



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΠΟΛΙΤΙΣΜΙΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΣ



Πτυχιακή Εργασία του

Λεβέντη Γεώργιου

A.M: 1312007082

Θέμα :

« Ανάπτυξη Αλγορίθμου Βέλτιστης Ασαφούς Συσταδοποίησης για την Εκπαίδευση Νευρωνικών Δικτύων Ακτινικής Βάσης »

Τριμελής Επιτροπή

Τσεκούρας Γεώργιος, Επίκουρος Καθηγητής (Επιβλέπων)

Μπαντιμαρούδης Φιλήμων, Αναπληρωτής Καθηγητής

Παυλογεωργάτος Γεράσιμος, Επίκουρος Καθηγητής

Μυτιλήνη
Ιούλιος 2011

Πίνακας Περιεχομένων

1	ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	6
1.1	Εισαγωγή.....	6
1.2	Πως δομείται ο Νευρώνας	7
1.3	Το Μοντέλο McCulloch – Pitts	8
1.4	Λειτουργία ενός Νευρωνικού Δικτύου	9
1.5	Εκπαίδευση Νευρωνικών Δικτύων.....	10
1.5.1	Μάθηση με Επίβλεψη (<i>supervised learning</i>)	10
1.5.2	Ενισχυτική Μάθηση (<i>reinforcement learning</i>)	11
1.5.3	Μάθηση χωρίς Επίβλεψη (<i>unsupervised learning</i>)	11
2	ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΑΚΤΙΝΙΚΗΣ ΒΑΣΗΣ.....	13
2.1	Αλγόριθμος Συσταδοποίησης C – Means.....	15
2.2	Αλγόριθμος Ασαφούς Συσταδοποίησης Fuzzy C – Means.....	17
3	ΔΕΙΚΤΕΣ ΑΣΑΦΟΥΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ.....	20
3.1	Δείκτης Xie – Beni.....	20
3.2	Δείκτης SV	21
4	Ο ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ	24
4.1	Βήμα 1: Ασαφή Συσταδοποίηση Εισόδου – Εξόδου.....	26
4.2	Βήμα 2: Βέλτιστη ασαφής συσταδοποίηση	27
4.3	Βήμα 3: Δίκτυο εκτίμησης των παραμέτρων	29
5	Αριθμητικά παραδείγματα	30
5.1	Προσέγγιση δισδιάστατης συνάρτησης.....	30
5.2	Προσέγγιση στατικής δισδιάστατης συνάρτησης.....	32
5.3	Η Τρισδιάστατη Συνάρτηση	34
5.4	Abalone Δεδομένα	37
6	ΣΥΜΠΕΡΑΣΜΑΤΑ	40
	ΒΙΒΛΙΟΓΡΑΦΙΑ	41

Ευχαριστίες

Από τη θέση αυτή, θα ήθελα να εκφράσω τις ευχαριστίες στους γονείς μου και στην αδελφή μου για την υλική και ηθική υποστήριξη που μου παρείχαν σε όλα τα χρόνια των σπουδών μου.

Επιπλέον, ευχαριστώ τον κ. Τσεκούρα Γεώργιο, Επίκουρο Καθηγητή του τμήματος Πολιτισμικής Τεχνολογίας και Επικοινωνίας του Πανεπιστημίου Αιγαίου, καθώς επίσης και τον κ. Νείρο Αντώνιο, υποψήφιο διδάκτορα για τη βοήθεια και την καθοδήγηση που μου προσέφεραν κατά την εκπόνηση της παρούσας πτυχιακής εργασίας.

Περίληψη

Αυτή η πτυχιακή εργασία προτείνει μια νέα μέθοδο [29] που συνδυάζει ασαφή συσταδοποίηση και τη βέλτιστη ασαφή συσταδοποίηση για τον αποτελεσματικό σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης. Πρώτα, θα εφαρμόσουμε την fuzzy c-means στο χώρο εισόδου-εξόδου με σκοπό να γίνει η προ-επεξεργασία των διαθέσιμων στοιχείων. Οι προκύπτουσες ομάδες 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 σταθμισμένης (weighted) fuzzy c-means. 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.

Abstract

This project proposes a new method [29] that combines input-output fuzzy clustering and optimal fuzzy clustering for the efficient design of radial basis function neural networks. We first apply the fuzzy c -means in the product (i.e. input-output) space to pre-process the available data. The resulting clusters are projected on the input space. The corresponding cluster centers are considered as a new data set which is further clustered by means of optimal fuzzy clustering in terms of the weighted fuzzy c -means. To accomplish this task we develop a new cluster validity index, which is used to identify the appropriate number of RBF hidden nodes. The algorithm is successfully implemented in well-known data sets where its performance is tested and evaluated.

1 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

1.1 Εισαγωγή

Το αντικείμενο των τεχνητών νευρωνικών δικτύων γνωρίζει ραγδαία ανάπτυξη τα τελευταία 25 χρόνια και αποτελεί πλέον ένα ευρύ και αυτόνομο επιστημονικό πεδίο, που σχετίζεται με το γενικότερο πλαίσιο της τεχνητής νοημοσύνης και των ευφυών συστημάτων. Τα τεχνητά νευρωνικά δίκτυα είναι προγράμματα για υπολογιστές που προσομοιώνουν τη βιολογική οργάνωση και τη λειτουργία των βιολογικών νευρώνων. Βασικό τους πλεονέκτημα είναι η ευπλαστικότητα, όπως συμβαίνει με τα εγκεφαλικά μας κύτταρα έτσι τα τεχνητά νευρωνικά δίκτυα δε χρειάζεται να επαναπρογραμματιστούν αν αλλάξει το περιβάλλον. Επιπλέον μπορούν να "μαθαίνουν" από μόνα τους αυτό που πρέπει να υπολογίσουν, χάρη σε ειδικά προγράμματα που σταδιακά διορθώνουν τα λάθη τους καθώς μεταβάλλεται η κατάσταση.

Πιο συγκεκριμένα, Νευρωνικό δίκτυο ονομάζεται ένα κύκλωμα διασυνδεδεμένων νευρώνων. Αυτοί οι νευρώνες ονομάζονται *τεχνητοί νευρώνες*, οπότε στη συγκεκριμένη περίπτωση πρόκειται για ένα αφηρημένο αλγοριθμικό κατασκευάσμα, το οποίο εμπίπτει στον τομέα της υπολογιστικής νοημοσύνης, όταν στόχος του νευρωνικού δικτύου είναι η επίλυση κάποιου υπολογιστικού προβλήματος, ή της υπολογιστικής νευροεπιστήμης.

Κίνητρο για τη μελέτη του νευρώνα και των νευρωνικών δικτύων είναι η ελπίδα ανακάλυψης ενός νέου υπολογιστικού μοντέλου βασισμένου σε δικτυακή δομή παρόμοια με αυτή του εγκεφάλου . Αυτή η καινούργια υπολογιστική πλατφόρμα, η οποία είναι γνωστή και ως Connectionist Model θα είναι πιο κατάλληλη για ανάπτυξη ευφυών αλγορίθμων και διαδικασιών σχετιζόμενων με τη μάθηση, με τη νοημοσύνη, με τη μνήμη κ.λ.π . [1].

Ιστορικά, η μελέτη των τεχνητών νευρωνικών δικτύων ξεκίνησε το 1943 από τον ψυχίατρο Warren McCulloch και τον μαθηματικό Walter Pitts. Στο άρθρο τους «A logical calculus and the ideas immanent in the nervous activity», περιγράφεται ένας λογικός υπολογισμός των νευρωνικών δικτύων που συνέδεσε τις μελέτες της νευροφυσιολογίας και της μαθηματικής λογικής. Το αρχικό τους μοντέλο νευρώνα είχε την υπόθεση ότι ακολουθούσε ένα νόμο της μορφής «όλα ή τίποτα» (all-or-none). Με έναν ικανοποιητικό αριθμό απλών μονάδων και συναπτικών βαρών, τα οποία είχαν τοποθετηθεί κατάλληλα και λειτουργούσαν συγχρονισμένα, οι McCulloch και Pitts απέδειξαν ότι ένα δίκτυο τόσο

συγκροτημένο θα μπορούσε να υπολογίσει οποιαδήποτε υπολογίσιμη συνάρτηση.

Άρα όπως διαπιστώνουμε, καθ' αυτό τον τρόπο «γεννήθηκαν» οι αρχές των νευρωνικών δικτύων και της τεχνητής νοημοσύνης .

