$ με τα δίκτυα CFA, CFC, and SFC που υπάρχουν στη διεθνή βιβλιογραφία. Τα αποτελέσματα που αναφέρονται σε αυτόν τον πίνακα είναι ιδιαίτερα πειστικά, δεδομένου ότι σε όλες τις περιπτώσεις, εκτός από την περίπτωση του SFC δικτύου, η προτεινόμενη μέθοδος ξεπέρασε όλες τις υπόλοιπες. Τέλος, ο πίνακας 6.3 παρέχει τις παραμέτρους του προτύπου ενός δικτύου με 10 κρυφούς κόμβους που παράγονται από τον αλγόριθμο για $\theta = 0.5$. Η αντίστοιχη απόδοση ήταν ίση με 0.5337.

Πίνακας 6.3. Ένα δίκτυο που παράγεται από τον προτεινόμενο αλγόριθμο για $c = 10$
και $\theta = 0.5$

i	v_i	σ_i	w_i	
1	-0.623	3.598	3.759	-5.676
2	-3.306	-4.156	2.863	8.672
3	3.135	0.480	4.201	-7.859
4	-3.700	0.751	3.111	-12.653
5	-3.742	3.716	3.325	9.263
6	0.135	-3.471	3.980	-11.491
7	3.160	3.649	4.113	4.764
8	-0.483	-0.066	3.904	9.947
9	-3.741	-1.879	2.863	6.674
10	3.627	-3.205	4.216	9.013

Παράδειγμα 2: Τρισδιάστατη συνάρτηση

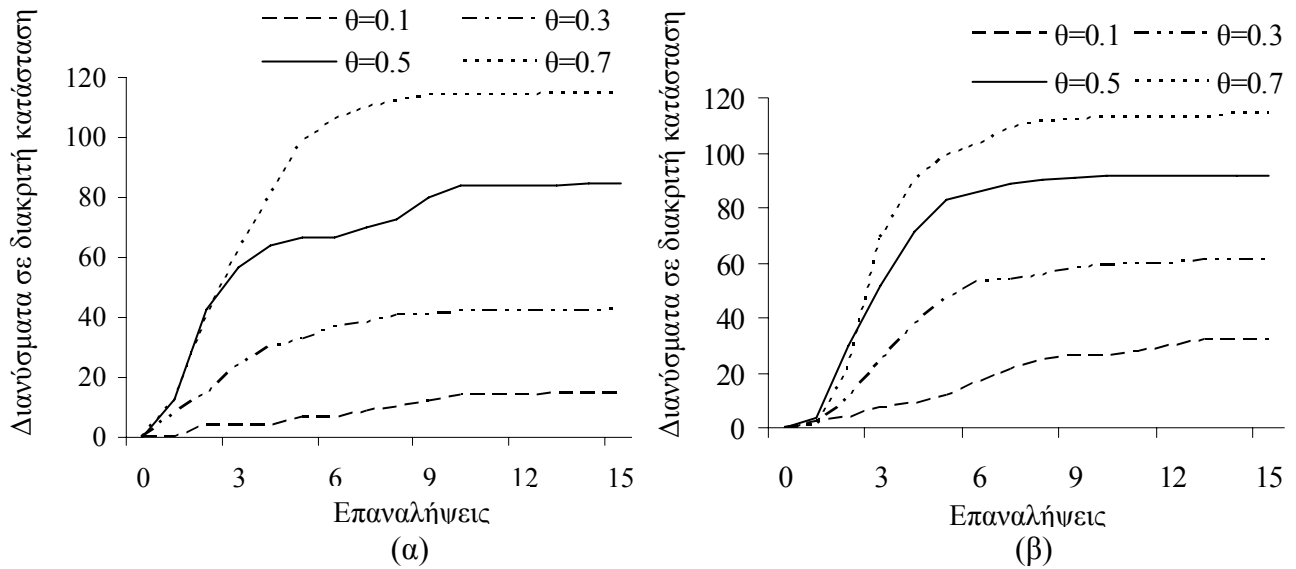
Σε αυτό το παράδειγμα, εξετάζεται η ακόλουθη τρισδιάστατη μη γραμμική συνάρτηση,

$$y = f(x_1, x_2, x_3) = (1 + x_1^{-2} + x_2^{-1.5} + x_3^{-1.0})^2 \text{ με } x_1, x_2, x_3 \in [1, 5].$$

200 ζευγάρια δεδομένων εισόδου-εξόδου εισήχθησαν και διαιρέθηκαν τυχαία σε σύνολο δεδομένων εκπαίδευσης που περιέχουν το 60% από το σύνολο δεδομένων συμπεριλαμβανομένου του υπολοίπου 40% των δεδομένων δοκιμής.

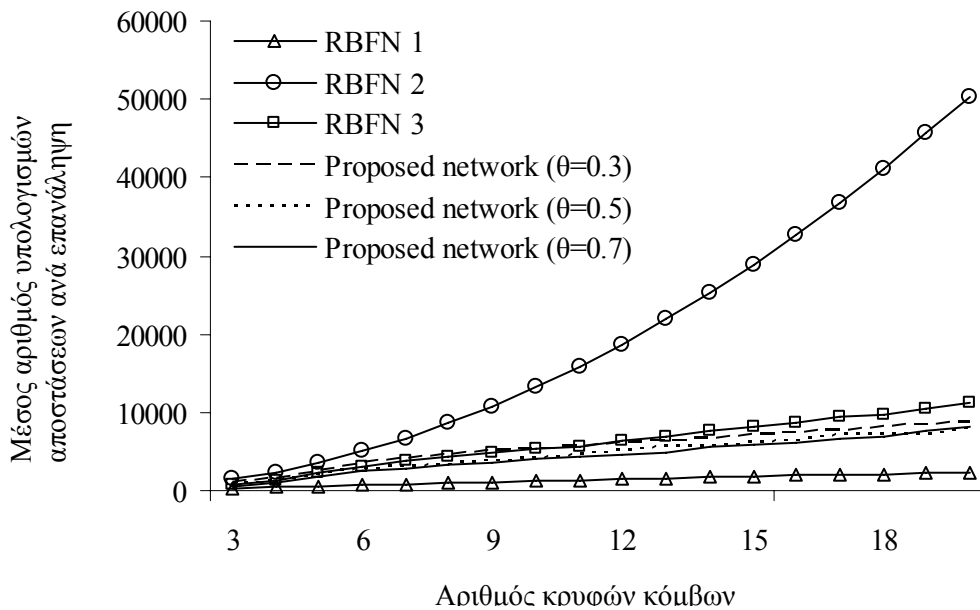
Ο δείκτης απόδοσης ήταν η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE):

$$RMSE = \sqrt{\sum_{k=1}^n (y_k - \tilde{y}_k)^2 / n}.$$



Σχήμα 6.9. Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού επανάληψης για: (α) $c=10$, και (β) $c=20$

Το σχήμα 6.9 παρουσιάζει τα ποσοστά μετάβασης από ασαφή (fuzzy mode) σε διακριτή κατάσταση (crisp mode) για τις διαφορετικές τιμές των παραμέτρων θ και c , όπου μπορούμε να δούμε παρόμοια αποτελέσματα με εκείνα που λαμβάνονται στο προηγούμενο παράδειγμα. Τα υπολογιστικά αποτελέσματα των τεσσάρων δικτύων παρουσιάζονται στο σχήμα 6.10,



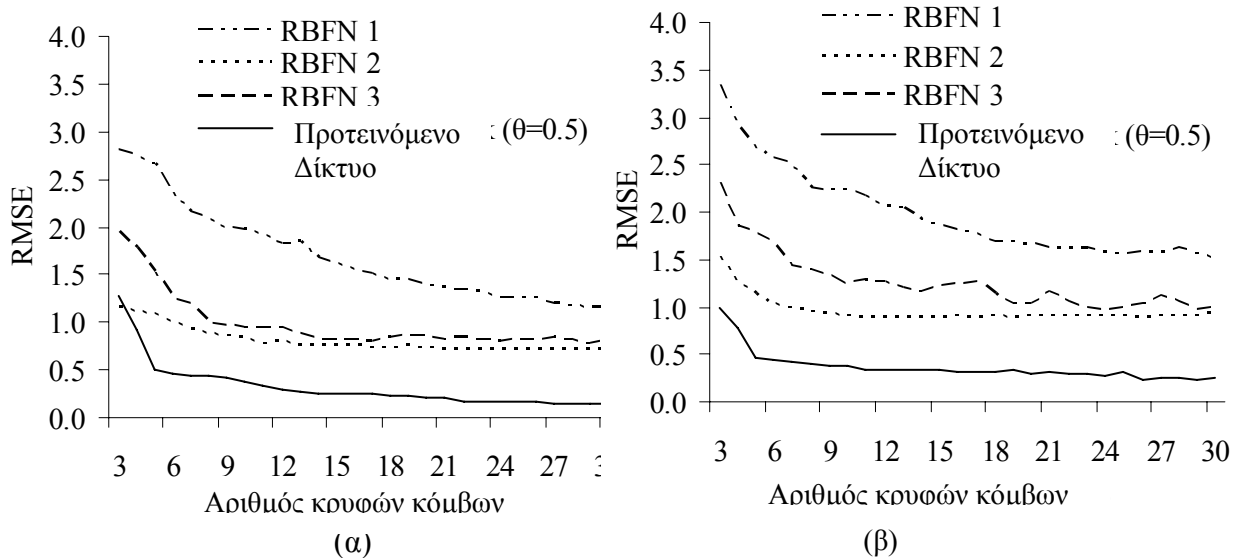
Σχήμα 6.10. Μέσος αριθμός υπολογισμών αποστάσεων ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 2.

Πίνακας 6.4. Μέσες τιμές RMSE και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης

c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
10	1.978 ± 0.1063	0.828 ± 0.0284	0.936 ± 0.1255	0.387 ± 0.0363	0.379 ± 0.0404	0.393 ± 0.0438
20	1.391 ± 0.1089	0.733 ± 0.0186	0.853 ± 0.0448	0.212 ± 0.0166	0.209 ± 0.0213	0.211 ± 0.0115
30	1.143 ± 0.0955	0.702 ± 0.0342	0.815 ± 0.0601	0.142 ± 0.0126	0.139 ± 0.0102	0.156 ± 0.0148

Πίνακας 6.5. Μέσες τιμές RMSE και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα δοκιμής

c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
10	2.224 ± 0.1083	0.908 ± 0.0356	1.233 ± 0.0428	0.355 ± 0.0268	0.368 ± 0.0382	0.356 ± 0.0431
20	1.662 ± 0.1063	0.899 ± 0.0222	1.028 ± 0.0459	0.323 ± 0.0537	0.279 ± 0.0445	0.302 ± 0.0461
30	1.516 ± 0.1241	0.930 ± 0.0285	0.982 ± 0.0649	0.286 ± 0.0421	0.249 ± 0.0361	0.278 ± 0.058

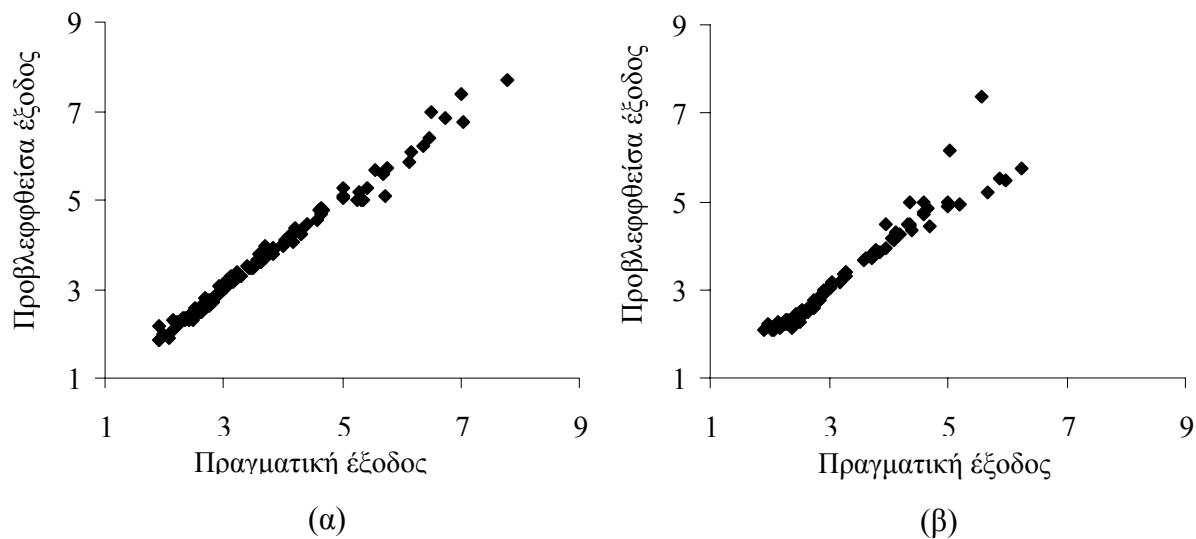


Σχήμα 6.11. RMSE μέσες τιμές που λαμβάνονται από τον προτεινόμενο αλγόριθμο ($\theta = 0.5$), το RBFN 1, το RBFN 2 και το RBFN 3 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής.

Οι τιμές RMSE και οι αντίστοιχες διαφορές που λαμβάνονται από τα τέσσερα δίκτυα για τα δεδομένα εκπαίδευσης και δοκιμής συνοψίζονται στους πίνακες 6.4 και 6.5, αντίστοιχα. Τα αποτελέσματα αυτών των πινάκων επεκτείνονται στο σχήμα 6.11, όπου οι μέσες τιμές RMSE των τεσσάρων δικτύων συγκρίνονται άμεσα για τους

διάφορους αριθμούς κρυφών κόμβων. Σαφώς, ο προτεινόμενος αλγόριθμος επιτυγχάνει καλύτερο RMSEs και στα δεδομένα εκπαίδευσης και στα δεδομένα δοκιμής .

Το σχήμα 6.12 παρουσιάζει τα γραφήματα διασποράς μεταξύ των προβλεφθέντων και αρχικών δειγμάτων εξόδου για τα δύο σύνολα δεδομένων. Αυτός ο αριθμός ελέγχει τις ικανότητες προσέγγισης και γενίκευσης του δικτύου μας. Η προσομοίωση για αυτήν την περίπτωση διευθετήθηκε για $c = 25$ κρυφούς κόμβους και $\theta = 0.3$, και η απόδοση υπολογίστηκε ως 0.168 (δεδομένα εκπαίδευσης) και 0.278 (δεδομένα δοκιμής).



Σχήμα 6.12. Γραφήματα διασποράς για $c = 25$ και $\theta = 0.3$ μεταξύ της αρχικής και προβλεφθείσας εξόδου: (α) δεδομένα εκπαίδευσης, (β) δεδομένα δοκιμής.

Πίνακας 6.6. Συγκριτική ανάλυση από την άποψη των μέσων τιμών RMSE

Αριθμός κρυφών κόμβων	Park et al.[71]		Προτεινόμενη μέθοδος ($\theta = 0.5$)	
	Δεδομένα εκπαίδευσης	Δεδομένα δοκιμής	Δεδομένα εκπαίδευσης	Δεδομένα δοκιμής
6	0.2734	0.5055	0.4611	0.4436
10	0.2564	0.3601	0.3792	0.3687
12	0.0681	0.4952	0.2925	0.3338
15	0.1490	0.3684	0.2562	0.3413
24	0.0973	0.2544	0.1749	0.2597
30	0.1881	0.4586	0.1396	0.2494

Τέλος, ο πίνακας 6.6 συγκρίνει τα αποτελέσματα που επιτεύχθηκαν εδώ και τα αντίστοιχα αποτελέσματα που βρίσκονται στην αναφορά [71]. Εξετάζοντας το σύνολο δεδομένων εκπαίδευσης, η μέθοδος που αναπτύχθηκαν από τον Park [71] είχε καλύτερα αποτελέσματα από την προτεινόμενη μέθοδο, εκτός από την περίπτωση για $c = 30$. Εντούτοις, ο προτεινόμενος αλγόριθμος επιτυγχάνει την καλύτερη απόδοση όσον αφορά το σύνολο δεδομένων δοκιμής, το οποίο είναι ένα πολύ σημαντικό ζήτημα.

Παράδειγμα 3: Μονοδιάστατη συνθετική συνάρτηση

Σε αυτό το παράδειγμα, εξετάζεται η ακόλουθη τρισδιάστατη μη γραμμική συνάρτηση,

$$y = 0.6 \eta\mu(\pi x) + 0.3 \eta\mu(3\pi x) + 0.1 \eta\mu(5\pi x) \text{ με } x \in [-1, 1].$$

Όπως και στις αναφορές [70,75], δημιουργήθηκαν 200 ζευγάρια δεδομένων εισόδου-εξόδου. Το σύνολο των δεδομένων διαιρέθηκε τυχαία σε σύνολο δεδομένων εκπαίδευσης που περιέχουν το 60% από το σύνολο δεδομένων συμπεριλαμβανομένου του υπολοίπου 40% των δεδομένων δοκιμής.

Πίνακας 6.7. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης.

c	RBFN 1	RBFN 2	Προτεινόμενος αλγόριθμος		
			$\mathcal{G} = 0.3$	$\mathcal{G} = 0.5$	$\mathcal{G} = 0.7$
4	$0.0318 \pm 2.1E-5$	$0.0278 \pm 2.5E-4$	$0.0228 \pm 7.1E-3$	$0.0208 \pm 4.3E-3$	$0.0238 \pm 4.1E-3$
6	$0.0348 \pm 1.5E-4$	$0.0279 \pm 4.5E-4$	$0.0205 \pm 2.2E-3$	$0.0203 \pm 5.2E-4$	$0.0208 \pm 1.3E-3$
8	$0.0231 \pm 6.9E-4$	$0.0174 \pm 3.3E-3$	$0.0192 \pm 3.2E-4$	$0.0171 \pm 1.5E-4$	$0.0193 \pm 9.5E-4$
10	$0.0177 \pm 5.6E-3$	$0.0111 \pm 4.9E-3$	$0.0062 \pm 2.4E-4$	$0.0062 \pm 3.7E-5$	$0.0066 \pm 3.9E-4$
12	$0.0103 \pm 1.9E-3$	$0.0053 \pm 1.2E-3$	$0.0019 \pm 1.4E-4$	$0.0019 \pm 1.1E-4$	$0.0024 \pm 5.2E-4$

Πίνακας 6.8. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής.

c	RBFN 1	RBFN 2	Προτεινόμενος αλγόριθμος		
			$\mathcal{G} = 0.3$	$\mathcal{G} = 0.5$	$\mathcal{G} = 0.7$
4	$0.0469 \pm 7.2E-5$	$0.0423 \pm 4.1E-4$	$0.0302 \pm 7.2E-3$	$0.0308 \pm 4.1E-3$	$0.0329 \pm 4.4E-3$
6	$0.0602 \pm 5.3E-4$	$0.0439 \pm 2.5E-3$	$0.0327 \pm 1.9E-3$	$0.0325 \pm 9.1E-4$	$0.0326 \pm 1.3E-3$
8	$0.0225 \pm 1.1E-3$	$0.0309 \pm 5.9E-3$	$0.0210 \pm 3.8E-4$	$0.0185 \pm 2.1E-4$	$0.0207 \pm 1.5E-3$
10	$0.0338 \pm 5.2E-3$	$0.0213 \pm 7.2E-3$	$0.0063 \pm 4.0E-4$	$0.0063 \pm 7.7E-5$	$0.0067 \pm 4.3E-4$
12	$0.0252 \pm 2.4E-3$	$0.0124 \pm 3.2E-3$	$0.0023 \pm 1.5E-3$	$0.0025 \pm 8.5E-4$	$0.0032 \pm 1.3E-3$

Πίνακας 6.9. Συγκριτική ανάλυση από την άποψη του MSE.

Μοντέλο	Αρ. κόμβων	MSE στα δεδομένα εκπαίδευσης	MSE στα δεδομένα δοκιμής
Pedrycz et al. [75]	≤ 33	0.0029 ± 0.001	0.0052 ± 0.004
	≤ 33	0.0033 ± 0.001	0.0035 ± 0.002
Park et al. [70]	6	$0.0045 \pm 6.2E-4$	0.0281 ± 0.014
Προτεινόμενο RBF ($\vartheta = 0.5$)	6	$0.0203 \pm 5.2E-4$	$0.0326 \pm 1.3E-3$
	9	$0.0061 \pm 4.5E-6$	$0.0062 \pm 1.2E-5$
	12	$0.0019 \pm 1.1E-4$	$0.0025 \pm 8.5E-4$

Πίνακας 6.10. Ένα δίκτυο που παράγεται από τον προτεινόμενο αλγόριθμο για $c = 12$ και $\theta = 0.5$

i	v_i	σ_i	w_i
1	-0.1742	0.5505	-767.10382
2	0.1743	0.7829	131210.7606
3	-0.0136	0.6576	15253.4538
4	-0.3919	0.4054	22.3588
5	0.7131	1.1421	-55358.2547
6	-0.6224	0.2517	-1.3866
7	0.5488	1.0325	73242.9523
8	0.2707	0.8472	-90990.2072
9	0.0862	0.7242	-72854.9244
10	0.3970	0.9313	-14465.5592
11	-0.8783	0.1065	-0.5235
12	0.8978	1.2652	14709.1521

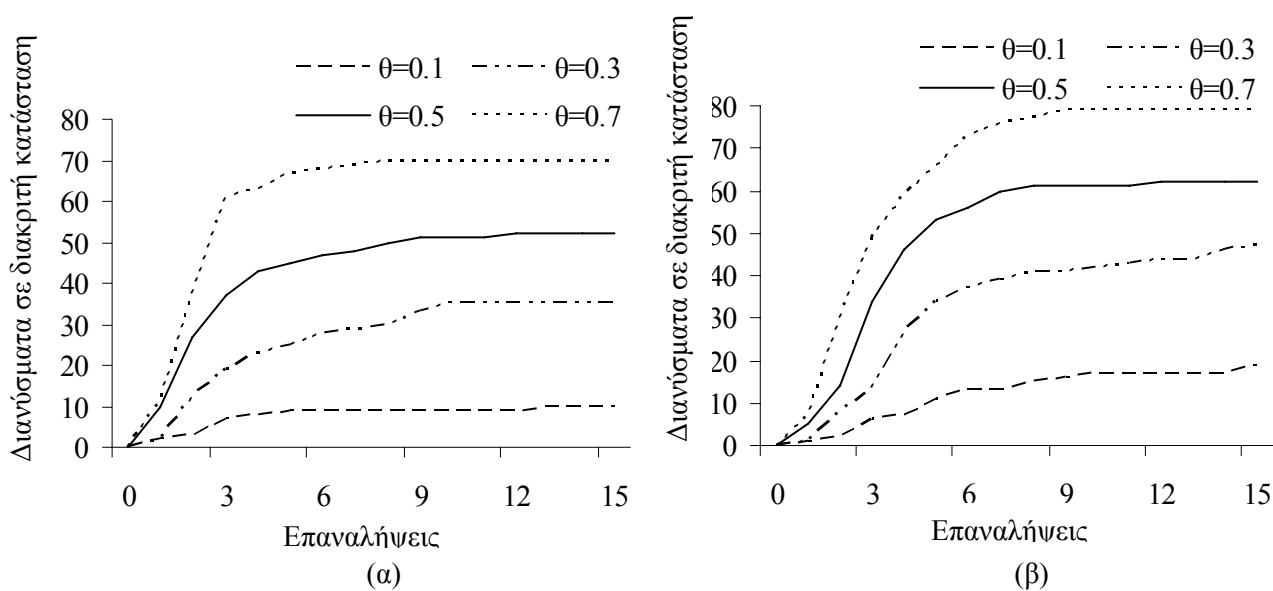
Οι τιμές MSE και οι αντίστοιχες διαφορές που λαμβάνονται από τα τέσσερα δίκτυα για τα δεδομένα εκπαίδευσης και δοκιμής συνοψίζονται στους πίνακες 6.6 και 6.7, αντίστοιχα. Σαφώς, ο προτεινόμενος αλγόριθμος επιτυγχάνει καλύτερο MSE και στα δεδομένα εκπαίδευσης και στα δεδομένα δοκιμής.

Ο πίνακας 6.9 συγκρίνει τις αποδόσεις του δικτύου για $\theta = 0.5$ με τα δίκτυα των Pedrycz et al. [75] και Park et al. [70] που υπάρχουν στη διεθνή βιβλιογραφία. Τα αποτελέσματα που αναφέρονται σε αυτόν τον πίνακα είναι ιδιαίτερα πειστικά, δεδομένου ότι σε όλες τις περιπτώσεις, η προτεινόμενη μέθοδος ξεπέρασε όλες τις υπόλοιπες.

