

**Πανεπιστήμιο Αιγαίου**  
**Τμήμα Πολιτισμικής Τεχνολογίας και Επικοινωνίας**

**Μελέτη της Επίπτωσης της Διασποράς των Συναρτήσεων Βάσης**  
**στο Σχεδιασμό Νευρωνικών Δικτύων Συναρτήσεων Ακτινικής**  
**Βάσης**

Πτυχιακή εργασία  
της

**Δρακουλάκη Ιωάννας**

A.M.: 1312005025

*Επιβλέπων: Τσεκούρας Γεώργιος, Επίκουρος Καθηγητής*  
*Συνεπιβλέποντες: Παυλογεωργάτος Γεράσιμος, Επίκουρος Καθηγητής*  
*Καλλονιάτης Χρήστος, Λέκτορας*

**Ιούνιος 2010**

## ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ.....	7
1. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ.....	10
1.1. Εισαγωγή.....	10
1.2. Δομή Νευρώνα .....	10
1.3. Συναρτήσεις Ενεργοποίησης.....	11
1.3.1. Γραμμική Συνάρτηση (Linear function) .....	12
1.3.2. Συνάρτηση Κατωφλιού ή Βήματος (Threshold Function).....	12
1.3.3. Κατά Τμήματα Γραμμική Συνάρτηση (Piecewise-Linear function).....	13
1.3.4. Σιγμοειδής Συνάρτηση (Sigmoid Function).....	14
1.3.5. Υπερβολική Εφαπτομένη (Hyperbolic Tangent Function).....	14
1.3.6. Γκαουσιανή Συνάρτηση (Gaussian Function).....	14
1.4. Αρχιτεκτονική Δικτύων.....	15
1.4.1. Δίκτυα Ενός Στρώματος.....	15
1.4.2. Δίκτυα Πολλών Στρωμάτων.....	15
1.4.3. Επαναλαμβανόμενα Δίκτυα (Recurrent Networks).....	17
1.5. Εποπτευόμενη και Μη-εποπτευόμενη Μάθηση.....	18
1.5.1. Μάθηση με επίβλεψη.....	19
1.5.2. Μάθηση χωρίς επίβλεψη.....	20
1.6. Δίκτυο Perceptron.....	21
1.6.1. Εκπαίδευση Δικτύου Perceptron .....	22
1.7. Πολυστρωματικό Perceptron .....	24
1.7.1. Εκπαίδευση Πολυστρωματικού Perceptron.....	24
1.8. Δίκτυα Kohonen.....	25
1.9. Εφαρμογές Τεχνητών Νευρωνικών Δικτύων.....	27
2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΑΚΤΙΝΙΚΗΣ ΒΑΣΗΣ.....	29
2.1. Εισαγωγή.....	29
2.2. Συναρτήσεις Ακτινικής Βάσης.....	29
2.3. Δίκτυα Ακτινικής Βάσης.....	30
2.4. Στρατηγικές Μάθησης.....	31
2.4.1. Σταθερά Κέντρα που Επιλέγονται Τυχαία .....	31
2.4.2. Αυτό-Οργανωμένη Επιλογή των Κέντρων .....	32
2.4.3. Επιβλεπόμενη Επιλογή των Κέντρων .....	33

2.4.4. Πλήρης Παρεμβολή με Κανονικοποίηση .....	34
2.5. Σύγκριση Multi Layer Perceptron και Radial Basis Function.....	35
3. ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ.....	36
3.1 Ο Αλγόριθμος των c-Μέσων (ή αλγόριθμος Lloyd).....	36
3.2 Αλγόριθμος Fuzzy c-Means.....	40
3.2.1 Βασικές Έννοιες Ασαφούς Λογικής.....	40
3.2.2 Ο Αλγόριθμος των Ασαφών c-Μέσων (Fuzzy c-Means).....	43
4. ΑΛΓΟΡΙΘΜΟΙ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ .....	46
4.1. Πρώτος Αλγόριθμος.....	46
4.2. Δεύτερος Αλγόριθμος.....	47
4.3. Τρίτος Αλγόριθμος.....	47
5. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ.....	48
5.1. Πρώτη Σύγκριση.....	48
5.2. Δεύτερη Σύγκριση.....	62
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	67
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	70

## **Ευχαριστίες**

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

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

## Περίληψη

Σε αυτή την πτυχιακή, εξετάστηκε ο τρόπος με τον οποίο ο υπολογισμός της διασπορά της γκαουσιανής συνάρτησης, επηρεάζει την απόδοση με ενός νευρωνικού δικτύου συνάρτησης ακτινικής βάσης. Για τον σκοπό αυτό, επιλέχθηκαν αρχικά τρεις μέθοδοι, οι οποίες βασίζονται στη συσταδοποίηση δεδομένων. Με τη χρήση των μεθόδων αυτών εκπαιδεύτηκαν τα αντίστοιχα νευρωνικά δίκτυα ακτινικής βάσης και μελετήθηκε ποσοτικά η επίδραση της διασποράς των συναρτήσεων ακτινικής βάσης στο μέσο τετραγωνικό σφάλμα που προέκυπτε σε κάθε περίπτωση. Σε όλες τις πειραματικές προσομοιώσεις χρησιμοποιήθηκαν δύο σύνολα δεδομένων, το ένα για την εκπαίδευση του νευρωνικού και το άλλο για τον έλεγχο της αξιοπιστίας του. Η σύγκριση μεταξύ των τριών μεθόδων έγινε με βάση το μέσο τετραγωνικό σφάλμα και για τα δύο παραπάνω σύνολα δεδομένων. Οι μέθοδοι συσταδοποίησης που λήφθηκαν υπόψη είναι ο αλγόριθμος c-means και ο fuzzy c-means.

## **Abstract**

Herein, we analyze in details the impact of the variance of the radial basis functions in the efficient design of Radial Basis Functions (RBF) Neural Networks. We selected three well-known training methods that are based on cluster analysis. We used these methods to train three RBF networks and, then, to quantitatively study the impact of the variances on the mean square error (MSE). In all simulation experiments we utilized two data sets. The first one was employed to train the RBF network and the second one to test its validity and reliability. The comparison between the three methods took place in terms of the MSE values obtained by the two data sets. The clustering algorithms were the c-means and the fuzzy c-means method.

## ΕΙΣΑΓΩΓΗ

Η πληροφορική είναι μια επιστήμη που εξελίσσεται με ραγδαίους ρυθμούς. Η τεχνολογία της αλλάζει συνεχώς τα τελευταία χρόνια και γίνονται έρευνες προκειμένου τα υπολογιστικά συστήματα που κατασκευάζονται να γίνονται όλο και πιο πολύπλοκα. Πλέον οι υπολογιστές μπορούν να επεξεργάζονται δεδομένα με τέτοιο τρόπο ώστε σε ορισμένες περιπτώσεις ο τρόπος με τον οποίο μαθαίνουν τείνει να αντιγράψει την ανθρώπινη συμπεριφορά.

Επειδή ο ανθρώπινος εγκέφαλος έχει μοναδικές ικανότητες στην ταχύτητα και τον τρόπο που επεξεργάζεται τα ερεθίσματα που δέχεται οι επιστήμονες προσπαθούν τα τελευταία πενήντα περίπου χρόνια να εντάξουν αυτές τις ικανότητες στους υπολογιστές. Αναπτύχθηκε έτσι ο τομέας της Τεχνητής Νοημοσύνης (TN), η οποία ορίζεται ως η επιστήμη των υπολογιστών που ασχολείται με τη σχεδίαση ευφύων υπολογιστικών συστημάτων. Μια ιδιαίτερη προσέγγιση στη TN είναι τα νευρωνικά δίκτυα. Ένας τεχνητός νευρώνας έχει παρόμοια δομή με αυτή του βιολογικού (Βλαχάβας, 2006).

Για να λειτουργήσουν οι τεχνητοί νευρώνες χρειάζεται να διεγερθούν. Αυτό γίνεται με τη χρήση των συναρτήσεων ενεργοποίησης. Μερικές από τις συναρτήσεις ενεργοποίησης είναι η βηματική, η σιγμοειδής, η υπερβολική εφαπτομένη, η συνάρτηση κατωφλιού και άλλες.

Τα τεχνητά νευρωνικά δίκτυα μαθαίνουν έπειτα από εκπαίδευση η οποία πραγματοποιείται με συγκεκριμένους αλγόριθμους που βασίζονται σε μάθηση υπό εκπαίδευση και σε αυτό-οργανούμενη μάθηση. Στην μάθηση υπό εκπαίδευση χρησιμοποιούνται δεδομένα που έχουν γνωστή δομή για να μάθει το δίκτυο ενώ στην αυτό-οργανούμενη μάθηση ο αλγόριθμος χρησιμοποιεί τυχαία αρχικοποίηση (Zurada, 1992).

Στην παρούσα εργασία, μελετήθηκαν τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης (Radial Basis Function Neural Networks - RBFNN). Τα νευρωνικά δίκτυα 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 RBFNN, 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-Means και ο 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.

Ο σκοπός της παρούσης εργασίας είναι η μελέτη της επίπτωσης της διασποράς των συναρτήσεων βάσης στο σχεδιασμό νευρωνικών δικτύων συναρτήσεων ακτινικής βάσης και συγκεκριμένα της γκαουσιανής συνάρτησης που είναι το πιο χαρακτηριστικό παράδειγμα τέτοιου τύπου συνάρτησης. Ουσιαστικά πραγματοποιήθηκε μια σύγκριση ανάμεσα σε τρεις μεθόδους μάθησης νευρωνικού συναρτήσεων ακτινικής βάσης με σκοπό να εξεταστεί για ποια τιμή μιας ορισμένης παραμέτρου, το νευρωνικό εκφέρει καλύτερα αποτελέσματα στις τιμές του μέσου τετραγωνικού σφάλματος τόσο στη διαδικασία εκπαίδευσης του νευρωνικού, όσο και στη δοκιμή του κατά πόσο ικανοποιητικά πραγματοποιήθηκε η εκπαίδευση. Οι παράμετροι που επιλέχθηκαν για να εξεταστεί το ερώτημα, επηρεάζουν την σχέση που υπολογίζει τη διασπορά των συναρτήσεων βάσης.

Ο αλγόριθμος που χρησιμοποιήθηκε στην πρώτη μέθοδο, διαμέριζε το σύνολο των δεδομένων του με τον αλγόριθμο c-Means, υπολόγιζε την ευκλείδεια απόσταση ανάμεσα στα κέντρα και τα δεδομένα, την οποία έπειτα χρησιμοποιούσε για να υπολογίσει τη διασπορά της συνάρτησης βάσης και τα ελάχιστα τετράγωνα. Τέλος εμφάνιζε το μέσο τετραγωνικό σφάλμα της εκπαίδευσης και της δοκιμής για όλους τους κρυφούς νευρώνες. Η παράμετρος που άλλαζε για να συγκριθούν τα αποτελέσματα ήταν ο αριθμός των κοντινότερων γειτόνων.

Η δεύτερη μέθοδος διαμέριζε τα δεδομένα με τη χρήση του αλγορίθμου Fuzzy c-Means, υπολόγιζε την ευκλείδεια απόσταση ανάμεσα στα κέντρα και τα δεδομένα, και τη συμμετοχή του κάθε δεδομένου στα προκαθορισμένα κέντρα. Έπειτα χρησιμοποιούσε τη συνάρτηση συμμετοχής για να υπολογίσει τα ελάχιστα τετράγωνα και τέλος υπολόγιζε το μέσο τετραγωνικό σφάλμα για τους νευρώνες εκπαίδευσης και δοκιμής. Η παράμετρος της οποίας οι τιμές μεταβάλλονταν ήταν η παράμετρος ασάφειας.

Στην τρίτη μέθοδο, τα δεδομένα συσταδοποιήθηκαν πάλι με τον αλγόριθμο Fuzzy c-Means, η απόσταση ανάμεσα στα κέντρα των συναρτήσεων ακτινικής βάσης και των δεδομένων βρέθηκε με την χρήση της ευκλείδειας απόστασης και η συμμετοχή των δεδομένων στις διάφορες συστάδες με τη συνάρτηση συμμετοχής. Έπειτα υπολογίστηκε η διασπορά των συναρτήσεων βάσης, που στη συγκεκριμένη περίπτωση ήταν η γκαουσιανή, και τέλος τα ελάχιστα τετράγωνα. Και σε αυτή τη μέθοδο τα αποτελέσματα που τυπώνονται είναι το μέσο τετραγωνικό σφάλμα των δεδομένων εκπαίδευσης και δοκιμής, και άλλαζαν οι τιμές της παραμέτρου ασάφειας.