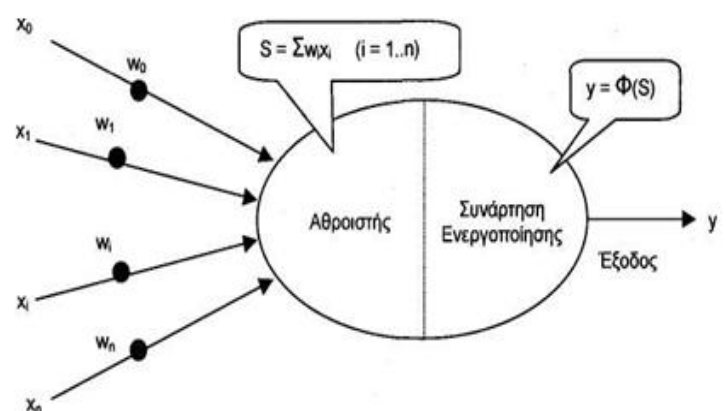
1.2 Πως δομείται ο Νευρώνας

Στη προηγούμενη ενότητα δώσαμε τον ορισμό του Νευρωνικού Δικτύου, το οποίο όπως είπαμε είναι ένα κύκλωμα, που αποτελείται από διασυνδεδεμένους – μεταξύ τους - νευρώνες. Πιο συγκεκριμένα σ' αυτή την ενότητα θα προσπαθήσουμε να αναλύσουμε τη δομή ενός νευρώνα, προκειμένου να καθιστούμε ικανοί να αντιληφθούμε τη σημασία που έχουν για ένα νευρωνικό δίκτυο.

Ο Τεχνητός Νευρώνας (artificial neuron) είναι ένα υπολογιστικό μοντέλο τα μέρη του οποίου σχετίζονται άμεσα με αυτά του βιολογικού νευρώνα. Όπως απεικονίζεται στην Εικόνα 1.1 , ένας τεχνητός νευρώνας δέχεται κάποια σήματα εισόδου $x_1, x_2, x_3, \dots, x_n$ τα οποία σε αντίθεση με τους ηλεκτρικούς παλμούς του εγκεφάλου, αντιστοιχούν σε συνεχείς μεταβλητές.

Σε αντιστοιχία με το βιολογικό νευρώνα, οι συνάψεις είναι τα σημεία ένωσης μεταξύ των διακλαδώσεων του άξονα (πύλη εξόδου) ενός νευρώνα και των δεντριτών (πύλες εισόδου) από άλλους νευρώνες. Αντίθετα, όμως, με μια σύναψη στον εγκέφαλο, το συναπτικό βάρος ενός τεχνητού νευρώνα ανήκει σε ένα εύρος που περιλαμβάνει τόσο αρνητικές όσο και θετικές τιμές [1, 3].

Το σώμα του τεχνητού νευρώνα χωρίζεται σε δύο μέρη, τον αθροιστή (adder) και τη συνάρτηση ενεργοποίησης ή κατωφλίου (activation ή threshold function). Ο αθροιστής προσθέτει τα σήματα εισόδου που είναι τοποθετημένα από τις αντίστοιχες συνάψεις του νευρώνα.



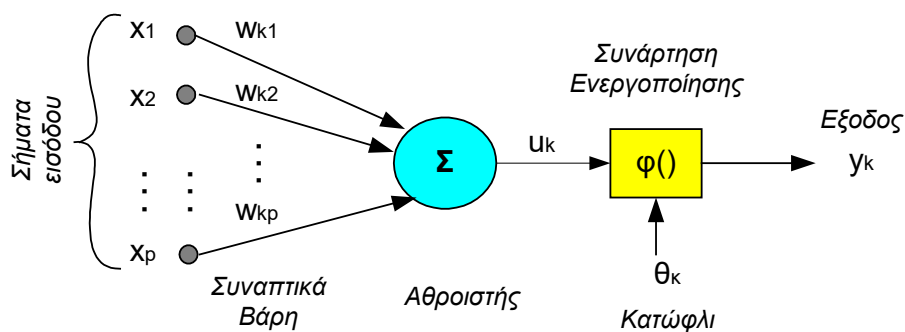
Εικόνα 1.1: Δομή ενός Τεχνητού Νευρώνα

Οι λειτουργίες που περιγράφονται στο τμήμα του αθροιστή συγκροτούν ένα γραμμικό συνδυασμό και παράγουν τη ποσότητα S . Επίσης, η συνάρτηση ενεργοποίησης ή κατωφλίου είναι ένα μη γραμμικό φίλτρο, το οποίο διαμορφώνει τη τελική τιμή του σήματος εξόδου y , σε συνάρτηση της ποσότητας S .

1.3 Το Μοντέλο McCulloch – Pitts

Όπως προείπαμε η μελέτη των τεχνητών νευρωνικών δικτύων ξεκίνησε το 1943 από τον ψυχίατρο Warren McCulloch και τον μαθηματικό Walter Pitts, οι οποίοι περιέγραψαν ένα απλό μοντέλο της δραστηριότητας του νευρώνα [2].

Η εργασία αυτή θεωρείται ο πρόδρομος των νευρωνικών δικτύων. Οι συγγραφείς απέδειξαν ότι τα νευρωνικά δίκτυα μπορούσαν να χρησιμοποιηθούν στην προσέγγιση αριθμητικών ή λογικών συναρτήσεων προτείνοντας παράλληλα την πρώτη δομή τεχνητού νευρώνα. Σύμφωνα με την πρότασή τους, η έξοδος του νευρώνα έχει τιμή ένα, εάν το σταθμισμένο άθροισμα των εισόδων είναι μεγαλύτερο από μια κατώτατη τιμή, διαφορετικά έχει τιμή μηδέν.



Εικόνα 1.2: Το μοντέλο των McCulloch και Pitts για το νευρώνα

1.4 Λειτουργία ενός Νευρωνικού Δικτύου

Όπως προαναφέραμε, το νευρωνικό δίκτυο είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους (νευρώνες, νευρώνια), διασυνδεδεμένους μεταξύ τους. Είναι εμπνευσμένο από το Κεντρικό Νευρικό Σύστημα (ΚΝΣ), το οποίο προσπαθεί να προσομοιώσει.

Οι νευρώνες είναι τα δομικά στοιχεία του δικτύου. Κάθε τέτοιος κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές (είτε από άλλους νευρώνες, είτε από το περιβάλλον), επιτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο. Η εν λόγω έξοδος είτε κατευθύνεται στο περιβάλλον, είτε τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου. Υπάρχουν τρεις τύποι νευρώνων: οι νευρώνες εισόδου, οι νευρώνες εξόδου και οι υπολογιστικοί νευρώνες ή κρυμμένοι νευρώνες. Οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

Εάν x_{ki} είναι η i -οστή είσοδος του k νευρώνα, w_{ki} : το i -οστό συναπτικό βάρος του k νευρώνα και $\phi(\cdot)$ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του k νευρώνα δίνεται από την εξίσωση:

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

Στον k -οστό νευρώνα υπάρχει ένα συναπτικό βάρος w_{k0} με ιδιαίτερη σημασία, το οποίο καλείται πόλωση ή κατώφλι (bias, threshold). Η τιμή της εισόδου του είναι πάντα η μονάδα, $x_{k0} = 1$. Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε ο νευρώνας παραμένει ανενεργός. Η ιδέα προέκυψε από τα βιολογικά νευρικά κύτταρα.

Όπως είναι φανερό, οι αριθμοί οι οποίοι συναποτελούν το διάνυσμα εισόδου (κάθε στοιχείο του διανύσματος τροφοδοτείται κατά τη

λειτουργία του δικτύου σε έναν νευρώνα εισόδου), αλλά και οι αριθμοί οι οποίοι συναποτελούν το διάνυσμα εξόδου (κάθε στοιχείο του οποίου εμφανίζεται, μετά το πέρας του ολικού υπολογισμού, σε έναν νευρώνα εξόδου), περιγράφουν χαρακτηριστικά του προς επίλυση προβλήματος. Συνήθως αυτό που μας ενδιαφέρει είναι το δίκτυο να απεικονίζει με ορθό τρόπο διανύσματα εισόδου σε κατάλληλα διανύσματα εξόδου, το πρόβλημα δηλαδή είναι η υλοποίηση μίας συνάρτησης πολλαπλών μεταβλητών, κατά κανόνα περίπλοκης και με άγνωστο ακριβή τύπο. Τέτοιες απεικονίσεις έχουν εφαρμογή σε ποικιλία τομέων της επιστήμης και της τεχνολογίας, αφού λειτουργούν ως αριθμητικά μοντέλα για πολλά διαφορετικά ζητήματα. Το ίδιο δίκτυο μπορεί να υλοποιήσει άπειρες διαφορετικές απεικονίσεις, μία για κάθε διαφορετική επιλογή συνόλου συναπτικών βαρών.

1.5 Εκπαίδευση Νευρωνικών Δικτύων

Μια από τις πιο βασικές ιδιότητες των Νευρωνικών Δικτύων είναι η ικανότητά τους για εκπαίδευση. Η εκπαίδευση αυτή επιτυγχάνεται μέσω της ανταλλαγής τιμών και βαρών, που αποσκοπεί στη βαθμιαία σύλληψη της πληροφορίας, η οποία στη συνέχεια θα είναι διαθέσιμη προς ανάκτηση. Υπάρχουν, βέβαια, πολλοί αλγόριθμοι που η εφαρμογή τους έχει στόχο την προσαρμογή των τιμών των βαρών ενός Τεχνητού Νευρωνικού Δικτύου. Όλες οι μέθοδοι μάθησης μπορούν να καταταχθούν σε τρεις κατηγορίες : τη Μάθηση με Επίβλεψη (supervised learning), την Ενισχυτική Μάθηση (reinforcement learning) και τη Μάθηση χωρίς Επίβλεψη (unsupervised learning) [4].