Τέλος, ο πίνακας 6.10 παρέχει τις παραμέτρους του προτύπου ενός δικτύου με 12 κρυφούς κόμβους που παράγονται από τον αλγόριθμο για $\theta=0.5$. Η αντίστοιχη απόδοση ήταν ίση με 0.0019 για τα δεδομένα εκπαίδευσης και 0.0025 για τα δεδομένα δοκιμής.

Παράδειγμα 4: Διαδικασία ωρίμανσης αποστάγματος κρασιού

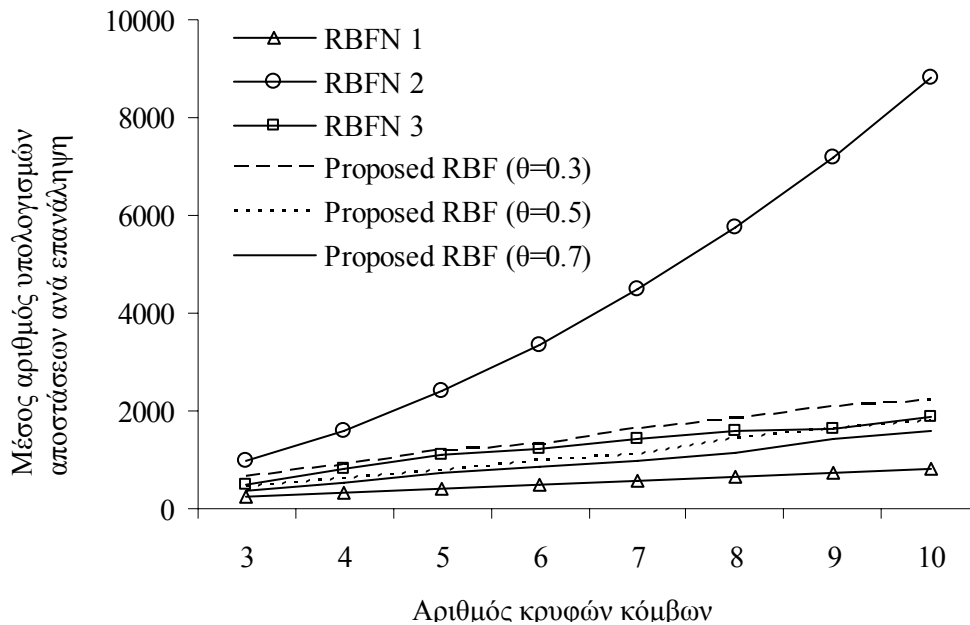
Σε αυτό το παράδειγμα εφαρμόζεται η προτεινόμενη μέθοδος σε ένα πραγματικό πρόβλημα ζώης αποκαλούμενο διαδικασία ωρίμανσης αποστάγματος. Φρεσκο-αποσταγμένα οινοπνεύματα όπως τα αποστάγματα, τα ουίσκι και τα ρούμια κρασιού αποθηκεύονται συνήθως μέσα σε δρύινα βαρέλια για αρκετά έτη για να βελτιώσουν το άρωμα και την γεύση τους.



Σχήμα 6.13. Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού επανάληψης για: (α) $c = 5$, και (β) $c = 10$.

Κατά τη διάρκεια αυτής της διαδικασίας, η οποία είναι γνωστή και ως διαδικασία ωρίμανσης αποστάγματος, διάφορα ξύλινα συστατικά εξάγονται και πολλές χημικές αντιδράσεις πραγματοποιούνται. Λόγω του μεγάλου αριθμού παραγόντων που έχουν επίπτωση στην διαδικασία ωρίμανσης, οι ακριβείς μηχανισμοί δεν έχουν γίνει κατανοητοί εντελώς [79]. Επιπλέον, η απουσία ενός αξιόπιστου χημικού δείκτη για να δείξει την πρόοδο της διαδικασίας ωρίμανσης εισάγει ορισμένα προβλήματα σχετικά με

το σχηματισμό του τελικού μίγματος που ελέγχει τη στρατηγική της γραμμής παραγωγής. Επομένως, η ποιοτική εκτίμηση είναι συνήθως βασισμένη στην οργανοληπτική δοκιμή των εμπειρογνομόνων των δειγμάτων από τα βαρέλια. Στο παρόν παράδειγμα, μελετάμε τη διαδικασία ωρίμανσης των παλαιών αποσταγμάτων κρασιού.



Σχήμα 6.14. Μέσος αριθμός υπολογισμών αποστάσεων ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα 3.

Ο προτεινόμενος αλγόριθμος χρησιμοποιήθηκε για να σχεδιάσει τα δίκτυα ένα RBF για την πρόβλεψη της ποιότητας του αρώματος των οινοπνευμάτων αποστάγματος κρασιού. Οι μεταβλητές εισόδου και εξόδου του συστήματος είναι:

$x_1 =$ Χρήση βαρελιών: ο αριθμός ξαναγεμισμάτων για κάθε βαρέλι (μεταβλητή εισόδου)

$x_2 =$ Ηλικία βαρελιών: η περίοδος χρήσης κάθε βαρελιού στα έτη (μεταβλητή εισόδου)

$x_3 =$ Ηλικία αποστάγματος: η περίοδος κάθε αποστάγματος που περιλαμβάνεται σε ένα συγκεκριμένο βαρέλι στα έτη (μεταβλητή εισόδου)

$y =$ Άρωμα είναι το άρωμα του ωριμασμένου αποστάγματος (μεταβλητή παραγωγής)

Το πλήρες σύνολο δεδομένων μπορεί να βρεθεί στην αναφορά [80]. Εντούτοις, για λόγους ευκολίας τα 120 ζευγάρια δεδομένων εισόδου-εξόδου που χρησιμοποιούνται σε αυτό το πείραμα δίνονται στο πίνακα 6.13.

Πίνακας 6.11. NRMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης

c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
3	2.880 ± 0.0237	0.983 ± 0.0045	1.075 ± 0.0001	0.970 ± 0.0295	0.959 ± 0.0114	0.960 ± 0.0913
4	2.581 ± 0.0531	0.976 ± 0.0032	0.971 ± 0.0153	0.822 ± 0.0197	0.821 ± 0.0260	0.821 ± 0.0291
5	2.268 ± 0.0777	0.921 ± 0.0182	0.896 ± 0.0295	0.635 ± 0.0263	0.576 ± 0.0269	0.577 ± 0.0255
6	2.228 ± 0.0673	0.775 ± 0.0248	0.845 ± 0.0347	0.508 ± 0.0139	0.477 ± 0.0137	0.480 ± 0.0187
7	2.227 ± 0.0945	0.710 ± 0.0577	0.791 ± 0.0492	0.474 ± 0.0102	0.464 ± 0.0100	0.473 ± 0.0142
8	2.203 ± 0.1114	0.660 ± 0.0455	0.771 ± 0.0321	0.462 ± 0.0812	0.476 ± 0.0954	0.474 ± 0.0112
9	2.142 ± 0.1117	0.638 ± 0.0451	0.761 ± 0.0335	0.486 ± 0.0771	0.480 ± 0.0898	0.495 ± 0.0939
10	2.126 ± 0.1421	0.621 ± 0.0576	0.787 ± 0.0422	0.498 ± 0.0523	0.493 ± 0.0755	0.502 ± 0.0843

Πίνακας 6.12. NRMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα δοκιμής

c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
3	3.062 ± 0.3322	1.323 ± 0.0024	1.084 ± 0.0173	1.275 ± 0.0568	1.265 ± 0.1413	1.259 ± 0.1113
4	2.828 ± 0.4680	1.338 ± 0.0005	1.044 ± 0.0211	1.083 ± 0.1480	1.079 ± 0.1876	1.101 ± 0.1841
5	2.791 ± 0.1407	1.308 ± 0.0175	1.084 ± 0.1155	0.796 ± 0.1617	0.779 ± 0.1625	0.763 ± 0.1028
6	2.741 ± 0.1011	1.169 ± 0.0370	0.978 ± 0.0605	0.646 ± 0.0910	0.671 ± 0.0734	0.691 ± 0.1054
7	2.702 ± 0.1174	1.096 ± 0.0498	0.896 ± 0.0869	0.612 ± 0.0743	0.637 ± 0.0841	0.658 ± 0.0823
8	2.643 ± 0.1228	1.039 ± 0.0325	0.843 ± 0.0823	0.608 ± 0.0625	0.636 ± 0.0819	0.647 ± 0.0971
9	2.606 ± 0.1271	1.015 ± 0.0397	0.822 ± 0.0641	0.587 ± 0.0479	0.602 ± 0.0536	0.626 ± 0.0723
10	2.559 ± 0.0972	1.007 ± 0.0341	0.807 ± 0.0544	0.590 ± 0.0499	0.599 ± 0.0496	0.625 ± 0.0469

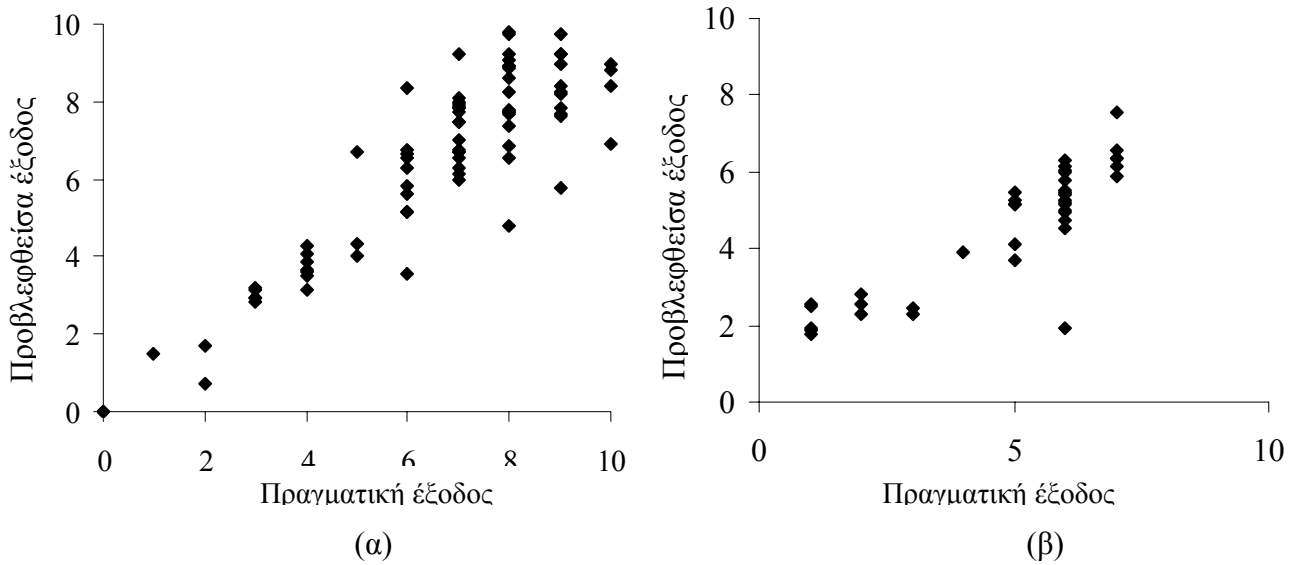
Για να εκτελεστούν οι προσομοιώσεις διαιρέθηκε το σύνολο δεδομένων σε υποσύνολο δεδομένων εκπαίδευσης που περιείχαν τα 80 πρώτα δείγματα και ένα υποσύνολο δοκιμής που αποτελείται από τα υπόλοιπα 40 δείγματα. Επιλέχθηκε το NRMSE ως δείκτης απόδοσης. Το σχήμα 6.13 επεξηγεί πώς η μετάβαση από ασαφή (fuzzy) σε διακριτή (crisp) κατάσταση επηρεάζεται από την παράμετρο θ για τις διάφορες τιμές του αριθμού κρυφών κόμβων.

Πίνακας 6.13. Σύνολο δεδομένων για τη διαδικασία ωρίμανσης αποστάγματος κρασιού.

x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3	y	x_1	x_2	x_3	y
9	12.65	6.14	6	8	21.01	7.38	7	3	16.00	1.38	3	9	16.83	3.57	6
8	20.44	6.14	6	4	21.96	7.48	8	4	21.01	1.47	3	9	19.87	3.57	6
9	16.07	6.24	6	9	21.01	7.48	7	5	23.86	1.47	3	4	9.80	3.57	7
8	23.48	6.24	5	6	24.05	7.48	7	2	19.49	1.57	3	6	19.11	3.67	6
4	22.72	6.33	8	2	16.64	7.48	9	8	20.44	1.66	3	7	9.80	3.67	6
5	16.07	6.53	7	6	20.82	7.76	7	1	14.93	1.76	4	8	20.44	3.67	6
10	23.67	6.62	6	8	24.05	7.86	7	1	21.77	1.95	4	1	12.65	3.67	7
8	16.26	6.62	7	4	21.01	7.95	8	1	12.08	1.95	5	1	9.61	3.67	7
6	22.72	6.62	6	10	24.62	8.05	8	9	20.44	2.14	4	4	20.63	3.67	6
3	9.80	6.62	8	2	24.05	8.05	9	7	16.64	2.14	4	8	9.80	3.86	6
8	23.29	3.95	6	1	21.01	8.05	9	5	22.34	2.14	4	7	13.22	3.86	6
6	10.37	3.95	6	9	9.80	8.34	9	4	16.26	2.14	4	5	9.80	3.86	7
3	13.03	3.95	7	5	9.61	8.34	10	0	9.04	2.33	6	2	18.73	3.86	7
4	21.77	3.95	7	4	9.04	8.34	10	8	20.44	2.42	4	10	24.62	3.95	5
0	13.98	4.14	8	0	9.80	8.34	10	9	13.60	2.52	5	4	12.46	2.52	5
7	9.42	5.09	8	2	16.83	8.34	10	4	9.42	0.24	0	1	3.98	2.71	6
9	9.80	5.19	6	9	9.61	6.62	7	5	14.93	0.24	0	8	22.72	2.71	4
5	9.61	5.29	8	7	9.80	6.72	7	1	8.85	0.43	2	5	15.12	2.90	5
9	20.44	5.38	6	1	9.42	6.72	9	1	11.89	0.52	2	5	9.04	3.09	6
7	13.03	5.38	8	9	13.22	6.81	7	4	12.08	0.52	1	3	18.92	3.09	5
3	13.60	5.38	8	5	9.80	6.81	7	7	12.46	0.62	1	7	13.03	3.19	5
0	9.80	5.38	9	5	13.60	6.81	8	6	19.49	0.62	1	10	17.21	3.28	5
1	23.29	5.38	8	3	9.61	6.81	8	10	13.98	0.62	1	2	18.73	3.28	6
0	14.93	5.38	9	7	13.41	6.91	7	6	12.08	1.00	2	7	9.80	3.38	6
6	20.25	5.76	7	1	14.36	6.91	9	2	22.34	1.09	1	3	9.80	3.38	6
3	17.02	5.76	8	9	17.02	7.00	7	8	16.83	1.19	1	1	9.61	3.38	7
1	20.63	5.76	9	5	16.83	7.00	7	7	9.80	1.28	3	8	12.65	3.48	6
0	12.84	5.86	9	3	13.03	7.10	9	3	22.34	1.28	2	7	16.64	3.48	6
5	13.41	5.95	8	3	17.02	7.10	8	5	9.80	1.38	3	3	16.64	3.48	6
4	19.49	5.95	7	8	16.64	7.38	8	8	22.72	1.38	2	2	21.39	3.48	6

Το σχήμα 6.14 επεξηγεί τις υπολογιστικές προσπάθειες των τεσσάρων δικτύων. Οι πίνακες 6.11 και 6.12 συγκρίνουν τα τέσσερα δίκτυα για τους διάφορους αριθμούς κρυφών κόμβων. Για το προτεινόμενο δίκτυο το καλύτερο αποτέλεσμα στην περίπτωση των δεδομένων εκπαίδευσης αναφέρεται για $c = 8$ και $\theta = 0.3$, ενώ στα δεδομένα δοκιμής για $c = 10$ και $\theta = 0.3$.

Τέλος, το σχήμα 6.15 παρουσιάζει τα γραφήματα διασποράς μεταξύ των αρχικών και προβλεφθέντων αποτελεσμάτων για $c = 10$ και $\theta = 0.3$. Οι τιμές NRMSE για αυτόν τον αριθμό ήταν: 0.4761 (δεδομένα εκπαίδευσης) και 0.5099 (δεδομένα δοκιμής).

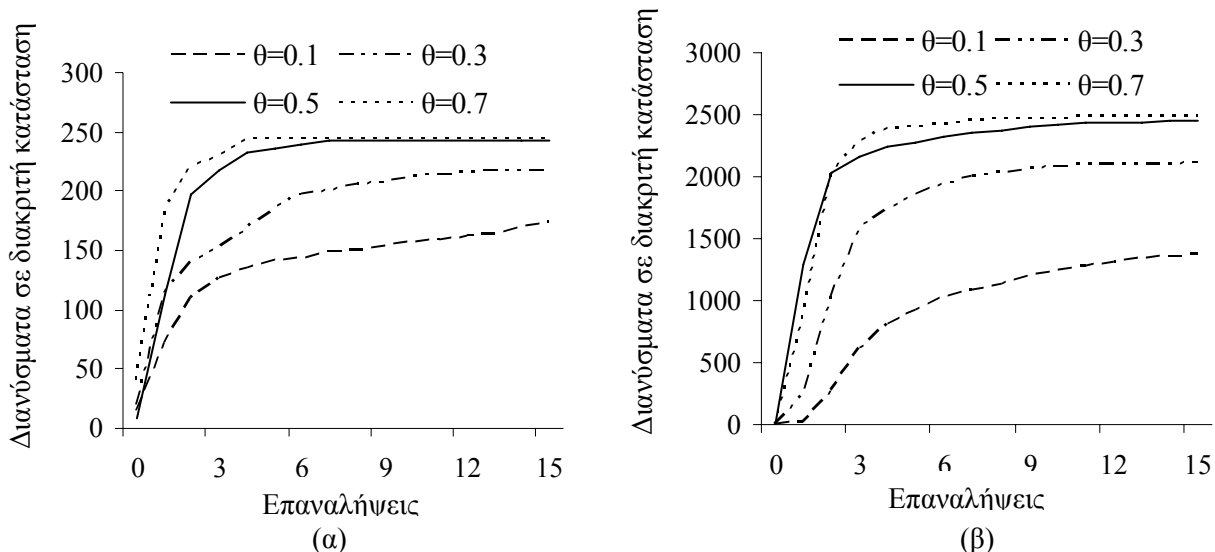


Σχήμα 6.15. Γραφήματα διασποράς για $c = 10$ και $\theta = 0.3$ μεταξύ της αρχικής και προβλεφθείσας εξόδου: (α) δεδομένα εκπαίδευσης, (β) δεδομένα δοκιμής

Παράδειγμα 5: Στοιχεία Μηχανικής Μάθησης

Σε αυτό το παράδειγμα, θεωρούμε δύο σύνολα δεδομένων από το Αρχείο Φύλαξης Μηχανικής Μάθησης (Repository Machine Learning) και συγκεκριμένα, τα δεδομένα Auto MPG και τα δεδομένα Abalone. Το σύνολο δεδομένων Auto MPG περιλαμβάνει 392 ζεύγη εισόδου-εξόδου (μετά την αφαίρεση 6 ελλιπών περιπτώσεων), όπου ο χώρος εισόδου περιλαμβάνει 7 εισόδους (δηλαδή τον αριθμό των κυλίνδρων, μετατόπιση, ιπποδύναμη, το βάρος, επιτάχυνση, έτος κατασκευής, προέλευση) και μία έξοδο (μίλια ανά γαλόνι-mpg). Το σύνολο δεδομένων Abalone περιλαμβάνει 4177 ζεύγη εισόδου-εξόδου (με 8 εισόδους). Όλα τα πειράματα διεξήχθησαν με το διαχωρισμό των διαθέσιμων δεδομένων σε ένα σύνολο δεδομένων εκπαίδευσης (60% του συνόλου των δεδομένων) και το σύνολο δεδομένων δοκιμής (40% των συνολικών δεδομένων). Τα τέσσερα δίκτυα συγκρίθηκαν όσον αφορά την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (RMSE). Και για τα δύο σύνολα δεδομένων, με τη μορφή μετάβασης από ασαφή σε διακριτή λειτουργία δίνεται στο σχήμα 6.16 για $c = 10$ και διάφορες τιμές της παραμέτρου θ .

Οι πίνακες 6.14 και 6.15 εμφανίζουν τις επιδόσεις των δικτύων όσον αφορά τα δεδομένα εκπαίδευσης και δοκιμής, για τις δύο πειραματικές περιπτώσεις.



Σχήμα 6.16. Διανύσματα που μεταφέρονται στην διακριτή κατάσταση (crisp mode) ως συνάρτηση του αριθμού επανάληψης με $c=10$ για: (α) τα δεδομένα Auto MPG, και (β) τα δεδομένα Abalone.

Πίνακας 6.14. RMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης και δοκιμής (παράδειγμα Auto MPG)

	c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta=0.3$)	Προτεινόμενο Δίκτυο ($\theta=0.5$)	Προτεινόμενο Δίκτυο ($\theta=0.7$)
Εκπαίδευση	4	3.806 ± 0.016	3.054 ± 0.288	3.036 ± 0.323	2.529 ± 0.257	2.524 ± 0.303	2.531 ± 0.230
	6	3.072 ± 0.019	2.603 ± 0.135	2.808 ± 0.098	2.484 ± 0.086	2.520 ± 0.072	2.531 ± 0.094
	8	2.886 ± 0.028	2.601 ± 0.004	2.612 ± 0.065	2.484 ± 0.135	2.473 ± 0.063	2.498 ± 0.143
	10	2.720 ± 0.030	2.585 ± 0.014	2.523 ± 0.052	2.446 ± 0.054	2.465 ± 0.046	2.469 ± 0.048
	12	2.631 ± 0.006	2.512 ± 0.028	2.508 ± 0.096	2.412 ± 0.103	2.403 ± 0.080	2.419 ± 0.092
Δοκιμή	4	7.699 ± 0.022	7.159 ± 0.108	7.173 ± 0.253	7.189 ± 0.306	7.195 ± 0.457	7.197 ± 0.287
	6	7.410 ± 0.010	7.174 ± 0.107	7.191 ± 0.189	7.140 ± 0.137	7.134 ± 0.102	7.150 ± 0.126
	8	7.222 ± 0.052	7.127 ± 0.001	7.100 ± 0.073	7.087 ± 0.068	7.093 ± 0.052	7.098 ± 0.073
	10	7.158 ± 0.104	7.111 ± 0.112	7.025 ± 0.102	7.032 ± 0.078	7.023 ± 0.096	7.043 ± 0.088
	12	7.108 ± 0.062	7.078 ± 0.086	7.039 ± 0.052	6.908 ± 0.054	6.894 ± 0.057	6.913 ± 0.063

Οι πίνακες 6.16 και 6.17 συγκρίνουν την προτεινόμενη μέθοδο με το δίκτυο SFC [89] και το δίκτυο Pedrycz et al. [75], για τα δεδομένα Auto MPG και τα δεδομένα Abalone, αντίστοιχα. Στον πίνακα 6.16, η προτεινόμενη μέθοδος αποδίδει καλύτερα στα δεδομένα εκπαίδευσης και χειρότερα στην περίπτωση των δεδομένων δοκιμής. Από την άλλη πλευρά, ο πίνακας 6.17 δείχνει άμεσα την ανωτερότητα του αλγορίθμου μας σε όλες τις περιπτώσεις.