Στη συνέχεια πραγματοποιήθηκαν τα πειράματα, όπου δόθηκαν σε κάθε μία από τις τρεις μεθόδους τρία διαφορετικά σύνολα δεδομένων. Για κάθε σύνολο δεδομένων ο κάθε αλγόριθμος έτρεξε δέκα φορές. Από τα αποτελέσματα που προέκυψαν, δηλαδή τις τιμές του μέσου τετραγωνικού σφάλματος, υπολογίστηκε ο μέσος όρος τους, ο οποίος τελικά χρησιμοποιήθηκε



για να φτιαχτούν τα γραφήματα όπου θα φαίνονταν οι διαφορές στα τελικά αποτελέσματα για τις διάφορες τιμές των παραμέτρων που άλλαζαν.

Έτσι, για τον πρώτο αλγόριθμο δημιουργήθηκαν τα γραφήματα του μέσου τετραγωνικού σφάλματος, ξεχωριστά οι τιμές της εκπαίδευσης και της δοκιμής, για διάφορες τιμές του αριθμού των κοντινότερων γειτόνων και το ίδιο έγινε και για το δεύτερο και τρίτο αλγόριθμο για τις διάφορες τιμές της παραμέτρου ασάφειας. Επιπλέον, εξετάστηκαν οι τρεις μέθοδοι μεταξύ τους για να συγκριθεί ποιο εκ των τριών νευρωνικών δικτύων μαθαίνει καλύτερα, εκφέροντας τα βέλτιστα αποτελέσματα του μέσου τετραγωνικού σφάλματος, δηλαδή αυτά με τις χαμηλότερες τιμές και την ταχύτερη σύγκλιση. Τα γραφήματα οργανώθηκαν με τέτοιο τρόπο ώστε να γίνεται εμφανής για κάθε πειραματική μέθοδο, και για τα τρία σύνολα δεδομένων εισόδου που εξετάστηκαν, πως συμπεριφέρθηκαν τα αποτελέσματα του μέσου τετραγωνικού σφάλματος για όλους τους κρυφούς νευρώνες.

Το πειραματικό κομμάτι υλοποιήθηκε στο περιβάλλον του MatLab.

Το περιεχόμενο της παρούσας εργασίας έχει οργανωθεί ως εξής:

Στο πρώτο κεφάλαιο, γίνεται μια εισαγωγή στα νευρωνικά δίκτυα, όπου περιγράφεται η βασική δομή του τεχνητού νευρώνα, αναφέρονται οι συναρτήσεις ενεργοποίησης και εξηγείται η δομή των νευρωνικών δικτύων πρόσθιας τροφοδότησης με ένα και με πολλά στρώματα, καθώς και ο τρόπος με τον οποίο λειτουργεί η μάθηση με και χωρίς επίβλεψη. Ακόμα, περιγράφονται ορισμένα από τα βασικά νευρωνικά δίκτυα, και τέλος αναφέρονται μερικές από τις εφαρμογές των τεχνητών νευρωνικών δικτύων.

Στο δεύτερο κεφάλαιο περιγράφονται ορισμένες έννοιες σχετικές με τα νευρωνικά δίκτυα συναρτήσεων ακτινικής βάσης. Πιο συγκεκριμένα, αναφέρονται οι συναρτήσεις ακτινικής βάσης, τα αντίστοιχα δίκτυα και οι στρατηγικές μάθησης που χρησιμοποιούνται από αυτά. Επιπλέον, γίνεται μια σύγκριση ανάμεσα στα νευρωνικά δίκτυα ακτινικής βάσης και τα πολυστρωματικά perceptron.

Στο τρίτο κεφάλαιο, αναλύονται οι αλγόριθμοι που χρησιμοποιήθηκαν για να εκτελεστούν τα πειράματα.

Στο τέταρτο κεφάλαιο παρουσιάζονται τα γραφήματα των αποτελεσμάτων και γίνεται μια σύγκριση μεταξύ αυτών.

Στο πέμπτο κεφάλαιο, εξηγούνται τα αποτελέσματα που προέκυψαν από την παρούσα έρευνα.

# 1. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## 1.1 Εισαγωγή

Η μελέτη των τεχνητών νευρωνικών δικτύων, ή όπως αλλιώς αναφέρονται ως νευρωνικά δίκτυα, ξεκίνησε από την αναγνώριση ότι ο ανθρώπινος εγκέφαλος υπολογίζει με έναν εντελώς διαφορετικό τρόπο από αυτό του συμβατικού ηλεκτρονικού υπολογιστή (Haykin, 1999). Βασικό κίνητρο για τη μελέτη των νευρώνων αποτελεί η ανακάλυψη ενός νέου υπολογιστικού μοντέλου που να βασίζεται σε μια δικτυακή δομή όμοια με αυτή του εγκεφάλου (Διαμαντάρας, 2007). Ο εγκέφαλος είναι ένας υψηλά πολύπλοκος, μη-γραμμικός και παράλληλος υπολογιστής. Μία από τις ικανότητές του είναι να οργανώνει τα δομικά συστατικά του, γνωστά ως νευρώνες, έτσι ώστε να πραγματοποιεί συγκεκριμένους υπολογισμούς, όπως η αναγνώριση προτύπων, η αντίληψη και ο έλεγχος κίνησης, σε πολύ ταχύτερους ρυθμούς ακόμα και από τον πιο γρήγορο ψηφιακό υπολογιστή που υπάρχει σήμερα. Για παράδειγμα, ο εγκέφαλος συνήθως χρειάζεται κατά προσέγγιση 100-200 ms για να ολοκληρώσει μια από τις λειτουργίες του σαν αυτή της αναγνώρισης ενός οικείου προσώπου. Σε έναν συμβατικό υπολογιστή, λειτουργίες λιγότερο πολύπλοκες από αυτή μπορεί να ολοκληρωθούν σε μέρες (Haykin, 1999).

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. Καθ' αυτό τον τρόπο «γεννήθηκαν» οι αρχές των νευρωνικών δικτύων και της τεχνητής νοημοσύνης (Haykin, 1999).

## 1.2 Δομή Νευρώνα

Ο νευρώνας είναι μια μονάδα που επεξεργάζεται πληροφορία το οποίο είναι βασικό στη λειτουργία ενός νευρωνικού δικτύου.

Ένα από τα βασικά δομικά στοιχεία ενός τεχνητού νευρώνα είναι ένα σύνολο συνάψεων ή συνδέσμων (set of synapses or connecting links), καθένα από τα οποία χαρακτηρίζεται από ένα

βάρος ή δύναμη. Πιο συγκεκριμένα, ένα σήμα  $x_j$  στην είσοδο της σύναψης  $j$  που συνδέεται με τον νευρώνα  $k$ , πολλαπλασιάζεται με το συναπτικό βάρος  $w_{kj}$ . Σε αντιστοιχία με το βιολογικό νευρώνα, οι συνάψεις είναι τα σημεία ένωσης μεταξύ των διακλαδώσεων του άξονα (πύλη εξόδου) ενός νευρώνα και των δενδριτιών (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 (Διαμαντάρας, 2007) (Haykin, 1999).

Εκτός από τα εισερχόμενα σήματα και τα αντίστοιχα βάρη, ο νευρώνας έχει κάποιο βάρος  $w_0$  που ονομάζεται πόλωση (bias) ή παράγοντας προδιάθεσης του νευρώνα. Αυτό το βάρος διαφέρει από τα υπόλοιπα στο ότι επιδρά συνεχώς με μια τιμή εισόδου  $x_0 = 1$ . Ο όρος αυτός δεν αποδίδεται στο εσωτερικό του νευρώνα. Πρόκειται για ένα εξωτερικό ερέθισμα το οποίο προστίθεται μαζί με τα υπόλοιπα εισερχόμενα σήματα. Ορισμένες φορές χρησιμοποιείται για να καθοριστεί έμμεσα και δυναμικά η θέση της συνάρτησης ενεργοποίησης στο καρτεσιανό επίπεδο (Βλαχάβας, 2006).

Ένα δεύτερο δομικό στοιχείο είναι ο αθροιστής (adder). Ο αθροιστής προσθέτει τα σήματα εισόδου που είναι τοποθετημένα από τις αντίστοιχες συνάψεις του νευρώνα. Οι λειτουργίες που περιγράφονται στο τμήμα του αθροιστή συγκροτούν ένα γραμμικό συνδυασμό.

Άλλο δομικό στοιχείο είναι η συνάρτηση ενεργοποίησης (activation function). Μια συνάρτηση ενεργοποίησης καλείται και συνάρτηση στοιχείου επικοινωνίας (processing element function) ή συνάρτηση καταστολής (squashing function). Η συνάρτηση αυτή εφαρμόζεται στο σύνολο των τιμών των εισόδων οι οποίες έρχονται μέσα από τα τόξα εισόδου. Μια συνάρτηση ενεργοποίησης καλείται και κανόνας διέγερσης. Όταν η είσοδος σε ένα νευρώνα είναι αρκετά μεγάλη, αυτός διεγείρεται, στέλνοντας ένα ηλεκτρικό σήμα στο σύνδεσμο εξόδου του (Dunham, 2004). Η συνάρτηση ενεργοποίησης περιορίζει το επιτρεπτό εύρος (permissible amplitude) της εξόδου του νευρώνα σε μια πεπερασμένη αξία (Haykin, 1999).

### 1.3 Συναρτήσεις Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης, δηλώνονται ως  $\varphi(v)$ , ορίζουν την έξοδο του νευρώνα σε όρους του επαγομένου τοπικού πεδίου  $v$ . Παρακάτω παρουσιάζονται μερικοί από τους πιο βασικούς τύπους συναρτήσεων ενεργοποίησης.

### 1.3.1. Γραμμική Συνάρτηση (Linear function)

Μια γραμμική συνάρτηση ενεργοποίησης παράγει μια γραμμική τιμή εξόδου βάσει της εισόδου της. Παρακάτω είναι μια τυπική συνάρτηση ενεργοποίησης

$$\varphi(v) = v$$

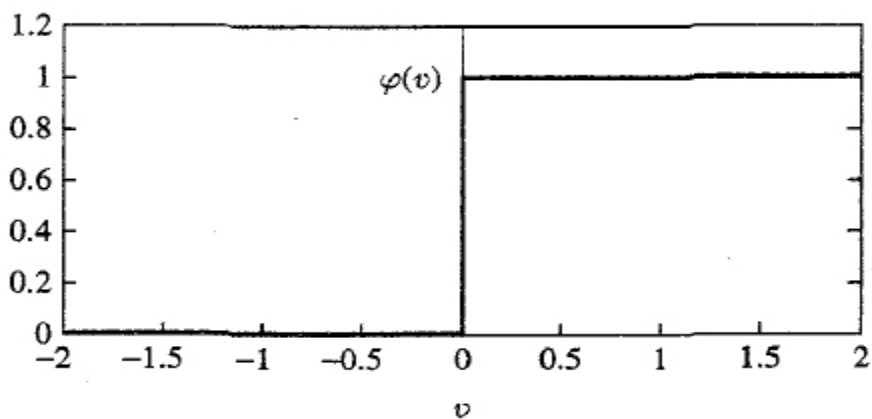
Με τη γραμμική συνάρτηση, η τιμή της εξόδου δεν έχει όρια σε σχέση με ελάχιστες ή μέγιστες τιμές (Dunham, 2004).

### 1.3.2. Συνάρτηση Κατωφλιού ή Βήματος (Threshold Function)

Η τιμή της εξόδου είναι είτε 1 είτε 0 και αυτό εξαρτάται από το άθροισμα των γινομένων των τιμών της εισόδου και των αντίστοιχων βαρών τους

$$\varphi(v) = \begin{cases} 1 & \text{αν } v \geq 0 \\ 0 & \text{αν } v < 0 \end{cases} \quad (1.1)$$

Οι δυαδικές τιμές της εξόδου μπορεί επίσης να είναι 1 ή -1. Όταν οι τιμές εξόδου της συνάρτησης είναι 1 ή -1 η συνάρτηση βήματος έχει τη διπολική μορφή της (Διαμαντάρας, 2007). Εναλλακτικά, η τιμή 1 μπορεί να αντικατασταθεί από οποιαδήποτε σταθερά. Μια παραλλαγή είναι η γραμμική συνάρτηση κατωφλιού (Dunham, 2004).