1.5.1 Μάθηση με Επίβλεψη (supervised learning)

Στην επιβλεπόμενη μάθηση υποθέτουμε την παρουσία ενός δασκάλου κατά τη διάρκεια της διαδικασίας της εκπαίδευσης. Έτσι κάθε δείγμα που χρησιμοποιείται για την εκπαίδευση του δικτύου (δείγμα εκπαίδευσης - training pattern) αποτελείται από ένα δείγμα εισόδου (input pattern) και ένα δείγμα επιθυμητής εξόδου (desired or target output pattern). Κατά τη διάρκεια της εκπαίδευσης γίνεται σύγκριση μεταξύ της εξόδου που υπολογίζει το δίκτυο (δίνοντας του σαν είσοδο το δείγμα εισόδου) και της επιθυμητής εξόδου (όπως αυτή καθορίζεται από το δείγμα επιθυμητής εξόδου) προκειμένου να καθοριστεί το λάθος. Το λάθος στη συνέχεια χρησιμοποιείται για να μεταβληθούν οι ελεύθερες

παράμετροι του δικτύου (δηλαδή τα βάρη και τα κατώφλια) έτσι ώστε να βελτιωθεί η απόδοσή του. Ο πίνακας των ελεύθερων παραμέτρων του δικτύου αρχικοποιείται είτε σε μηδέν είτε σε μικρές τυχαίες τιμές. Στη συνέχεια παρουσιάζουμε στο δίκτυο τα διανύσματα εκπαίδευσης x^p , $p=1, 2, \dots$, ένα κάθε φορά και υπολογίζουμε την έξοδο y^p του δικτύου. Το διάνυσμα εξόδου που υπολογίζει το δίκτυο συγκρίνεται με το διάνυσμα επιθυμητής εξόδου t^p και καθορίζεται το λάθος του δικτύου $e^p = y^p - t^p$. Η μέση τιμή του τετραγώνου των λαθών που παράγονται κατά την εκπαίδευση ενός νευρωνικού δικτύου μετά το πέρασμα όλων των παραδειγμάτων εκπαίδευσης είναι το Μέσο Τετραγωνικό Σφάλμα (Mean Square Error - MSE). Στη συνέχεια το MSE χρησιμοποιείται, μέσω κάποιων υπολογισμών και ανάδρασης, για τη διόρθωση της τιμής των ελεύθερων παραμέτρων του δικτύου προκειμένου να μειωθεί το λάθος για κάθε δείγμα εκπαίδευσης. Η διαδικασία της διόρθωσης της τιμής των ελεύθερων παραμέτρων επαναλαμβάνεται έως ότου οι τιμές του συγκλίνουν σε ένα σύνολο τιμών τέτοιο ώστε το δίκτυο να είναι ικανό να επιτελέσει το έργο για το οποίο εκπαιδεύεται.

1.5.2 Ενισχυτική Μάθηση (reinforcement learning)

Στην ενισχυτική μάθηση υποθέτουμε και πάλι την ύπαρξη δασκάλου, αλλά σε αντίθεση με την επιβλεπόμενη μάθηση, δεν παρέχουμε στο δίκτυο τη σωστή απάντηση (δηλαδή το επιθυμητό διάνυσμα εξόδου). Αντί αυτού, δίνουμε μια ένδειξη για το αν η έξοδος που υπολογίζεται είναι σωστή ή λάθος. Το δίκτυο θα πρέπει, στη συνέχεια, να χρησιμοποιήσει αυτήν την πληροφορία, που του παρέχει ο δάσκαλος, προκειμένου να βελτιώσει την απόδοσή του. Τυπικά, δίνεται μια ανταμοιβή, ενισχύοντας τα βάρη των κόμβων που δίνουν σωστή απάντηση και μια τιμωρία ελαττώνοντας τις τιμές των βαρών σε αυτούς τους κόμβους που δίνουν λάθος απάντηση.

1.5.3 Μάθηση χωρίς Επίβλεψη (unsupervised learning)

Στην μη-επιβλεπόμενη μάθηση δεν υπάρχει κάποιος δάσκαλος για να εφοδιάσει το δίκτυο με την σωστή απάντηση, δηλαδή τα δείγματα εκπαίδευσης αποτελούνται μόνο από τα δείγματα εισόδου και δεν περιέχουν δείγματα επιθυμητής εξόδου. Στην περίπτωση αυτή, λοιπόν, το σύστημα πρέπει να μάθει ανακαλύπτοντας και προσαρμόζοντας τον εαυτό του σε κάποια δομικά χαρακτηριστικά των διανυσμάτων εισόδου, αυτό γίνεται ανακαλύπτοντας κάποιες στατιστικές κανονικότητες και ομαδοποιήσεις των δειγμάτων εισόδου. Ένα τέτοιο είδος μάθησης επιτυγχάνεται με την ενίσχυση επιλεγμένων βαρών προκειμένου το

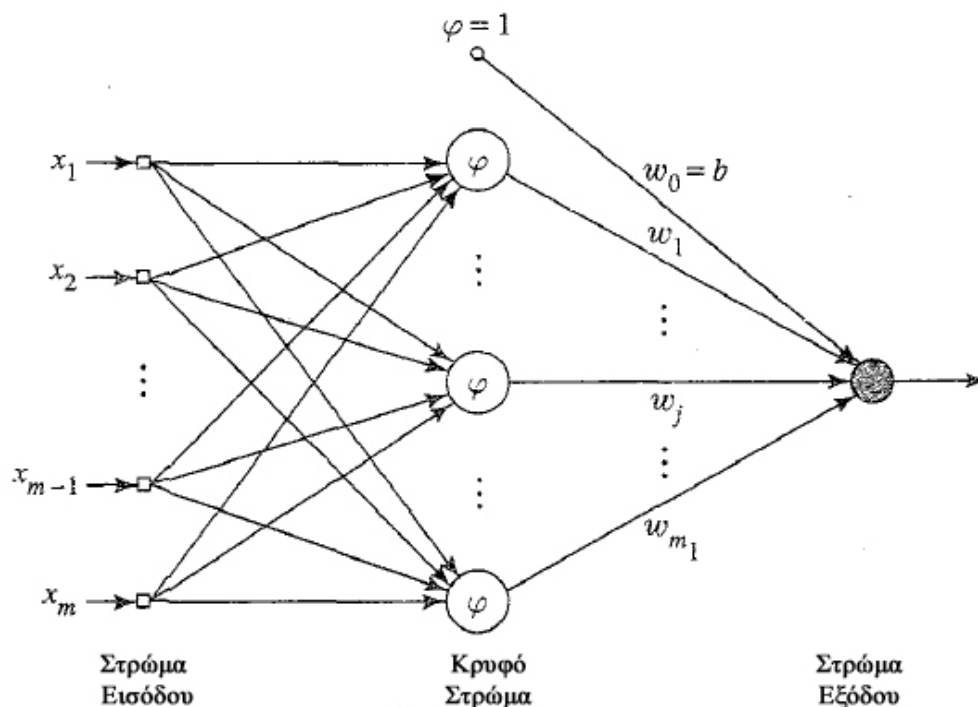
διάνυσμα εξόδου να ταιριάζει σε κεντρικά πρωτότυπα δείγματα εκπαίδευσης που είναι αντιπροσωπευτικά ενός συνόλου από παρόμοια δείγματα.

Από τις τρεις γενικές μεθοδολογίες εκπαίδευσης που προαναφέρθηκαν αυτές που χρησιμοποιούνται περισσότερο είναι η επιβλεπόμενη και η μη-επιβλεπόμενη μάθηση ενώ η ενισχυτική μάθηση χρησιμοποιείται σπάνια.

2 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΑΚΤΙΝΙΚΗΣ ΒΑΣΗΣ

Τα νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (radial basis function neural networks ή RBF NN) αποτελούν μία συγκεκριμένη αρχιτεκτονική που ανήκει στα δίκτυα πρόσθιας τροφοδότησης και παρουσιάζει κάποια πλεονεκτήματα έναντι των άλλων νευρωνικών δικτύων.

Η δομή του νευρωνικού δικτύου ακτινικής συνάρτησης βάσης παρουσιάζεται στην Εικόνα 2.1 Το δίκτυο αποτελείται από τρεις στοιβάδες: τη στοιβάδα εισόδου, μία κρυφή στοιβάδα και μία στοιβάδα εξόδου. Η στοιβάδα εισόδου διανέμει τα δεδομένα εισόδου στους νευρώνες της κρυφής στοιβάδας ως έχουν, δηλαδή χωρίς αυτά να πολλαπλασιάζονται με κάποια συναπτικά βάρη. Σημειώνεται ότι το δίκτυο RBF είναι πλήρως συνδεδεμένο. Σε κάθε νευρώνα της κρυφής στοιβάδας αντιστοιχεί ένα κέντρο διάστασης ίσης με τον αριθμό μεταβλητών εισόδου, ενώ μια ακτινική συνάρτηση χρησιμοποιείται για το μη γραμμικό μετασχηματισμό των δεδομένων εισόδου.