Πίνακας 6.15. RMSE μέσες τιμές και οι αντίστοιχες διακυμάνσεις (variances) για τα δεδομένα εκπαίδευσης και δοκιμής (παράδειγμα Abalone)

	c	RBFN 1	RBFN 2	RBFN 3	Προτεινόμενο Δίκτυο ($\theta = 0.3$)	Προτεινόμενο Δίκτυο ($\theta = 0.5$)	Προτεινόμενο Δίκτυο ($\theta = 0.7$)
Εκπαίδευση	10	2.856 ± 0.237	2.578 ± 0.026	2.558 ± 0.023	2.455 ± 0.055	2.449 ± 0.030	2.476 ± 0.080
	20	2.706 ± 0.175	2.317 ± 0.023	2.517 ± 0.098	2.109 ± 0.058	2.126 ± 0.072	2.113 ± 0.046
	30	2.567 ± 0.110	2.288 ± 0.027	2.454 ± 0.065	1.934 ± 0.050	1.887 ± 0.063	1.953 ± 0.043
Δοκιμή	10	2.939 ± 0.490	2.807 ± 0.111	2.727 ± 0.153	2.678 ± 0.130	2.685 ± 0.257	2.723 ± 0.096
	20	2.823 ± 0.108	2.414 ± 0.038	2.330 ± 0.189	2.259 ± 0.063	2.269 ± 0.102	2.269 ± 0.072
	30	2.625 ± 0.088	2.349 ± 0.019	2.237 ± 0.073	2.111 ± 0.067	2.011 ± 0.052	2.123 ± 0.057

Πίνακας 6.16. Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Auto MPG

Αρ. κρυφών κόμβων	SFC [89]		Προτεινόμενη Μέθοδος ($\theta = 0.5$)	
	Δεδομένα Εκπαίδευσης	Δεδομένα Δοκιμής	Δεδομένα Εκπαίδευσης	Δεδομένα Δοκιμής
6	2.75	6.77	2.52	7.13
8	2.71	6.61	2.47	7.09
10	2.74	6.57	2.46	7.02
12	2.68	6.52	2.40	6.89

Πίνακας 6.17. Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Abalone

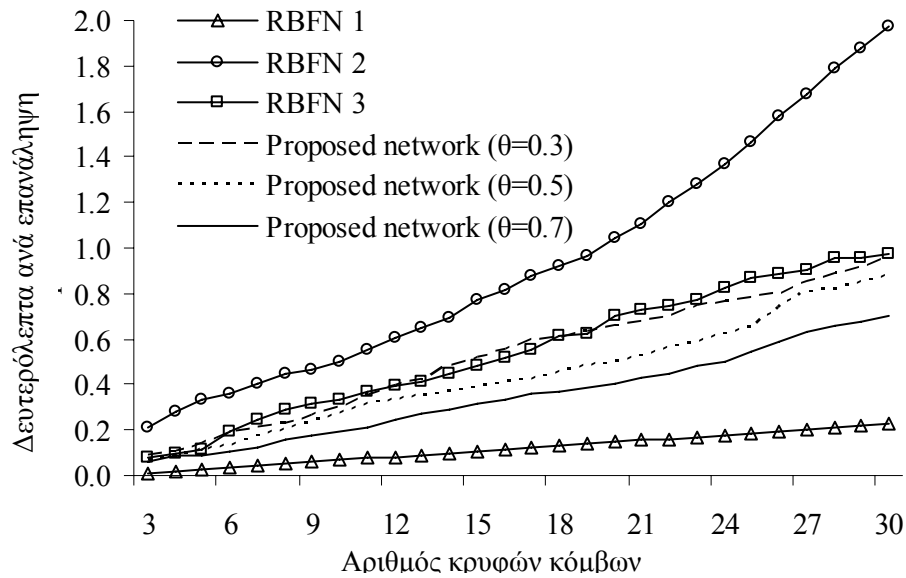
Μοντέλο	Αριθμός κρυφών κόμβων	RMSE
Pedrycz et al. [75]	(Δεδομένα $m=2$)	30
	Εκπαίδευσης) $m=4$	30
	(Δεδομένα $m=2$)	30
	Δοκιμής) $m=4$	30
Προτεινόμενη μέθοδος ($\theta = 0.5$)	(Δεδομένα Εκπαίδευσης)	30
	(Δεδομένα Δοκιμής)	30

Όπως και στο [89], για το παράδειγμα Auto MPG, πραγματοποιήσαμε μία επιπλέον πειραματική προσομοίωση χρησιμοποιώντας μόνο δύο μεταβλητές (δηλαδή, την υποδύναμη και την επιτάχυνση). Ο πίνακας 6.18 δείχνει τα συγκριτικά αποτελέσματα της προσομοίωσης. Και πάλι, ο πίνακας αυτός επιβεβαιώνει την ακριβή συμπεριφορά του δικτύου μας. Παρατηρήστε ότι όταν το $\theta = 0.5$ τότε το προτεινόμενο δίκτυο

υπερτερεί τόσο από την μέθοδο SFC όσο και από την μέθοδο CFC σε όλες τις πειραματικές περιπτώσεις.

Πίνακας 6.18. Συγκριτική ανάλυση ως προς τις RMSE μέσες τιμές για τα δεδομένα Auto MPG (2 εισόδους).

Μέθοδος	Αριθμός κρυφών κόμβων	Δεδομένα Εκπαίδευσης	Δεδομένα Δοκιμής
CFC [72]	12	5.67	6.70
	6	3.22	6.61
SFC [89]	8	3.04	6.49
	10	3.01	6.42
	12	2.89	6.16
	6	3.22	6.59
Προτεινόμενο δίκτυο ($\theta=0.3$)	8	3.11	6.46
	10	2.94	6.28
	12	2.90	6.24
Προτεινόμενο δίκτυο ($\theta=0.5$)	6	3.20	6.57
	8	2.98	6.27
	10	2.93	6.20
Προτεινόμενο δίκτυο ($\theta=0.7$)	12	2.85	6.11
	6	3.24	6.61
	8	3.15	6.43
	10	2.94	6.27
	12	2.93	6.27



Σχήμα 6.17. Δευτερόλεπτα ανά επανάληψη ως συνάρτηση του αριθμού κρυφών κόμβων για το παράδειγμα δεδομένων Abalone.

Το γεγονός ότι το σύνολο δεδομένων Abalone είναι ένα μεγάλο σύνολο δεδομένων που μας δίνει τη δυνατότητα να μετρηθεί ποσοτικά ο χρόνος σε δευτερόλεπτα ανά επανάληψη που απαιτείται από τους τέσσερις αλγόριθμους. Για την εκτέλεση της προσομοίωσης τρέξαμε όλους τους αλγόριθμους χρησιμοποιώντας μια μηχανή υπολογιστή με επεξεργαστή διπλού πυρήνα στα 2.13 GHz, και τα αποτελέσματα δίνονται στο σχήμα 6.17. Ο αριθμός αυτός υποστηρίζει σθεναρά τη θεωρητική υπολογιστική ανάλυση πολυπλοκότητας και σαφώς μπορούμε να διακρίνουμε την ομοιότητα με το σχήμα 6.7.

6.4 Συμπεράσματα

Στο παρόν κεφάλαιο έχει αναπτυχθεί ένας υβριδικός αλγόριθμος συσταδοποίησης για την αποτελεσματική σχεδίαση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Υπάρχουν δύο βασικές απόψεις σχεδίου της προτεινόμενης μεθοδολογίας. Ο πρώτος αφορά τον τρόπο που καθορίζεται η δομή των συστάδων στο πολυδιάστατο χώρο των χαρακτηριστικών γνωρισμάτων (feature space). Για να ολοκληρωθεί αυτός ο στόχος, χρησιμοποιείται μια στρατηγική μετάβασης από την ασαφή κατάσταση (fuzzy mode), όπου κάθε διάνυσμα εκπαίδευσης ανήκει σε πολλές συστάδες με διαφορετικούς βαθμούς συμμετοχής, στην διακριτή κατάσταση (crisp mode), όπου κάθε διάνυσμα εκπαίδευσης ανήκει μόνο σε μια συστάδα. Αυτή η μετάβαση πραγματοποιείται μέσω των αναλυτικών όρων που εξάγονται από την ελαχιστοποίηση μιας εξειδικευμένης αντικειμενικής συνάρτησης (objective function). Το τελικό αποτέλεσμα είναι ένας υβριδικός διαμερισμός όπου οι ασαφείς και οι διακριτοί όροι λαμβάνονται υπόψη ταυτόχρονα. Ο ανταγωνισμός μεταξύ των υβριδικών συστάδων πραγματοποιείται όσον αφορά τις κοινές ασαφείς περιοχές τους και τις διακριτές περιοχές τους, επίσης. Έτσι φαίνεται ότι το προτεινόμενο σχέδιο συσταδοποίησης συνδυάζει τα πλεονεκτήματα της διακριτής και της ασαφούς συσταδοποίησης, παράγοντας μια αξιόπιστη εκτίμηση των κέντρων των συναρτήσεων βάσης. Η δεύτερη άποψη αφορά την εκτίμηση των πλατών (widths) των πυρήνων. Για τον υπολογισμό του πλάτους της συνάρτησης βάσης χρησιμοποιήσαμε ένα συγκεκριμένο τύπο που θεωρεί την υποκείμενη κατανομή των δεδομένων της κάθε υβριδικής συστάδας. Αυτή η απόσταση λαμβάνει υπόψη την υποκείμενη δομή

δεδομένων κάθε συστάδας όπως αυτή η δομή αποκαλύπτεται από τον υβριδικό διαμερισμό και χρησιμοποιείται για να υπολογίσει αποτελεσματικά τα πλάτη (widths) των πυρήνων. Η απόδοση της προτεινόμενης μεθοδολογίας αξιολογήθηκε μέσω της εφαρμογής της, σε πέντε περιπτώσεις δοκιμής. Διάφορα συγκριτικά αποτελέσματα έδειξαν ότι ο προτεινόμενος αλγόριθμος είναι σε θέση να δημιουργήσει ακριβή νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης.

Ανάπτυξη μεθόδου Κοκκοποίησης Πληροφοριών για την Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης

7.1 Εισαγωγή

Η κοκκοποίηση πληροφοριών φανερώνεται στην κατασκευή των θεωριών των κόκκων και των διαδικασιών που περιλαμβάνουν τους κόκκους, οι οποίοι εξαρτώνται από τις συλλήψεις που συνδέονται με τη διαδικασία διαχωρισμού (portioning) ενός συνόλου δεδομένων σε διάφορα ευπροσδιόριστα σύνολα όπως υποσύνολα (subsets), κλάσεις (classes) και συστάδες (clusters) [3,4]. Η έννοια του κόκκου, εντούτοις, υπερβαίνει την έννοια του συνόλου, δεδομένου ότι θεωρείται ως μορφή κάτω από την οποία διάφορα αρχικά μετρήσιμα αντικείμενα (π.χ. δεδομένα) συλλαμβάνονται ως ενιαίο αντικείμενο που μπορεί να αναγνωριστεί, κατέχοντας οποιοδήποτε βαθμό πολυπλοκότητας. Επιπλέον, ένας κόκκος θεωρείται ότι κατέχει τις εσωτερικές ιδιότητες που δείχνουν τις αλληλεπιδράσεις μεταξύ των αντικειμένων που ανήκουν σε αυτό, καθώς και τις εξωτερικές ιδιότητες καθορίζοντας την αλληλεπίδρασή του με άλλους κόκκους [4,75]. Έτσι φαίνεται ότι ένας κόκκος μπορεί να αντιμετωπιστεί και ως συλλογή (δηλ. σύνολο) των μεμονωμένων αντικειμένων που χαρακτηρίζονται από τις εσωτερικές ιδιότητές του και συνολικά που χαρακτηρίζονται από τις εξωτερικές ιδιότητές του. Λαμβάνοντας υπόψη την πρακτική εφαρμογή, υπάρχει μια σειρά υπολογιστικών πλαισίων κάτω από την ομπρέλα των οποίων, η κοκκοποίηση πληροφοριών μπορεί να αναπτυχθεί αποτελεσματικά, να εφαρμοστεί και να αξιολογηθεί. Τα πλαίσια αυτά περιλαμβάνουν σύνολα και ανάλυση διαστημάτων, ασαφή σύνολα και σύνολα πιθανοτήτων [3,4,75]. Η μελέτη αυτή έχει σκοπό να ασχοληθεί με την ασαφή θεωρία που ως κατάλληλη

εννοιολογική και αλγοριθμική πλατφόρμα για τη δημιουργία και εφαρμογή κόκκων πληροφορίας κατά τη διαδικασία εκπαίδευσης των δικτύων συναρτήσεων ακτινικής βάσης (RBF).

Εξαιτίας της απλότητας και τις ικανότητες προσέγγισης, τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (RBF) έχουν εφαρμοστεί και αξιολογηθεί κάτω από διάφορες μορφές σε ένα ευρύ πεδίο εφαρμογών [23,30,32,48,51,97]. Μέχρι τώρα, η κοκκοποίηση πληροφοριών δεν έχει περιληφθεί εντατικά στην κατασκευή των νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης, αν και η δυνατότητα αυτής της προοπτικής φαίνεται να είναι ευνοϊκή. Οι κύριες εξελίξεις που εμφανίζονται στη σύνδεση της κοκκοποίησης πληροφοριών και στον σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης, εμπίπτουν στη σφαίρα των νευρο-ασαφών (neuro-fuzzy) τεχνικών [5,30,48,53,58,69,97]. Αυτές οι τεχνικές εκμεταλλεύονται τα οφέλη της θεωρίας της ασαφούς συσταδοποίησης (fuzzy clustering) στο σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Κατά την εφαρμογή της ασαφούς κοκκοποίησης στο σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης, μπορούμε να διακρίνουμε δύο φάσεις [79]. Η πρώτη φάση εξετάζει την κοκκοποίηση των αριθμητικών δεδομένων, και η δεύτερη ενσωματώνει τους παραχθέντες ασαφείς κόκκους στη διαδικασία εκπαίδευσης. Να αναγνωρίσουμε ότι οι ασαφείς κόκκοι συμβάλλουν κατά τρόπο απλό στον υπολογισμό των κέντρων των συναρτήσεων ακτινικής βάσης. Εντούτοις, είναι εξίσου σημαντικός και ο προσδιορισμός των αντίστοιχων πλατών (widths). Οι προσανατολισμένες στην ασαφή κοκκοποίηση προσεγγίσεις λύνουν αυτό το πρόβλημα με την έκφραση των συναρτήσεων βάσης από την άποψη των ασαφών συναρτήσεων συμμετοχής, όπως καθορίζονται μέσω της διαδικασίας κοκκοποίησης [71,72,74,76]. Το κύριο εμπόδιο αυτής της τεχνικής είναι ότι μας επιτρέπει να εξετάσουμε μόνο έναν τύπο συναρτήσεων βάσης, αν και αυτό μπορεί να μην είναι πάντα αρμόζον [31]. Αφ' ετέρου, οι κλασσικές προσεγγίσεις περιλαμβάνουν μόνο τις θέσεις των κέντρων συστάδων [6,60,83], το οποίο άμεσα υπονοεί ότι η υποκείμενη δομή των δεδομένων δεν λαμβάνεται υπόψη.

Στο παρόν κεφάλαιο, παρέχεται μια διαφορετική πτυχή όσον αφορά αυτό το πρόβλημα. Συγκεκριμένα, έχουμε αναπτύξει μια συνάρτηση απόστασης μεταξύ των ασαφών κόκκων που συνδυάζει τους διαφορετικούς τύπους πληροφοριών. Το βασικό

σημείο είναι ότι χρησιμοποιούμε αυτήν την συνάρτηση προκειμένου να αντιμετωπιστεί κάθε κόκκος ως ένα αντικείμενο και έπειτα για να υπολογίσουμε αποτελεσματικά τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης. Οι κόκκοι πληροφοριών δημιουργούνται σε τρία στάδια. Το πρώτο στάδιο καθορίζει μια προσανατολισμένη προς το χρήστη (user-defined) ασαφή κοκκοποίηση (fuzzy granulation) για κάθε σύνολο δεδομένων. Το δεύτερο στάδιο κοκκοποίησης χρησιμοποιείται κυρίως για να προεπεξεργαστεί τα διαθέσιμα δεδομένα εκπαίδευσης. Για να επιτευχθεί πρακτικά αυτός ο στόχος χρησιμοποιείται ο ασαφής αλγόριθμος διαμερισμού (fuzzy partition) [95] για να ομαδοποιήσει όλα τα δεδομένα εισόδου εκπαίδευσης σε διάφορες συστάδες. Το τρίτο στάδιο κοκκοποίησης χρησιμοποιεί τη μέθοδο των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM), η οποία παράγει μια πιο αφηρημένη κοκκοποίηση του χώρου των χαρακτηριστικών γνωρισμάτων (feature space) με περαιτέρω επεξεργασία των αποτελεσμάτων του προηγούμενου βήματος. Το τελικό αλγοριθμικό σχήμα προϋποθέτει μια σύμπραξη (synergy) μεταξύ των ασαφών συνόλων και των νευρωνικών δικτύων.

7.2 Μερικές νέες έννοιες στη ασαφή κοκκοποίηση πληροφοριών

Σε αυτή την παράγραφο, εισάγονται νέες έννοιες στην ασαφή κοκκοποίηση πληροφοριών και βασισμένες σε αυτές, αναπτύσσεται μια μετρική απόσταση (metric distance) μεταξύ των ασαφών κόκκων. Πριν συνεχίσουμε με τη λεπτομερή μαθηματική ανάλυση, αναθεωρούμε την έννοια του συνυπολογισμού μεταξύ δύο ασαφών κόκκων όπως πρότεινε ο Pedrycz [75].

Εξετάζουμε ένα σύνολο δεδομένων $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, το οποίο καθορίζεται στο p - διάστατο χώρο χαρακτηριστικών γνωρισμάτων (feature space) $X = X_1 \times X_2 \times \dots \times X_p \subset \mathfrak{R}^p$, όπου X_j είναι το σύνολο δεδομένων στην j - οστή διάσταση ($1 \leq j \leq p$). Επιπλέον, ορίζουμε ως G_i και G_t δύο ασαφείς κόκκοι του \mathbf{X} . Κατόπιν, ο βαθμός συνυπολογισμού του G_i στο G_t όσον αφορά ένα συγκεκριμένο διάνυσμα $\mathbf{x} \in \mathbf{X}$ είναι,

$$(G_i \Rightarrow G_t)(\mathbf{x}) = \begin{cases} 1 & \text{εάν } u_i(\mathbf{x}) \leq u_t(\mathbf{x}) \\ 1 - u_i(\mathbf{x}) + u_t(\mathbf{x}) & \text{αλλιώς} \end{cases} \quad (7.1)$$

όπου $u_i(\mathbf{x})$ και $u_t(\mathbf{x})$ είναι οι αντίστοιχοι βαθμοί συμμετοχής (membership degrees). Παρατηρήστε ότι $0 \leq (G_i \Rightarrow G_t)(\mathbf{x}) \leq 1$. Κατόπιν, ο βαθμός συνυπολογισμού του G_i στο G_t όσον αφορά το σύνολο X είναι,

$$\|G_i \subset G_t\| = \frac{1}{\aleph(X)} \sum_{\mathbf{x}_k \in X} (G_i \Rightarrow G_t)(\mathbf{x}_k)$$

όπου $\aleph(X) = N$ είναι η πληθικότητα συνόλου (cardinality) του X . Τέλος, ο βαθμός ισότητας μεταξύ G_i και G_t είναι,

$$\|G_i = G_t\| = \frac{1}{\aleph(X)} \sum_{\mathbf{x}_k \in X} \min\{(G_i \Rightarrow G_t)(\mathbf{x}_k), (G_t \Rightarrow G_i)(\mathbf{x}_k)\} \quad (7.2)$$

Είναι προφανέστερο ότι οι κόκκοι G_i και G_t είναι ίσοι εάν και μόνο εάν ο βαθμός ισότητας μεταξύ τους ισούται με μονάδα, δηλ. $\|G_i = G_t\| = 1$.

Γενικά, υπάρχουν τρεις τρόποι να παραχθούν οι ασαφείς κόκκοι από ένα σύνολο δεδομένων [3,75]. Κατ' αρχάς, χρησιμοποιώντας τις προσανατολισμένες προς το χρήστη (user defined) μεθόδους, οι οποίες πραγματοποιούνται από το χρήστη (user) / τον εμπειρογνώμονα (expert). Τέτοιο είδος ασαφών κοκκοποιήσεων μπορεί να κατασκευαστεί με τον καθορισμό μιας σειράς ασαφών συνόλων σε κάθε σύνολο δεδομένων. Δεύτερον, χρησιμοποιώντας τις αλγοριθμικές προσεγγίσεις βελτιστοποίησης που έρχονται υπό μορφή ανεπίβλεπτων μηχανισμών εκπαίδευσης, όπως η ασαφής συσταδοποίηση (fuzzy clustering) ή η ασαφής διανυσματική κβαντοποίηση (fuzzy vector quantization – FVQ) (βλ ενότητα 2.7). Τρίτον, συνδυάζοντας τους ανωτέρω δύο τρόπους. Οι προσανατολισμένες προς το χρήστη προσεγγίσεις κωδικοποιούν τις πληροφορίες χαμηλότερων επιπέδων που είναι πολύ συγκεκριμένες και παράγονται από τη διαίσθηση των εμπειρογνομώνων και τη δομή των δεδομένων σε κάθε διάσταση, επίσης. Η δομή των δεδομένων σε κάθε διάσταση είναι εύκολα ορατή και εάν συνδυαστεί με την ειδική γνώση εμείς μπορεί να δημιουργήσουμε μια πολύ αποτελεσματική κοκκοποίηση του συνόλου των δεδομένων. Αφ' ετέρου, οι ασαφείς κόκκοι που παράγονται με τη ασαφή συσταδοποίηση (fuzzy clustering) κωδικοποιούν τις πληροφορίες υψηλού επιπέδου, οι

οποίες είναι περισσότερο αφηρημένες. Η πολυπλοκότητα αυτής της αφαίρεσης μπορεί να αντιπροσωπευθεί και να αποκαλυφθεί από την άποψη των χονδρών και λεπτών κόκκων (σχέσεων) [4,75].

Η βασική ιδέα αυτού του τμήματος εμπίπτει στη σφαίρα των ανωτέρω συλλήψεων. Συγκεκριμένα, θεωρήστε ότι έχουμε δημιουργήσει μια προσανατολισμένη προς το χρήστη ασαφή κοκκοποίηση (fuzzy granulation) σε κάθε πεδίο ορισμού X_j ($1 \leq j \leq p$) με το διαμελισμό της q_j στους ασαφείς αριθμούς $A_j^1, A_j^2, \dots, A_j^{q_j}$. Αυτή η κοκκοποίηση πρέπει να ικανοποιήσει ορισμένα κριτήρια. Μια λεπτομερής ανάλυση σε αυτό το θέμα αναπτύσσεται σε επόμενη ενότητα, στο παρόν κεφάλαιο. Εντούτοις, για περιγραφικούς λόγους αυτού του τμήματος, πρέπει να παρέχουμε τον ακόλουθο ορισμό, αποκαλούμενος ως *δ-επιπέδου κριτήριο κάλυψης* (*δ-level criterion coverage*) [75].

Ορισμός 1. Η προσανατολισμένη προς το χρήστη (user-defined) κοκκοποίηση (granulation) σε ολόκληρο το πεδίο ορισμού (universe of discourse) X_j ($1 \leq j \leq p$) ικανοποιεί το *δ-επιπέδου κριτήριο κάλυψης* (*δ-level criterion coverage*) εάν $\forall x \in X_j, \exists A_j^l : A_j^l(x) > \delta$ με $\delta \in (0,1]$ και $1 \leq l \leq q_j$.

Επιπλέον, υποθέστε ότι έχει εφαρμοστεί ανεξάρτητα μια αλγοριθμική κοκκοποίηση βελτιστοποίησης (παραδείγματος χάριν ασαφής συσταδοποίηση) στο διαθέσιμο σύνολο δεδομένων και έχει παραχθεί ένα σύνολο c κόκκων $\mathbf{G} = \{G_1, G_2, \dots, G_c\}$. Κατόπιν, βρίσκουμε ένα υβριδικό σχέδιο, το οποίο εμφανίζεται να είναι εύκαμπτο υπό την έννοια ότι έχουμε ταυτόχρονη πρόσβαση στις χαμηλού επιπέδου πληροφορίες που κωδικοποιούνται από τον προσανατολισμένο προς το χρήστη (user-defined) ασαφή διαμερισμό (fuzzy partition) και τις υψηλού επιπέδου πληροφορίες που κωδικοποιούνται με την ασαφή συσταδοποίηση (fuzzy clustering). Ο τελευταίος στόχος είναι να συνδυαστούν αυτού του δύο τύπου πληροφορίες προκειμένου να κατασκευαστεί μια μετρική απόσταση (metric distance) μεταξύ δύο κόκκων. Αυτή η απόσταση είναι βασισμένη στη δομή κάθε κόκκου όπως αυτή η δομή αποκαλύπτεται από την άποψη του συνδυασμού αυτών των δύο τύπων πληροφοριών.