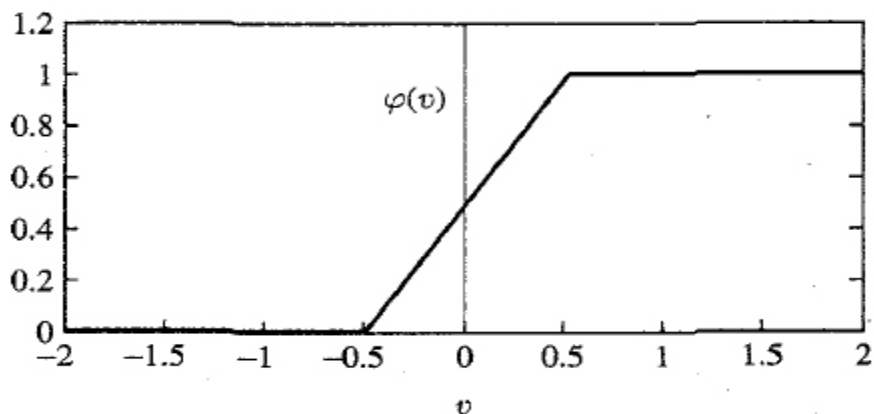
Εικόνα 1.1: Συνάρτηση Κατωφλιού ή Βήματος

### 1.3.3. Κατά Τμήματα Γραμμική Συνάρτηση (Piecewise-Linear function)

Ο τύπος της κατά τμήματα γραμμικής συνάρτησης είναι

$$\varphi(v) = \begin{cases} 1 & \text{αν } v \geq +\frac{1}{2} \\ v & \text{αν } +\frac{1}{2} > v > -\frac{1}{2} \\ 0 & \text{αν } v \leq -\frac{1}{2} \end{cases} \quad (1.2)$$

όπου ο παράγοντας αύξησης μέσα στη γραμμική περιοχή θεωρείται ότι είναι ενότητα. Η τιμή αυτού του τύπου συνάρτησης αυξάνεται σταδιακά από μια χαμηλή τιμή σε μια υψηλή τιμή. Αυτή η μορφή συνάρτησης ενεργοποίησης μπορεί να θεωρηθεί και ως μια προσέγγιση σε ένα μη-γραμμικό ενισχυτή (Dunham, 2004) (Haykin, 1999).



Εικόνα 1.2: Κατά Τμήματα Γραμμική Συνάρτηση

### 1.3.4. Σιγμοειδής Συνάρτηση (Sigmoid Function)

Η σιγμοειδής συνάρτηση, της οποίας το γράφημα έχει σχήμα s, είναι μακράν ο πιο κοινός τύπος συνάρτησης ενεργοποίησης που χρησιμοποιείται στην κατασκευή των τεχνητών νευρωνικών δικτύων. Ορίζεται σαν μια γνησίως αύξουσα συνάρτηση που ισορροπεί ανάμεσα στη γραμμική και τη μη-γραμμική συμπεριφορά. Έχει την παρακάτω μορφή:

$$\varphi(v) = \frac{1}{1 + e^{-v}} \quad (1.3)$$

Σε αντίθεση με τη συνάρτηση κατωφλιού, όπου παίρνει τις τιμές 0 ή 1, η σιγμοειδής συνάρτηση μπορεί να λάβει τιμές σε όλο το πεδίο από 0 έως 1. Επιπλέον η σιγμοειδής συνάρτηση παραγωγίζεται ενώ η συνάρτηση κατωφλιού όχι (Dunham, 2004).

### 1.3.5. Υπερβολική Εφαπτομένη (Hyperbolic Tangent Function)

Η συνάρτηση υπερβολικής εφαπτομένης είναι μια παραλλαγή της σιγμοειδούς συνάρτησης

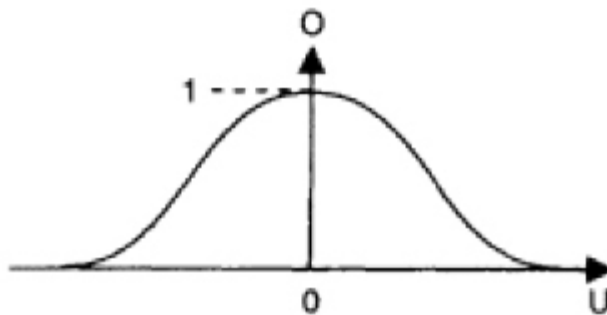
$$\varphi(v) = \tanh(v) = \frac{e - e^{-v}}{e + e^{-v}} \quad (1.4)$$

Αυτή η συνάρτηση έχει έξοδο με κέντρο το μηδέν, η οποία μπορεί να τη βοηθήσει στη μάθηση. (Dunham, 2004).

### 1.3.6. Γκαουσιανή Συνάρτηση (Gaussian Function)

Η γκαουσιανή συνάρτηση είναι μια καμπύλη σε σχήμα καμπάνας και παίρνει τιμές εξόδου στο διάστημα [0,1]. Μια τυπική γκαουσιανή συνάρτηση είναι η (Dunham, 2004) (Haykin, 1999):

$$\varphi(v) = e^{-\frac{v^2}{\sigma^2}} \quad (1.5)$$



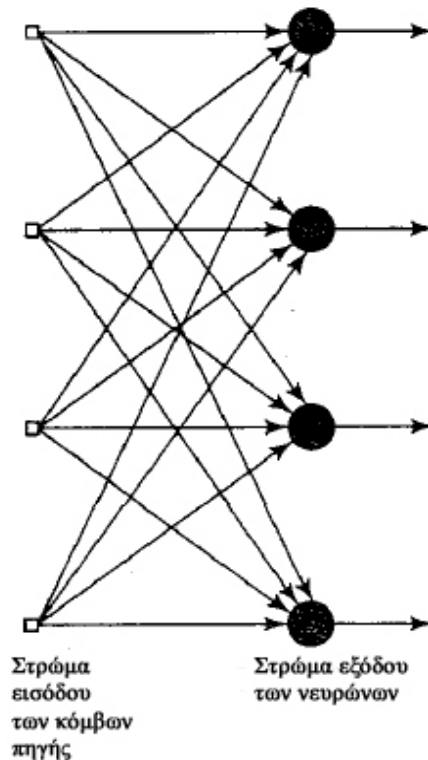
Εικόνα 1.3: Γκαουσιανή Συνάρτηση

## 1.4. 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.

### 1.4.1. Δίκτυα Ενός Στρώματος

Σε ένα νευρωνικό δίκτυο που αποτελείται από στρώματα, 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 feedforward. Η μορφή του φαίνεται στο παρακάτω σχήμα (Haykin, 1999).



Εικόνα 1.4: Δίκτυο Ενός Στρώματος

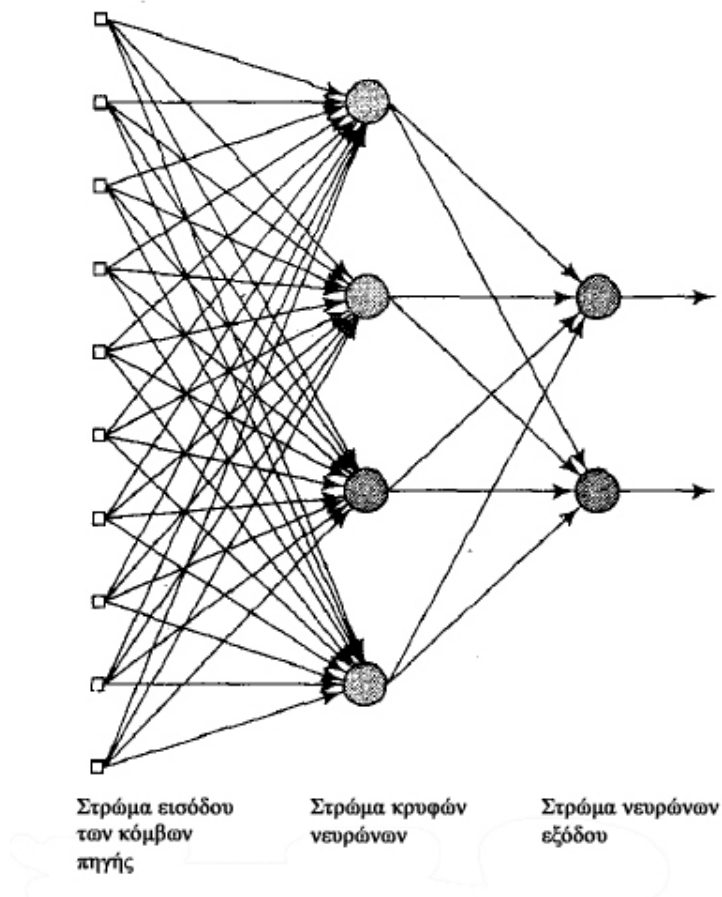
### 1.4.2. Δίκτυα Πολλών Στρωμάτων

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.



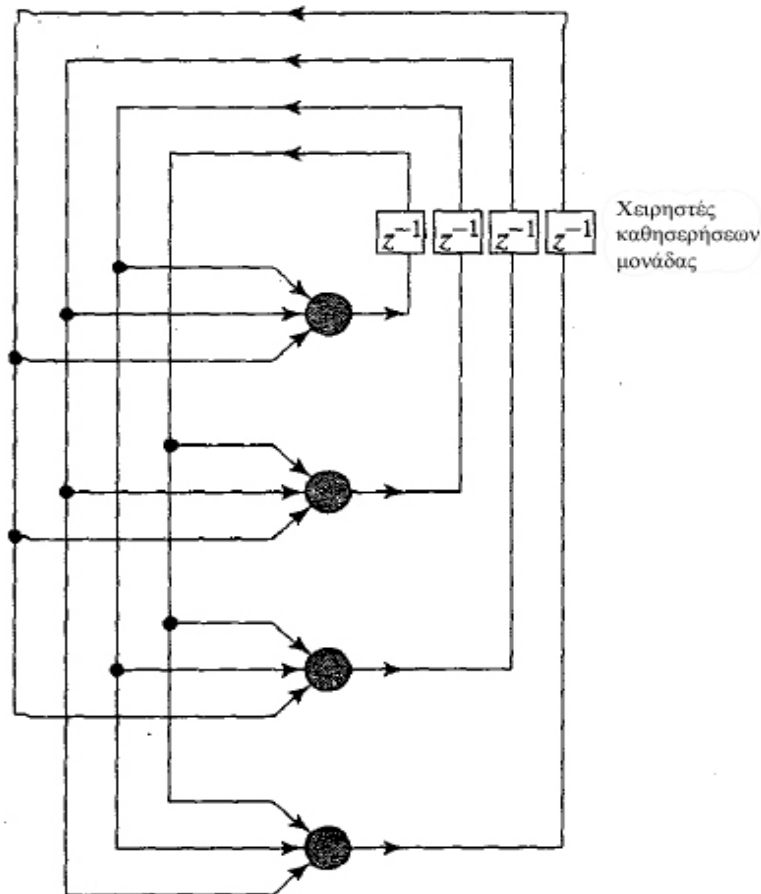


Εικόνα 1.5: Δίκτυο Πολλών Στρωμάτων

Το παραπάνω σχήμα απεικονίζει τη μορφή ενός δικτύου 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) (Haykin, 1999).

### 1.4.3. Επαναλαμβανόμενα Δίκτυα (Recurrent Networks)

Ένα επαναλαμβανόμενο νευρωνικό δίκτυο ξεχωρίζει από ένα 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. Στη δομή του δικτύου του σχήματος, δεν υπάρχουν αυτό-τροφοδοτούμενοι βρόγχοι. Αυτό σημαίνει ότι η έξοδος του νευρώνα δεν τροφοδοτεί πίσω την είσοδο του ίδιου νευρώνα (Haykin, 1999).



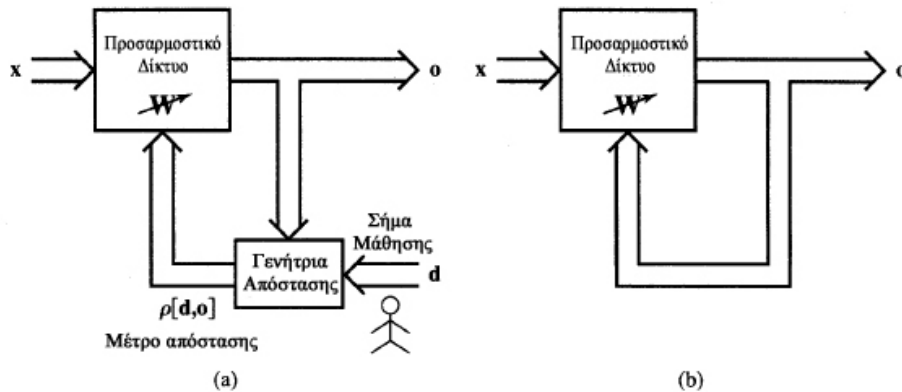
Εικόνα 1.6: Δομή Επαναλαμβανόμενου δικτύου

### 1.5. Εποπτευόμενη και Μη-εποπτευόμενη Μάθηση

Υπό την έννοια της μάθησης σε ένα δίκτυο, αναφέρεται μια διαδικασία που υποχρεώνει ένα δίκτυο να αποδώσει μια συγκεκριμένη απόκριση σε μια σαφή είσοδο. Η μάθηση είναι απαραίτητη όταν η πληροφορία σχετικά με τις εισόδους/ εξόδους είναι άγνωστη ή ελλιπής εκ των προτέρων, έτσι ώστε να μην είναι δυνατή η σχεδίαση του δικτύου. Η πλειοψηφία των νευρωνικών δικτύων πρέπει να εκπαιδευτούν με μία μέθοδο εκπαίδευσης που να είναι εποπτευόμενη ή μη-εποπτευόμενη.