Εικόνα 2.1: Τυπική Αρχιτεκτονική ενός δικτύου ακτινικής βάσης (RBF Network)

Ο σχεδιασμός των νευρωνικών δικτύων ακτινικής βάσης συνάρτησης (RBF) βασίζεται στην εκτίμηση του αριθμού των κρυφών κόμβων και στις αντίστοιχες παραμέτρους, όπως τα κέντρα, τα πλάτη και τα βάρη διασύνδεσης. Το σύνολο δεδομένων εισόδου-εξόδου (input-output data), είναι $\{(\mathbf{x}_k; y_k)\}_{k=1}^N$, όπου N είναι ο αριθμός των δειγμάτων εκπαίδευσης, όπου \mathbf{x}_k είναι το k -οστό διάνυσμα εισόδου και όπου y_k το k -οστό διάνυσμα εξόδου. Επιλέγουμε συναρτήσεις - βάσης (basis functions) Γκαουσιανού (Gaussian) τύπου.

$$g_i(\mathbf{x}_k) = \exp\left(-\frac{\|\mathbf{x}_k - \mathbf{v}_i\|^2}{\sigma_i^2}\right) \quad (2.1)$$

Όπου $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c \in \mathcal{R}^p$ είναι τα κέντρα των συναρτήσεων-βάσης και τα αντίστοιχα πλάτη (widths). Η εκτιμώμενη έξοδος του δικτύου χρησιμοποιεί την ακόλουθη γραμμική συνάρτηση παλινδρόμησης (linear regression functional),

$$f(\mathbf{x}_k) = \tilde{y}_k = \sum_{i=1}^c w_i g_i(\mathbf{x}_k) \quad (2.2)$$

Όπου w_i είναι το βάρος σύνδεσης / συνδετικό βάρος (connection weight) του i -οστού κρυφού κόμβου (hidden node). Επίσης, η διαδικασία εκπαίδευσης ενός RBF δικτύου έχει ως στόχο να αποκτήσουμε ένα σύνολο τιμών για τις παραμέτρους \mathbf{v}_i, σ_i , καθώς επίσης και για τα συνδετικά βάρη w_i . Κατ'αρχήν, οι τιμές αυτών των παραμέτρων αντανακλούν τις χωρικές ιδιότητες των δεδομένων εκπαίδευσης, καθώς επίσης και τις σχέσεις εισόδου-εξόδου (input-output relations).

Δεν αναπτυσσόμαστε σε περαιτέρω λεπτομέρειες σχετικά με τη βελτιστοποίηση, αλλά εμείς επικεντρωνόμαστε κατά κύριο λόγο στην εξαγωγή ενός συνόλου τιμών για το \mathbf{v}_i , δεδομένου ότι τα πλάτη και τα συνδετικά βάρη συνήθως υπολογίζονται με πρότυπες μεθόδους. Στην υπάρχουσα βιβλιογραφία μπορεί κανείς να συναντήσει τη χρήση της συσταδοποίησης για να προσδιοριστούν τα στοιχεία του κέντρου (center elements) [8, 9, 10, 11]. Δύο από τις πιο ευρέως χρησιμοποιούμενες τεχνικές συσταδοποίησης είναι η c -means [12, 13, 14] και ο αλγόριθμος fuzzy c -means [15, 16]. Όσον αφορά το σχεδιασμό των δικτύων RBF, η ανάλυση διασποράς (cluster analysis) μπορεί να γίνει με δύο

διαφορετικούς τρόπους . Πιο συγκεκριμένα ο πρώτος αφορά την συσταδοποίηση του χώρου εισόδου (input space) και ο δεύτερος την συσταδοποίηση του χώρου εισόδου-εξόδου (input-output space).

Η πρώτη προσέγγιση αποτελεί μια γρήγορη και προφανής διαδικασία, ενώ η δεύτερη φαίνεται να είναι πιο αποτελεσματική, για να λάβει υπόψη πληροφορίες σχετικά με τη δομή των δεδομένων στο χώρο εξόδου, κάτι που είναι μια κρίσιμη προϋπόθεση σε συστήματα που βασίζονται στη παλινδρόμηση, όπως τα δίκτυα RBF. Ως εκ τούτου, οι τοποθεσίες των κέντρων επηρεάζονται όχι μόνο από την κατανομή των εισόδων αλλά και από τις αποκλίσεις του δείγματος εξόδου. Η ιδέα της συνένωσης της εξόδου προς το διάνυσμα εισόδου κατά τη διαδικασία της συσταδοποίησης έχει, ανεξάρτητα, εξεταστεί από διάφορους συγγραφείς [17, 18, 19].

Έτσι η εκπαίδευση ενός RBF δικτύου περιλαμβάνει δύο στάδια. Στο πρώτο που γίνεται χωρίς επίβλεψη καθορίζονται το κέντρο και το εύρος της συνάρτησης ενεργοποίησης για κάθε κόμβο του κρυφού στρώματος. Στο δεύτερο που γίνεται με επίβλεψη καθορίζονται τα βάρη των συνδέσεων μεταξύ κρυφού και στρώματος εξόδου.

2.1 Αλγόριθμος Συσταδοποίησης C – Means

Για το πρώτο στάδιο της χωρίς επίβλεψης εκπαίδευσης έχουν προταθεί διάφορες διαδικασίες εκπαίδευσης. Μια από τις πιο δημοφιλείς και αυτή που χρησιμοποιούμε στη συνέχεια είναι η εκπαίδευση με τη χρήση του αλγορίθμου C - Means.

Ο αλγόριθμος C – Means είναι ένας διαχωριστικός αλγόριθμος , ο οποίος είναι βασισμένος σε ένα πρότυπο. Πιο συγκεκριμένα, κάθε συστάδα συσχετίζεται με ένα κεντρικό σημείο (centroid) , αναλυτικότερα, κάθε σημείο ανατίθεται στη συστάδα με το κοντινότερο κεντρικό σημείο. Αξιοσημείωτο είναι το γεγονός ότι ο αριθμός των ομάδων, C, είναι η είσοδος στον αλγόριθμο συσταδοποίησης.

Η βασική δομή του αλγορίθμου συσταδοποίησης C – Means παρουσιάζεται αναλυτικότερα στα παρακάτω βήματα :

- 1: Επιλογή C σημείων ως τα αρχικά κεντρικά σημεία.
- 2: **Repeat.**
- 3: Ανάθεση όλων των αρχικών σημείων στο κοντινότερο τους από τα C κεντρικά σημεία.

- 4: Επανα-υπολογισμός του κεντρικού σημείου κάθε συστάδας.
- 5: **Until** τα κεντρικά σημεία να μην αλλάζουν.

Με άλλα λόγια, αξίζει να αναφέρουμε κάποιες παρατηρήσεις σχετικά με τη δομή, καθώς επίσης και με τη λειτουργία του αλγορίθμου συσταδοποίησης C-Means. Αυτές είναι οι εξής :

- 1) Τα αρχικά κεντρικά σημεία συνήθως επιλέγονται τυχαία. Επίσης, οι συστάδες που παράγονται διαφέρουν από το ένα τρέξιμο του αλγορίθμου στο άλλο.
- 2) Η εγγύτητα των σημείων υπολογίζεται με βάση κάποια απόσταση που εξαρτάται από το είδος των σημείων, στα παραδείγματα θα θεωρήσουμε την Ευκλείδεια απόσταση. Επειδή η απόσταση υπολογίζεται συχνά πρέπει να είναι σχετικά απλή.
- 3) Το κεντρικό σημείο είναι (συνήθως) το μέσο (mean) των σημείων της συστάδας (το οποίο μπορεί να μην είναι ένα από τα δεδομένα εισόδου).

Πώς όμως υπολογίζουμε τα κέντρα των συστάδων ;

Η ζητούμενη απάντηση δίνεται από τις παρακάτω μαθηματικές πράξεις και σχέσεις :

$$J = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|\bar{x}_k - \bar{v}_i\|^2 \quad (2.3)$$

$$\begin{aligned} \frac{\partial J}{\partial \bar{v}_i} = 0 &\Rightarrow \sum_{k=1}^n u_{ik} (-2)(\bar{x}_k - \bar{v}_i) = 0 \Rightarrow \\ \sum_{k=1}^n u_{ik} \bar{x}_k - \sum_{k=1}^n u_{ik} \bar{v}_i &= 0 \Rightarrow \\ \bar{v}_i &= \frac{\sum_{k=1}^n u_{ik} \bar{x}_k}{\sum_{k=1}^n u_{ik}} \end{aligned} \quad (2.4)$$

Στις προαναφερθείσες σχέσεις εμφανίζεται μια συνάρτηση, η οποία συμβολίζεται με u_{ik} , η συγκεκριμένη συνάρτηση ονομάζεται *Συνάρτηση Συμμετοχής* (Membership Function) και υπολογίζεται βάσει της παρακάτω σχέσης :

$$u_{ik} = \begin{cases} 1, & \text{αν } \|\bar{x}_k - \bar{v}_i\| = \min_{1 \leq j \leq c} \{\|\bar{x}_k - \bar{v}_j\|\} \\ 0, & \text{αλλιώς} \end{cases} \quad (2.5)$$

2.2 Αλγόριθμος Ασαφούς Συσταδοποίησης Fuzzy C – Means

Ο Fuzzy c-means (FCM) είναι ένας αλγόριθμος βάσει του οποίου γίνεται συσταδοποίηση (clustering), το οποίο επιτρέπει ένα διάνυσμα εισόδου να ανήκει σε μία ή σε περισσότερες της μίας συστάδες (clusters). Πρέπει να σημειωθεί ότι άλλοι αλγόριθμοι (πχ C-means) επιτρέπουν σε ένα διάνυσμα εισόδου να ανήκει σε ένα και μόνο ένα cluster.