Απεικονίζεται το πρόβλημα από την άποψη ενός τυχαίου διανύσματος δεδομένων $\mathbf{x}_k = [x_{k1}, \dots, x_{kj}, \dots, x_{kp}]^T \in \mathbf{X}$. Ο βαθμός συμμετοχής \mathbf{x}_k στον κόκκο G_i ($1 \leq i \leq c$) συμβολίζεται ως $u_i(\mathbf{x}_k)$, και ο βαθμός συμμετοχής του j -τιμήματος στο ασαφές σύνολο A_j^l όπως $\mu_j^l(x_{kj})$ με $1 \leq l \leq q_j$. Κατόπιν, οι ακόλουθοι δύο ορισμοί αναφέρονται στη μερική και συνολική συμβολή (*contribution*) του ασαφές συνόλου (fuzzy set) A_j^l στον κόκκο (granule) G_i .

Ορισμός 2. Η μερική συμβολή (*contribution*) του ασαφούς συνόλου (fuzzy set) A_j^l στον ασαφή κόκκο (fuzzy granule) G_i ορίζεται όσον αφορά το πεδίο ορισμού (*universe of discourse*) $\mathbf{x}_k \in \mathbf{X}$, και δίνεται ως ($1 \leq i \leq c$, $1 \leq j \leq p$, $1 \leq l \leq q_j$, $1 \leq k \leq N$),

$$\Gamma_{i,j,l,k} = \mu_j^l(x_{kj}) \cdot u_i(\mathbf{x}_k) \quad (7.3)$$

Ένας αποτελεσματικότερος τρόπος να οριστεί η συνάρτηση $\Gamma_{i,j,l,k}$, που διατηρεί επίσης τη γενικότητα του ορισμού 2, είναι ο επόμενος,

$$\Gamma_{i,j,l,k} = \begin{cases} u_i(\mathbf{x}_k), & \text{εάν } \mu_j^l(x_{kj}) = \max_{1 \leq r \leq q_j} \{\mu_j^r(x_{kj})\} \\ 0, & \text{αλλιώς} \end{cases} \quad (7.4)$$

Σε αυτό το κεφάλαιο, επιλέχθηκε η εξ. (7.3) για να υπολογιστεί η ποσότητα $\Gamma_{i,j,l,k}$. Έχει εξεταστεί επίσης η εξ. (7.4) σε διάφορες προσομοιώσεις που δεν αναφέρονται στο παρόν κεφάλαιο, από τις οποίες αποδείχθηκε ότι αυτή η επιλογή αποδίδει εξίσου καλά όπως και η πρώτη.

Ορισμός 3. Η συνολική συμβολή (*contribution*) του ασαφούς συνόλου (fuzzy set) A_j^l στον ασαφή κόκκο (fuzzy granule) G_i ορίζεται όσον αφορά ολόκληρο το πεδίο ορισμού (*universe of discourse*) $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ και δίνεται ως ($1 \leq i \leq c$, $1 \leq j \leq p$, $1 \leq l \leq q_j$),

$$Z_{i,j,l} = \sum_{k=1}^N \Gamma_{i,j,l,k} \quad (7.5)$$

Με βάση την ανωτέρω ανάλυση, οι επόμενοι ορισμοί εισάγουν την έννοια της ανομοιότητας μεταξύ δύο ασαφών κόκκων (fuzzy granules).

Ορισμός 4. Η ανομοιότητα (dissimilarity) μεταξύ δύο κόκκων (granules) G_i και G_t όσον αφορά το ασαφές σύνολο (fuzzy set) A_j^l ορίζεται ως $(1 \leq i \leq c, 1 \leq t \leq c, 1 \leq j \leq p, 1 \leq l \leq q_j)$,

$$\Delta Z_{i,t,j,l} = |Z_{i,j,l} - Z_{t,j,l}| \quad (7.6)$$

Ορισμός 5. Η ανομοιότητα (dissimilarity) μεταξύ δύο κόκκων (granules) G_i και G_t όσον αφορά ολόκληρο το πεδίο ορισμού (universe of discourse) X_j ορίζεται ως $(1 \leq i \leq c, 1 \leq t \leq c, 1 \leq j \leq p)$,

$$D_j(G_i, G_t) = \sum_{l=1}^{q_j} \Delta Z_{i,t,j,l} \quad (7.7)$$

Ορισμός 6. Η ανομοιότητα (dissimilarity) μεταξύ δύο κόκκων (granules) G_i και G_t όσον αφορά το χώρο των χαρακτηριστικών γνωρισμάτων (feature space) $X = X_1 \times X_2 \times \dots \times X_p$ δίνεται από την επόμενη εξίσωση $(1 \leq i \leq c, 1 \leq t \leq c)$,

$$D(G_i, G_t) = \sum_{j=1}^p D_j(G_i, G_t) \quad (7.8)$$

Κάνοντας όλους τους απαραίτητους υπολογισμούς και αντικαταστάσεις η εξ. (7.8) λαμβάνει τη παρακάτω μορφή,

$$D(G_i, G_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} \sum_{k=1}^N \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \quad (7.9)$$

Η αναλυτική απόδειξη της εξ. (7.9) παρουσιάζεται παρακάτω.

Από τις εξ. (7.7) και (7.8) παίρνουμε,

$$D(G_i, G_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} \Delta Z_{i,t,j,l} \quad (7.10)$$

Λαμβάνοντας υπόψη την εξ. (7.6), η εξ. (7.10) γίνεται,

$$D(G_i, G_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} |Z_{i,j,l} - Z_{t,j,l}| \quad (7.11)$$

Χρησιμοποιώντας τις εξ. (7.3) και (7.5) μπορούμε να τροποποιήσουμε την εξ. (7.11) ως,

$$D(G_i, G_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} \left| \sum_{k=1}^N \Gamma_{i,j,l,k} - \sum_{k=1}^N \Gamma_{t,j,l,k} \right| = \sum_{j=1}^p \sum_{l=1}^{q_j} \left| \sum_{k=1}^N \mu_j^l(x_{kj}) \cdot u_i(\mathbf{x}_k) - \sum_{k=1}^N \mu_j^l(x_{kj}) \cdot u_t(\mathbf{x}_k) \right| \quad (7.12)$$

η οποία μπορεί να ξαναγραφεί ως,

$$D(G_i, G_t) = \sum_{j=1}^p \sum_{l=1}^{q_j} \sum_{k=1}^N \left| \mu_j^l(x_{kj}) u_i(\mathbf{x}_k) - \mu_j^l(x_{kj}) u_t(\mathbf{x}_k) \right| \quad (7.13)$$

Αφού $\mu_j^l(x_{kj}) > 0$, η εξ. (7.13) δίνει την εξ. (7.9).

Θεώρημα 1. Έστω ότι έχει δημιουργηθεί μια προσανατολισμένη προς το χρήστη κοκκοποίηση σε κάθε πεδίο ορισμού X_j ($1 \leq j \leq p$) που ικανοποιεί το δ -επιπέδου κριτήριο κάλυψης (δ -level criterion coverage) που δίνεται στον **ορισμό 1**. Επίσης, έστω ότι έχει παραχθεί ένας αλγόριθμος βελτιστοποίησης κοκκοποίησης $\mathbf{G} = \{G_1, G_2, \dots, G_c\}$ του χώρου των χαρακτηριστικών γνωρισμάτων (feature space) $X = X_1 \times X_2 \times \dots \times X_p$. Τότε, η συνάρτηση $D(G_i, G_t)$ εξ. (7.9) είναι μια μετρική απόσταση (metric distance), και το διάστημα (\mathbf{G}, D) είναι ένας διακριτός μετρικός χώρος (discrete metric space).

Απόδειξη

Για να καταστήσουμε ολόκληρη την προσέγγιση πρακτικότερη, θεωρούμε το σύνολο δεικτών $I = \{1, 2, \dots, c\}$ και καθιερώνουμε μια bijective αντιστοιχία μεταξύ των στοιχείων του συνόλου I και των στοιχείων του συνόλου \mathbf{G} . Επομένως, το \mathbf{G} μπορεί να αντιμετωπισθεί ως μια οικογένεια με ευρετήριο των συνόλων. Για να δείξουμε ότι η

συνάρτηση ανομοιότητας D που δίνεται από την εξ. (7.9) είναι μια μετρική συνάρτηση απόστασης (metric distance function), πρέπει να δείξουμε ότι ικανοποιεί τις ακόλουθες ιδιότητες,

$$(P1) \quad D(G_i, G_t) \geq 0, \quad \forall i, t \in I$$

$$(P2) \quad D(G_i, G_t) = 0 \Leftrightarrow G_i = G_t, \quad \forall i, t \in I$$

$$(P3) \quad D(G_i, G_t) = D(G_t, G_i), \quad \forall i, t \in I$$

$$(P4) \quad D(G_i, G_t) \leq D(G_i, G_r) + D(G_r, G_t), \quad \forall i, r, t \in I$$

Αφού $\mu_j^l(x_{kj}) \in [0, 1]$ και $|u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \geq 0$ ($\forall i, t \in I$), καταλήγουμε στο συμπέρασμα ότι η συνάρτηση D ικανοποιεί την ιδιότητα (P1).

Για να αποδείξουμε την ιδιότητα (P2), χρησιμοποιούμε το συνυπολογισμό και το βαθμό ισότητας μεταξύ δύο ασαφών κόκκων που δίνονται στις εξισώσεις (7.1) και (7.2), αντίστοιχα.

Υποθέτοντας ότι $D(G_i, G_t) = 0$, η εξ. (7.9) μπορεί να ξαναγραφεί ως,

$$D(G_i, G_t) = \sum_{k=1}^N \sum_{j=1}^p \sum_{l=1}^{q_j} \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \quad (7.14)$$

Αφού $\forall \mathbf{x}_k \in \mathbf{X}: \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \geq 0$ και $D(G_i, G_t) = 0$, από την εξ. (7.14) προκύπτει ότι πρέπει να είναι,

$$\sum_{j=1}^p \sum_{l=1}^{q_j} \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| = 0, \quad \forall k \in \{1, 2, \dots, N\} \quad (7.15)$$

Αλλά η ποσότητα $|u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)|$ είναι ανεξάρτητη από τους δείκτες j και l . Επομένως, η (7.15) δίνει,

$$|u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \sum_{j=1}^p \sum_{l=1}^{q_j} \mu_j^l(x_{kj}) = 0 \quad (7.16)$$

Υπό τον όρο ότι η προσανατολισμένη προς το χρήστη (user-defined) κοκκοποίηση σε κάθε πεδίο ορισμού (universe of discourse) ικανοποιεί το δ -κριτήριο κάλυψης προκύπτει ότι $\forall x_{kj} \in X_j, \exists A_j^l : A_j^l(x) > \delta$ με $\delta \in (0,1]$. Επομένως,

$$\sum_{j=1}^p \sum_{l=1}^{q_j} \mu_j^l(x_{kj}) \neq 0, \quad \forall k \in \{1, 2, \dots, N\} \quad (7.17)$$

Συνδυάζοντας τις εξισώσεις (7.16) και (7.17) εύκολα επαληθεύουμε ότι $|u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| = 0$ και έτσι $\forall \mathbf{x}_k \in \mathbf{X} u_i(\mathbf{x}_k) = u_t(\mathbf{x}_k)$. Εισάγοντας το τελευταίο αποτέλεσμα στη σχέση (7.1), καταλήγουμε στο συμπέρασμα ότι για κάθε $\mathbf{x}_k \in \mathbf{X}$, οι επόμενες δύο εξισώσεις ισχύουν : $(G_i \Rightarrow G_t)(\mathbf{x}_k) = 1$ και $(G_t \Rightarrow G_i)(\mathbf{x}_k) = 1$. Επομένως,

$$\min\{(G_i \Rightarrow G_t)(\mathbf{x}_k), (G_t \Rightarrow G_i)(\mathbf{x}_k)\} = 1, \quad \forall \mathbf{x}_k \in \mathbf{X} \quad (7.18)$$

Με αντικατάσταση της εξίσωσης (7.18) στην εξίσωση (7.2) και δεδομένης της σχέσης $\aleph(\mathbf{X}) = N$ προκύπτει ότι,

$$\|G_i = G_t\| = \frac{1}{\aleph(\mathbf{X})} \sum_{\mathbf{x}_k \in \mathbf{X}} \min\{(G_i \Rightarrow G_t)(\mathbf{x}_k), (G_t \Rightarrow G_i)(\mathbf{x}_k)\} = \frac{1}{N} N = 1$$

Κατά συνέπεια $G_i = G_t$.

Αντιθέτως, εάν $G_i = G_t$ έπεται ότι $\|G_i = G_t\| = 1$, που έχει σαν αποτέλεσμα,

$$\sum_{\mathbf{x}_k \in \mathbf{X}} \min\{(G_i \Rightarrow G_t)(\mathbf{x}_k), (G_t \Rightarrow G_i)(\mathbf{x}_k)\} = N \quad (7.19)$$

Αφού $0 \leq (G_i \Rightarrow G_t)(\mathbf{x}_k) \leq 1$ και $0 \leq (G_t \Rightarrow G_i)(\mathbf{x}_k) \leq 1$, η εξίσωση (7.19) μπορεί να ισχύει μόνο όταν η εξίσωση (7.18) ισχύει. Κατά συνέπεια, συνδυάζοντας τις ανωτέρω δύο ανισότητες με την εξίσωση (7.18) παίρνουμε,

$$(G_i \Rightarrow G_t)(\mathbf{x}_k) = 1 \quad (7.20)$$

και

$$(G_t \Rightarrow G_i)(\mathbf{x}_k) = 1 \quad (7.21)$$

Παρεμβάλλοντας την (7.20) στην εξ. (7.1) ολοκληρώνουμε ότι $u_i(\mathbf{x}_k) \leq u_t(\mathbf{x}_k)$. Ομοίως, συνδυάζοντας την εξ. (7.21) με την εξ. (7.1) προκύπτει ότι $u_t(\mathbf{x}_k) \leq u_i(\mathbf{x}_k)$. Επομένως $u_i(\mathbf{x}_k) = u_t(\mathbf{x}_k)$. Αντικαθιστώντας αυτά τα αποτελέσματα στην εξ. (7.9) προκύπτει ότι $D(G_i, G_t) = 0$.

Η ιδιότητα (P3) μπορεί εύκολα να επαληθευθεί αφού:

$$|u_i(\mathbf{x}_k) - u_t(\mathbf{x}_k)| = |u_t(\mathbf{x}_k) - u_i(\mathbf{x}_k)|.$$

Τέλος, η ιδιότητα (P4) αποδεικνύεται ως εξής,

$$\begin{aligned} D(C_i, C_t) &= \sum_{j=1}^p \sum_{l=1}^{q_j} \sum_{k=1}^N \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_r(\mathbf{x}_k) + u_r(\mathbf{x}_k) - u_t(\mathbf{x}_k)| \leq \\ & \sum_{j=1}^p \sum_{l=1}^{q_j} \sum_{k=1}^N \mu_j^l(x_{kj}) |u_i(\mathbf{x}_k) - u_r(\mathbf{x}_k)| + \sum_{j=1}^p \sum_{l=1}^{q_j} \sum_{k=1}^N \mu_j^l(x_{kj}) |u_r(\mathbf{x}_k) - u_t(\mathbf{x}_k)| = \\ & = D(G_i, G_r) + D(G_r, G_t) \end{aligned}$$

Τα ανωτέρω ισχύουν $\forall i, r, t \in I$.

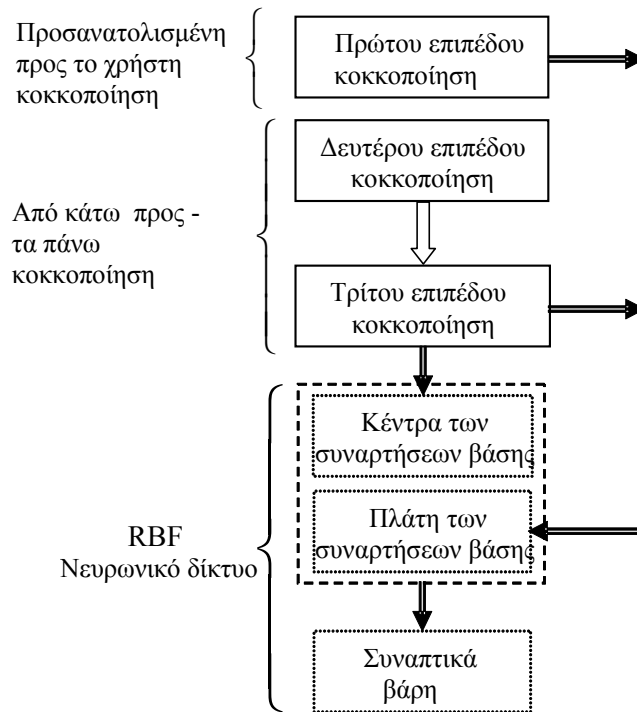
Για αυτόν τον λόγο, το συμπέρασμα ότι ο χώρος (\mathbf{G}, D) είναι ένα διανυσματικός χώρος είναι προφανές δεδομένου ότι το σύνολο $\mathbf{G} = \{G_1, G_2, \dots, G_c\}$ είναι πεπερασμένο. Αυτό ολοκληρώνει την απόδειξη του θεωρήματος 1.

■

7.3 Εκπαίδευση Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής Βάσης (RBFNN) χρησιμοποιώντας Ασαφείς Κόκκους

Ο αλγόριθμος εκπαίδευσης περιλαμβάνει τρία επίπεδα κοκκοποίησης, και η βασική δομή της απεικονίζεται στο σχήμα 7.1. Στο πρώτο επίπεδο παράγουμε μια προσανατολισμένη προς το χρήστη κοκκοποίηση σε κάθε πεδίο ορισμού, το οποίο υποχρεώνεται για να ικανοποιήσει ορισμένες απαιτήσεις. Τα τελευταία δύο επίπεδα συνθέτουν μια από κάτω προς τα επάνω διαδικασία κοκκοποίησης και είναι αλγοριθμικές διαδικασίες βελτιστοποίησης. Η χρήση του δεύτερου επιπέδου είναι κυρίως για να προεπεξεργαστούν

τα δεδομένα προκειμένου να αποκαλυφθούν μερικές από τις υποκείμενες δομές τους. Εντούτοις, πρέπει επίσης να κατέχει μερικές ιδιότητες. Πρέπει να είναι λιγότερο ευαίσθητο στην αρχικοποίηση. Η ασαφής συσταδοποίηση (fuzzy clustering) έχει αποδειχθεί ένα αξιόπιστο εργαλείο όσον αφορά αυτήν την πτυχή.



Σχήμα 7.1. Βασική τοπολογία του προτεινόμενου αλγορίθμου.

Εδώ, εφαρμόζουμε έναν ασαφή αλγόριθμο διαμερισμού (fuzzy partition) για να ομαδοποιήσει όλα τα δεδομένα εισόδου εκπαίδευσης σε διάφορες συστάδες. Η εφαρμογή του ασαφούς διαμερισμού (fuzzy partition) λαμβάνει ένα αρχικό διαμερισμό του συνόλου δεδομένων, ο οποίος (διαμερισμός) επεξεργάζεται περαιτέρω από τη μέθοδο των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) [91] στο τρίτο επίπεδο κοκκοποίησης. Ο τρόπος που κάθε ένα από τα ανωτέρω επίπεδα έχει επιπτώσεις στην εκτίμηση παραμέτρων του νευρωνικού δικτύου RBF απεικονίζεται επίσης στο σχήμα 7.1. Όπως φαίνεται εύκολα, τα κέντρα των συναρτήσεων πυρήνων είναι ένα απλό αποτέλεσμα της από κάτω προς τα επάνω (bottom-up) διαδικασίας κοκκοποίησης. Αφ' ετέρου, τα αντίστοιχα πλάτη (widths) καθορίζονται με το συνδυασμό των πληροφοριών

που κωδικοποιούνται στο πρώτο και τρίτο επίπεδο κοκκοποίησης. Αυτός ο στόχος υλοποιείται από την εξ. (7.9).

7.3.1 Πρώτο επίπεδο κοκκοποίησης

Η προσανατολισμένη προς το χρήστη (user-defined) κοκκοποίηση (granulation) συνδέεται έντονα με το πλαίσιο της έννοιας γνώσης (cognition concept) [74]. Το πλαίσιο της γνώσης (cognition) είναι το αποτέλεσμα μιας διαδικασίας κοκκοποίησης, σύμφωνα με την οποία μια οικογένεια των επονομαζόμενων ασαφών συνόλων καθορίζεται πέρα από ολόκληρο το πεδίο ορισμού (universe of discourse). Μπορούμε να πούμε αρκετά σχετικά με το ότι ένα πλαίσιο της γνώσης (cognition) είναι ένα κωδικοβιβλίο (codebook) των οντοτήτων κάθε ένα από το οποίο ορίζεται με μια εννοιολογική ετικέτα όπως *μικρός*, *μεγάλος*, *συνδύτερος*, κ.λπ. Για να καταστήσουμε την επίσημη περιγραφή συμβατή με την ορολογία που χρησιμοποιείται σε αυτό το κεφάλαιο, θεωρούμε τον χώρο των χαρακτηριστικών γνωρισμάτων (feature space) $X = X_1 \times X_2 \times \dots \times X_p$. Το πλαίσιο της γνώσης F_j πέρα από ολόκληρο το πεδίο ορισμού (universe of discourse) X_j ($1 \leq j \leq p$) αντιπροσωπεύεται όπως: $F_j = \{A_j^1, A_j^2, \dots, A_j^{q_j}\}$, όπου A_j^l ($1 \leq l \leq q_j$) είναι ένα ασαφές σύνολο (fuzzy set), και q_j ο αριθμός των ασαφών συνόλων (fuzzy sets).

Υπάρχουν δύο σημαντικές απαιτήσεις που πρέπει να ικανοποιηθούν κατά το σχεδιασμό ενός πλαισίου της γνώσης (cognition) [74]:

α) *δ-επιπέδου κριτήριο κάλυψης (δ-level criterion coverage)*: Η λειτουργία αυτού του κριτηρίου περιγράφηκε στον ορισμό 1.

β) *Σημασιολογική αρτιότητα (semantic soundness)*: Αφορά τη διαχωριστικότητα (interpretability) της προκύπτουσας κοκκοποίησης [57]. Αν και η έννοια είναι περίπλοκη, υπάρχουν μερικές πολύ χρήσιμες οδηγίες. Κατ' αρχάς, κάθε ασαφές σύνολο πρέπει να είναι μονοκόρυφο (unimodal) και κανονικό (normal). Δεύτερον, ένας ορισμένος επικαλυπτόμενος βαθμός μεταξύ των ζευγαριών των γειτονικών ασαφών συνόλων πρέπει να ληφθεί υπόψη. Τρίτον, ο αριθμός q_j πρέπει να κρατηθεί χαμηλός. Οι χαρακτηριστικές τιμές που έχουν εξεταστεί επιτυχώς πέρα από ένα ευρύ φάσμα των εφαρμογών είναι $q_j = 7 \pm 2$ [3, 4, 74].

Η πρώτη απαίτηση αναμένεται να συμβαδίζει με τα διαθέσιμα πειραματικά δεδομένα στον υψηλότερο βαθμό, και παρέχει τη μεγιστοποίηση κάλυψης ολόκληρου του πεδίου ορισμού (universe of discourse). Ο δεύτερος επιτρέπει την κωδικοποίηση της καθορισμένης με σαφήνεια σημασιολογίας, και έρχεται υπό μορφή ελαχιστοποίησης της διάδοσης των ασαφών συνόλων (fuzzy sets) που συμμετέχουν στο πλαίσιο της γνώσης (cognition).