Η μάθηση με ανατροφοδότηση είτε από τον εκπαιδευτή είτε από το περιβάλλον, είναι πιο τυπική για νευρωνικά δίκτυα. Αυτή η μάθηση λέγεται και στοιχειώδης (incremental) και συνήθως εκτελείται σε βήματα. Η γενική ιδέα της ανατροφοδότησης έχει βασικό ρόλο στην εκπαίδευση του δικτύου. Με την ευρύτερη έννοια μπορεί να γίνει κατανοητό σαν μια εισαγωγή ενός προτύπου συσχετίσεων στο μονοπάτι αιτίας – αποτελέσματος. Ξεχωρίζουν δύο

διαφορετικοί τύποι μάθησης, εκπαίδευση με επίβλεψη και χωρίς επίβλεψη. Στο παρακάτω σχήμα, φαίνονται τα διαγράμματα και των δύο τεχνικών εκπαίδευσης.



Εικόνα 1.7: (a) Μάθηση με επίβλεψη, (b) Μάθηση χωρίς επίβλεψη

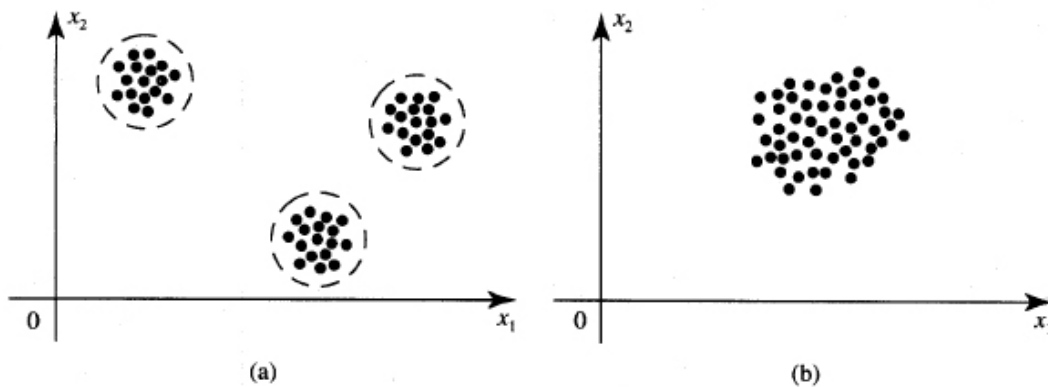
### 1.5.1. Μάθηση με επίβλεψη

Στην εποπτευόμενη μάθηση, γίνεται η υπόθεση ότι σε κάθε χρονική στιγμή όταν εφαρμόζεται η είσοδος, η επιθυμητή απόκριση  $d$  του συστήματος παρέχεται από τον «δάσκαλο». Η λειτουργία απεικονίζεται στο σχήμα 1.7(a). Η απόσταση  $\rho[d,o]$  ανάμεσα στην πραγματική και την επιθυμητή απόκριση εξυπηρετεί ως μια μέθοδος μέτρησης σφάλματος και χρησιμοποιείται για να διορθώσει τις παραμέτρους του δικτύου εξωτερικά. Εφόσον τα βάρη θεωρούνται ευπροσάρμοστα, ο «δάσκαλος» μπορεί να εκτελέσει ένα σχέδιο ανταμοιβής-και-τιμωρίας για να προσαρμόσει τον πίνακα βαρών του δικτύου  $W$ . Για παράδειγμα, στη μάθηση συσταδοποίησης των προτύπων εισόδου ή καταστάσεων με γνωστές αντιδράσεις, το σφάλμα μπορεί να χρησιμοποιηθεί για να τροποποιήσει τα βάρη έτσι ώστε το σφάλμα να μειωθεί. Αυτή η μέθοδος μάθησης είναι από τις πιο συνηθισμένες. Επίσης χρησιμοποιείται σε πολλές καταστάσεις φυσικής μάθησης. Ένα σύνολο από πρότυπα εισόδου και εξόδου το οποίο καλείται σύνολο εκπαίδευσης είναι απαραίτητο για τη συγκεκριμένη διαδικασία μάθησης.

Τυπικά, η μάθηση με επίβλεψη επιβραβεύει τις συσταδοποιήσεις ή τις συσχετίσεις που είναι ακριβείς, ενώ τιμωρεί εκείνες που αποφέρουν ανακριβείς αποκρίσεις. Ο «δάσκαλος» υπολογίζει το σφάλμα αρνητικής κλίσης (negative error gradient director) και μειώνει το σφάλμα αντίστοιχα. Σε πολλές περιπτώσεις, οι εισοδοί, οι έξοδοι και η εκτιμώμενη κλίση είναι ντετερμινιστικά, παρόλα αυτά, η ελαχιστοποίηση του σφάλματος προέρχεται όλων των τυχαίων κανονικοποιήσεων. Σαν αποτέλεσμα, οι περισσότεροι αλγόριθμοι εποπτευόμενης μάθησης μειώνονται στη στοχαστική ελαχιστοποίηση του σφάλματος σε έναν πολυδιάστατο χώρο βαρών.

### 1.5.2. Μάθηση χωρίς επίβλεψη

Η εικόνα 1.7(b) δείχνει το διάγραμμα της μάθησης χωρίς επίβλεψη. Στη μη-εποπτευόμενη μάθηση, η επιθυμητή απόκριση δεν είναι γνωστή, άρα δεν μπορούν να χρησιμοποιηθούν σαφείς πληροφορίες σφάλματος για να βελτιωθεί η συμπεριφορά του δικτύου. Εφόσον δεν υπάρχει διαθέσιμη πληροφορία για την ορθότητα ή ανακρίβεια των αποκρίσεων, η μάθηση πρέπει να έρχεται εις πέρας βάσει των παρατηρήσεων των αποκρίσεων των εισόδων, στις οποίες δεν υπάρχει γνώση ή είναι σχετικά ασήμαντη. Για παράδειγμα, η μη-εποπτευόμενη μάθηση μπορεί εύκολα να έχει ως αποτέλεσμα στο να βρει το όριο ανάμεσα σε δύο κλάσεις των προτύπων εισόδου όπως φαίνεται στην εικόνα 1.8.



Εικόνα 1.8: Δισδιάστατα πρότυπα: (a) Συσταδοποιημένα, (b) Συστάδες χωρίς εμφανή διαχωρισμό

Στην εικόνα 1.8, τα όρια των συστάδων είναι βασισμένα σε μεγάλα και αντιπροσωπευτικά δείγματα εισόδων και η θέση τους θεωρείται ευνοϊκή. Οι κατάλληλοι μηχανισμοί αυτό-προσαρμογής βαρών πρέπει να εμπεριέχονται στο εκπαιδευμένο δίκτυο, διότι δεν είναι διαθέσιμες εξωτερικές οδηγίες σχετικά με τις πιθανές συστάδες. Ένας κανόνας πιθανής προσαρμογής δικτύου είναι ο εξής: Ένα πρότυπο που προστίθεται στη συστάδα πρέπει να βρίσκεται πιο κοντά στο κέντρο της ίδιας συστάδας παρά στο κέντρο μιας οποιαδήποτε άλλης.

Οι αλγόριθμοι μη εποπτευόμενης μάθησης χρησιμοποιούν πρότυπα τα οποία συνήθως είναι περιττά και ακατέργαστα δεδομένα που δεν έχουν ετικέτες σε σχέση με την συμμετοχή των κλάσεων τους ή τους συνδέσμούς τους. Σε αυτή τη μέθοδο μάθησης, το δίκτυο πρέπει να ανακαλύψει μόνο του όλα τα πιθανά πρότυπα που υπάρχουν, τις ομοιογένειές του, τις ξεχωριστές ιδιότητες κ.τ.λ. Καθώς το δίκτυο εντοπίζει όλα αυτά, υποβάλλεται σε αλλαγή των παραμέτρων, οι οποίες καλούνται αυτό-οργάνωση (self-organization).

Η τεχνική της μάθησης με επίβλεψη συχνά χρησιμοποιείται για να υλοποιηθεί συσταδοποίηση ως τη μη επιβλεπόμενη κατηγοριοποίηση των αντικειμένων δίχως την παροχή

πληροφοριών για τις ακριβείς κατηγορίες. Αυτού του είδους η μάθηση αντιστοιχεί στην ελάχιστη εκ των προτέρων διαθέσιμη πληροφορία. Ορισμένες από τις πληροφορίες σχετικά με τον αριθμό των συστάδων, ή η ομοιότητα εναντίον της ανομοιότητας των προτύπων, μπορεί να είναι εξυπηρετικές για αυτή τη μέθοδο εκπαίδευσης. Εν τέλει, η μάθηση συχνά δεν είναι δυνατή σε ένα περιβάλλον χωρίς επίβλεψη, όπως θα ήταν αληθές για την περίπτωση b της εικόνας 1.8, όπου οι κατηγορίες των προτύπων δεν είναι ευδιάκριτες ακόμα και στον άνθρωπο.

Η μη επιβλεπόμενη μάθηση καλείται και μάθηση χωρίς δάσκαλο. Αυτή η ορολογία δεν θεωρείται ακριβής, γιατί η μάθηση χωρίς δάσκαλο δεν είναι δυνατό να πραγματοποιηθεί. Παρόλα αυτά, ο δάσκαλος σαν οντότητα δεν χρειάζεται να εμπλέκεται σε κάθε στάδιο της εκπαίδευσης. Πρέπει να θέτει τους στόχους στη μέθοδο μη επιβλεπόμενης μάθησης (Zurada, 1992).

## 1.6. 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.

$$v = \sum_{i=1}^n w_i x_i - \theta, \quad (1.6)$$

$$y = \varphi(v)$$

Η συνάρτηση μεταφοράς απεικονίζει το διάνυσμα εισόδου  $x = [x_1, x_2, \dots, x_n]^T$  στην έξοδο  $y$ . Οι παράμετροι  $w_1, w_2, \dots, w_n$  είναι τα συναπτικά βάρη. Η παράμετρος  $\theta$  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  $\theta$ :

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

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

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

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

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  $\theta$  (Διαμαντάρας, 2007).

### 1.6.1. 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  $d^{(p)}$  για το πρότυπο  $p$  διαφέρει από την έξοδο του δικτύου  $y = f(\mathbf{w}(k-1)^T \mathbf{x}^{(p)})$  όπου  $\mathbf{w}(k-1)$  είναι το επαυξημένο διάνυσμα των συναπτικών βαρών μετά την επανάληψη  $k-1$ . Όταν υπάρχει σφάλμα, τότε η διόρθωση των βαρών γίνεται προσθέτοντας ή αφαιρώντας ένα ποσοστό του προτύπου  $\mathbf{x}^{(p)}$ . Συγκεκριμένα, αν κατά την επανάληψη  $k$ , εισάγεται το πρότυπο  $p$  τότε ο κανόνας της διόρθωσης είναι:  $\mathbf{w}(k) = (\mathbf{w}(k-1) + \beta(d^{(p)} - y)\mathbf{x}^{(p)})$  όπου  $\mathbf{w}(k)$  είναι το επαυξημένο διάνυσμα συναπτικών βαρών μετά την επανάληψη  $k$ . 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  $\mathbf{w}(k) = (\mathbf{w}(k-1) + \beta(d^{(p)} - y)\mathbf{x}^{(p)})$  γράφεται εναλλακτικά ως εξής:

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

$$w_0(k) = (w_0(k-1) + \beta(d^{(p)} - \gamma)).$$

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

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

(1.7)

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

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

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

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

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

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

- Αν  $d^{(p)} = 1$  τότε η διέγερση  $u_k^{(p)}$  αυξάνεται πλησιάζοντας το επιθυμητό  $y = 1$ , ενώ
- Αν  $d^{(p)} = 0$  τότε η διέγερση  $u_k^{(p)}$  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 (Διαμαντάρας, 2007).

## 1.7. 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  $\varphi(u) = \frac{1}{1+e^{-u}}$  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 εφαπτομένη  $\tanh(u) = \frac{e^{-u}-e^u}{e^{-u}+e^u}$ , η οποία έχει την παρακάτω απλή σχέση με τη σιγμοειδή συνάρτηση  $\tanh(u) = 2\varphi(2u) - 1$ .