Ο FCM προτάθηκε από τον Dunn το 1973, βελτιώθηκε στην τελική του μορφή από τον Bezdek το 1981 και χρησιμοποιείται κυρίως για εφαρμογές αναγνώρισης προτύπων. Η υλοποίηση του βασίζεται στην ελαχιστοποίηση της παρακάτω συνάρτησης κόστους:

$$J_m = \sum_{k=1}^N \sum_{i=1}^n (u_{ik})^m \| \mathbf{x}_k - \mathbf{v}_i \|^2 \quad (2.6)$$

υπό τον περιορισμό

$$\sum_{i=1}^n u_{ik} = 1 \quad (2.7)$$

όπου

i : αναπαριστά το κάθε μια συστάδα (cluster) (συνολικά n)

k : αναπαριστά το κάθε ένα διάνυσμα εισόδου (συνολικά N)

m : ένας οποιοσδήποτε πραγματικός αριθμός μεγαλύτερος του 1

u_{ik} : ο βαθμός συμμετοχής του x_k στο cluster i (οι τιμές που παίρνει είναι μεταξύ 0 – δεν ανήκει καθόλου σε αυτό το cluster – έως 1 – ανήκει μόνο σε αυτό το cluster)

x_k : είναι το k -οστό διάνυσμα εισόδου (έστω διάστασης d)

c_i : είναι το κέντρο του cluster i (είναι και αυτό διάστασης d)

$\|*\|$: οποιαδήποτε νόρμα (ευκλείδεια απόσταση) εκφράζει την συνάφεια μεταξύ των διανυσμάτων εισόδου και κέντρου του cluster.

Η ασαφής κατάτμηση (fuzzy partitioning) υλοποιείται μέσω της επαναληπτικής βελτιστοποίησης της συνάρτησης κόστους (2.6) ενώ σε κάθε επανάληψη αλλάζουν οι τιμές των βαθμών συμμετοχής (u_{ik}) αλλά και των κέντρων των συστάδων (v_i), οι οποίες δίνονται από τις παρακάτω σχέσεις :

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|_A}{\|x_k - v_j\|_A} \right)^{\frac{2}{m-1}}} \quad (2.8)$$

$$v_i = \frac{\sum_{k=1}^N u_{ik}^m x_k}{\sum_{k=1}^N u_{ik}^m} \quad (2.9)$$

Η επαναληπτική διαδικασία θα διακοπεί όταν

$$\max(|u_{ik}^{(w+1)} - u_{ik}^{(w)}|) < \varepsilon \quad (2.10)$$

(όπου ε είναι μια μεταβλητή που αντιπροσωπεύει το κριτήριο τερματισμού και παίρνει τιμές μεταξύ 0 και 1 και w είναι ο αριθμός των επαναλήψεων) ή όταν ο αλγόριθμος εκτελέσει τον μέγιστο αριθμό επαναλήψεων (epochs).

Η διαδικασία που περιγράφηκε παραπάνω συγκλίνει σε ένα τοπικό ελάχιστο (που αποκαλείται και saddle point) της τιμής του J_m (βλέπε 2.6).

Η αλγοριθμική αναπαράσταση του FCM παρουσιάζεται παρακάτω:

1. Στο πρώτο βήμα γίνεται η αρχικοποίηση του πίνακα που περιέχει τους βαθμούς συμμετοχής των διανυσμάτων εισόδου στα clusters. Δίνονται δηλαδή τιμές στον πίνακα $U = [u_{ik}]$ (τιμές στον πίνακα U για την επανάληψη 0, δηλαδή $U^{(0)}$).
2. Για την w επανάληψη γίνεται ο υπολογισμός των κέντρων των συστάδων $Y^{(w)} = [v_i]$ βάση των τιμών του πίνακα βαθμών συμμετοχής $U^{(w)}$. Ο υπολογισμός αυτός γίνεται από την 2.9.
3. Γίνεται ενημέρωση των πινάκων $U^{(w)}$ και $U^{(w+1)}$ βάση της 2.8.
4. Αν ικανοποιείται η 2.10 (αν δηλαδή το ε είναι μικρότερο από το ορισθέν κριτήριο τερματισμού) τότε ο αλγόριθμος σταματάει. Διαφορετικά επιστρέφει στο βήμα 2 για να κάνει καινούργια επανάληψη.

Σε αυτό το σημείο θα πρέπει να διευκρινιστούν κάποια θέματα:

- ✓ Όπως έχει ήδη αναφερθεί παραπάνω, το κάθε διάνυσμα εισόδου ανήκει σε μια συστάδα με την έννοια ότι έχει ένα αντίστοιχο βαθμό συμμετοχής, που είναι ο αντίστοιχος δείκτης του πίνακα συμμετοχής των διανυσμάτων εισόδου. Όπως θα περίμενε κανείς τα αθροίσματα των γραμμών αυτού του πίνακα είναι ίσα με 1.
- ✓ Ο αλγόριθμος αυτός (FCM) δεν βρίσκει τον αριθμό των συστάδων που υπάρχουν σε ένα σετ διανυσμάτων εισόδου. Ο αριθμός αυτός είναι ένα από τα δεδομένα του αλγορίθμου.

3 ΔΕΙΚΤΕΣ ΑΣΑΦΟΥΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

3.1 Δείκτης Xie – Beni

Ο δείκτης Xie-Beni, για το σταθμισμένο (weighted) αλγόριθμο Ασαφών c-Μέσων (Fuzzy C - Means) υπολογίζονται από την ακόλουθη εξίσωση:

$$S_{XB} = \frac{\sum_{k=1}^n \sum_{i=1}^c \rho_k(u_{ik})^p \|a^k - u_i\|^2}{n \min_{i \neq j} \left\{ \|u_i - u_j\|^2 \right\}} \quad (3.1)$$

Η ανωτέρω εξίσωση συνδυάζει το ακόλουθο μέτρο πυκνότητας:

$$\pi = \frac{\sum_{k=1}^n \sum_{i=1}^c \rho_k(u_{ik})^p \|a^k - u_i\|^2}{n} \quad (3.2)$$

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

$$s = \min_{i \neq j} \left\{ \|u_i - u_j\|^2 \right\} \quad (3.3)$$

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

3.2 Δείκτης SV

Η βέλτιστη ασαφής συσταδοποίηση αφορά τον προσδιορισμό του (βέλτιστου) αριθμού των συστάδων, που παρέχουν μια συγκεχυμένη κατάτμηση με τις παρακάτω ιδιότητες:

- (α) στοιχεία που ανήκουν στην ίδια συστάδα είναι όσο το δυνατόν όμοια, και
- (β) τα δεδομένα που ανήκουν σε διαφορετικές συστάδες είναι όσο το δυνατόν ανόμοια γίνεται.

Παρά το γεγονός ότι η ασαφής (fuzzy) C-Means μπορεί να εντοπίσει ομοιότητες μέσα σε σύνολο δεδομένων, δεν μπορεί να εμφανίσει τη βέλτιστη ασαφή συσταδοποίηση, δεδομένου ότι προϋποθέτει μια εκ-των προτέρων γνώση του αριθμού των συστάδων. Σε γενικές γραμμές, η βέλτιστη ασαφή συσταδοποίηση είναι ένα πρόβλημα εύρεσης βέλτιστου αριθμού συστάδων και όσον αφορά την fuzzy c-means, η μέθοδος αυτή αναφέρεται στην ανάπτυξη μιας αξιόπιστης συνάρτησης (δείκτης) που θα εφαρμόζεται στα αποτελέσματα του προβλήματος βελτιστοποίησης, και θα οδηγήσει στις καλύτερες δυνατές λύσεις συσταδοποίησης. Ο δείκτης εγκυρότητας της ασαφούς c-means, που προτείνεται στην παρούσα ενότητα χρησιμοποιεί πολλές ιδέες που παρουσιάζονται στην βιβλιογραφία [5,6,7], και αναλύεται στις επόμενες παραγράφους.