7.3.2 Δεύτερο επίπεδο κοκκοποίησης

Σε αυτό το επίπεδο χρησιμοποιείται ο αλγόριθμος ασαφούς διαμερισμού [95] (fuzzy partition) για να ομαδοποιήσει όλα τα δεδομένα εισόδου εκπαίδευσης σε διάφορες συστάδες. Ο αλγόριθμος περιγράφεται εν συντομία ως εξής:

Υποθέστε ότι μας δίνεται ένα σύνολο N διανυσμάτων δεδομένων εισόδου $X = \{x_1, x_2, \dots, x_N\} \subset R^m$. Στο i -οστό σύνολο δεδομένων X_i ($i = 1, 2, \dots, m$) καθορίζουμε ένα ασαφή διαμερισμό, διαμερίζοντας ομοιόμορφα σε q_i συμμετρικά τριγωνικά ασαφή σύνολα, με τις συναρτήσεις συμμετοχής της μορφής

$$A(x) = \begin{cases} 1 - \frac{|x - \alpha|}{\delta\alpha}, & \text{αν } x \in [\alpha - \delta\alpha, \alpha + \delta\alpha] \\ 0 & , \text{ αλλιώς} \end{cases} \quad (7.22)$$

όπου α είναι το κέντρο του τριγώνου στο οποίο ο βαθμός συμμετοχής ορίζεται ως μονάδα, και $\delta\alpha$ είναι το μισό πλάτος. Το σύνολο τιμών της i -οστής γλωσσικής μεταβλητής εισόδου ορίζεται ως

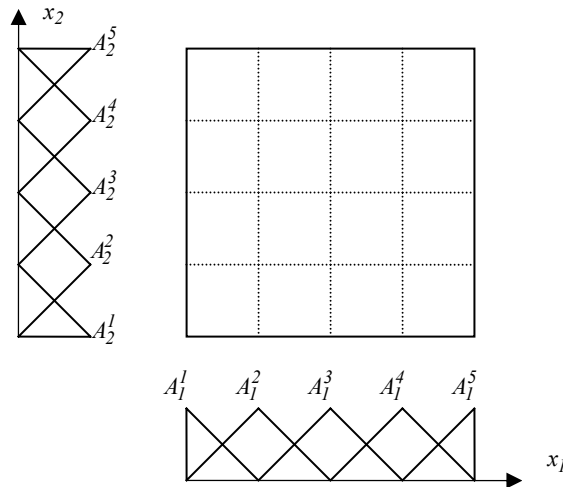
$$T_i = \{A_i^1, A_i^2, \dots, A_i^{q_i}\}, \quad 1 \leq i \leq m \quad (7.23)$$

Σαν φυσικό αποτέλεσμα, οι ασαφής διαμερισμοί στους καθολικούς χώρους εισόδου καθορίζουν ένα ασαφή διαμερισμό στο χώρο εισόδου X διαιρώντας τον σε K ασαφής υποχώρους A^1, A^2, \dots, A^K , με

$$K = \prod_{i=1}^m q_i \quad (7.24)$$

Ο j -οστός ασαφής υποχώρος A^j ($1 \leq j \leq K$) λαμβάνεται ως συνδυασμός m ιδιαίτερων ασαφών συνόλων $A_i^1, A_i^2, \dots, A_i^m$ με $A_i^j \in T_i$ ($i=1,2,\dots,m$) και μπορεί να αναπαρασταθεί όπως:

$$A^j = [A_1^j, A_2^j, \dots, A_m^j]^T \quad (7.25)$$



Σχήμα 7.2. Ασαφής διαμερισμός στο δισδιάστατο χώρο εισόδου.

Το σχήμα 7.2 παρουσιάζει τον ασαφή διαμερισμό που καθορίζεται στο δισδιάστατο χώρο εισόδου, όπου οι χώροι εισόδου x_1 και x_2 χωρίζονται ομοιόμορφα σε πέντε συμμετρικά τριγωνικά ασαφή σύνολα.

Το βασικό βήμα σε όλο το σχέδιο της μηχανής συμπεράσματος είναι να αναπτυχθεί ένας τύπος προκειμένου να καθοριστεί η πολυδιάστατη συνάρτηση συμμετοχής ενός διανύσματος δεδομένων εισόδου $x_k = [x_{k1}, x_{k2}, \dots, x_{km}]^T$ στον l -ασαφές υποχώρο A_l ($l=1,2,\dots,L$).

Από την εξ. (7.22), προκύπτει ότι ένα ασαφές σύνολο A μπορεί να περιγραφεί πλήρως από το κέντρο του και το μισό πλάτος

$$A = \{\alpha, \delta\alpha\} \quad (7.26)$$

Υιοθετώντας την έκφραση (7.25), και καθορίζοντας τα διανύσματα

$$a^l = [a_1^l, a_2^l, \dots, a_m^l]^T, \quad \delta a^l = [\delta a_1^l, \delta a_2^l, \dots, \delta a_m^l]^T \quad (7.27)$$

ο υποχώρος A^l στην εξ. (7.26) μπορεί να γραφτεί όπως

$$A^l = \{a^l, \delta a^l\} \quad (7.28)$$

Η έκφραση (7.28) δείχνει πως ο υποχώρος A^l μπορεί να αντιμετωπισθεί ως παραλληλόγραμμο στο χώρο των εισόδων, το κέντρο του οποίου είναι το διάνυσμα a^l και οι πλευρές του ορίζονται από το διάνυσμα δa^l . Μετά από αυτήν την ανάλυση, η πολυδιάστατη συνάρτηση συμμετοχής (membership function) $A^l(x_k)$ ορίζεται όπως [94,95]

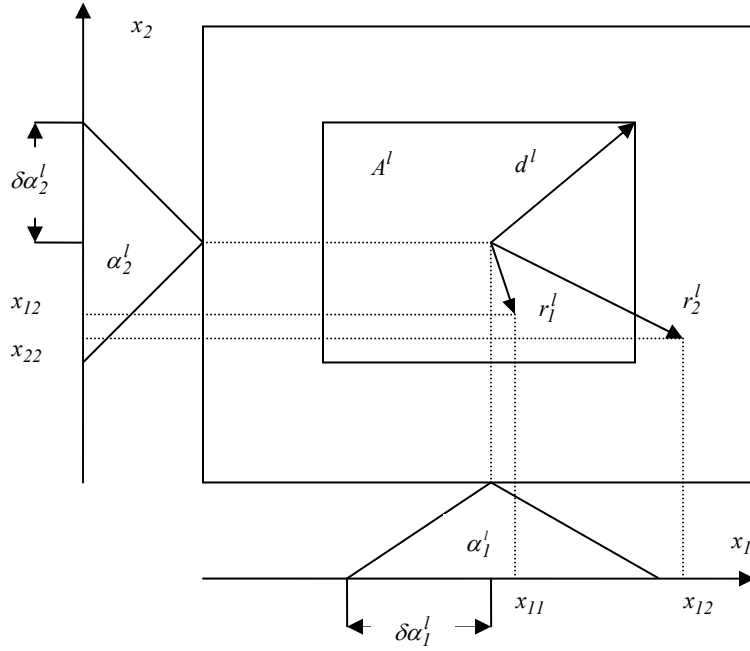
$$A^l(x_k) = 1 - rd^l(x_k) \quad (7.29)$$

όπου $rd^l(x_k)$ είναι η ευκλείδεια (euclidean) σχετική απόσταση μεταξύ του A^l και του τρέχοντος διανύσματος δεδομένων εισόδου x_k , το οποίο δίνεται από τον εξής τύπο [94,95]

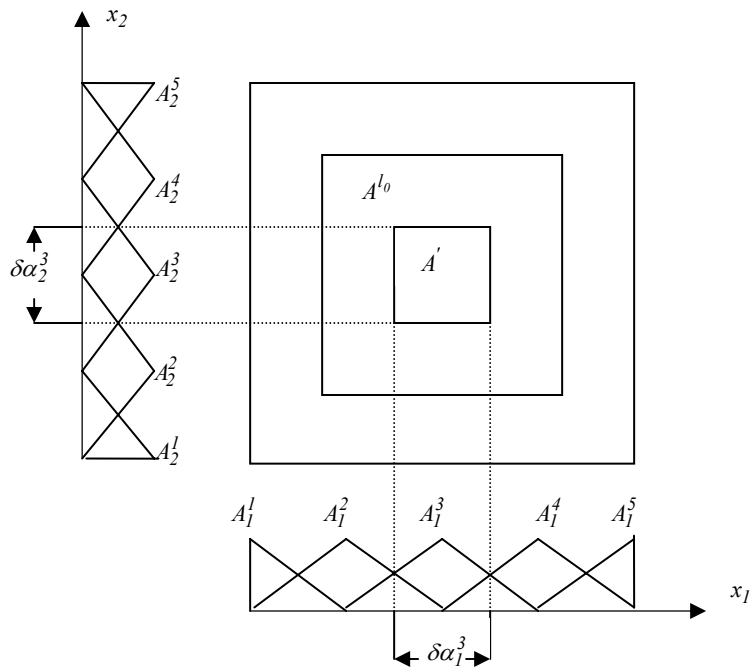
$$rd^l(x_k) = \begin{cases} \frac{\left[\sum_{i=1}^m (\alpha_i^l - x_{ki})^2 \right]^{1/2}}{\left[\sum_{i=1}^m (\delta \alpha_i^l)^2 \right]^{1/2}}, & \text{αν } \left[\sum_{i=1}^m (\alpha_i^l - x_{ki})^2 \right]^{1/2} \leq \left[\sum_{i=1}^m (\delta \alpha_i^l)^2 \right]^{1/2} \\ 1 & , \text{ αλλιώς} \end{cases} \quad (7.30)$$

Το σχήμα 7.3 δείχνει την παραπάνω γεωμετρική ερμηνεία στον δισδιάστατο χώρο των εισόδων. Το σχήμα 7.4 δείχνει διαγραμματικά στον δισδιάστατο χώρο γιατί η παράμετρος θ δεν πρέπει να πάρει τιμές μικρότερες ή ίσες του 0.5. Ας υποθέσουμε ότι στο βήμα 3 του αλγορίθμου 7.2, ο ασαφής υποχώρος A^{l_0} , ο οποίος με βάση τις σχέσεις (7.25) και (7.28) ορίζεται ως, $A^{l_0} = [A_1^3 \ A_2^3]^T = \left\{ [a_1^3 \ a_2^3]^T, [\delta a_1^3 \ \delta a_2^3]^T \right\}$ είναι αυτός με τη μικρότερη σχετική Ευκλείδεια απόσταση $rd^{l_0}(x_k)$ από το διάνυσμα εισόδου x_k . Στο ίδιο σχήμα, ο υποχώρος A' είναι ένα υποσύνολο του A^{l_0} , το οποίο ορίζεται ως,

$$A' = \left\{ [a_1^3 \ a_2^3]^T, \left[\frac{\delta a_1^3}{2} \quad \frac{\delta a_2^3}{2} \right]^T \right\}$$



Σχήμα 7.3 Γεωμετρική ερμηνεία της υπόθεσης του l ασαφές υποχώρου και της σχετικής Ευκλείδειας απόστασης στον δισδιάστατο χώρο.



Σχήμα 7.4 Προσδιορισμός της ελάχιστης τιμής της παραμέτρου θ στον δισδιάστατο χώρο των εισόδων.

Αλγόριθμος 7.1: Προσδιορισμός του πιο κοντινού ασαφές υποχώρου σε ένα δεδομένο διάνυσμα εισόδου

Βήμα 1) Υποθέστε ότι μας δίνεται ένα διάνυσμα δεδομένων εισόδου x_k με $x_k \in R^m$.

Διαμερίστε όλα τα πεδία ορισμού εισόδου σε ασαφή σύνολα και καθορίστε τους αντίστοιχους όρους συνόλων. Παραδείγματος χάριν στο σχήμα 7.3, όπου ένα σύστημα με δύο εισόδους απεικονίζεται, οι όροι συνόλων για τις μεταβλητές x_1 , x_2 είναι

$$T_1 = \{A_1^1, A_1^2, A_1^3, A_1^4, A_1^5\}$$

$$T_2 = \{A_2^1, A_2^2, A_2^3, A_2^4, A_2^5\}$$

Βήμα 2) Καθορίστε τις τιμές συμμετοχής $x_{k2}, x_{k1}, \dots, x_{km}$ στα αντίστοιχα ασαφή σύνολα.

Παραδείγματος χάριν στο σχήμα 7.3 (α), το x_{k1} έχει βαθμό συμμετοχής 0.7 στο A_1^2 , 0.3 στο A_1^3 και μηδέν βαθμούς σε όλα τα άλλα ασαφή σύνολα. Ομοίως, στη σχήμα 7.3 (β) το x_{k2} έχει βαθμό συμμετοχής 0.75 στο A_2^1 , 0.25 στο A_2^2 και μηδέν βαθμούς συμμετοχής σε όλα τα άλλα ασαφή σύνολα.

Βήμα 3) Ορίστε $x_{k2}, x_{k1}, \dots, x_{km}$ στα ασαφή σύνολα με το μέγιστο βαθμό συμμετοχής.

Παραδείγματος χάριν, στο σχήμα 7.3 (α), το x_{k1} ορίζεται στο A_1^2 και στο σχήμα 7.3 (β) το x_{k2} ορίζεται στο A_2^1 .

Βήμα 4) Για τα δεδομένα εισόδου x_k παράγεται ο ασαφής υποχώρος A^{j_0} που δίνεται ως συνδυασμός των ασαφών συνόλων που επιλέγονται στο βήμα 3.

Παραδείγματος χάριν στο σχήμα 7.3, $A^{j_0} = [A_1^2, A_2^1]^T$.

Η μέθοδος WANG [98], είναι πολύ απλή και ισχυρή, αλλά παράγει έναν μεγάλο αριθμό ασαφών κανόνων. Θα δείξουμε ότι μπορούμε σημαντικά να μειώσουμε τον αριθμό ασαφών κανόνων του συστήματος με τη χρησιμοποίηση της κοντινότερης έννοιας γειτόνων.

Αλγόριθμος 7.2: Ασαφής διαμερισμός του χώρου εισόδου σε διάφορες συστάδες

Βήμα 1) Υποθέστε ότι μας δίνονται τα ζευγάρια δεδομένων εισόδου – εξόδου της μορφής (x_k, y_k) ($k = 1, 2, \dots, N$). Θέστε $L = 0$ και επέλεξε μία τιμή για το $\theta \in (0.5, 1]$.

Βήμα 2) Πάρτε το πρώτο διάνυσμα δεδομένων εισόδου x_1 και εφαρμόστε τον αλγόριθμο 7.1 για να παραγάγετε τον πρώτο ασαφή υποχώρο $A^1 = \{a^1, \delta a^1\}$. Επίσης θέστε $L = 1$.

Βήμα 3) Υποθέστε ότι $k - 1$ τα διανύσματα δεδομένων εισόδου έχουν εξεταστεί και L ασαφής υποχώροι έχουν παραχθεί, με $1 < L < k - 1$.

Εισάγετε το k -οστό διάνυσμα δεδομένων εισόδου x_k . Χρησιμοποιώντας την εξίσωση (7.30), υπολογίστε τις ευκλείδειες (euclidean) σχετικές αποστάσεις $rd^l(x_k)$ ($l = 1, 2, \dots, L$) μεταξύ x_k και όλους τους L ασαφής υποχώρους που έχουν δημιουργηθεί. Υποθέστε ότι η ελάχιστη απόσταση $rd^{l_0}(x_k)$ αντιστοιχεί στον ασαφή υποχώρο $A^{l_0} = \{a^{l_0}, \delta a^{l_0}\}$. Εάν η ανισότητα

$$rd^{l_0}(x_k) < \theta \quad (7.31)$$

είναι αληθής πήγαινε στο βήμα 5, αλλιώς πήγαινε στο βήμα 4.

Βήμα 4) Εφαρμόστε τον αλγόριθμο 7.1 για να δημιουργήσετε τον πιο κοντινό ασαφή υποχώρο όσον αφορά το τρέχον διάνυσμα x_k και θέσε $L = L + 1$.

Βήμα 5) Εάν $k = N$ πηγαίνετε στο βήμα 6, αλλιώς θέστε το επόμενο διάνυσμα δεδομένων εισόδου και επιστρέψτε στο βήμα 3.

Βήμα 6) Για $k = 1, 2, \dots, N$ και $l = 1, 2, \dots, L$, χρησιμοποιήστε την εξίσωση (7.30) για να υπολογιστούν οι ευκλείδειες (euclidean) σχετικές αποστάσεις $rd^l(x_k)$.

Βήμα 7) Για $k = 1, 2, \dots, N$ και $l = 1, 2, \dots, L$, χρησιμοποιήστε την εξίσωση (7.29) για να υπολογίσετε την συνάρτηση συμμετοχής $A^l(x_k)$.

Ο αλγόριθμος αρχίζει με το πρώτο σημείο δεδομένων x_1 και παράγει τον πρώτο ασαφή υποχώρο, με την εφαρμογή του αλγορίθμου 7.1. Για $i = 1, 2, \dots, N$, επιλέγει το ασαφές σύνολο που ορίζει το μέγιστο βαθμό συμμετοχής $x_i(1)$. Ο συνδυασμός των N

ασαφών συνόλων, διατυπώνει τον ασαφή υποχώρο που περιγράφει καλύτερα το διάνυσμα εισόδου x_l εξ. (7.27).

Για όλα τα υπόλοιπα διανύσματα εισόδου x_k , $k = 2, 3, \dots, K$, ο αλγόριθμος καθορίζει τον ασαφή υποχώρο που είναι το πιο κοντά, μεταξύ των L υποχώρων που έχουν δημιουργηθεί, στο διάνυσμα εισόδου x_k . Πιο συγκεκριμένα, ο αλγόριθμος υπολογίζει τις Ευκλείδειες (euclidean) σχετικές αποστάσεις $(l = 1, 2, \dots, L) rd^l(x_k)$ μεταξύ x_k και τους L ασαφής υποχώρους που έχουν δημιουργηθεί μέχρι τώρα. Εάν το ελάχιστο αυτών των αποστάσεων είναι μεγαλύτερο από το n , κατόπιν το x_k δεν περιγράφεται αρκετά από οποιοδήποτε από τους ασαφής υποχώρους, και ο αλγόριθμος παράγει νέο υποχώρο, με την εφαρμογή του αλγορίθμου 7.1. Εάν η ανισότητα (7.31) δεν ικανοποιείται, ο ασαφής υποχώρος, που έχει δημιουργηθεί ήδη, μπορεί να οριστεί στο συγκεκριμένο διάνυσμα εισόδου. Επομένως, δεν υπάρχει καμία ανάγκη να δημιουργηθεί νέος υποχώρος και ο αλγόριθμος προχωράει στο επόμενο διάνυσμα εκπαίδευσης.

Ο ανωτέρω αλγόριθμος λαμβάνει ένα διαμερισμό (partition) του συνόλου X σε L συστάδες C_1, C_2, \dots, C_L . Οι ασαφής πληθικότητες συνόλου (cardinalities) αυτών των συστάδων δίνονται από την επόμενη εξίσωση,

$$\aleph(C_r) = \sum_{k=1}^N A^l(x_k) \quad (1 \leq r \leq L)$$

Επομένως, το σχετικό μέγεθος της συστάδας C_r είναι,

$$\rho_r = \frac{\aleph(C_r)}{\sum_{i=1}^L \aleph(C_i)} \quad (1 \leq r \leq L) \quad (7.32)$$

Το σχετικό μέγεθος παρέχει ένα βάρος σημασίας της συστάδας στο γενικό διαμερισμό.

7.3.3 Τρίτο επίπεδο κοκκοποίησης

Σε αυτό το βήμα χρησιμοποιούμε την μέθοδο των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) [91] για να παραχθεί μια πιο αφηρημένη κοκκοποίηση του συνόλου

δεδομένων. Η μέθοδος των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) κατέχουν μια ενδιαφέρουσα έμφυτη ιδιότητα: υποθέτει ότι στα δεδομένα ορίζονται τα βάρη, το άθροισμα των οποίων ισούται με μονάδα. Αυτά τα βάρη απεικονίζουν τη σχετική σημασία των δεδομένων, και θεωρούνται ως γνωστές παράμετροι στη διαδικασία βελτιστοποίησης. Αυτό είναι μια πολύ ελκυστική ιδιότητα, δεδομένου ότι επιτρέπει τη συγκέντρωση των κόκκων αντί των καθαρών αριθμητικών δεδομένων. Συγκεκριμένα, μπορούμε να αντιπροσωπεύσουμε τους κόκκους που παράγονται στο προηγούμενο επίπεδο όπως $C_r = (\mathbf{a}_r, \rho_r)$ με $1 \leq r \leq n$ όπου $n = L$ είναι ο αριθμός των συστάδων που δημιουργήθηκε από το προηγούμενο επίπεδο. Κατόπιν, η μέθοδος των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) θα παράσχει ένα διαμερισμό (partition) του συνόλου $\{(\mathbf{a}_1, \rho_1), (\mathbf{a}_2, \rho_2), \dots, (\mathbf{a}_n, \rho_n)\}$ σε ένα σύνολο από c -κόκκους $c < n$ $\mathbf{G} = \{G_1, G_2, \dots, G_c\}$. Η αντικειμενική συνάρτηση είναι,

$$J_\rho = \sum_{r=1}^n \sum_{i=1}^c \rho_r (h_{ir})^m \|\mathbf{a}_r - \mathbf{v}_i\|^2 \quad (7.33)$$

όπου \mathbf{v}_i ($1 \leq i \leq c$) είναι το κέντρο (πρωτότυπο) του κόκκου G_i , h_{ir} είναι ο βαθμός συμμετοχής $C_r = (\mathbf{a}_r, \rho_r) \in G_i$, και $m \in (1, \infty)$ είναι ένας παράγοντας για να ρυθμίσει το βαθμό συμμετοχής σταθμίζοντας την επίδραση. Το πρόβλημα είναι να ελαχιστοποιηθεί η αντικειμενική συνάρτηση κόστους J_ρ κάτω από τον περιορισμό,

$$\sum_{i=1}^c h_{ir} = 1 \quad \forall r$$

Τα πρωτότυπα (κέντρα) και οι αντίστοιχες συναρτήσεις συμμετοχής που λύνουν αυτό το πρόβλημα βελτιστοποίησης βρίσκονται εύκολα να είναι,

$$\mathbf{v}_i = \frac{\sum_{r=1}^n \rho_r (h_{ir})^m \mathbf{a}_r}{\sum_{r=1}^n \rho_r (h_{ir})^m}, \quad (1 \leq i \leq c) \quad (7.34)$$

$$h_{ir} = 1 / \left(\sum_{t=1}^c \left(\frac{\|\mathbf{a}_r - \mathbf{v}_i\|}{\|\mathbf{a}_r - \mathbf{v}_t\|} \right)^{2/(m-1)} \right) \quad (1 \leq i \leq c, 1 \leq r \leq n) \quad (7.35)$$

Οι εξισώσεις (7.34) και (7.35) αποτελούν μια επαναληπτική διαδικασία βελτιστοποίησης, η οποία ελαχιστοποιεί τελικά τη συνάρτηση J_ρ .

7.3.4 Εκτίμηση παραμέτρων του δικτύου RBF

Το αποτέλεσμα του προηγούμενου βήματος είναι ένα σύνολο c κόκκων $\mathbf{G} = \{G_1, G_2, \dots, G_c\}$, οι αντιπρόσωποι των οποίων είναι τα διανύσματα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$. Αυτό που θέλουμε εδώ είναι να καθιερωθεί μια σχέση μεταξύ των συνόλων \mathbf{G} και \mathbf{X} . Αυτή η σχέση μπορεί να καθοριστεί εάν ορίσουμε το βαθμό συμμετοχής του k -διανύσματος εκπαίδευσης \mathbf{x}_k ($1 \leq k \leq N$) στον i -κόκκο G_i όπως φαίνεται παρακάτω,

$$u_{ik} = 1 / \sum_{t=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_t\|} \right)^{2/(m-1)}, \quad (1 \leq i \leq c, 1 \leq k \leq N) \quad (7.36)$$

όπου $m \in (1, \infty)$ είναι η παράμετρος ασάφειας. Για αυτόν τον λόγο, η αντιστοιχία μεταξύ των επιτευχθέντων αποτελεσμάτων μέχρι τώρα και η ανάλυση που πραγματοποιήθηκε στην παράγραφο 7.2 είναι προφανείς. Είμαστε τώρα έτοιμοι να παρέχουμε μια λεπτομερή παρουσίαση του τρόπου που υπολογίζουμε τις παραμέτρους σχεδιασμού του δικτύου συναρτήσεων ακτινικής βάσης (RBF). Αρχικά, ο αριθμός των κρυφών κόμβων είναι ίσος με το c , ενώ τα κέντρα για τις συναρτήσεις πυρήνων είναι τα διανύσματα $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c$.