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 (Διαμαντάρας, 2007).

### 1.7.1. 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



εκπαίδευση σε οποιοδήποτε νευρωνικό δίκτυο. Χαρακτηριστικό της εκπαίδευσης ενός 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 αλγορίθμους εκπαίδευσης MLP είναι ο back-propagation.

Ο αλγόριθμος back-propagation προτάθηκε από τον Paul Werbos τη δεκαετία του 1970. 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.

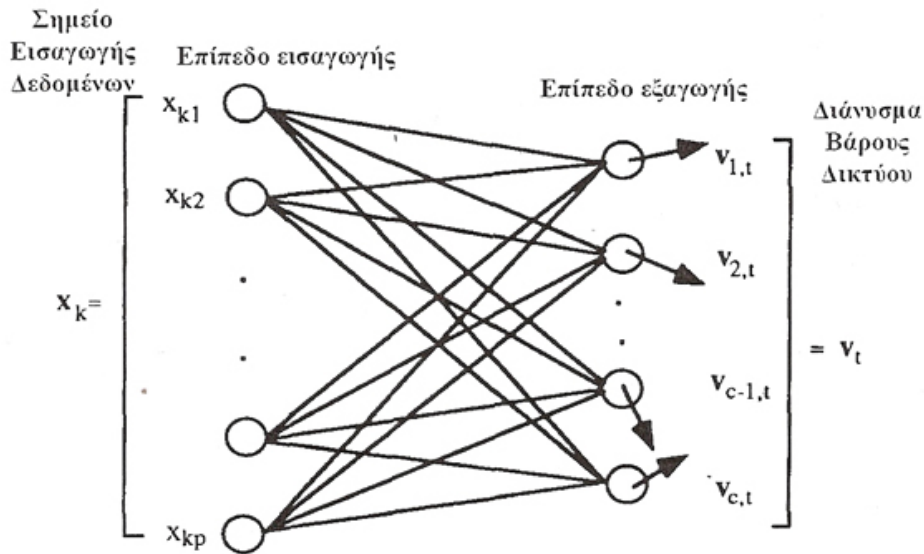
Βήμα 5: Ανατίθεται στους νευρώνες ένας συντελεστής ευθύνης για το τοπικό σφάλμα. Όσο μεγαλύτερος είναι ο συντελεστής βάρους τόσο μεγαλύτερη είναι και η υπευθυνότητα στους νευρώνες.

Βήμα 6: Ο αλγόριθμος επιστρέφει στο βήμα 3, έχοντας ως σφάλμα τους συντελεστές ευθύνης της προηγούμενης επανάληψης.

## 1.8. Δίκτυα Kohonen

Η συσταδοποίηση δικτύων του Kohonen (Kohonen, 1984) 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 επίπεδο εισόδου και ένα επίπεδο 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.



Εικόνα 1.10: Δομή Δικτύου Kohonen

### Αλγόριθμος Δικτύου Kohonen

Βήμα 1: Ορίζεται ο αριθμός των βαρών  $c$  και  $\epsilon > 0$ , κάποια μικρή σταθερά.

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

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

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

B. Ταξινόμηση της  $\{d_{tk}^2\}$  με σειρά από το μικρότερο στο μεγαλύτερο:  $d_{1k}^2 \leq d_{2k}^2 \leq \dots \leq d_{ck}^2$ .

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

D. Αν  $|N_c| = n_c$  ενημερώνονται οι  $(n_c - 1)$  δεσμοί οι οποίοι είναι πιο κοντά στο  $\mathbf{x}_k$  στο B:  $\mathbf{v}_{t,c} = \mathbf{v}_{t,c-1} + \alpha_{tk,t}(\mathbf{x}_k - \mathbf{v}_{t,c-1})$ ,  $t = 2, 3, \dots, (n_c - 1)$ .

E. Επόμενο  $k$ .

Βήμα 4: Υπολογισμός του  $E_c = \|\mathbf{v}_c - \mathbf{v}_{t,c}\|^2 = \sum_t \|\mathbf{v}_{t,c} - \mathbf{v}_{t,c-1}\|^2$ .

Βήμα 5: Αν  $E_c \leq \epsilon$  τότε ο αλγόριθμος σταματάει, διαφορετικά οι ρυθμοί μεταβολής  $\{\alpha_{tk,t}\}$  αλλάζουν και ενημερώνεται η γειτονιά  $N_c$ . Επόμενο  $t$ .

Χαρακτηριστικά του αλγόριθμου Kohonen είναι:

- Είναι σειριακός – ένα  $\mathbf{x}_k$  τροφοδοτείται στο δίκτυο, και κάποια βάρη ενημερώνονται αμέσως.

- Η παράμετρος  $\alpha$  έχει συνταχθεί στο  $L_k$  και  $L$ . Αυτό γίνεται εξαιτίας του ότι οι τιμές του  $\alpha$  που χρησιμοποιήθηκαν στην ενημέρωση των βαρών διανέμονται στο  $N_{\alpha}$ , κεντραρισμένες στο νικητή, το  $v_{1,t}$ , και η σύνταξη του  $w_1$  είναι μια συνάρτηση του  $x_k$ . Επιπλέον, το  $\alpha$  συνήθως μειώνεται με το χρόνο ( $t$ ) για να εξαναγκάσει το  $E_t \leq s$  (καθαρά  $E_t \rightarrow 0$  όπως  $a_{i,k,t} \rightarrow 0$ ). Αυτό είναι μια τεχνητή στρατηγική τερματισμού, στην οποία ο κανόνας ενημέρωσης εγγυάται την εξάρτηση των διαδοχικών επαναλήψεων μετά από αρκετό χρόνο, ακόμα και 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  $\{v_t\} \rightarrow \{v_s\}$  για  $\{a\} \rightarrow 0$ , όπου το  $v_s$  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.

## 1.9. Εφαρμογές Τεχνητών Νευρωνικών Δικτύων

Τα τεχνητά νευρωνικά δίκτυα είναι δημοφιλή σε προβλήματα που περιέχουν μη προβλέψιμες λειτουργίες και τα οποία δεν είναι πλήρως κατανοητά. Τέτοιου είδους προβλήματα σχετίζονται με κατηγοριοποίηση (classification), αναγνώριση (recognition), αποτίμηση (assessment) και πρόβλεψη (forecasting/prediction).

Πιο συγκεκριμένα, εφαρμογές που σχετίζονται με την κατηγοριοποίηση παρατηρούνται στον τομέα της ιατρικής, όπως κατηγοριοποίηση εικόνων που προέρχονται από εξετάσεις υπερήχων, ηλεκτροκαρδιογραφήματα, τεστ Παπανικολάου. Σε αυτές τις περιπτώσεις τα νευρωνικά δίκτυα κάνουν αρχικά μια πρώτη διάγνωση επιταχύνοντας την όλη διαδικασία, και στη συνέχεια οι γιατροί εξετάζουν τις περιπτώσεις που θεωρούνται μη-φυσιολογικές. Παραδείγματα υπάρχουν και στον τομέα της άμυνας, με κατηγοριοποίηση εικόνων προερχόμενων από συσκευές όπως radar και sonar. Εφαρμόζονται επιπλέον στη γεωργία όπου πραγματοποιείται έλεγχος καλλιεργειών σε συνδυασμό με δορυφορικά συστήματα τηλε-επισκόπησης, αλλά και στην οικονομία και τον τομέα των επιχειρήσεων με την κατηγοριοποίηση των πελατών βάσει των αγοραστικών τους συνηθειών.

Εφαρμογές σχετικές με την αναγνώριση χρησιμοποιούνται σε τράπεζες, για τη γνησιότητα υπογραφής και των τραπεζογραμμάτων, στην πληροφορική και τις τηλεπικοινωνίες για αναγνώριση ήχου, εικόνας και γραπτού κειμένου, είτε χειρόγραφο είτε τυπωμένο.

Η αποτίμηση συναντάται στον τομέα της άμυνας, παρακολούθηση στόχων, στην ασφάλεια, εντοπισμός κίνησης, ταύτιση δακτυλικών αποτυπωμάτων και ανάλυση εικόνας σε συστήματα επιτήρησης. Επίσης στη μηχανολογία με την παρακολούθηση, επιθεώρηση και τον έλεγχο των προϊόντων.

Τέλος, δραστηριότητες που σχετίζονται με την πρόβλεψη, χρησιμοποιούνται στην οικονομία για την πρόβλεψη ισοτιμίας νομισμάτων, βραχυπρόθεσμη εκτίμηση τιμών μετοχών και πρόβλεψη πωλήσεων, στη γεωργία για την πρόβλεψη της παραγωγής με τη χρήση δορυφορικών εικόνων, και στην μετεωρολογία για την πρόβλεψη του καιρού (Βλαχάβας, 2006).

## 2. ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΑΚΤΙΝΙΚΗΣ ΒΑΣΗΣ

### 2.1. Εισαγωγή

Το κίνητρο για τη μελέτη μεθόδων όπου χρησιμοποιούνται συναρτήσεις ακτινικής βάσης ξεκίνησε με την ιδέα να ερευνηθεί η σχεδίαση ενός νευρωνικού δικτύου που να προσεγγίζει προβλήματα σε χώρους μεγάλων διαστάσεων και να παρέχει καλύτερη κατηγοριοποίηση των δεδομένων εκπαίδευσης. Επιπλέον, η γενίκευση είναι ανάλογη στη χρήση της πολυδιάστατης επιφάνειας για να παρεμβάλει τα δεδομένα που χρησιμοποιούνται για δοκιμές. Στα πλαίσια των νευρωνικών δικτύων, οι κρυφοί κόμβοι παρέχουν ένα σύνολο συναρτήσεων που συγκροτούν μια αυθαίρετη βάση για τα πρότυπα εισόδου (διανύσματα) όταν αυτά εκτείνονται στον κρυφό χώρο. Αυτές οι συναρτήσεις ονομάζονται συναρτήσεις ακτινικής βάσης. Οι συναρτήσεις ακτινικής βάσης χρησιμοποιήθηκαν αρχικά για την επίλυση του προβλήματος παρεμβολής πολλών πραγματικών μεταβλητών. Οι πρώτες έρευνες ξεκίνησαν το 1985 από τον Powell. Η μέθοδος αυτή είναι πλέον ένα από τα σημαντικότερα πεδία στην αριθμητική ανάλυση (Haykin, 1999).

### 2.2. Συναρτήσεις Ακτινικής Βάσης

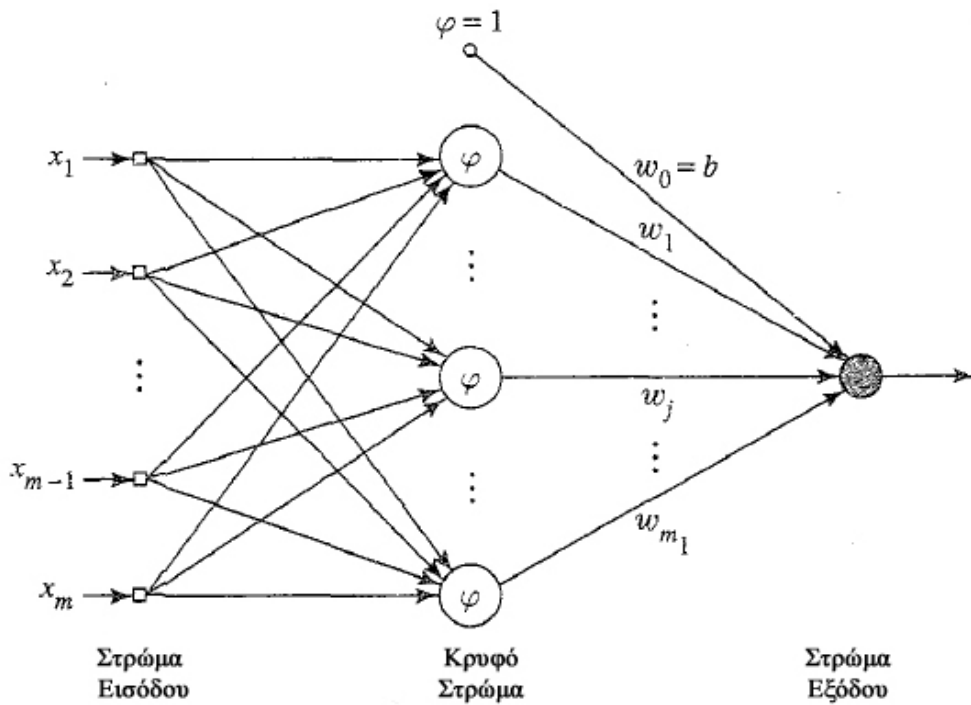
Μια ακτινική συνάρτηση ή συνάρτηση ακτινικής βάσης (radial basis function – RBF) είναι μια κατηγορία συναρτήσεων των οποίων η τιμή μειώνεται (ή αυξάνεται) με την απόσταση από ένα κεντρικό σημείο. Η γκαουσιανή συνάρτηση ενεργοποίησης που περιγράφεται σε προηγούμενη ενότητα είναι μια συνάρτηση ακτινικής βάσης με κεντρικό σημείο τα 0. Μια RBF έχει το σχήμα της γκαουσιανής συνάρτησης και ένα RBF δίκτυο είναι συνήθως ένα νευρωνικό δίκτυο με τρία στρώματα. Το στρώμα εισόδου χρησιμοποιείται για να εισάγει δεδομένα. Μια γκαουσιανή συνάρτηση ενεργοποίησης χρησιμοποιείται στο κρυμμένο στρώμα, ενώ μια γραμμική συνάρτηση ενεργοποίησης χρησιμοποιείται στο στρώμα εξόδου. Ο στόχος είναι να έχουμε τους κρυμμένους κόμβους να μαθαίνουν να αποκρίνονται μόνο σε ένα υποσύνολο της εισόδου, εκείνο στο οποίο η γκαουσιανή συνάρτηση είναι κεντραρισμένη. Αυτό συνήθως επιτυγχάνεται μέσω της εποπτευόμενης μάθησης. Όταν χρησιμοποιούνται συναρτήσεις RBF σαν συναρτήσεις ενεργοποίησης στο κρυμμένο στρώμα, οι κόμβοι πρέπει να είναι ευαίσθητοι σε ένα υποσύνολο των τιμών εισόδου (Dunham, 2004).