Η διακύμανση (σ_i) και η ασαφής πληθικότητα συνόλου (fuzzy cardinality) (n_i) της i -οστής στοιβάδας αντίστοιχα δίνονται [5,6],

$$\sigma_i = \sum_{k=1}^N (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2, \quad 1 \leq i \leq c \quad (3.4)$$

$$n_i = \sum_{k=1}^N u_{ik}, \quad 1 \leq i \leq c \quad (3.5)$$

Οι Bensaid et al. [6] όρισαν την πυκνότητα (π_i) της i -οστής στοιβάδας ως το λόγο μεταξύ της διακύμανσης και της ασαφούς πληθικότητας συνόλου,

$$\pi_i = \frac{\sigma_i}{n_i}, \quad 1 \leq i \leq c \quad (3.6)$$

Η συνολική πυκνότητα (π) των ασαφών c -καταμήσεων (c -partition) δίνεται ως το άθροισμα των π_i όλων των συστάδων [7],

$$\pi = \sum \frac{\sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|_A^2}{n_i} \quad (3.7)$$

Προκειμένου να καθοριστεί ο ασαφής διαχωρισμός των ασαφών c -καταμήσεων, κάποια σημείωση (notation) εισάγεται. Κατά πρώτο λόγο, εμείς χρησιμοποιούμε το σύνολο των διανυσμάτων,

$$[z_1, z_2, \dots, z_c, z_{c+1}]^T = [v_1, v_2, \dots, v_c, \bar{X}] \quad (3.8)$$

όπου $\bar{X} = \sum_{k=1}^n \frac{X_k}{n}$ είναι ο μεγάλος μέσος όρος του X . Τότε το i -οστό διάνυσμα z_i θεωρείται ως το κέντρο του ασαφούς συνόλου, το οποίο αποτελείται από τα εναπομείναντα διανύσματα z_j ($1 \leq j \leq c+1, j \neq i$), και η συνάρτηση μελών (membership function) του δίνεται ως εξής,

$$\mu_{ij} = \frac{1}{\sum_{\substack{l=1 \\ l \neq j}}^{c+1} \left(\frac{\|z_j - z_l\|_A}{\|z_j - z_i\|_A} \right)^w}, \quad 1 \leq i \leq c+1, \quad 1 \leq j \leq c+1, j \neq i \quad (3.9)$$

όπου $w \in (0, \infty)$ είναι ένας παράγοντας για να ρυθμίσουμε το αποτέλεσμα της στάθμισης του μ_{ij} . Με βάση την (3.9) ορίζουμε την ασαφή απόκλιση (fuzzy deviation) (dev_{ij}) μεταξύ z_i και z_j από την επόμενη εξίσωση,

$$dev_{ij} = (\mu_{ij})^{(2+w)/2w} \|z_j - z_i\|_A, \\ 1 \leq i \leq c+1, 1 \leq j \leq c+1, j \neq i \quad (3.10)$$

Σύμφωνα με την παραπάνω σημειογραφία, ο βαθμός συμμετοχής μ_{ij} του j -οστού διανύσματος (z_j) στο i -οστό διάνυσμα (z_i) υπολογίζεται σχετικά με τα υπόλοιπα διανύσματα στην (3.8) και μιας που το \bar{X} είναι σταθερά, η ασαφής απόκλιση (dev_{ij}) στην (3.10) παρέχεται μια εκτίμηση κατά πόσο κοντά ή πόσο μακριά είναι ένα κέντρο συστάδας από ένα άλλο, λαμβάνοντας υπόψη τις θέσεις των άλλων κέντρων συστάδων. Επομένως, το σύνολο όλων των $dev_{ij} (1 \leq i \leq c+1, 1 \leq j \leq c+1, j \neq i)$ δίνει ένα μέτρο του πόσο καλά διαχωρίζονται τα συμπλέγματα. Με βάση την παραπάνω ανάλυση, ορίζουμε την έννοια του ασαφούς διαχωρισμού (fuzzy separation) (s) της ασαφούς c -κατάτμησης ως το άθροισμα των τετραγώνων όλων των dev_{ij} ,

$$s = \sum_{i=1}^{c+1} \sum_{\substack{j=1 \\ j \neq i}}^{c+1} (dev_{ij})^2 \quad (3.11)$$

Τέλος, ο δείκτης εγκυρότητας δίνεται ως ο λόγος μεταξύ της συνολικής πυκνότητας (π), και του ασαφούς διαχωρισμού (s):

$$s_V = \frac{\sum_{i=1}^c \frac{\sum_{k=1}^n (u_{ik})^m \|x_k - v_i\|_A^2}{n_i}}{\sum_{i=1}^{c+1} \sum_{\substack{j=1 \\ j \neq i}}^{c+1} (\mu_{ij})^{(2+w)/w} \|z_j - z_i\|_A^2} \quad (3.12)$$

Ο συντελεστής w καθορίζει την ασάφεια του τμήματος διαχωρισμού, και η τιμή του επηρεάζει άμεσα την απόδοση του δείκτη εγκυρότητας.

4 Ο ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΑΛΓΟΡΙΘΜΟΣ ΕΚΠΑΙΔΕΥΣΗΣ

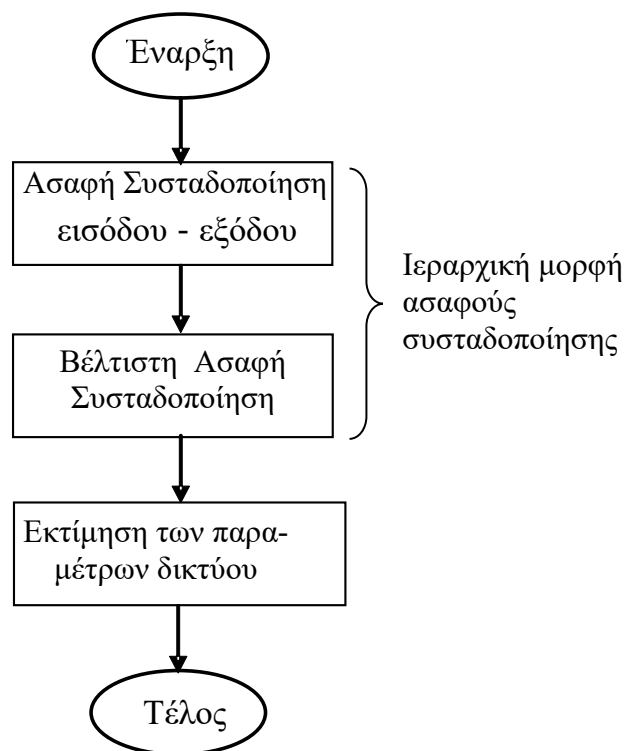
Ο προτεινόμενος αλγόριθμος χρησιμοποιεί τρεις βαθμίδες εκπαίδευσης και το διάγραμμα ροής του απεικονίζεται στην Εικόνα 4.1 . Το Βήμα 1 εφαρμόζει την fuzzy c-means στο χώρο των χαρακτηριστικών γνωρισμάτων εισόδου-εξόδου (input-output feature space), προκειμένου να προ-επεξεργαστεί τα δεδομένα. Στο Βήμα 2, τα παραγόμενα κέντρα διασποράς προβάλλονται στο χώρο εισόδου (input space) και θεωρούνται ως το νέο σύνολο δεδομένων, το οποίο συγκεντρώνεται περαιτέρω (further clustered) μέσω της βέλτιστης ασαφούς συσταδοποίησης , αποκτώντας έτσι τον αριθμό των κρυφών κόμβων (hidden nodes). Η βέλτιστη ασαφή συσταδοποίηση βασίζεται στην εφαρμογή της σταθμισμένης ασαφούς c-means (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 σταθμισμένης ασαφούς c-means (weighted fuzzy c-means). Ακόμα μπορούμε να χρησιμοποιήσουμε την κλασική fuzzy c-means για να επεξεργαστούμε τα κέντρα της διασποράς (cluster centers) που παράγονται από την προ-επεξεργαστική μονάδα (pre-processing unit).

Προφανώς, η εφαρμογή της fuzzy c-means 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 fuzzy c-means 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 εκδοχής της fuzzy c-means [20].



Εικόνα 4.1. Το διάγραμμα ροής του προτεινόμενου αλγόριθμου.

4.1 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 ασαφή συστάδες. Δεδομένου ότι η συνάρτηση μέλους (membership function) του κ -οστού διανύσματος στην l -οστή συστάδα συμβολίζεται με $u_l(\mathbf{z}_k)$ και η παράμετρος ασάφειας ως m , η fuzzy c -means ελαχιστοποιεί την ακόλουθη αντικειμενική συνάρτηση (objective function):

$$J_m = \sum_{k=1}^N \sum_{l=1}^n (u_l(\mathbf{z}_k))^m \|\mathbf{z}_k - \mathbf{a}_l\|^2 \quad (4.1)$$

βάσει του επόμενου περιορισμού,

$$\sum_{l=1}^n u_l(\mathbf{z}_k) = 1 \quad \forall k \quad (4.2)$$

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

$$\mathbf{a}_l = \frac{\sum_{k=1}^N (u_l(\mathbf{z}_k))^m \mathbf{z}_k}{\sum_{k=1}^N (u_l(\mathbf{z}_k))^m} \quad (4.3)$$

$$u_l(\mathbf{z}_k) = \frac{1}{\sum_{j=1}^n \left(\frac{\|\mathbf{z}_k - \mathbf{a}_l\|}{\|\mathbf{z}_k - \mathbf{a}_j\|} \right)^{2/(m-1)}} \quad (4.4)$$

Στη συνέχεια, η fuzzy c -means αποτελείται από επαναληπτικές χρήσεις της εφαρμογής των δύο παραπάνω εξισώσεων.