Για την μετρική απόσταση (metric distance) εξ. (7.9) επισημαίνουμε τις ακόλουθες παρατηρήσεις. Κατ' αρχάς, δεν θεωρεί τις κεντρικές θέσεις αλλά μάλλον τους βαθμούς συμμετοχής των δεδομένων εκπαίδευσης, απεικονίζοντας κατ' αυτό τον τρόπο την πραγματική ελλοχεύουσα κατανομή τους. Δεύτερον, εξετάζει τη δομή δεδομένων χωριστά σε κάθε πεδίο ορισμού από την άποψη της προσανατολισμένης προς το χρήστη (user defined) κοκκοποίησης και τελικά, λαμβάνει υπόψη τις βασικές πληροφορίες που διαφορετικά θα διακρινόντουσαν πολύ δύσκολα. Τέλος, είναι μια μετρική απόσταση (metric distance) και επομένως μεταχειρίζεται κάθε κόκκο ως ένα αντικείμενο. Κατά συνέπεια, βασισμένοι στην ανωτέρω ανάλυση και χρησιμοποιώντας την αντίστοιχη μαθηματική ορολογία, μπορούμε να χρησιμοποιήσουμε την εξ. (7.9) για να

υπολογίσουμε τις αποστάσεις μεταξύ των ζευγαριών των κόκκων $D(G_i, G_t)$ με $(i \neq t)$.

Κατόπιν, για τον i -οστό κόκκο ορίζουμε την ελάχιστη απόστασή της,

$$D_{\min}(G_i) = \min_{1 \leq t \leq c, t \neq i} \{D(G_i, G_t)\}, \quad \forall i \in \{1, 2, \dots, c\} \quad (7.37)$$

Για αυτόν τον λόγο, προτείνεται να υπολογιστεί το πλάτος (width) της i -οστής συνάρτησης ακτινικής βάσης ως εξής,

$$\sigma_i = \beta D_{\min}(G_i) \quad (7.38)$$

όπου $\beta > 0$ είναι μια ευρηματική (heuristic) παράμετρος.

Για να υπολογιστούν τα συναπτικά βάρη (synaptic weights) ορίζεται ο $N \times c$ πίνακας $\mathbf{H} = [g_i(\mathbf{x}_k)]$ όπου $g_i(\mathbf{x}_k)$ δίνεται στην εξ. (1.8), ο $N \times 1$ πίνακας εξόδου $\mathbf{Y} = [y_k]$, και ο $c \times 1$ πίνακας $\mathbf{W} = [w_i]$ όπου w_i είναι το συναπτικό βάρος του i -οστού κόμβου.

Ο στόχος είναι να ελαχιστοποιηθεί η ακόλουθη συνάρτηση σφάλματος,

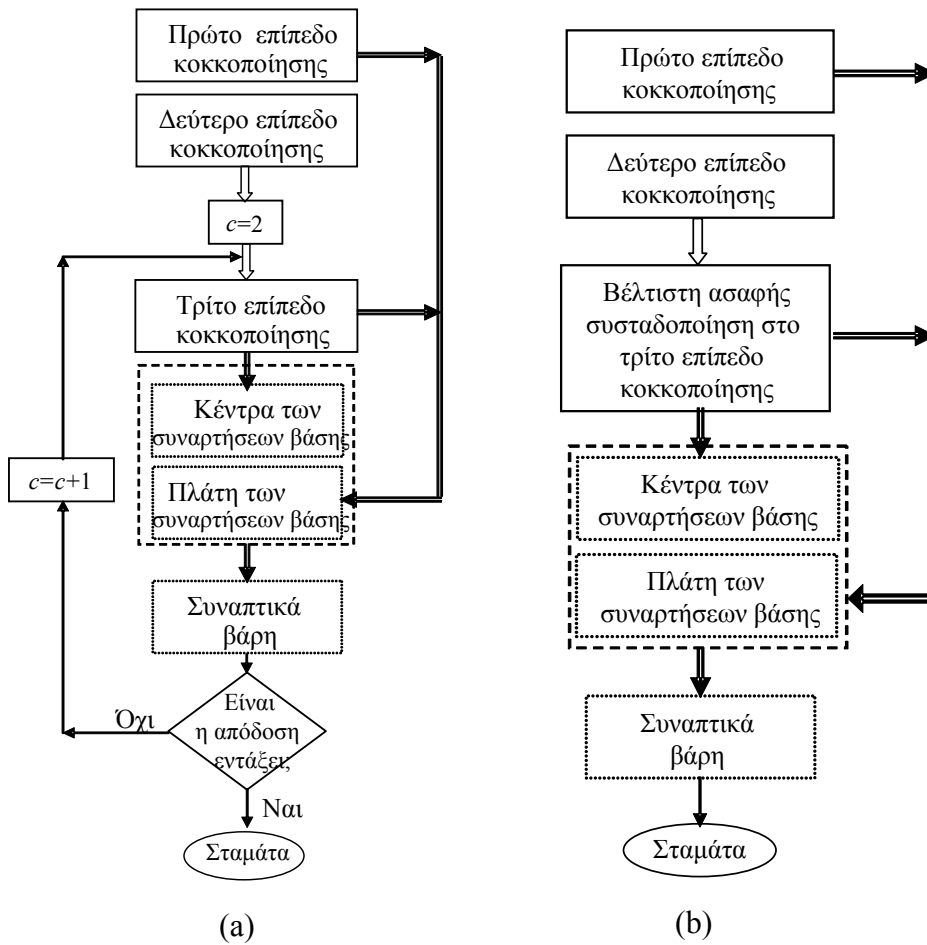
$$E(\mathbf{W}) = \|\tilde{\mathbf{Y}} - \mathbf{Y}\|^2 \quad (7.39)$$

όπου $\tilde{\mathbf{Y}} = \mathbf{H}\mathbf{W}$ είναι ο πίνακας των κατ' εκτίμηση αποτελεσμάτων. Για να επιτευχθεί αυτός ο στόχος, υιοθετήθηκε η πολύ γνωστή μέθοδος των ελαχίστων τετραγώνων (least squares method),

$$\mathbf{W} = [\mathbf{H}^T \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{Y} \quad (7.40)$$

7.3.5 Περαιτέρω σχόλια στον αλγόριθμο

Το πρώτο σημαντικό ζήτημα που εξετάζεται σε αυτήν την υποενότητα είναι ο υπολογισμός του κατάλληλου αριθμού των κρυφών κόμβων. Γενικά, υπάρχουν δύο πλατφόρμες βασισμένες στην ασαφή συσταδοποίηση για να κάνουν αυτό [5,7]. Σύμφωνα με τον πρώτο, ο αλγόριθμος ξεκινάει με δύο κρυφούς κόμβους και αυξάνεται βαθμιαία αυτός ο αριθμός μέχρι η απόδοση του δικτύου να βρίσκεται μέσα σε μια προ-διευκρινισμένη ακρίβεια (pre-specified accuracy).



Σχήμα 7.5. Δύο τροποποιήσεις του αλγορίθμου έτσι ώστε να μπορεί να χρησιμοποιηθεί για τον υπολογισμό του κατάλληλου αριθμού κρυφών κόμβων.

Ο δεύτερος είναι βασισμένος στη χρήση της βέλτιστης ασαφής συσταδοποίησης (optimal fuzzy clustering). Ο προτεινόμενος αλγόριθμος μπορεί εύκολα να τροποποιηθεί για να ικανοποιήσει και τις δύο από τις ανωτέρω πλατφόρμες όπως καταδεικνύεται στο σχήμα 7.5.

7.4 Μελέτη προσομοίωσης

Ο προτεινόμενος αλγόριθμος συγκρίνεται με δύο διαφορετικά δίκτυα RBF. Ο πρώτος, αποκαλούμενος RBFN 1, είναι ένα τυποποιημένο δίκτυο RBF με τις συναρτήσεις ακτινικής βάσης τα κέντρα των οποίων υπολογίζονται από τον αλγόριθμο Ασαφών c -Μέσων (βλέπε παράγραφο 2.3) και η μορφή αυτών των συναρτήσεων είναι η ίδια στις

ασαφείς συναρτήσεις συμμετοχής που δίνονται από την εξίσωση (7.26). Το δεύτερο δίκτυο, αποκαλούμενο RBFN 2, χρησιμοποιεί τον αλγόριθμο των Ασαφών c -Μέσων (βλέπε παράγραφο 2.3) για να υπολογιστούν τα κέντρα των συναρτήσεων ακτινικής βάσης και την εξ. (1.10) για να υπολογιστούν τα πλάτη (widths) των γκαουσιανών συναρτήσεων (gaussian functions) σε 1. Σε όλα τα πειράματα, διαιρέσαμε τυχαία τα δεδομένα εκπαίδευσης (60%) και τα δεδομένα δοκιμής (40%). Τέλος, όλες οι προσομοιώσεις διευθετήθηκαν για $m = 2$ και για 25 διαφορετικές ενάρξεις (δηλ. 25 τρεξίματα).

Παράδειγμα 1: Μονοδιάστατη συνθετική συνάρτηση

Η συνάρτηση είναι η:

$$y = 0.6 \eta\mu(\pi x) + 0.3 \eta\mu(3\pi x) + 0.1 \eta\mu(5\pi x) \quad \text{με } x \in [-1, 1].$$

Όπως και στις αναφορές [70,75], δημιουργήθηκαν 200 ζευγάρια δεδομένων εισόδου-εξόδου. Καθορίστηκε η παράμετρος $\beta = 0.025$ και δημιουργήθηκε μια κοκκοποίηση του συνόλου των δεδομένων σε 9 ομοιόμορφα διανεμημένους συμμετρικούς τριγωνικούς ασαφείς αριθμούς. Εφαρμόζοντας τον αλγόριθμο 7.2 για $\vartheta = 0.51$ στο δεύτερο επίπεδο του αλγορίθμου κοκκοποίησης, αποκτήσαμε $n = 94$ συστάδες. Η συγκριτική ανάλυση πραγματοποιήθηκε από την άποψη του μέσου-τετραγωνικού-σφάλματος (MSE).

Πίνακας 7.1. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης.

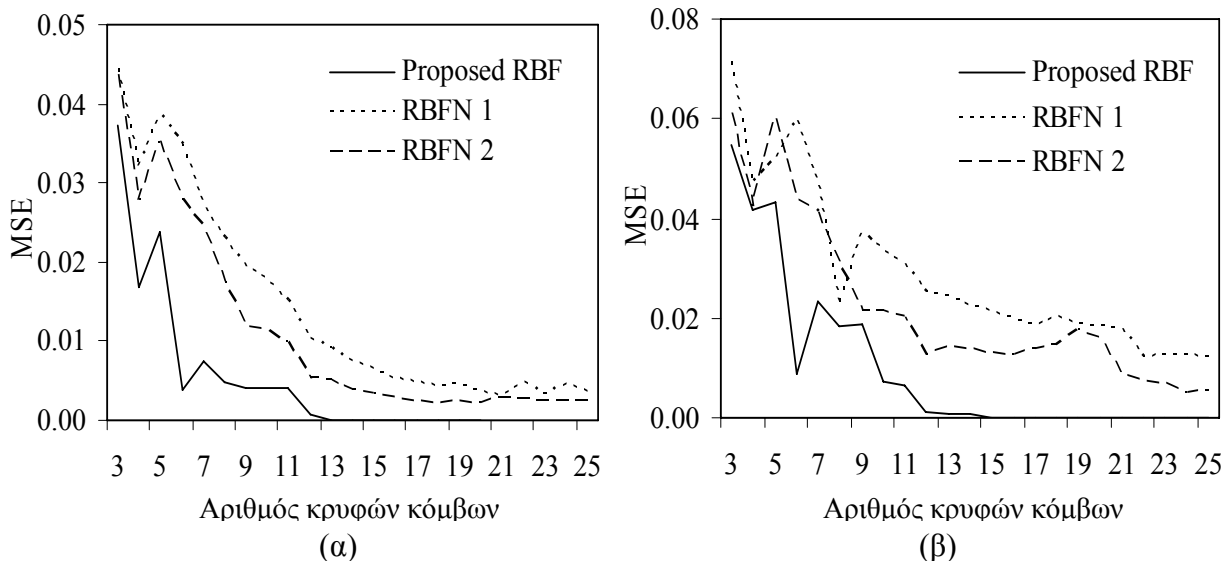
c	RBFN 1	RBFN 2	<u>Προτεινόμενος αλγόριθμος</u> $\vartheta = 0.51$
4	$0.0318 \pm 2.1\text{E-}5$	$0.0278 \pm 2.5\text{E-}4$	$0.0168 \pm 4.3\text{E-}3$
6	$0.0348 \pm 1.5\text{E-}4$	$0.0279 \pm 4.5\text{E-}4$	$0.0037 \pm 5.2\text{E-}4$
8	$0.0231 \pm 6.9\text{E-}4$	$0.0174 \pm 3.3\text{E-}3$	$0.0048 \pm 1.5\text{E-}4$
10	$0.0177 \pm 5.6\text{E-}3$	$0.0111 \pm 4.9\text{E-}3$	$0.0040 \pm 3.7\text{E-}5$
12	$0.0103 \pm 1.9\text{E-}3$	$0.0053 \pm 1.2\text{E-}3$	$0.0007 \pm 1.1\text{E-}4$
14	$0.0074 \pm 1.1\text{E-}3$	$0.0037 \pm 5.5\text{E-}4$	$5.8\text{E-}5 \pm 8.4\text{E-}6$
16	$0.0053 \pm 1.2\text{E-}3$	$0.0028 \pm 8.5\text{E-}4$	$7.9\text{E-}6 \pm 3.1\text{E-}6$
18	$0.0044 \pm 8.2\text{E-}4$	$0.0022 \pm 5.4\text{E-}4$	$8.1\text{E-}6 \pm 5.8\text{E-}6$
20	$0.0037 \pm 5.2\text{E-}4$	$0.0021 \pm 4.9\text{E-}4$	$1.4\text{E-}6 \pm 7.6\text{E-}7$

Πίνακας 7.2. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής.

c	RBFN 1	RBFN 2	Προτεινόμενος αλγόριθμος
			$\mathcal{G} = 0.51$
4	$0.0469 \pm 7.2E-5$	$0.0423 \pm 4.1E-4$	$0.0416 \pm 4.1E-3$
6	$0.0602 \pm 5.3E-4$	$0.0439 \pm 2.5E-3$	$0.0089 \pm 9.1E-4$
8	$0.0225 \pm 1.1E-3$	$0.0309 \pm 5.9E-3$	$0.0185 \pm 2.1E-4$
10	$0.0338 \pm 5.2E-3$	$0.0213 \pm 7.2E-3$	$0.0073 \pm 7.7E-5$
12	$0.0252 \pm 2.4E-3$	$0.0124 \pm 3.2E-3$	$0.0010 \pm 8.5E-4$
14	$0.0225 \pm 1.7E-3$	$0.0136 \pm 3.4E-3$	$5.7E-4 \pm 7.7E-5$
16	$0.0198 \pm 1.4E-3$	$0.0126 \pm 3.7E-3$	$5.9E-5 \pm 2.9E-5$
18	$0.0204 \pm 2.3E-3$	$0.0145 \pm 3.8E-3$	$4.0E-5 \pm 3.8E-5$
20	$0.0183 \pm 1.9E-3$	$0.0161 \pm 2.5E-3$	$2.2E-5 \pm 4.5E-6$

Πίνακας 7.3. Συγκριτική ανάλυση από την άποψη του MSE.

Μοντέλο	Αρ. κόμβων	MSE στα δεδομένα εκπαίδευσης	MSE στα δεδομένα δοκιμής
Pedrycz et al. [75]	≤ 33	0.0029 ± 0.001	0.0052 ± 0.004
	≤ 33	0.0033 ± 0.001	0.0035 ± 0.002
Park et al. [70]	6	$0.0045 \pm 6.2E-4$	0.0281 ± 0.014
Προτεινόμενο RBF ($\mathcal{G} = 0.51$)	6	$0.0037 \pm 5.2E-4$	$0.0089 \pm 9.1E-4$
	15	$6.8E-6 \pm 4.5E-6$	$4.2E-5 \pm 1.2E-5$
	20	$1.4E-6 \pm 7.6E-7$	$2.2E-5 \pm 4.5E-6$



Σχήμα 7.6. MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\mathcal{G} = 0.51$), του RBFN 1, και RBFN 2 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής.

Οι πίνακες 7.1 και 7.2 παρουσιάζουν τις τιμές MSE και τις αντίστοιχες διαφορές που λαμβάνονται από τα τρία δίκτυα για το δεδομένα εκπαίδευσης και δοκιμής, αντίστοιχα. Αυτοί οι πίνακες άμεσα δείχνουν ότι η προτεινόμενη μέθοδος επιτυγχάνει σαφώς την καλύτερη απόδοση. Επιπλέον, μπορεί να ειπωθεί ότι ο προτεινόμενος αλγόριθμος παρουσιάζει μια ανεπαίσθητη συμπεριφορά όσον αφορά την επιλογή ϑ , όπου μπορεί να γίνει αντιληπτό, τα ελαφρώς καλύτερα αποτελέσματα για $\vartheta = 0.51$.

Το σχήμα 7.6 απεικονίζει την πλήρη σύγκριση μεταξύ των τριών δικτύων από την άποψη των μέσων τιμών σφάλματος MSE έναντι του αριθμού κρυφών κόμβων. Φαίνεται ότι το RBFN 2 παρουσιάζει καλύτερα αποτελέσματα από το RBFN 1, μια παρατήρηση που υποστηρίζεται έντονα στα ακόλουθα παραδείγματα, επίσης. Μπορούμε να σκεφτούμε ότι, δεδομένου ότι και τα δύο δίκτυα χρησιμοποιούν τον αλγόριθμο Ασαφών c -Μέσων (FCM) για να υπολογιστούν τα κέντρα πυρήνων, τα πλάτη (widths) βελτιστοποιούνται καλύτερα με τη χρησιμοποίηση της εξ. (1.10). Τέλος, ο πίνακας 7.3 παρέχει μια συγκριτική ανάλυση μεταξύ του προτεινόμενου αλγορίθμου και άλλων προσεγγίσεων που υπάρχουν στη διεθνή βιβλιογραφία. Σύμφωνα με αυτόν τον πίνακα η προτεινόμενη μέθοδος παρέχει τα καλύτερα αποτελέσματα σε κάθε περίπτωση. Αξίζει να σημειωθεί ότι, ο αλγόριθμος είναι σε θέση να υπολογίσει τα ακριβή αποτελέσματα που χρησιμοποιούν έναν μικρό αριθμό κρυφών κόμβων.

Παράδειγμα 2: Τρισδιάστατη συνάρτηση

Σε αυτό το παράδειγμα, εξετάζεται η ακόλουθη τρισδιάστατη συνάρτηση,

$$y = f(x_1, x_2, x_3) = (1 + x_1^{-2} + x_2^{-1.5} + x_3^{-1.0})^2 \quad \text{με} \quad x_1, x_2, x_3 \in [1, 5].$$

Όπως συνέβει και στην αναφορά [71], δημιουργήθηκε ένα σύνολο 200 ζευγαριών δεδομένων εισόδου-εξόδου. Η ασαφής κοκκοποίηση σε κάθε πεδίο ορισμού αποτελέστηκε από 9 συμμετρικούς τριγωνικούς ασαφείς αριθμούς, και τέθηκε η παράμετρος $\beta = 1$. Εφαρμόζοντας τον αλγόριθμο 7.2 για $\vartheta = 0.51$ στο δεύτερο επίπεδο του αλγορίθμου κοκκοποίησης, αποκτήσαμε $n = 70$ συστάδες.

Ο δείκτης απόδοσης ήταν η τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος (Root Mean Square Error - RMSE),

$$RMSE = \sqrt{\sum_{k=1}^N (y_k - \tilde{y}_k)^2 / N}$$

Πίνακας 7.4. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του RMSE για τα δεδομένα εκπαίδευσης.

c	RBFN 1	RBFN 2	Προτεινόμενος Αλγόριθμος $\varrho = 0.51$
5	$0.434 \pm 1.1E-6$	$0.555 \pm 1.4E-7$	0.199 ± 0.0013
10	0.394 ± 0.0205	0.362 ± 0.0541	0.166 ± 0.0131
15	0.376 ± 0.0127	0.242 ± 0.0171	0.124 ± 0.0102
20	0.354 ± 0.0244	0.212 ± 0.0294	0.097 ± 0.0088
25	0.318 ± 0.0309	0.191 ± 0.0198	0.072 ± 0.0047
30	0.326 ± 0.0212	0.197 ± 0.0273	0.060 ± 0.0034

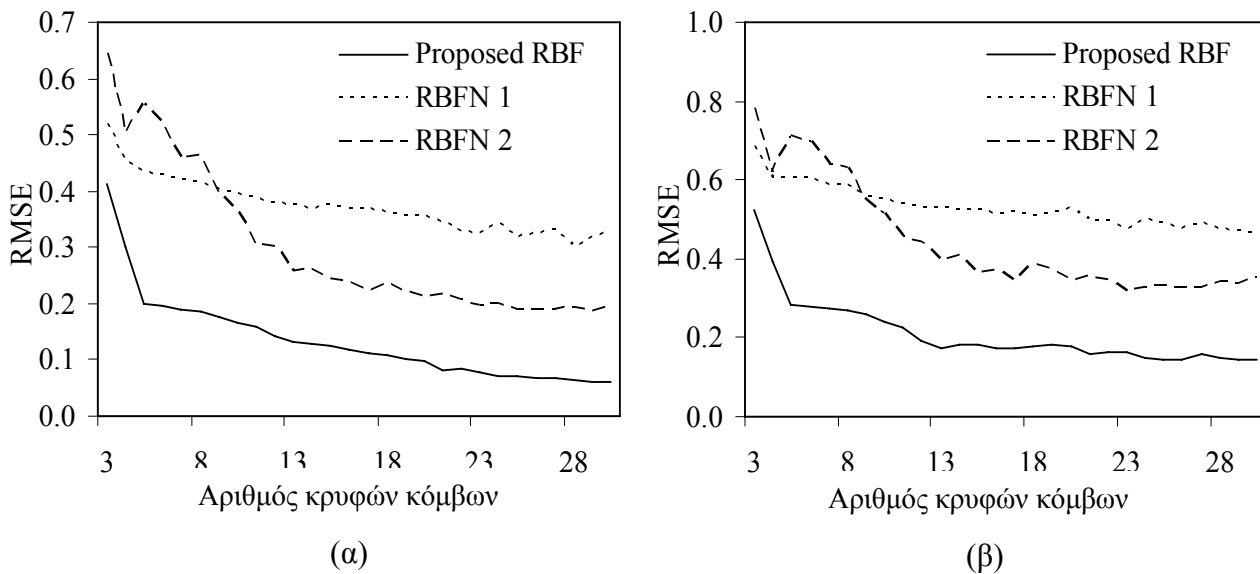
Πίνακας 7.5. Συγκριτική ανάλυση των τριών δικτύων από την άποψη RMSE για τα δεδομένα δοκιμής.

c	RBFN 1	RBFN 2	Προτεινόμενος Αλγόριθμος $\varrho = 0.51$
5	$0.605 \pm 7.9E-8$	$0.713 \pm 1.2E-6$	0.282 ± 0.0031
10	0.552 ± 0.0304	0.514 ± 0.0805	0.238 ± 0.0249
15	0.525 ± 0.0237	0.366 ± 0.0292	0.182 ± 0.0140
20	0.529 ± 0.0191	0.344 ± 0.0324	0.180 ± 0.0115
25	0.492 ± 0.0436	0.331 ± 0.0280	0.146 ± 0.0159
30	0.466 ± 0.0270	0.351 ± 0.0575	0.142 ± 0.0099

Οι τιμές RMSE και οι αντίστοιχες διαφορές που λαμβάνονται από τα τρία δίκτυα αναφέρονται στους πίνακες 7.4 και 7.5. Το σχήμα 7.7 παρουσιάζει πλήρη σύγκριση μεταξύ των τριών δικτύων για τους διάφορους αριθμούς κρυφών κόμβων. Τέλος, ο πίνακας 7.6 συγκρίνει τα επιτευχθέντα αποτελέσματα από τον προτεινόμενο αλγόριθμο με άλλα αποτελέσματα που υπάρχουν στη διεθνή βιβλιογραφία. Σύμφωνα με αυτόν τον πίνακα, η προτεινόμενη μέθοδος λαμβάνει τις καλύτερες τιμές για τη δοκιμή RMSE, ένα γεγονός που δείχνει άμεσα τις ανώτερες ικανότητες σχεδιασμού του. Όσον αφορά τα δεδομένα εκπαίδευσης, εκτός από την περίπτωση όπου $c = 12$, ο αλγόριθμος επιτυγχάνει την καλύτερη απόδοση, επίσης.