Οι συναρτήσεις ακτινικής βάσης (radial basis function) χαρακτηρίζονται από μια συνάρτηση  $f$  όπου υπάρχει κάποιο διάνυσμα  $c$  το οποίο λέγεται κέντρο (center ή centroid) και η τιμή της συνάρτησης εξαρτάται μόνο από την απόσταση του  $x$  από το κέντρο (Διαμαντάρας, 2007).

$$f(x) = f(\|x - c\|) \tag{2.1}$$

### 2.3. Δίκτυα Ακτινικής Βάσης

Η δομή ενός νευρωνικού δικτύου ακτινικής βάσης (radial basis function network –RBF) στην πιο βασική της μορφή, αποτελείται από τρία επίπεδα με εντελώς διαφορετικούς ρόλους το καθένα. Το στρώμα εισόδου (input layer) αποτελείται από μονάδες αισθητήρων (source nodes – sensory units) οι οποίες συνδέουν το δίκτυο με το περιβάλλον του. Το δεύτερο επίπεδο, το μόνο κρυφό στρώμα στο δίκτυο, εφαρμόζει έναν μη-γραμμικό μετασχηματισμό από το στρώμα εισόδου στο κρυφό στρώμα (hidden layer). Στις περισσότερες εφαρμογές το κρυφό στρώμα είναι of high dimensionality. Το στρώμα εξόδου (output layer) είναι γραμμικό και παρέχει την αντίδραση του δικτύου στο πρότυπο ενεργοποίησης (σήμα) που εφαρμόζεται στο στρώμα εισόδου (Haykin, 1999).



Εικόνα 2.1: Νευρωνικό Δίκτυο Ακτινικής Βάσης - Radial Basis Function Network

Τα δίκτυα ακτινικής βάσης έχουν τις ρίζες τους στη μαθηματική θεωρία της προσέγγισης συναρτήσεων και στην ουσία υλοποιούν μια συνάρτηση παρεμβολής (interpolation function), η οποία προσεγγίζει την τιμή μιας συνάρτησης σε κάποιο σημείο ως το μέσο όρο (με βάρη) των τιμών της συνάρτησης σε κοντινά σημεία.

Τα δίκτυα RBF είναι νευρωνικά δίκτυα feed forward με ένα μόνο κρυμμένο επίπεδο, του οποίου οι κρυμμένοι νευρώνες υπολογίζουν μια ειδική συνάρτηση του διανύσματος εισόδου (Λύκας, 2008).

## 2.4. Στρατηγικές Μάθησης

Η διαδικασία μάθησης που επιχειρείται από ένα δίκτυο συνάρτησης ακτινικής βάσης (RBF) ανεξάρτητα από το θεωρητικό υπόβαθρο, μπορεί να εικονιστεί όπως παρακάτω. Τα γραμμικά βάρη που σχετίζονται με τη μονάδα(ες) εξόδου τείνουν να εξελίσσονται σε διαφορετική χρονική κλίμακα σε σύγκριση με τις μη-γραμμικές συναρτήσεις ενεργοποίησης των κρυφών μονάδων. Άρα, καθώς οι συναρτήσεις ενεργοποίησης του κρυφού στρώματος εξελίσσονται αργά σύμφωνα με τη στρατηγική μη-γραμμικής βελτιστοποίησης, τα βάρη του στρώματος εξόδου προσαρμόζονται γρήγορα μέσω μιας στρατηγικής γραμμικής βελτιστοποίησης. Αυτό που είναι αξιοσημείωτο είναι ότι τα διαφορετικά επίπεδα του δικτύου RBF εκτελούν διαφορετικές ενέργειες, και άρα είναι λογικό να ξεχωρίζει η βελτιστοποίηση του κρυφού στρώματος από το στρώμα εξόδου του δικτύου χρησιμοποιώντας διαφορετικές τεχνικές, οι οποίες μπορεί να λειτουργούν σε διαφορετικές χρονικές κλίμακες.

Υπάρχουν διάφορες τεχνικές μάθησης που μπορούν να χρησιμοποιηθούν στη σχεδίαση ενός δικτύου RBF, και οι οποίες εξαρτώνται από τον τρόπο με τον οποίο έχουν καθοριστεί τα κέντρα των συναρτήσεων ακτινικής βάσης. Παρακάτω γίνεται λεπτομερής αναφορά αυτών των προσεγγίσεων (Haykin, 1999).

### 2.4.1. Σταθερά Κέντρα που Επιλέγονται Τυχαία (*Fixed Centers Selected at Random*)

Η πιο απλή προσέγγιση μάθησης, είναι να θεωρηθούν σταθερές συναρτήσεις ακτινικής βάσης που να ορίζουν τις συναρτήσεις ενεργοποίησης των κρυφών μονάδων. Οι θέσεις των κέντρων μπορούν να έχουν επιλεγεί τυχαία από το σύνολο των δεδομένων εκπαίδευσης. Αυτή η μέθοδος χρησιμοποιείται όταν τα δεδομένα εκπαίδευσης είναι κατανεμημένα με τέτοιο τρόπο ώστε να αντιπροσωπεύουν το υπάρχον πρόβλημα. Για τις συναρτήσεις ακτινικής βάσης, μπορεί να χρησιμοποιηθεί μια ισοτροπική γκαουσιανή συνάρτηση της οποίας η τυπική απόκλιση να είναι σταθερή σε σχέση με το εύρος των κέντρων. Συγκεκριμένα, μια (κανονικοποιημένη) συνάρτηση ακτινικής βάσης με κέντρο στο  $t_i$  ορίζεται ως

$$G(\|x - t_i\|^2) = \exp\left(-\frac{m_1}{d_{\max}^2} * (\|x - t_i\|^2)\right), i = 1, 2, \dots, m_1 \quad (2.2)$$

Όπου  $m_1$  είναι ο αριθμός των κέντρων και  $d_{\max}$  είναι η μέγιστη απόσταση ανάμεσα στα επιλεγμένα κέντρα. Στην πραγματικότητα, η τυπική απόκλιση, δηλαδή το εύρος, όλων των γκαουσιανών συναρτήσεων είναι σταθερό στο  $\sigma = \frac{d_{\max}}{\sqrt{2m_1}}$ .

Αυτός ο τύπος εξασφαλίζει ότι οι μεμονωμένες συναρτήσεις ακτινικής βάσης δεν είναι πολύ μυτερές ή πολύ επίπεδες, καταστάσεις των γκαουσιανών συναρτήσεων που πρέπει να αποφεύγονται. Σαν εναλλακτική στη σχέση  $\sigma = \frac{d_{\max}}{\sqrt{2m_1}}$  μπορούν να χρησιμοποιηθούν μεμονωμένα κλιμακωτά κέντρα με ευρύτερα πλάτη σε περιοχές με μικρότερη πυκνότητα δεδομένων, όπου απαιτείται πειραματισμός με τα δεδομένα εκπαίδευσης.

Οι μονοί παράμετροι που χρειάζεται να είναι γνωστοί για αυτή την προσέγγιση είναι τα γραμμικά πλάτη στο επίπεδο εξόδου του δικτύου. Μια απλή διαδικασία για να υλοποιηθεί είναι να χρησιμοποιηθεί η ψευδοαντίστροφη μέθοδος. Συγκεκριμένα, γίνεται  $\mathbf{w} = \mathbf{G}^+ \mathbf{d}$  όπου  $\mathbf{d}$  είναι η επιθυμητή απόκριση του διανύσματος στο σύνολο εκπαίδευσης. Ο πίνακας  $\mathbf{G}^+$  είναι ο ψευδοαντίστροφος του πίνακα  $\mathbf{G}$ , ο οποίος ορίζεται ως  $\mathbf{G} = \{g_{ji}\}$  όπου

$$g_{ji} = \exp\left(-\frac{m_1}{d^2} * (\|x_j - t_i\|^2)\right), j = 1, 2, \dots, N \text{ και } i = 1, 2, \dots, m_1 \quad (2.3)$$

όπου  $x_j$  είναι το  $j$ -οστό διάνυσμα εισόδου του συνόλου εκπαίδευσης (Haykin, 1999).

#### 2.4.2. Αυτό-Οργανωμένη Επιλογή των Κέντρων (Self-Organized Selection of Centers)

Βασικό πρόβλημα με τη μέθοδο των σταθερών κέντρων που περιγράφεται προηγουμένως, είναι το γεγονός ότι μπορεί να χρειάζεται ένα μεγάλο σύνολο εκπαίδευσης για ένα ικανοποιητικό επίπεδο επίδοσης. Ένας τρόπος για να ξεπεραστεί αυτός ο περιορισμός είναι να χρησιμοποιηθεί μι διαδικασία υβριδικής μάθησης, που να περιλαμβάνει δύο στάδια. Πρώτον, το αυτό-οργανωμένο στάδιο εκπαίδευσης, σκοπός του οποίου είναι να υπολογίσει τις κατάλληλες θέσεις για τα κέντρα των συναρτήσεων ακτινικής βάσης στο κρυφό στρώμα, και δεύτερον, το στάδιο μάθησης με επίβλεψη, το οποίο συμπληρώνει το σχέδιο του δικτύου υπολογίζοντας τα γραμμικά βάρη του στρώματος εξόδου. Αν και η διαδικασία batch μπορεί να χρησιμοποιηθεί για να εφαρμόσει αυτά τα δύο στάδια μάθησης είναι προτιμότερο να χρησιμοποιηθεί μια προσαρμοστική προσέγγιση (iterative=επαναληπτικό).

Για την αυτό-οργανωμένη διαδικασία μάθησης, είναι απαραίτητος ένας αλγόριθμος συσταδοποίησης που να διαμοιράζει το ήδη υπάρχον σύνολο δεδομένων σε υποκατηγορίες, κάθε



μια από τις οποίες χρειάζεται να είναι όσο το δυνατόν πιο ομοιογενής. Ένας τέτοιος αλγόριθμος είναι ο c-means, ο οποίος τοποθετεί τα κέντρα των συναρτήσεων ακτινικής βάσης στις περιοχές του χώρου εισόδου X, όπου υποβάλλονται σημαντικά δεδομένα.

Έχοντας αναγνωρίσει τα κέντρα των γκαουσιανών συναρτήσεων ακτινικής βάσης, και το κοινό τους πλάτος χρησιμοποιώντας τον c-Means ή την ενισχυμένη εκδοχή του, το επόμενο και τελικό στάδιο είναι η διαδικασία υβριδικής μάθησης, προκειμένου να υπολογιστούν τα βάρη του στρώματος εξόδου. Μια απλή μέθοδος για αυτό τον υπολογισμό είναι ο αλγόριθμος των ελάχιστων μέσων τετραγώνων (Least-mean-squares – LMS). Το διάνυσμα των σημάτων εξόδου που δημιουργήθηκε από τις κρυφές μονάδες, αποτελεί το διάνυσμα εισόδου στον αλγόριθμο LMS. Επίσης, ο αλγόριθμος συσταδοποίησης c-Means των κρυφών μονάδων και ο αλγόριθμος LMS των μονάδων εξόδου μπορούν να κάνουν τους δικούς τους υπολογισμούς με παράλληλο τρόπο, ώστε να επιταχυνθεί η διαδικασία εκπαίδευσης (Haykin, 1999).