4.2 Βήμα 2: Βέλτιστη ασαφής συσταδοποίηση

Σε αυτό το βήμα, τα κέντρα διασποράς (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 διάστημα εισόδου (input space),

$$\mathbf{v}_l = \underset{X}{proj}(\mathbf{a}_l) \quad (4.5)$$

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

$$r_l = \aleph(U^l) / \sum_{j=1}^n \aleph(U^j) \quad (4.6)$$

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

$$\aleph(U^l) = \sum_{k=1}^N u_l(\mathbf{z}_k) \quad (1 \leq l \leq n) \quad (4.7)$$

Τότε, η αντικειμενική συνάρτηση (objective function) της σταθμισμένης fuzzy c-means είναι [14],

$$J_r = \sum_{l=1}^n \sum_{i=1}^c r_l (h_i(\mathbf{v}_l))^q \|\mathbf{v}_l - \mathbf{v}_i\|^2 \quad (4.8)$$

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 διασποράς 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,

$$\sum_{i=1}^c h_i(\mathbf{v}_l) = 1 \quad \forall l \quad (4.9)$$

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

$$\mathbf{v}_i = \frac{\sum_{l=1}^n r_l (h_l(\mathbf{v}_i))^q \mathbf{v}_l}{\sum_{l=1}^n r_l (h_l(\mathbf{v}_i))^q} \quad (4.10)$$

$$h_i(\mathbf{v}_l) = \frac{1}{\sum_{j=1}^c \left(\frac{\|\mathbf{v}_l - \mathbf{v}_j\|}{\|\mathbf{v}_l - \mathbf{v}_i\|} \right)^{2/(q-1)}} \quad (4.11)$$

Η (4.10) και η (4.11) αποτελούν μια επαναληπτική διαδικασία βελτιστοποίησης. Για να εκτελέσουμε τη βέλτιστη ασαφή συσταδοποίηση εφαρμόζουμε το δείκτη εγκυρότητας της εξίσωσης:

$$VI = \frac{\pi}{s} = \frac{\sum_{i=1}^c \frac{\sum_{k=1}^N (u_i(\mathbf{x}_k))^m \|\mathbf{x}_k - \mathbf{v}_i\|^2}{n_i}}{\min_{1 \leq i \leq c} \{D_{\min}(P_i)\}} \quad (4.12)$$

σε σχέση με τα κέντρα διασποράς \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,

$$u_i(\mathbf{x}_k) = 1 / \sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{\frac{2}{m-1}} \quad (4.13)$$

4.3 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 διασποράς \mathbf{v}_i ($1 \leq i \leq c$). Για να υπολογίσουμε τα πλάτη, που απασχολούν τη μέθοδο, η οποία αναπτύχθηκε από τον Νείρο και τον Τσεκούρα στο [21],

$$\sigma_i = \frac{2 * d_{max}^i}{3} \quad (1 \leq i \leq c) \quad (4.14)$$

όπου d_{max} δίνεται στην επόμενη εξίσωση,

$$d_{max}^i = \{ \max_{1 \leq k \leq N} \{ \|\mathbf{x}_k - \mathbf{v}_i\|^2 \} : \mathbf{x}_k \in C_i \text{ ώστε } u_i(\mathbf{x}_k) \geq \theta \} \quad (4.15)$$

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

Τελικά, τα συνδετικά βάρη υπολογίζονται με πρότυπο την προσέγγιση των ελαχίστων τετραγώνων.

5 Αριθμητικά παραδείγματα

Στην ενότητα αυτή αξιολογούμε την προτεινόμενη μέθοδο και συγκρίνουμε τις επιδόσεις της με δύο διαφορετικά δίκτυα. Η πρώτη, που ονομάζεται RBFN 1, είναι ένα πρότυπο δίκτυο RBF, το οποίο χρησιμοποιεί τον αλγόριθμο c-means για τον προσδιορισμό των κέντρων του πυρήνα στην (1), ενώ τα αντίστοιχα πλάτη ορίζονται ίσα με 1. Το δεύτερο δίκτυο, που ονομάζεται RBFN 2 έχει εκπαιδευτεί από τον αλγόριθμο fuzzy c-means και οι ακτινικές συναρτήσεις βάσεων συμπίπτουν με τις λειτουργίες συμμετοχής των αντιστοίχων ασαφών συστάδων. Όλες οι προσομοιώσεις πραγματοποιήθηκαν με τη χρήση 10 διαφορετικών αρχικοποιήσεων (δηλαδή 10 αναλύσεις για κάθε πειραματική περίπτωση) για την fuzzy c-means, η παράμετρος ασάφειας ορίστηκε ίση με 2. Για την εκτέλεση των πειραμάτων για τον προτεινόμενο αλγόριθμο, χρησιμοποιήσαμε το Xie-Beni (XB) δείκτη [23], το δείκτη SV που αναπτύχθηκε στο [22], και τον προτεινόμενο αλγόριθμο, και οι πειραματικές περιπτώσεις είναι τρία γνωστά σύνολα δεδομένων, και η πειραματική ανάλυση δίνεται μέσα στις επόμενες υποενότητες.

5.1 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 21x 21 σημεία ισομερώς κατανομημένα στο διάστημα εισόδου. Τα τρία δίκτυα συγκρίθηκαν όσον αφορά την ομαλοποιημένη μέση τετραγωνική απόκλιση (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 παραγόμενων στοιχείων. 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 fuzzy c-means επιλέχθηκε ως $n = 30$.

Πίνακας 1. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα προσέγγισης δισδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB Δείκτης [23]	SV Δείκτης [22]	Προτεινόμενος Δείκτης
2	0.00208647	0.94510584	0.009190
3	0.00157383	0.34411418	0.007174
4	0.00137195	0.21514754	0.005514
5	0.00095763	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 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 3, 8 και 9 αντίστοιχα. Η προκύπτουσα ομαλοποιημένη μέση τετραγωνική απόκλιση (NRMSE), για τους τρεις δείκτες απεικονίζεται στον Πίνακα 2. Επιπλέον, αυτός ο πίνακας συγκρίνει τις επιδόσεις των δικτύων που δημιουργούνται την παρούσα χρονική στιγμή με τις επιδόσεις των άλλων μεθόδων που υπάρχουν στη βιβλιογραφία.

Πίνακας 2. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές NRMSE.

Μοντέλο	Αριθμός κρυφών κόμβων	NRMSE	
CFC [24]	6	0.838 ± 0.073	
	9	0.812 ± 0.100	
	12	0.823 ± 0.056	
CFA [24]	6	0.812 ± 0.044	
	8	0.713 ± 0.016	
	9	0.708 ± 0.041	
Το μοντέλο μας	10	0.705 ± 0.035	
	XB δείκτης	3	$0.914 \pm 1.2E-6$
	SV δείκτης	8	$0.754 \pm 1.0E-4$
	Προτεινόμενος δείκτης	9	$0.703 \pm 1.2E-5$

5.2 Προσέγγιση στατικής δισδιάστατης συνάρτησης

Η λειτουργία αυτή περιγράφεται από την ακόλουθη εξίσωση,

$$y = (1 + x_1^{-2} + x_2^{-1.5}), \quad 1 \leq x_1, x_2 \leq 5 \quad (5.1)$$

Χρησιμοποιώντας την (5.1) δημιουργήσαμε 50 στοιχεία εισόδου - εξόδου.

Ο αρχικός αριθμός των συστάδων για την fuzzy c-means επιλέχθηκε ως $n = 30$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας για τη fuzzy c-means και για τη σταθμισμένη fuzzy c-means ήταν $m=2$ και $q=2$, αντιστοίχως.

Πίνακας 3. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της Προσέγγισης στατικής δισδιάστατης συνάρτησης.

Αριθμός Συστάδων	XB δείκτης [23]	SV δείκτης [22]	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

Ο Πίνακας 3 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες ισχύος. 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 μέση τετραγωνική απόκλιση (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 4. Επιπλέον, ο πίνακας αυτός συγκρίνει τις επιδόσεις των δικτύων που δημιουργήσαμε την παρούσα χρονική στιγμή με τις επιδόσεις των άλλων μεθόδων που υπάρχουν στη βιβλιογραφία.

Πίνακας 4. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE.

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE
Lee et al. [26]		8	0.0148
Kim et al. [25]		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	$0.0738 \pm 1.3E-8$
	SV δείκτης	13	$0.0318 \pm 2.6E-5$
	Προτεινόμενος δείκτης	10	$0.0306 \pm 1.0E-4$

5.3 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 εισόδου-εξόδου ζεύγη δεδομένων και τυχαία τα διαιρέσαμε σε σύνολα δεδομένων εκπαίδευσης εμπεριέχοντας το 60 % και σύνολα δεδομένων δοκιμών, τα οποία εμπεριέχουν το υπόλοιπο 40 % των δεδομένων. Ο αρχικός αριθμός των συστάδων για την fuzzy c-means επιλέχθηκε ως $n=30$. Σε όλες τις προσομοιώσεις οι παράμετροι ασάφειας ήταν: $m=2$ και $q=2$, αντίστοιχα.

Πίνακας 5. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα της Συνθετικής 3d Συνάρτησης.