Πίνακας 7.6. Συγκριτική ανάλυση από την άποψη των μέσων τιμών RMSE.

Αριθμός κρυφών κόμβων	Park et al. [71]		Προτεινόμενη Μέθοδος	
	Δεδομένα εκπαίδευσης	Δεδομένα δοκιμής	Δεδομένα εκπαίδευσης	Δεδομένα δοκιμής
6	0.2734	0.5055	0.1960	0.2789
10	0.2564	0.3601	0.1667	0.2382
12	0.0681	0.4952	0.1437	0.1941
15	0.1490	0.3684	0.1238	0.1817
24	0.0973	0.2544	0.0714	0.1480
30	0.1881	0.4586	0.0601	0.1425



Σχήμα 7.7. MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\vartheta = 0.51$), του RBFN 1, και RBFN 2 ως συνάρτηση του αριθμού κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής.

Παράδειγμα 3: Διαδικασία ωρίμανσης αποστάγματος κρασιού

Σε αυτό το παράδειγμα εφαρμόζεται η προτεινόμενη μέθοδος σε ένα πραγματικό πρόβλημα ζώης αποκαλούμενο διαδικασία ωρίμανσης αποστάγματος κρασιού. Τα πρόσφατα αποσταγμένα οινοπνεύματα όπως τα αποστάγματα, τα ούισκι και τα ρούμια κρασιού αποθηκεύονται συνήθως στα δρύινα βαρέλια για αρκετά έτη για να βελτιώσουν το άρωμα και την προτίμησή τους. Λόγω του μεγάλου αριθμού παραγόντων που έχουν επίπτωση στην διαδικασία ωρίμανσης, οι ακριβείς μηχανισμοί δεν έχουν γίνει κατανοητοί εντελώς. Επιπλέον, η απουσία ενός αξιόπιστου χημικού δείκτη για να δείξει

την πρόοδο της διαδικασίας ωρίμανσης εισάγει ορισμένα προβλήματα. Επομένως, η ποιοτική εκτίμηση είναι συνήθως βασισμένη στην οργανοληπτική δοκιμή των εμπειρογνομόνων των δειγμάτων από τα βαρέλια. Ο προτεινόμενος αλγόριθμος χρησιμοποιήθηκε για να σχεδιαστεί ένα δίκτυο συναρτήσεων ακτινικής βάσης (RBF) για την πρόβλεψη της ποιότητας του αρώματος των οινοπνευμάτων αποστάγματος κρασιού. Συγκεκριμένα, οι μεταβλητές εισόδου του συστήματος είναι: $x_1 = \text{Χρήση βαρελιών}$ είναι ο αριθμός ξαναγεμισμάτων για κάθε βαρέλι, $x_2 = \text{ηλικία βαρελιών}$ είναι η περίοδος χρήσης κάθε βαρελιού στα έτη, και $x_3 = \text{ηλικία αποστάγματος}$ είναι η περίοδος κάθε αποστάγματος που περιλαμβάνεται σε ένα συγκεκριμένο βαρέλι στα έτη.

Η μεταβλητή εξόδου είναι $y = \text{άρωμα}$, και είναι το άρωμα του ωριμασμένου αποστάγματος. Τα δεδομένα μπορούν να βρεθούν στην [79] βιβλιογραφική αναφορά. Εντούτοις, για λόγους ευκολίας τα 120 ζευγάρια δεδομένων εισόδου-εξόδου που χρησιμοποιούνται σε αυτό το κεφάλαιο δίνονται στον πίνακα 6.9. Για να εκτελεστεί το πείραμα, χωρίστηκε ομοιόμορφα όλο το σύνολο δεδομένων σε 7 συμμετρικά τριγωνικά ασαφή σύνολα και καθορίστηκε η παράμετρος $\beta = 4$. Εφαρμόζοντας τον αλγόριθμο 7.2 για $\vartheta = 0.51$ στο δεύτερο επίπεδο του αλγορίθμου κοκκοποίησης, αποκτήσαμε $n = 62$ συστάδες.

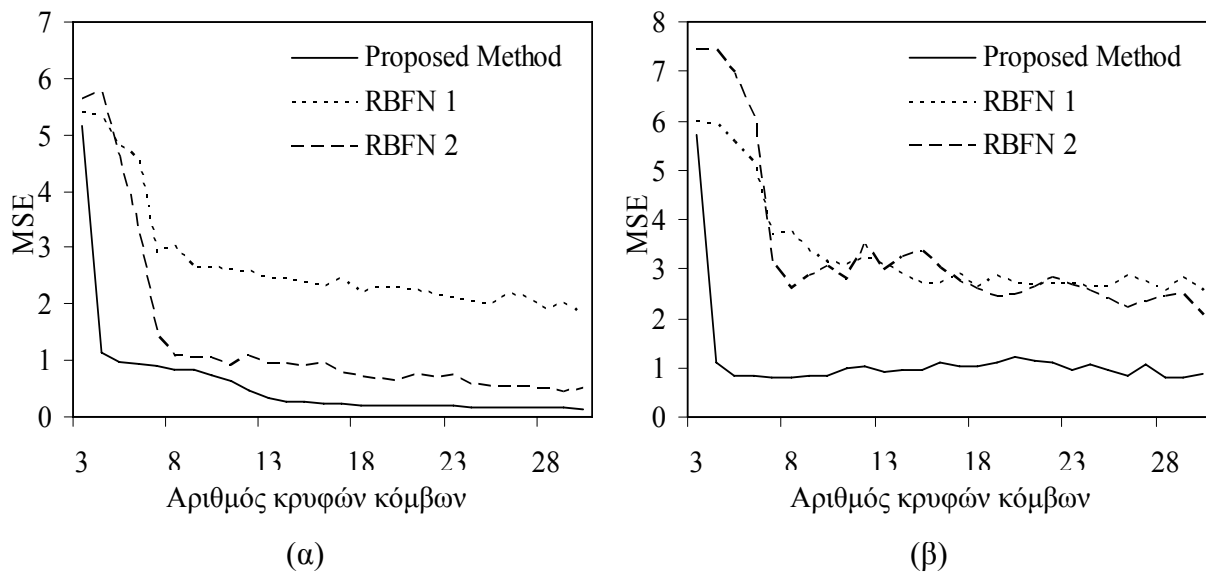
Ο δείκτης απόδοσης ήταν το μέσο τετραγωνικό σφάλμα (MSE). Οι πίνακες 7.7 και 7.8 απεικονίζουν τη συγκριτική ανάλυση μεταξύ των τριών δικτύων ως συνάρτηση του αριθμού κρυφών κόμβων, όπου μπορούμε να ανιχνεύσουμε παρόμοια αποτελέσματα με τα προηγούμενα παραδείγματα.

Πίνακας 7.7. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα εκπαίδευσης.

c	RBFN 1	RBFN 2	<u>Προτεινόμενος Αλγόριθμος</u> $\vartheta = 0.51$
5	4.825 ± 0.3783	4.615 ± 1.0777	0.968 ± 0.0042
10	2.649 ± 0.2063	1.052 ± 0.1737	0.744 ± 0.0911
15	2.363 ± 0.1154	0.904 ± 0.0923	0.257 ± 0.0539
20	2.268 ± 0.2419	0.631 ± 0.1578	0.196 ± 0.0063
25	1.986 ± 0.2894	0.538 ± 0.1846	0.175 ± 0.0074
30	1.805 ± 0.1791	0.504 ± 0.1733	0.146 ± 0.0087

Πίνακας 7.8. Συγκριτική ανάλυση των τριών δικτύων από την άποψη του MSE για τα δεδομένα δοκιμής.

c	RBFN 1	RBFN 2	Προτεινόμενος Αλγόριθμος
			$\vartheta = 0.51$
5	5.565 ± 0.1954	6.973 ± 1.9215	0.828 ± 0.0067
10	3.133 ± 0.4449	3.060 ± 0.5436	0.833 ± 0.0774
15	2.687 ± 0.3331	3.366 ± 0.7008	0.940 ± 0.1093
20	2.720 ± 0.2274	2.477 ± 0.6487	1.202 ± 0.1113
25	2.646 ± 0.1503	2.418 ± 0.4214	0.946 ± 0.1831
30	2.539 ± 0.2223	2.052 ± 0.3135	0.886 ± 0.1487



Σχήμα 7.8. MSE (μέση τιμή) του προτεινόμενου αλγορίθμου ($\vartheta = 0.51$), του RBFN 1 και RBFN 2 ως συνάρτηση του αριθμού κρυφών κόμβων για: (α) τα δεδομένα εκπαίδευσης, και (β) τα δεδομένα δοκιμής.

Τέλος, το σχ. 7.8 απεικονίζει τη συμπεριφορά του δείκτη απόδοσης για τα τρία δίκτυα, για τα δεδομένα εκπαίδευσης και δοκιμής, αντίστοιχα. Με βάση αυτόν το σχήμα μπορούμε να καταλήξουμε στο συμπέρασμα ότι η απόδοση του προτεινόμενου δικτύου, είναι πολύ καλύτερη στην περίπτωση των δεδομένων δοκιμής.

7.5 Συμπεράσματα

Στο παρόν κεφάλαιο, ερευνήθηκε πώς η κοκκοποίηση πληροφοριών (information granulation) μπορεί να περιληφθεί ρητά στο σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Για να ολοκληρωθεί αποτελεσματικά αυτός ο αλγόριθμος, αναπτύχθηκε μια νέα συνάρτηση απόστασης (distance function) μεταξύ των ασαφών κόκκων (fuzzy granules) που συνδυάζει τις πληροφορίες χαμηλότερων επιπέδων που κωδικοποιούνται με μια προσανατολισμένη προς το χρήστη (user-defined) διαδικασία κοκκοποίησης με τις πληροφορίες υψηλού επιπέδου που κωδικοποιούνται από μια από κάτω προς τα επάνω αλγοριθμική κοκκοποίηση βελτιστοποίησης του χώρου των χαρακτηριστικών γνωρισμάτων (feature space). Ολόκληρο το αλγοριθμικό σχήμα αποτελείται από τρία βήματα. Σε πρώτη φάση καθορίστηκε μια κοκκοποίηση σε ολόκληρο το πεδίο ορισμού (universe of discourse), ο οποίος απαιτήθηκε για να ικανοποιήσει ορισμένες ιδιότητες. Τα άλλα δύο βήματα ήρθαν υπό μορφή βελτιστοποιημένων ασαφών διαμερισμών. Συγκεκριμένα, το δεύτερο στάδιο βημάτων χρησιμοποίησε έναν ασαφή αλγόριθμο διαμερισμού (fuzzy partition) για να προεπεξεργαστεί τα δεδομένα εκπαίδευσης και για να λάβει ένα αρχικό σύνολο συστάδων. Στο τρίτο βήμα η μέθοδος των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) εφαρμόστηκε για να καθαρίσει τις άσχημα σκιαγραφημένες συστάδες που παρήχθησαν στο προηγούμενο βήμα και έτσι, για να συναγάγουν μια πιο αφηρημένη ασαφής κοκκοποίηση του διαστήματος χαρακτηριστικών γνωρισμάτων. Η εκτίμηση των παραμέτρων του δικτύου RBF ήταν αποκλειστικά βασισμένη στις πληροφορίες που παρήχθησαν στα ανωτέρω βήματα κοκκοποίησης. Η απόδοση της προτεινόμενης μεθοδολογίας αξιολογήθηκε μέσω της εφαρμογής της σε τρεις περιπτώσεις δοκιμής. Διάφορα συγκριτικά αποτελέσματα έδειξαν ότι το προτεινόμενο σχήμα είναι σε θέση να παράξει νευρωνικά δίκτυα RBF που είναι πολύ ακριβή, χρησιμοποιώντας έναν μικρό αριθμό κρυφών κόμβων.

Συζήτηση Αποτελεσμάτων και Μελλοντική Εργασία

8.1 Συμπεράσματα

Στην παρούσα διδακτορική διατριβή διεξήχθη συστηματική έρευνα και αποδείχθηκε ότι μπορεί να χρησιμοποιηθεί αποτελεσματικά η ασαφής συσταδοποίηση στην εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Συγκεκριμένα, αναπτύχθηκαν πέντε νέες καινοτόμοι μέθοδοι εκπαίδευσης νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Όλες οι προτεινόμενοι μέθοδοι συγκρίθηκαν πάνω σε πραγματικές συναρτήσεις και μοντέλα και αποδείχθηκε ότι επιφέρουν καλύτερα αποτελέσματα χρησιμοποιώντας μικρό αριθμό κρυφών κόμβων άρα και λιγότερων παραμέτρων. Αναλυτικά τα συμπεράσματα είναι τα εξής.

8.1.2 Συμπεράσματα κάθε μεθόδου

Στο 3ο κεφάλαιο, μια νέα μέθοδος [64] βασισμένη στην ασαφή συσταδοποίηση αναπτύχθηκε και αξιολογήθηκε όσον αφορά τη διαδικασία εκπαίδευσης νευρωνικών δικτύων RBF. Ο αλγόριθμος χρησιμοποιεί τρεις φάσεις εκπαίδευσης για να υπολογίσει τις κατάλληλες τιμές για τις παραμέτρους δικτύων. Στην πρώτη φάση εκτελεί μια δύο σταδίων ασαφή συσταδοποίηση (fuzzy clustering) κατά τέτοιο τρόπο ώστε η υποκείμενη δομή δεδομένων μπορεί να ανακαλυφθεί αποτελεσματικά. Στη δεύτερη φάση, τα βάρη σύνδεσης νευρώνων υπολογίζονται και τελικά, στην τρίτη φάση, όλες οι παράμετροι του νευρωνικού δικτύου βελτιστοποιούνται με την εφαρμογή του αλγορίθμου

οπισθοδιάδοσης. Ο προτεινόμενος αλγόριθμος προσφέρει δύο σημαντικά πλεονεκτήματα:

1. Η διαδικασία εκπαίδευσης ελέγχεται από τον αριθμό κόμβων. Δηλαδή σε κάθε επανάληψη ο αριθμός κρυφών κόμβων αυξάνεται μέχρι η απόδοση του νευρωνικού δικτύου να είναι αρκετά ακριβής. Αυτό το γεγονός άμεσα υπονοεί ότι η διαδικασία θα σταματήσει μόλις ανιχνεύεται ο μικρότερος αξιόπιστος αριθμός κόμβων διατηρώντας συγχρόνως το μικρότερο αριθμό επαναλήψεων.
2. Η εκτίμηση των παραμέτρων των κρυφών κόμβων δεν εξαρτάται από οποιαδήποτε αρχική τυχαία επιλογή των τιμών τους αλλά μάλλον είναι απλό αποτέλεσμα της διαδικασίας συσταδοποίησης.

Διάφορες προσομοιώσεις εκτελέστηκαν, οι οποίες έδειξαν ότι η μεθοδολογία μπορεί να είναι πολύ επιτυχής όταν εφαρμόζεται στη μοντελοποίηση συστημάτων.

Στο 4^ο κεφάλαιο, έγινε ενσωμάτωση της ιεραρχικής και βέλτιστης συσταδοποίησης για τον σχεδιασμό αποδοτικών νευρωνικών δικτύων RBF [66]. Η βασική ιδέα είναι να σχετίζεται άμεσα η τοπολογία του δικτύου με την υποκείμενη δομή δεδομένων, και να χρησιμοποιεί τη βέλτιστη ασαφή συσταδοποίηση (optimal fuzzy clustering) για να ανιχνεύσει αυτήν την δομή. Κατόπιν η ανάθεση είναι προφανής, δεδομένου ότι τα κέντρα συστάδων και οι ασαφείς διαφορές συστάδων αντιστοιχούν στα κέντρα (centers) και τα πλάτη (widths) των συναρτήσεων ακτινικής βάσης, αντίστοιχα. Τα συναπτικά βάρη (synaptic weights) υπολογίστηκαν από τον αλγόριθμο ελαχίστων τετραγώνων (least squares), ενώ ολόκληρο το σύνολο παραμέτρων (κέντρα, πλάτη, συναπτικά βάρη) βελτιστοποιήθηκε με τη βοήθεια του αλγορίθμου οπισθοδιάδοσης (back-propagation). Τέλος, οι προσομοιώσεις έδειξαν ότι η μέθοδος είναι σε θέση να λάβει την ακριβή απόδοση που χρησιμοποιεί έναν μικρό αριθμό κόμβων.

Στο 5^ο κεφάλαιο προτάθηκε μια τρίτη μέθοδος [63] που συνδυάζει ασαφή συσταδοποίηση και τη βέλτιστη ασαφή συσταδοποίηση για τον αποτελεσματικό σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Πρώτα, εφαρμόστηκε η μέθοδος Ασαφών *c*-Μέσων (FCM) στο χώρο εισόδου-εξόδου με σκοπό να γίνει η προεπεξεργασία των διαθέσιμων δεδομένων. Οι προκύπτουσες συστάδες προβάλλονται στο χώρο εισόδου. Τα αντίστοιχα κέντρα συστάδων θεωρούνται ως ένα νέο σύνολο δεδομένων, το οποίο στη συνέχεια αναλύεται περισσότερο, χρησιμοποιώντας τη

βέλτιστη μέθοδο της ασαφούς συσταδοποίησης από πλευράς της μεθόδου σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM). Για να επιτευχθεί αυτό, αναπτύξαμε ένα νέο δείκτη εγκυρότητας, ο οποίος χρησιμοποιείται για να προσδιορίσει τον κατάλληλο αριθμό των κρυφών RBF κόμβων. Ο αλγόριθμος έχει εφαρμοστεί με επιτυχία σε γνωστές σειρές δεδομένων, όπου οι επιδόσεις του έχουν δοκιμαστεί και αξιολογηθεί.

Στο 6^ο κεφάλαιο αναπτύχθηκε ένας υβριδικός αλγόριθμος συσταδοποίησης [65] για την αποτελεσματική εκπαίδευση νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Υπάρχουν δύο βασικές απόψεις σχεδίου της προτεινόμενης μεθοδολογίας. Ο πρώτος αφορά τον τρόπο που καθορίζουμε τη δομή των συστάδων στο πολυδιάστατο χώρο των χαρακτηριστικών γνωρισμάτων (feature space). Για να ολοκληρώσουμε αυτόν τον στόχο, χρησιμοποιήσαμε μια στρατηγική μετάβασης από την ασαφή κατάσταση (fuzzy mode), όπου κάθε διάνυσμα εκπαίδευσης ανήκει σε πολλές συστάδες με διαφορετικούς βαθμούς συμμετοχής, στην διακριτή κατάσταση, όπου κάθε διάνυσμα εκπαίδευσης ανήκει μόνο σε μια συστάδα. Αυτή η μετάβαση πραγματοποιείται μέσω των αναλυτικών όρων που εξάγονται από την ελαχιστοποίηση μιας εξειδικευμένης αντικειμενικής συνάρτησης (objective function). Το τελικό αποτέλεσμα είναι ένας υβριδικός διαμερισμός όπου οι ασαφείς και οι διακριτοί όροι λαμβάνονται υπόψη ταυτόχρονα. Ο ανταγωνισμός μεταξύ των υβριδικών συστάδων πραγματοποιείται όσον αφορά τις κοινές ασαφείς περιοχές τους και τις διακριτές περιοχές τους, επίσης. Έτσι φαίνεται ότι το προτεινόμενο σχέδιο συσταδοποίησης συνδυάζει τα πλεονεκτήματα της διακριτής και της ασαφούς συσταδοποίησης, παράγοντας μια αξιόπιστη εκτίμηση των κέντρων των συναρτήσεων βάσης. Η δεύτερη άποψη αφορά την εκτίμηση των πλατών (widths) των πυρήνων. Για να εξετάσουμε αυτό το ζήτημα, προτείνεται η μέθοδος που αναπτύχθηκε και χρησιμοποιήθηκε και στα δύο προηγούμενα κεφάλαια. Η απόδοση της προτεινόμενης μεθοδολογίας αξιολογήθηκε μέσω της εφαρμογής της, σε πέντε περιπτώσεις δοκιμής. Διάφορα συγκριτικά αποτελέσματα έδειξαν ότι ο προτεινόμενος αλγόριθμος είναι σε θέση να δημιουργήσει ακριβή νευρωνικά δίκτυα RBF.

Τέλος, στο 7^ο κεφάλαιο, ερευνήθηκε πώς η κοκκοποίηση πληροφοριών (information granulation) μπορεί να περιληφθεί ρητά στο σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης (RBF). Για να ολοκληρωθεί αποτελεσματικά αυτός ο αλγόριθμος, αναπτύχθηκε μια νέα συνάρτηση απόστασης (distance function)

μεταξύ των ασαφών κόκκων (fuzzy granules) που συνδυάζει τις πληροφορίες χαμηλότερων επιπέδων που κωδικοποιούνται με μια προσανατολισμένη προς το χρήστη (user-defined) διαδικασία κοκκοποίησης με τις πληροφορίες υψηλού επιπέδου που κωδικοποιούνται από μια από κάτω προς τα επάνω αλγοριθμική κοκκοποίηση βελτιστοποίησης του χώρου των χαρακτηριστικών γνωρισμάτων (feature space). Ολόκληρο το αλγοριθμικό σχήμα αποτελείται από τρία βήματα. Σε πρώτη φάση καθορίστηκε μια κοκκοποίηση σε ολόκληρο το πεδίο ορισμού (universe of discourse), ο οποίος απαιτήθηκε για να ικανοποιήσει ορισμένες ιδιότητες. Τα άλλα δύο βήματα ήρθαν υπό μορφή βελτιστοποιημένων ασαφών διαμερισμών. Συγκεκριμένα, το δεύτερο στάδιο βημάτων χρησιμοποίησε έναν ασαφή αλγόριθμο διαμερισμού (fuzzy partition) για να προεπεξεργαστεί τα δεδομένα εκπαίδευσης και για να λάβει ένα αρχικό σύνολο συστάδων. Στο τρίτο βήμα η μέθοδος των σταθμισμένων (weighted) Ασαφών c -Μέσων (FCM) εφαρμόστηκε για να καθαρίσει τις άσχημα σκιαγραφημένες συστάδες που παρήχθησαν στο προηγούμενο βήμα και έτσι, για να συναγάγουν μια πιο αφηρημένη ασαφής κοκκοποίηση του διαστήματος χαρακτηριστικών γνωρισμάτων. Η εκτίμηση των παραμέτρων του δικτύου RBF ήταν αποκλειστικά βασισμένη στις πληροφορίες που παρήχθησαν στα ανωτέρω βήματα κοκκοποίησης. Η απόδοση της προτεινόμενης μεθοδολογίας αξιολογήθηκε μέσω της εφαρμογής της σε τρεις περιπτώσεις δοκιμής. Διάφορα συγκριτικά αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος είναι σε θέση να μοντελοποιήσει νευρωνικά δίκτυα RBF τα οποία είναι πολύ ακριβή χρησιμοποιώντας έναν μικρό αριθμό κρυφών κόμβων.

8.1.2 Σύγκριση των προτεινόμενων μεθόδων

Στην παρούσα παράγραφο συγκρίνονται όλες τις προτεινόμενες μέθοδοι μεταξύ τους. Συγκεκριμένα, μελετήσαμε το παράδειγμα προσέγγισης μονοδιάστατης συνθετικής συνάρτησης εξ. (3.21) όσον αφορά το μέσο τετραγωνικό σφάλμα (MSE) για όλες τις προτεινόμενες μεθόδους. Στον πίνακα 8.1 αναλύονται τα αποτελέσματα της σύγκρισης τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα δοκιμής. Σύμφωνα με τον πίνακα αυτό μπορούμε να διαπιστώσουμε ότι η 4^η μέθοδος (κεφ. 6) επιφέρει τα καλύτερα αποτελέσματα MSE τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα

δοκιμής. Συγκεκριμένα, το μέσο τετραγωνικό σφάλμα (MSE) για τα δεδομένα εκπαίδευσης ήταν ίσο με 0.0018 και για τα δεδομένα δοκιμής ήταν 0.0025.

Πίνακας 8.1 Σύγκριση των προτεινόμενων μεθόδων για τα δεδομένα εκπαίδευσης και για τα δεδομένα δοκιμής.