### 2.4.3. Επιβλεπόμενη Επιλογή των Κέντρων

Στην τρίτη στρατηγική μάθησης, τα κέντρα των συναρτήσεων ακτινικής βάσης, και όλοι οι άλλοι ελεύθεροι παράμετροι του δικτύου υφίστανται μια επιβλεπόμενη διαδικασία μάθησης. Με άλλα λόγια το RBF δίκτυο λαμβάνει την πιο γενική του μορφή. Ένας αλγόριθμος που χρησιμοποιείται για μια τέτοια διαδικασία είναι αυτός της μάθησης διόρθωσης λάθους (error-correction learning), ο οποίος εφαρμόζεται με μια διαδικασία βαθμίδας καθόδου (gradient-descent) που αναπαριστά τη γενίκευση του αλγόριθμου Least Mean Square.

Το πρώτο βήμα στην ανάπτυξη αυτής της διαδικασίας μάθησης, είναι να οριστεί η στιγμιαία τιμή της συνάρτησης κόστους

$$E = \frac{1}{2} \sum_{j=1}^N e_j^2 \quad (2.4)$$

όπου το N είναι το μέγεθος του δείγματος εκπαίδευσης, και  $e_j$  είναι το σήμα σφάλματος που ορίζεται από:

$$e_j = d_j - F^*(x_j) = d_j - \sum_{i=1}^M w_i G(\|x_j - \mathbf{t}_i\|, c_i) \quad (2.5)$$

Η απαίτηση είναι να βρεθούν οι ελεύθεροι παράμετροι  $w_i, \epsilon_p$  και  $\Sigma^{-1}$  ( η τελευταία συσχετίζεται με τον κανόνα στάθμισης (norm-weighting) του  $C_i$ ) έτσι ώστε να ελαχιστοποιηθεί το E (Haykin, 1999).

#### 2.4.4. Πλήρης Παρεμβολή με Κανονικοποίηση (Strict Interpolation with Regularization)

Η πλήρης παρεμβολή με κανονικοποίηση είναι μια μέθοδος σχεδίασης δικτύων συναρτήσεων ακτινικής βάσης που συνδυάζει τα στοιχεία της θεωρίας κανονικοποίησης και της εκτίμησης παλινδρόμησης πυρήνα. Η μέθοδος περιλαμβάνει τη συνδυαστική χρήση των ακόλουθων στοιχείων.

1. Μια συνάρτηση ακτινικής βάσης  $G$ , η οποία με τη χρήση ιεράρχησης θεωρείται ως ο πυρήνας μιας (μέσης τετραγωνικής) σταθεράς, γνωστή ως υπολογισμός οπισθοδρόμησης.
2. Έναν διαγώνιο πίνακα norm-weighting εισόδων  $\Sigma^{-1}$  κοινό για όλα τα κέντρα με εισόδους  $\Sigma = \text{diag}(h_1, h_2, \dots, h_{m_0})$  όπου  $h_1, h_2, \dots, h_{m_0}$  είναι τα εύρη ζώνης κατά διάσταση ενός σταθερού υπολογισμού οπισθοδρόμησης με κλιμακωτό πυρήνα  $G$ , όπως διατυπώθηκε προηγουμένως, και  $m_0$  είναι η διάσταση του χώρου εισόδου.
3. Κανονικοποίηση με πλήρη παρεμβολή που περιλαμβάνει εκπαίδευση για τα γραμμικά βάρη.
4. Επιλογή της παραμέτρου κανονικοποίησης  $\lambda$  και των παραγόντων κλιμάκωσης  $\alpha_1, \alpha_2, \dots, \alpha_{m_0}$  των εισόδων. Όσον αφορά την επιλογή αυτών των παραμέτρων, ισχύει ότι όσο μεγαλύτερο είναι η τιμή του  $\lambda$ , τόσο μεγαλύτερος είναι και ο θόρυβος που φθείρει τις μετρήσεις των παραμέτρων. Επίσης όταν η συνάρτηση βάσης είναι μονοτροπικού πυρήνα, η παράμετρος  $\alpha_i$  είναι πιο ευαίσθητη στη συνολική έξοδο του δικτύου που σχετίζεται με την διάσταση της εισόδου. Αντίθετα, όταν είναι μεγάλη, τόσο λιγότερο είναι σχετική η έξοδος του δικτύου με τη διάσταση της εισόδου. Άρα η παράμετρος  $\alpha_i$  μπορεί να χρησιμοποιηθεί για να αξιολογήσει τη σχετική σημασία των μεταβλητών εισόδου και επομένως να υποδείξει ποιες μεταβλητές είναι κατάλληλες για μείωση της διαστατικότητας (Haykin, 1999).

## 2.5. Σύγκριση Multi Layer Perceptron και Radial Basis Function

Τόσο τα δίκτυα MLP όσο και τα δίκτυα RBF έχουν κοινό χαρακτηριστικό ότι υλοποιούν μη γραμμικές απεικονίσεις από το χώρο των εισόδων  $\mathbb{R}^d$  στο χώρο των εξόδων  $\mathbb{R}^p$ . Και τα δυο δίκτυα χαρακτηρίζονται από την ιδιότητα της παγκόσμιας προσέγγισης, επομένως μπορούν να προσεγγίσουν (με οποιαδήποτε ακρίβεια) οποιαδήποτε συνάρτηση και για όσο μεγάλο αριθμό νευρώνων χρησιμοποιηθεί. Ωστόσο, χαρακτηρίζονται και από σημαντικές διαφορές, τις οποίες πρέπει κανείς να λαμβάνει υπόψη όταν επιλέγει ποιο από τα δύο δίκτυα θα χρησιμοποιήσει.

Η θεμελιώδης διαφορά μεταξύ MLP και RBF βρίσκεται στη μορφή των συναρτήσεων  $h_i(x)$  του κρυμμένου επιπέδου. Κάθε κρυμμένος νευρώνας του MLP υπολογίζει το εσωτερικό γινόμενο των τιμών των εισόδων με το αντίστοιχο διάνυσμα βαρών, και το αποτέλεσμα περνάει από τη συνάρτηση ενεργοποίησης του νευρώνα. Στην ουσία κάθε κρυμμένη μονάδα στο MLP ορίζει την εξίσωση ενός υπερεπιπέδου, δηλαδή το MLP υλοποιεί συναρτήσεις ταξινόμησης διαχωρίζοντας το χώρο των προτύπων με υπερεπίπεδα και ορίζοντας τις περιοχές απόφασης ως τομές των υπερεπιπέδων αυτών. Επομένως, στον υπολογισμό μιας εξόδου συμμετέχουν όλες οι κρυμμένες μονάδες και για το λόγο αυτό λέμε ότι ένα δίκτυο MLP μαθαίνει χρησιμοποιώντας καταναμημένη αναπαράσταση, δηλαδή η γνώση σχετικά με την έξοδο κατανέμεται στις τιμές των βαρών όλων των κρυμμένων νευρώνων.

Αντίθετα τα δίκτυα RBF λειτουργούν δημιουργώντας τοπικές αναπαραστάσεις, δηλαδή η γνώση σχετικά με την έξοδο που αντιστοιχεί σε κάποια είσοδο  $x$  ενσωματώνεται στα βάρη εκείνων των κρυμμένων νευρώνων οι οποίοι περιλαμβάνουν το πρότυπο  $x$  στη σφαίρα επιρροής τους. Δηλαδή, στην περίπτωση των δικτύων RBF ο χώρος προτύπων διαιρείται σε περιοχές, καθεμιά από τις οποίες αντιστοιχεί σε μια μονάδα του κρυμμένου επιπέδου. Αυτό βρίσκεται σε πλήρη αντιδιαστολή με την περίπτωση του δικτύου MLP, όπου κάθε περιοχή ορίζεται ως η τομή πολλών υπερεπιπέδων, που καθένα ορίζεται από έναν κρυμμένο νευρώνα.

Άλλες διαφορές μεταξύ MLP και RBF είναι ότι στην περίπτωση του MLP είναι δυνατό να έχουμε περισσότερα από ένα κρυφά επίπεδα, ενώ τα δίκτυα RBF έχουν μόνο ένα κρυμμένο επίπεδο. Επίσης, τα δίκτυα MLP εκπαιδεύονται με ενιαίο τρόπο (δηλαδή ενημερώνονται όλα τα βάρη με τον ίδιο αλγόριθμο), ενώ στα δίκτυα RBF γίνεται συνήθως εκπαίδευση δύο σταδίων με διαφορετικούς αλγόριθμους σε κάθε στάδιο (Λύκας, 2008).

### 3. ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ

#### 3.1 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 ομαδοποίηση (Quick Clustering).

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

$$J_{c\text{-Means}} = \sum_{k=1}^c \sum_{i=1}^n R_{ik} \|x_i - v_k\|^2 \quad (3.1)$$

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  $\mu_{ik}$  είναι η συνάρτηση συμμετοχής του  $\mathbf{x}_k$  στην  $i$ -οστή συστάδα και δίνεται από την παρακάτω σχέση, που είναι γνωστή ως ο νόμος του κοντινότερου γείτονα,

$$\mu_{ik} = \begin{cases} 1, & \text{αν } \|\mathbf{x}_k - \mathbf{v}_i\|^2 = \min_j \{\|\mathbf{x}_k - \mathbf{v}_j\|^2\} \\ 0 & , \text{ αλλιώς} \end{cases} \quad (3.2)$$

Η ελαχιστοποίηση της  $I_{c\text{-Means}}$  επιτυγχάνεται αν υπολογίζουμε τα κέντρα με βάση την παρακάτω εξίσωση,

$$\mathbf{v}_i = \frac{\sum_{k=1}^n \mu_{ik} \mathbf{x}_k}{\sum_{k=1}^n \mu_{ik}} \quad (3.3)$$

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 (3.1) δίνεται στη συνέχεια.

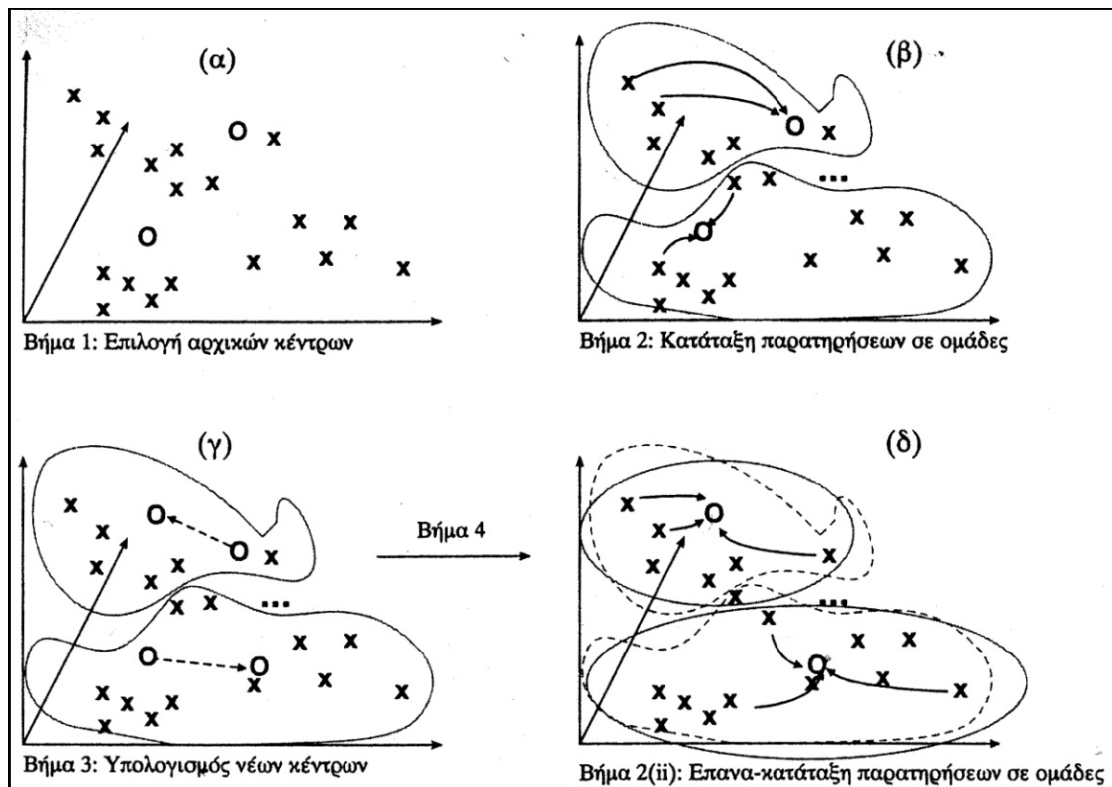
### **Αλγόριθμος c-Means**

Βήμα 1: Τυχαία επιλογή των αρχικών τιμών των κέντρων των συστάδων  $\mathbf{v}_i$  ( $1 \leq i \leq c$ ).