Αριθμός Συστάδων	XB δείκτης [23]	SV δείκτης [22]	Προτεινόμενος δείκτης
2	0.00208647	0.94510584	0.009190
3	0.00157383	0.34411418	0.007174
4	0.00137195	0.21514754	0.005514
5	0.00095763	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 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 8. Η προκύπτουσα μέση τετραγωνική απόκλιση (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 6 . Επιπλέον, ο πίνακας αυτός συγκρίνει τις επιδόσεις των δικτύων που δημιουργήσαμε την παρούσα χρονική στιγμή με τις επιδόσεις των άλλων μεθόδων που υπάρχουν στη βιβλιογραφία και με τα RBFN 1 και RBFN 2.

Πίνακας 6. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE.

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE
Park et. al. [27]	(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 Abalone Δεδομένα

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

Η προκύπτουσα μέση τετραγωνική απόκλιση (RMSE), για τους τρεις δείκτες απεικονίζεται στον πίνακα 7 , όπου μπορούμε εύκολα να ελέγξουμε ότι ο προτεινόμενος αλγόριθμος επιτυγχάνει το καλύτερο αποτέλεσμα με το μικρότερο αριθμό των κρυφών κόμβων. Ο Πίνακας 8 απεικονίζει τις βέλτιστες τιμές που λαμβάνονται από τους τρεις δείκτες ισχύος. Σύμφωνα με αυτόν τον πίνακα ο βέλτιστος αριθμός των συστάδων για τον XB [23], SV [22] και τον προτεινόμενο δείκτη ήταν 10, 20 και 14 αντίστοιχα .

Πίνακας 7. Συγκριτική ανάλυση όσον αφορά τις μέσες τιμές RMSE .

Μοντέλο		Αριθμός κρυφών κόμβων	RMSE	
Pedrycz et al. [28]	(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
	XB δείκτης (δεδομένα δοκιμής)		10	2.9338 ± 0.096
	SV δείκτης (δεδομένα εκπαίδευσης)		20	2.0825 ± 0.046
	SV δείκτης (δεδομένα δοκιμής)		20	2.2114 ± 0.120
	Προτεινόμενος δείκτης (δεδομένα εκπαίδευσης)		14	2.0226 ± 0.043
	Προτεινόμενος δείκτης (δεδομένα δοκιμής)		14	2.1511 ± 0.023

Πίνακας 8. Βέλτιστα αποτελέσματα συσταδοποίησης για το παράδειγμα Abalone.

Αριθμός Συστάδων	XB δείκτης [23]	SV δείκτης [22]	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

6 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στη παρούσα εργασία διερευνήσαμε πώς η ασαφή συσταδοποίηση εισόδου-εξόδου μπορεί να συνεργαστεί με τη βέλτιστη ασαφή συσταδοποίηση προκειμένου να εκπαιδεύσουν αποτελεσματικά νευρωνικά δίκτυα RBF.

Αρχικά, εφαρμόσαμε την ασαφή συσταδοποίηση στο διάστημα εισόδου-εξόδου, και αναμέναμε τα προκύπτοντα κέντρα διασποράς στο διάστημα εισόδου. Για κάθε κέντρο διασποράς εμείς αποδώσαμε ένα συντελεστή στάθμισης, ο οποίος σχετίζεται με το μέγεθος (δηλαδή πληθικότητας) της αντίστοιχης συστάδας . Εμείς, λοιπόν, αντιμετωπίσαμε τα κέντρα διασποράς μαζί με τα αντίστοιχα βάρη ως νέο σύνολο δεδομένων και πραγματοποιήσαμε τη βέλτιστη ασαφή συσταδοποίηση με την εφαρμογή της σταθμισμένης fuzzy c-means .

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.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Διαμαντάρας Κ., 2007, *Τεχνητά Νευρωνικά Δίκτυα*, Κλειδάριθμος, Αθήνα
- [2] McCulloch, W. and Pitts, W., “ *A logical calculus and the ideas immanent in the nervous activity* ”, *Bulletin of Mathematical Biophysics*, vol. 5 , pp. 115-133, 1943
- [3] Haykin, S., 1999, *Neural Networks - A Comprehensive Foundation*, Pearson Education, Delhi
- [4] Patterson D. (1996). *Artificial Neural Networks, Theory and Applications*. Prentice Hall.
- [5] Xie LX, Beni G. *A validity measure for fuzzy clustering*. *IEEE Trans PAMI* 1991;13:841–7.
- [6] Bensaid AM, Hall LO, Bezdek JC, Clarke LP, Silbiger ML, Arrington JA, Murtagh RF. *Validity-guided (re)clustering with applications to image segmentation*. *IEEE Trans Fuzzy Syst* 1996;4:112–23.
- [7] Zahid N, Abouelala O, Limouri M, Essaid A. *Unsupervised fuzzy clustering*. *Pattern Recogn Lett* 1999;20:123–9.
- [8] F. Behloul, B.P.F. Lelieveldt, A. Boudraa, and J.H.C. Reiber, *Optimal design of radial basis function neural networks for fuzzy-rule extraction in high dimensional data*, *Pattern Recognition*, 35, 2002, 659-675.
- [9] N.B. Karayiannis, *Reformulated radial basis neural networks trained by gradient descent*, *IEEE Transactions on Neural Networks*, 10(3), 1999 , 657–671.
- [10] M.B. Lucks, N. Oki, *A Radial Basis Network for Function Approximation*, *Proceedings of the 42nd Midwest Symposium on Circuits and Systems*, 1999, 1099–1101.
- [11] F. Schwenker, H.A. Kestler, G. Palm, *Three learning phases for radial-basis-function networks*, *Neural Networks*, 14, 2001, 439-458.
- [12] A.K. Jain, R.C. Dubes, *Algorithms for clustering data* (Englewood Cliffs, NJ: Prentice Hall, 1988).

- [13] D. Shia, D.S. Yeungb, J. Gaoc, *Sensitivity analysis applied to the construction of radial basis function networks*, Neural Networks, 18, 2005, 951–957.
- [14] J.K. Sing, D.K. Basu, M. Nasipuri, M. Kundu, *Improved k-means algorithm in the design of RBF neural networks*, International Conference on Convergent Technologies for Asia-Pacific Region 2, 2003, 841 – 845.
- [15] J.C. Bezdek., *Fuzzy Mathematics in Pattern Classification PhD Dissertation* (Ithaca, NY : Cornell University, 1973).
- [16] N.R Pal, J.C. Bezdek, *On cluster validity for the fuzzy c-means model*, IEEE Transactions on Fuzzy Systems,3(3),1995, 370 - 379 .
- [17] I. Cha and S. A. Kassam, *RBFN restoration of nonlinearly degraded images*, IEEE Transactions on Image Processing, 5, 1996, 964–975.
- [18] C.L. Chen, W.C. Chen, and F. Y. Chang, *Hybrid learning algorithm for Gaussian potential function networks*, Proc. Inst. Elect. Eng., pt. D, 140(6), 1993, 442–448.
- [19] Z. Uykan, C. Güzelis,, M.E. Çelebi, and H.N. Koivo, *Analysis of Input–Output Clustering for Determining Centers of RBFN*, IEEE Transactions on Neural Networks, 11(4), 2000, 851-858.
- [20] G. E. Tsekouras, *On the use of the weighted fuzzy c-means in fuzzy modelling*, Adv Eng Softw ,36 ,2005, 287–300.
- [21] A. D. Niros, and G. E.Tsekouras, *On training radial basis function neural networks using optimal fuzzy clustering*, 17th IEEE Mediterranean Conference on Control and Automation, 2009, 395 - 400.
- [22] G.E. Tsekouras, H. Sarimveis, *A new approach for measuring the validity of the fuzzy c-means algorithm*, Advances in Engineering Software, 35, 2004, 567–575.
- [23] LX Xie and G. Beni, *A validity measure for fuzzy clustering*. IEEE IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 1991, 841–847.
- [24] J. González, I. Rojas, H. Pomares, J. Ortega and A. Prieto, *A New Clustering Technique for Function Approximation*, IEEE Transactions on Neural Networks, 13(1), 2002, 132-142.

- [25] E. Kim, M. Park, S. Kim, M. Park, *A transformed input-domain approach to fuzzy modelling*, IEEE Transactions on Fuzzy Systems, 6(4),1998 ,596-604.
- [26] S.J. Lee, C.S. Ouyang, *A neuro-fuzzy system modelling with self constructing rule generation and hybrid SVD-based learning*, IEEE Transactions on Fuzzy Systems, 11, 2003, 341–53.
- [27] H.S. Park, W. Pedrycz, S.K. Oh, *Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields*, IEEE Transactions on Neural Networks, 20 (10), 2009, 1604-1616.
- [28] W. Pedrycz, H.S. Park, S.K. Oh, *A granular-oriented development of functional radial basis function neural networks*, Neurocomputing, 72, 2008, 420-435.
- [29] Antonios D. Niros, George Leventis, Dimitrios Tsolakis, George E. Tsekouras, and Michael Kenteris, *A Hierarchical Fuzzy Clustering Framework for Training RBF Networks*, Proceedings of the IASTED International Conference, Artificial Intelligence and Soft Computing (ASC 2011),2011, 248-254.