Μέθοδος	Αριθμός κρυφών κόμβων	MSE στα δεδομένα εκπαίδευσης	MSE στα δεδομένα δοκιμής
1 ^η (κεφ. 3)	14	0.060	0.064
2 ^η (κεφ. 4)	8	0.0058	0.0079
3 ^η (κεφ. 5)	12	0.0034	0.0036
4 ^η (κεφ. 6)	12	0.0018	0.0025
5 ^η (κεφ. 7)	6	0.0037	0.0089

8.2 Μελλοντική εργασία (Future work)

Σαν μελλοντική εργασία μπορεί να αναπτυχθούν οι παρακάτω νέες μέθοδοι:

1. Μέθοδοι υπολογισμού των πλατών (widths) των συναρτήσεων ακτινικής βάσης (radial basis functions).
2. Αλγόριθμοι υπολογισμού των συναπτικών βαρών (weights) χρησιμοποιώντας νέες μεθόδους.
3. Αλγόριθμοι περαιτέρω βελτιστοποίησης όλων των παραμέτρων του νευρωνικού δικτύου RBF όπως ο αλγόριθμος particle swarm optimization.

ΠΑΡΑΡΤΗΜΑ

Δημοσιεύσεις σε έγκυρα επιστημονικά περιοδικά και στα πρακτικά διεθνών συνεδρίων με κριτές:

Niros, D.A., Tsekouras, G. (2011) ‘A Novel Training Algorithm for RBF Neural Network Using a Hybrid Fuzzy Clustering Approach’, *Fuzzy Sets and Systems* (article in press).

Niros, D.A., Leventis, G., Tsolakis, D., Tsekouras, G. and Kenteris, M. (2011) ‘A Hierarchical Fuzzy Clustering Framework for Training RBF Networks’, *In the Proceedings of the IASTED International Conference, Artificial Intelligence and Soft Computing* (ASC 2011), pp. 248-254.

Tsekouras, G., Dartzentas, D., Drakoulaki, I., Niros, D.A. (2010) ‘Fast Fuzzy Vector Quantization’, *World Congress on Computational Intelligence* (WCCI), pp. 1439-1446.

Niros, D.A., Tsekouras, G. (2009) ‘On Training Radial Basis Function Neural Networks using Optimal Fuzzy Clustering’, *IEEE 17th Mediterranean Conference on Control & Automation*, pp. 395-400.

Niros, D.A., Tsekouras, G. (2008) ‘A Fuzzy Clustering Algorithm to Estimate the Parameters of Radial Basis Functions Neural Networks and its Application to System Modeling’, *Lecture Notes in Artificial Intelligence*, vol. 5138, pp. 194–204.

Tsekouras, G., Gavalas, D., Filios, S., Niros, D.A., Bafaloukas, G. (2008) ‘A Clustering Framework to Build Focused Web Crawlers for Automatic Extraction of Cultural Information’, *Lecture Notes in Artificial Intelligence*, vol. 5138, pp. 419–424.

BIBΛΙΟΓΡΑΦΙΑ

- [1] Alexandridis, A., Patrinos, P., Sarimveis, H., Tsekouras, G.E. (2005) ‘A two-stage evolutionary algorithm for variable selection in the development of RBF neural network models’, *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 2, pp. 149-162.
- [2] Babuska, R., Verbruggen, H. (2004) ‘Neuro-fuzzy methods for nonlinear system identification’, *Annual Reviews in Control*, vol. 27, pp. 73-85.
- [3] Bargiela, A., Pedrycz, W. (2008) ‘Toward a theory of granular computing for human-centered information processing’, *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 2, pp. 320-330,.
- [4] Bargiela, A., Pedrycz, W., Hirota, K. (2004) ‘Granular Prototyping in Fuzzy Clustering’, *IEEE Transactions on Fuzzy Systems*, vol. 12, no. 52, pp. 697-709.
- [5] Behloul, F., Lelieveldt, B.P.F., Boudraa, A., Reiber, J.H.C. (2002) ‘Optimal design of radial basis function neural networks for fuzzy-rule extraction in high dimensional data’, *Pattern Recognition*, vol. 35, pp. 659-675.
- [6] Benoudjit, N., Verleysen, M. (2003) ‘On the kernel widths in radial-basis function networks’, *Neural Processing Letters*, vol. 18, pp. 139-154.
- [7] Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P., Silbiger, M.L., Arrington, J.A., Murtagh, R.F. (1996) ‘Validity-guided (re)clustering with applications to image segmentation’, *IEEE Transactions on Fuzzy Systems*, vol. 4, pp. 112–23.
- [8] Bezdek, J.C. (1987) *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press.
- [9] Bezdek, J.C.. (1973) *Fuzzy Mathematics in Pattern Classification*. PhD Dissertation, Ithaka, New York: Cornell University.
- [10] Bezdek, J.C. (1974) ‘Cluster Validity with the Fuzzy Sets’, *Journal of Cybernetics*, vol. 3, pp. 58-73.
- [11] Bezdek, J.C. (1980) ‘A Convergence Theorem for the fuzzy ISODATA Clustering Algorithms’, *IEEE Transactions Pattern Anal. Machine Intelligence (PAMI)*, vol. 2, pp. 1-8.

- [12] Bezdek, J.C., Pal, D.K. (1992) *Fuzzy Models for Pattern Recognition - Methods that Search for Patterns in Data*, New York: IEEE Press.
- [13] Billings, S.A., Zheng, G.L. (1995) ‘Radial Basis Function Network Configuration Using Genetic Algorithms’, *Neural Networks*, vol. 8, pp. 877-890.
- [14] Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Oxford: Clarendon Press.
- [15] Broomhead, D., Lowe, D. (1998) ‘Multivariable functional interpolation and adaptive networks’, *Complex Systems*, vol. 2, pp. 321-355.
- [16] Cha, I. and Kassam, S. A. (1996) ‘RBFN restoration of nonlinearly degraded images’, *IEEE Transactions on Image Processing*, vol. 5, pp. 964–975.
- [17] Chen, C.L., Chen, W.C. and Chang, F. Y. (1993) ‘Hybrid learning algorithm for Gaussian potential function networks’, *Proc. Inst. Elect. Eng.*, vol. 140, no. 6, pp. 442–448.
- [18] Chen, S., Cowan, C.F.N., Grant, P. M. (1991) ‘Orthogonal least squares learning algorithm for radial basic function network’, *IEEE Transactions on Neural Networks*; vol. 2, pp. 302-309.
- [19] Chen, Y., Yang, B., Dong, J. and Abraham, A. (2005) ‘Time-series Forecasting using Flexible Neural Tree Model’, *Information Science*, vol. 174, pp. 219-235.
- [20] Chen, Y., Yang, B., Zhou, J. (2006) ‘Automatic Design of Hierarchical RBF Networks for System Identification’, pp. 1191–1195.
- [21] Cho, K.B., Wang, B.H. (1995) ‘Radial basis function based adaptive fuzzy systems their application to system identification and prediction’, *Fuzzy Sets and Systems*, vol. 83, pp. 325-339.
- [22] Dunham, M. (2005) *Data Mining: εισαγωγικά και προηγμένα θέματα εξόρυξης γνώσης από δεδομένα*, Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- [23] Er, M.J., Chen, W., Wu, S. (2005) ‘High-speed face recognition based on discrete cosine transform and RBF neural networks’, *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 679–691.
- [24] Figueiredo, M.A. (2000) ‘On Gaussian Radial Basis Function Approximations: Interpretation’, *Extensions and Learning Strategies*, pp. 618-621.

- [25] Filho, O.M.M., Ebecken, N.F.F. (2009) 'Visual RBF network design based on star coordinates', *Advances in Engineering Software*, vol. 40, no. 9, pp. 913-919.
- [26] Fisch, D., Hofmann, A., Sick, B. (2010) 'On the versatility of radial basis function neural networks: A case study in the field of intrusion detection', *Information Sciences*, vol. 180, no. 12, pp. 2421-2439.
- [27] Gholizadeh, S., Salajegheh, J., Salajegheh, E. (2009) 'An intelligent neural system for predicting structural response subject to earthquakes', *Advances in Engineering Software*, vol. 40, no.8, pp. 630-639.
- [28] Gonzalez, J., Rojas, I., Ortega, J., Pomares, H., Fernandez, J. & Diaz, A.F. (2003) 'Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation', *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1478–1495.
- [29] González, J., Rojas, I., Pomares, H., Ortega, J., Prieto, A. (2002) 'A New Clustering Technique for Function Approximation', *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 132-142.
- [30] Haddadnia, J., Faeza, K., Ahmadi, M. (2003) 'A fuzzy hybrid learning algorithm for radial basis function neural network with application in human face recognition', *Pattern Recognition*, vol. 36, pp. 1187 – 1202.
- [31] Harpham, C., Dawson, C.W. (2006) 'The effect of different basis functions on a radial basis function network for time series prediction: A comparative study', *Neurocomputing*, vol. 69, pp. 2161–2170.
- [32] Haykin, S. (1999) *Neural Networks - A Comprehensive Foundation*, Second Edition, New Jersey: Prentice Hall.
- [33] Honda, K., Ichihashi, H. (2005) 'A new approach to fuzzification of memberships in cluster analysis', *Lecture Notes in Artificial Intelligence*, vol. 3558, pp. 172-182. (Also in the Proceedings of the International Conference on Modelling Decisions for Artificial Intelligence (2005)).
- [34] Hoppner, Fr., Klawonn, Fr., Kruse, Fud., Runkler, Th. (1999) *Fuzzy Cluster Analysis*, Wiley.

- [35] Huang, D.S., Zhao, W.B. (2005) ‘Determining the centers of radial basis probabilistic neural networks by recursive orthogonal least square algorithms’, *Applied Mathematics and Computation*, vol. 162, pp. 461–473.
- [36] Jain, A.K., Dubes, R.C. (1988) *Algorithms for clustering data*, Englewood Cliffs New Jersey: Prentice Hall.
- [37] Jang, J.-S.R. Sun, C.-T. Mizutani, E. (1997) *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*, Upper Saddle River New Jersey: Prentice Hall.
- [38] Karayiannis, N. B. (1999) ‘Reformulated radial basis neural networks trained by gradient descent’, *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 657–671.
- [39] Karayiannis, N.B., Pai, P.I. (1995) ‘Fuzzy Vector Quantization Algorithms and Their Application in Image Compression’, *IEEE Transactions on Image Processing*, vol. 4, no. 9, pp. 1193-1201.
- [40] Katayama, R., Kajitani, Y., Kuwata, K., Nishida, Y. (1993) ‘Self generating radial basis function as neuro-fuzzy model and its application to nonlinear prediction of chaotic time series’, *2nd Int. Conf on Fuzzy Systems (FUZZIEEE’ 93)*, pp. 407-414.
- [41] Kim, D., Kim, C. (1997) ‘Forecasting time series with genetic fuzzy predictor ensembles’, *IEEE Transactions on Fuzzy Systems*, vol. 5, pp. 523-535.
- [42] Kim, E., Park, M., Kim, S., Park, M. (1998) ‘A transformed input-domain approach to fuzzy modelling’, *IEEE Transactions on Fuzzy Systems*, vol. 6, no. 4, pp. 596-604.
- [43] Klir, G.J. (1995) ‘Principles of uncertainty: What are they? Why do we need them?’, *Sets and Systems*, vol. 74, pp. 15-31.
- [44] Klir, G.J., Yuan, B. (1995) *Fuzzy Sets and Fuzzy Logic*, New Jersey: Prentice Hall.
- [45] Kohonen, T. (1988) *Self-organization and Associative Memory*, 3rd edition. New Jersey: Springer.

- [46] Kumar, R., Ganguli, R., Omkar, S.N. (2010) ‘Rotorcraft parameter estimation using radial basis function neural network’, *Applied Mathematics and Computation*, vol. 216, no. 2, pp. 584-597.
- [47] Lazaro, M., Santamaria, I., Pantaleon, C. (2003) ‘A new EM-based training algorithm for RBF networks’, *Neural Networks*, vol. 16, no. 1, pp. 69–77.
- [48] Lee, S.J., Ouyang, C.S. (2003) ‘A neuro-fuzzy system modeling with selfconstructing rule generation and hybrid SVD-based learning’, *IEEE Transactions on Fuzzy Systems*, vol. 11, pp. 341–53.
- [49] Leng, G., McGinnity, T.M., Prasad, G. (2006) ‘Design for self-organizing fuzzy neural networks based on genetic algorithms’, *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 6, pp. 755-766.
- [50] Leonard, J.A., Kramer, M.A. (1991) ‘Radial Basis Function Networks for Classifying Process Faults’, *IEEE Control Systems*, vol. 31.
- [51] Li, Y., Qiang, S., Zhuang, X., Kaynak, O. (2004) ‘Robust and adaptive back stepping control for nonlinear systems using RBF neural networks’, *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 693–701.
- [52] Linkens, D.A., Chen, M.Y. (1999) ‘Input selection and partition validation for fuzzy modelling using neural network’, *Fuzzy Sets Systems*, vol. 107, pp. 299–308.
- [53] Looney, C.G. (2009) ‘Fuzzy connectivity clustering with radial basis kernel functions’, *Fuzzy Sets and Systems*, vol. 160, pp. 1868–1885.
- [54] Lucks, M.B., Oki, N. (1999) ‘A Radial Basis Network for Function Approximation’, *In the Proceedings of the 42nd Midwest Symposium on Circuits and Systems*, pp. 1099-1101.
- [55] Luo, W., Karim, M.N., Morris, A., Martin, E.B. (1996) ‘Control Relevant Identification of a pH Wastewater Neutralisation Process using Adaptive Radial Basis Function Networks’, *Computers and Chemical Engineering*, S1017.
- [56] Meghabghab, G., Kandel, A. (2004) ‘Stochastic simulations of web search engines: RBF versus second-order regression models’, *Information Sciences*, vol. 159, no. 1-2, pp. 1-28.

- [57] Mencar, C., Fanelli, A.M. (2008) ‘Interpretability constraints for fuzzy information granulation’, *Information Sciences*, vol. 178, pp. 4585–4618.
- [58] Mitra, S., Basak, J. (2001) ‘FRBF: A fuzzy radial basis function network’, *Neural Computing and Applications*, vol. 10, pp. 244-252.
- [59] Miyamoto, S., Umayahara, S. (1998) ‘Fuzzy clustering by quadratic regularization’, *In the Proceedings of the IEEE International Conference on Fuzzy Systems*, pp. 1394-1399.
- [60] Moody, J., Darken, C.J. (1989) ‘Fast learning in networks of locally-tuned processing units’, *Neural Computation*, vol. 1, pp. 281–294.
- [61] Murata, J., Itoh, S., Hirasawa, K. (1999) ‘Size-Reducing RBF Networks’, *In the Proceedings of the International Joint Conference on Neural Networks*, vol. 2, pp. 1308-1312.
- [62] Narendra, K.S., Parthasarathy, K. (1990) ‘Identification and Control of Dynamical Systems Using Neural Networks’, *IEEE Transactions on Neural Networks*, vol. 1, no. 4.
- [63] Niros, D.A., Leventis, G., Tsolakis, D., Tsekouras, G. and Kenteris, M. (2011) ‘A Hierarchical Fuzzy Clustering Framework for Training RBF Networks’, *In the Proceedings of the IASTED International Conference, Artificial Intelligence and Soft Computing (ASC 2011)*, pp. 248-254.
- [64] Niros, A., Tsekouras, G. (2008) ‘A Fuzzy Clustering Algorithm to Estimate the Parameters of Radial Basis Functions Neural Networks and its Application to System Modeling’, *Lecture Notes in Artificial Intelligence*, vol. 5138, pp. 194–204.
- [65] Niros, D.A., Tsekouras, G. (2010) ‘A Novel Training Algorithm for RBF Neural Network Using a Hybrid Fuzzy Clustering Approach’, *Fuzzy Sets and Systems* (to appear).
- [66] Niros, D.A., Tsekouras, G. (2009) ‘On Training Radial Basis Function Neural Networks using Optimal Fuzzy Clustering’, *IEEE 17th Mediterranean Conference on Control & Automation*, pp. 395-400.
- [67] Pal, N.R., Bezdek, J.C. (1995) ‘On clustering validity for the fuzzy c-means model’ *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370-379.

- [68] Panda, S.S., Chakraborty, D., Pal, S.K. (2008) ‘Flank wear prediction in drilling using back propagation neural network and radial basis function network’, *Applied Soft Computing*, vol. 8, no. 2, pp. 858-871.
- [69] Panella, M., Gallo, A.S. (2005) ‘An input-output clustering approach to the synthesis of ANFIS networks’, *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 1, pp. 69-81.
- [70] Park, H.S., Oh, S.K., Kim, A.K. (2009) ‘Development of design strategy for RBF neural network with the aid of context-based FCM’, *Lecture Notes in Computer Science*, vol. 5551, pp. 156-163.
- [71] Park, H.S., Pedrycz, W., Oh, S.K. (2009) ‘Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields’, *IEEE Transactions on Neural Networks*, vol. 20, no. 10, pp. 1604-1616.
- [72] Pedrycz, W. (1998) ‘Conditional Fuzzy Clustering in the Design of Radial Basis Function Neural Networks’ *IEEE Transactions on Neural Networks*, vol. 9, no. 4, pp. 601-612.
- [73] Pedrycz, W. (2005) *Knowledge-based clustering : from data to information granules*, New Jersey: John Wiley & Sons Inc.
- [74] Pedrycz, W. (2008) *Fuzzy Sets as a User-Centric Processing Framework of Granular Computing*, In Handbook of Granular Computing, W. Pedrycz, A. Skowron., V. Kreinovich (Eds), pp. 97-139, John Wiley & Sons.
- [75] Pedrycz, W., Park, H.S, Oh, S.K. (2008) ‘A granular-oriented development of functional radial basis function neural networks’, *Neurocomputing*, vol. 72, pp. 420–435.
- [76] Pedrycz, W., Vukovich, G. (2001) ‘Granular neural networks’, *Neurocomputing*, vol. 36, pp. 205-224.
- [77] Pedrycz, W. (1984) ‘Construction of Fuzzy Relational Models’, *Cybernetics and Systems Research*, vol. 2, pp. 545-549.
- [78] Pedrycz, W. (1993) *Fuzzy Control and Fuzzy Systems*, Research Studies, New Jersey: Press Wiley.

- [79] Raptis, C.G., Siettos, C.I., Kiranoudis, C.T., Bafas, G.V. (2000) ‘Classification of aged wine distillates using fuzzy and neural network systems’, *Journal of Food Engineering*, vol. 46, pp. 267-275.
- [80] Rivas, V.M., Merelo, J.J., Castillo, P.A., Arenas, M.G., Castellano, J.G. (2004) ‘Evolving RBF neural networks for time-series forecasting with EvRBF’, *Information Sciences*, vol. 165, no. 3-4, pp. 207-220.
- [81] Roh, S.B., Ahn, T.C., Pedrycz, W. (2010) ‘The design methodology of radial basis function neural networks based on fuzzy K-nearest neighbors approach’, *Fuzzy Sets and Systems*, vol. 161, no. 13, pp. 1803-1822.
- [82] Saha, A., Keeler, J.D. (1989) ‘Algorithms for better representation and faster learning in radial basis function networks’, *In Advances in Neural Information Processing Systems*, pp. 482–489.
- [83] Sarimveis, H., Alexandridis, A., Tsekouras, G., Bafas, G. (2002) ‘A Fast and Efficient Algorithm for Training Radial Basis Function Neural Networks Based on a Fuzzy Partition of the Input Space’, *Ind. Eng. Chem. Research*, vol. 41, pp. 751-759.
- [84] Sarimveis, H., Doganis, P., Alexandridis, A. (2006) ‘A classification technique based on radial basis function neural networks’, *Advances in Engineering Software*, vol. 37, no. 4, pp. 218-221.
- [85] Sarimveis, S., Alexandridis, A., Bafas, G. (2003) ‘A fast training algorithm for RBF networks based on subtractive clustering’, *Neurocomputing*, vol. 51, pp. 501-505.
- [86] Schwenker, F., Kestler, H.A., Palm, G. (2001) ‘Three learning phases for radial-basis-functions networks’, *Neural Networks*, vol. 14, pp. 439-458.
- [87] Shi, D., Yeung, D.S., Gao, J. (2005) ‘Sensitivity analysis applied to the construction of radial basis function networks’, *Neural Networks*, vol. 18, pp. 951–957.
- [88] Sing, J.K., Basu, D.K., Nasipuri, M., Kundu, M. (2003) ‘Improved k-means algorithm in the design of RBF neural networks’, *International Conference on Convergent Technologies for Asia-Pacific Region*, vol. 2, pp. 841 – 845.

- [89] Staiano, A., Tagliaferri, R., Pedrycz, W. (2006) ‘Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering’, *Neurocomputing*, vol. 69, pp. 1570–1581.
- [90] Tikhonov, A., Arsenin, V. (1977) *Solution of ill-posed Problems*, Washington: DC.
- [91] Tsekouras, G. (2005) ‘On the use of the weighted fuzzy c-means in fuzzy modelling’, *Advances in Engineering Software*, vol. 36, pp. 287–300.
- [92] Tsekouras, G., Mamalis, A., Anagnostopoulos, C., Gavalas, D., Economou, D. (2008) ‘Improved batch fuzzy learning vector quantization for image compression’, *Information Sciences*, vol. 178, pp. 3895-3907.
- [93] Tsekouras, G.E., Sarimveis, H. (2004) ‘A new approach for measuring the validity of the fuzzy c-means algorithm’, *Advances in Engineering Software*, vol. 35, pp. 567–575.
- [94] Tsekouras, G., Sarimveis, H., Kavakli, E. and Bafas, B. (2005) ‘A hierarchical fuzzy-clustering approach to fuzzy modelling’, *Fuzzy Sets and Systems*, vol. 150, pp. 245-266.
- [95] Tsekouras, G. (2002) ‘Fuzzy Modeling Based on Ordinary Fuzzy Partitions and Nearest Neighbor Clustering’, *Journal of Intelligent and Robotic Systems*, vol. 43, pp. 255–282.
- [96] Uykan, Z., Güzelis, C., Çelebi, M.E. and Koivo, H.N. (2000) ‘Analysis of Input–Output Clustering for Determining Centers of RBFN’, *IEEE Transactions on Neural Networks*, vol. 11, no. 4, pp. 851-858.
- [97] Valls, J.M., Galvan, I.M. and Isasi, P. (2008) ‘Learning radial basis neural networks in a lazy way: A comparative study’, *Neurocomputing*, vol. 71, pp. 2529– 2537.
- [98] Wang, L. X. and Mendel, J. M. (1992) ‘Generating fuzzy rules from numerical data with applications’, *IEEE Transactions on Systems Man and Cybernetics*, vol. 22, no. 6, pp. 1414–1427.
- [99] Xie, L.X., Beni, G. (1991) ‘A validity measure for fuzzy clustering’, *IEEE Transactions Pattern Anal. Machine Intelligence (PAMI)*, vol 13, pp. 841–847.

- [100] Yeung, D.S., Chan, P.P.K., Ng, W.W.Y. (2009) ‘Radial Basis Function network learning using localized generalization error bound’, *Information Sciences*, vol. 179, no. 19, pp. 3199-3217.
- [101] Zadeh, L. A. (1965) ‘Fuzzy Sets’, *Information and Control*, pp.338-353.
- [102] Zahid, N., Abouelala, O., Limouri, M., Essaid, A. (1999) ‘Unsupervised fuzzy clustering’, *Pattern Recognition Letters*, vol. 20, pp. 123-129.
- [103] Zemouri, R., Racoceanu, D., Zerhouni, N. (2003) ‘Recurrent radial basis function network for time-series prediction’, *Engineering Applications of Artificial Intelligence*, vol. 16, pp. 453–463.
- [104] Zhao, Z.Q., Huang, D.S. (2007) ‘A mended hybrid learning algorithm for radial basis functions neural networks to improve generalization capability’, *Applied Mathematical Modelling*, vol. 31, pp. 1271–1281.
- [105] Zurada, J. M. (1992) *Introduction to Artificial Neural Systems*, St. Paul: West Publishing Company.
- [106] Βαζιργιάννης, Μ., Χαλκίδη, Μ. (2005) *Εξόρυξη από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό*, Αθήνα: Τυπωθήτω – Γιώργος Δαρδάνος.
- [107] Βλαχάβας, Ι., Βασιλειάδης, Ν., Κόκκορας, Φ., Σακελλαρίου, Η. (2006) *Τεχνητή Νοημοσύνη*, Αθήνα: Εκδόσεις Γκιούρδας.
- [108] Διαμαντάρας, Κ. (2007) *Τεχνητά Νευρωνικά Δίκτυα*, Αθήνα: Κλειδάριθμος.
- [109] Λύκας, Α. (2008) *Τεχνητά Νευρωνικά Δίκτυα – Εφαρμογές*, ΕΑΠ.