Βήμα 2: Προσδιορισμός των συναρτήσεων συμμετοχής με βάση την εξίσωση (3.2) και κατάταξη των διανυσματικών δεδομένων στις συστάδες.

Βήμα 3: Υπολογισμός των καινούργιων κέντρων των συστάδων (διανύσματα κωδικοποίησης) με βάση την εξίσωση (3.3).

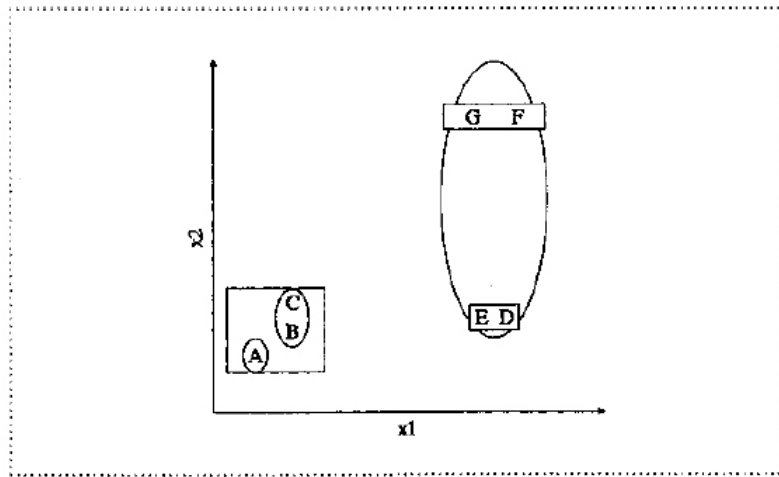
Βήμα 4: Προσδιορισμός των συναρτήσεων συμμετοχής με βάση την εξίσωση (3.2) και ανακατάταξη των διανυσματικών δεδομένων στις συστάδες.



Εικόνα 3.1: Επαναληπτική διαδικασία του αλγόριθμου c-Means.

Η εικόνα 3.1 δείχνει πώς δουλεύει ο αλγόριθμος 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 της εικόνας β. Στη συνέχεια, από όλες τις παρατηρήσεις που έχουν καταταχθεί υπολογίστηκαν τα νέα κέντρα των συστάδων. Έτσι, στην εικόνα 3.1 (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.



Εικόνα 3.2: 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 ομαδοποίηση. 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 της εικόνας 3.2 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 ομαδοποίηση 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 ομαδοποιήσεων 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 ομαδοποίηση.

Ακόμα, παρατηρείται πως η δυναμική του αλγορίθμου 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.

### 3.2 Αλγόριθμος Fuzzy c-Means

Πριν αναφερθούμε στον εν λόγω αλγόριθμο, κρίνεται απαραίτητο να αναλύσουμε τις βασικές έννοιες της ασαφούς λογικής.

#### 3.2.1 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 και του ασαφούς 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 είναι το ασαφές σύνολο. Η κύρια δομική διαφορά μεταξύ του ασαφούς συνόλου και του διακριτού συνόλου είναι ο τρόπος συμμετοχής των στοιχείων τους σε αυτά. Οι ορισμοί που ακολουθούν προσδιορίζουν τις δομές των δύο συνόλων.

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

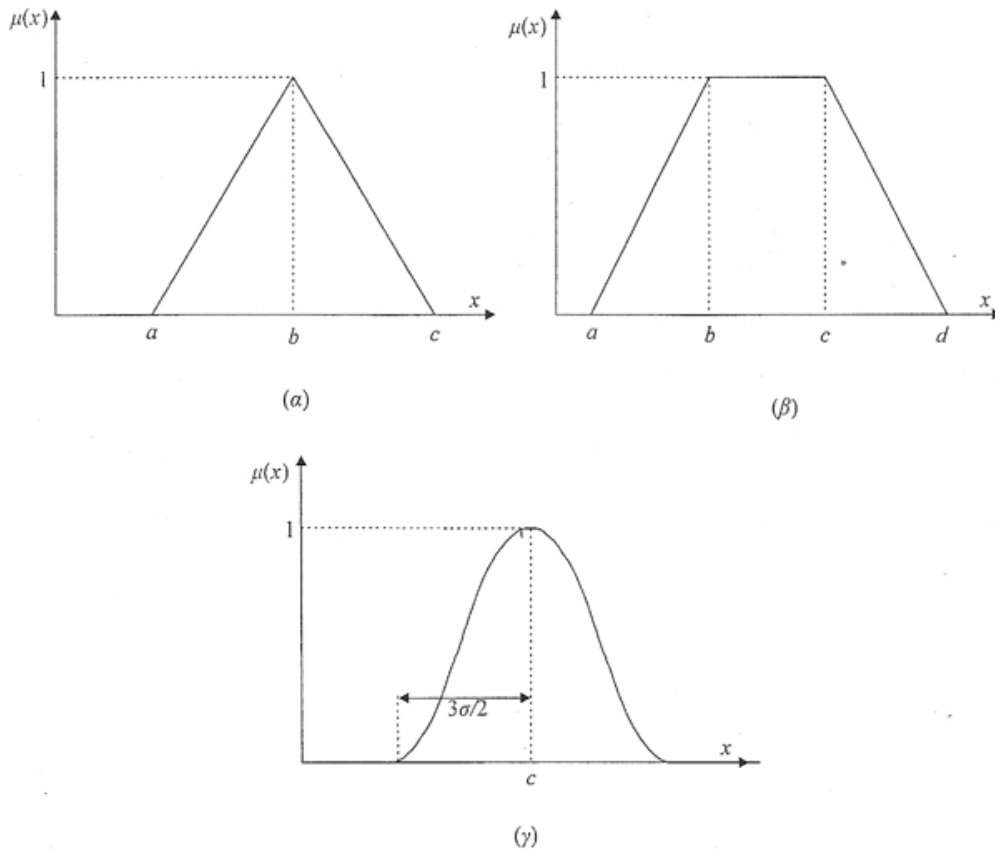
$$\mu_A = \begin{cases} 1, & \text{αν } x \in A \\ 0, & \text{αλλιώς} \end{cases} \quad \mu_B: U \rightarrow \{0,1\} \quad (3.4)$$

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



$$\mu_A = \begin{cases} f(x) \in [0, 1], & \text{αν } x \in A \\ 0 & , \text{ αλλιώς} \end{cases} \quad \text{με } \mu_A: U \rightarrow [0, 1] \quad (3.5)$$

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



Εικόνα 3.3: Ασαφή σύνολα με: (α) τριγωνική, (β) τραπεζοειδή και (γ) γκαουσιανή συνάρτηση συμμετοχής

- *Τριγωνική Συνάρτηση Συμμετοχής*

Η τριγωνική συνάρτηση συμμετοχής δίνεται από την παρακάτω σχέση,

$$\mu(x) = \begin{cases} 0 & , \text{ αν } x < a \\ \frac{x-a}{b-a}, & \text{ αν } a < x < b \\ -\frac{x-c}{c-b}, & \text{ αν } b < x < c \\ 0 & , \text{ αν } x \geq c \end{cases} \quad (3.6)$$

Όπως φαίνεται στην εικόνα 3.3(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,

$$\mu(x) = \begin{cases} 0 & , \text{αν } x < a \\ \frac{x-a}{b-a} & , \text{αν } a < x < b \\ 1 & , \text{αν } b < x < c \\ -\frac{x-d}{d-b} & , \text{αν } c < x < d \\ 0 & , \text{αν } x \geq d \end{cases} \quad (3.7)$$

Διαγραμματικά η τραπεζοειδής συνάρτηση συμμετοχής φαίνεται στην εικόνα 3.3(β)

- *Γκαουσιανή Συνάρτηση Συμμετοχής*

Η γκαουσιανή συνάρτηση συμμετοχής δίνεται από την παρακάτω σχέση,

$$\mu(x) = \exp\left[-\frac{(c-x)^2}{\sigma^2}\right] \quad (3.8)$$

Η εικόνα 3.3(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-Means θα οδηγήσει σε τοπικό ελάχιστο. Αυτό οφείλεται στο γεγονός ότι υπάρχει περίπτωση πολλά κέντρα συστάδων να μην κινηθούν καθόλου κατά τη διάρκεια εκτέλεση 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 (Klir 1995, Pedrycz 1984, Pedrycz 1993):

- 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».

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 (Bezdek and Pal, 1999). 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.

### 3.2.2 Ο Αλγόριθμος των Ασαφών c-Μέσων (Fuzzy c-Means)

Η πιο διαδεδομένη μέθοδος ασαφούς ανάλυσης συστάδων είναι ο αλγόριθμος Fuzzy c-Means (Bezdek, 1973). Ο αλγόριθμος αυτός είναι αντίστοιχος με τον κλασικό 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 (Bezdek, 1973) (Bezdek and Pal, 1992):

$$J_m = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \|x_k - v_i\|^2 \quad (3.9)$$

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  $v_i = [v_{i,1}, v_{i,2}, \dots, v_{i,p}]^T$  ( $1 \leq i \leq c$ ) είναι το διάνυσμα που αντιστοιχεί στο κέντρο της  $i$ -οστής συστάδας. Το κριτήριο  $J_m$  βελτιστοποιείται ως προς τις παραμέτρους  $u_{ik}$  και  $v_i$ . Το πρόβλημα είναι η ελαχιστοποίηση της  $J_m$  υπό τον ακόλουθο περιορισμό,

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \quad (3.10)$$

Πιο συγκεκριμένα η προσέγγιση έχει ως εξής:

Για σταθερές τιμές των  $u_{ik}$ , οι τιμές των  $v_i$ , οι οποίες ελαχιστοποιούν το  $J_m$  δίνονται από την εξίσωση (Bezdek, 1974):

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \quad (3.11)$$

Για σταθερές τιμές των  $v_i$ , 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 (3.10) είναι:

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_j\|^2}{\|x_k - v_i\|^2} \right)^{\frac{1}{m-1}}} \quad (1 \leq k \leq n, 1 \leq i \leq c) \quad (3.12)$$

Με βάση την παραπάνω ανάλυση ο αλγόριθμος Fuzzy c-Means διατυπώνεται ως εξής (Bezdek, 1973) (Bezdek and Pal, 1992):

### Αλγόριθμος Fuzzy c-Means

Έστω ένα σύνολο διανυσματικών δεδομένων  $X = \{x_1, x_2, \dots, x_n\}$  με  $x_k \in R^p$  ( $1 \leq k \leq n$ ), 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  $v_1, v_2, \dots, v_c$  και της παραμέτρου  $\epsilon$ .

Βήμα 2: Χρήση της εξίσωσης (3.12) για τον υπολογισμό των συναρτήσεων συμμετοχής  $u_{ik}$  ( $1 \leq k \leq n, 1 \leq i \leq c$ ).

Βήμα 3: Με βάση την εξίσωση (3.11) γίνεται προσδιορισμός των νέων τιμών για τα κέντρα των ασαφών υποομάδων  $v_1^{NEW}, v_2^{NEW}, \dots, v_c^{NEW}$ .

Βήμα 4: Αν  $\max_i \{\|v_i - v_i^{NEW}\|^2\} < \epsilon$  τότε ο αλγόριθμος σταματάει, αλλιώς θέτει  $v_i = v_i^{NEW}$  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 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 ομάδων. 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  $1/c$ , δηλαδή τα στοιχεία ανήκουν σε όλες τις υπό μελέτη συστάδες με τον ίδιο βαθμό πίστης (Βαζιργιάννης, 2005). Σε αυτή την περίπτωση όλα τα κέντρα συστάδων θα τραβηχτούν προς τα διανυσματικά δεδομένα πολύ αργά.

## 4. ΑΛΓΟΡΙΘΜΟΙ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ

### 4.1. Πρώτος Αλγόριθμος

Η πρώτη μέθοδος που χρησιμοποιήθηκε για το πείραμα, προτάθηκε από τους Moody και Darken το 1989. Σε αυτή τη μέθοδο χρησιμοποιήθηκε ο αλγόριθμος c-Means για την εύρεση των κέντρων των συναρτήσεων βάσης.

Η διασπορά των συναρτήσεων βάσης υπολογίστηκε από το μέσο όρο της απόστασης από κέντρο της συστάδας από τους  $\tau$  κοντινότερους γείτονές της. Δίνεται από τη σχέση:

$$\sigma_i = \frac{1}{\tau} \sqrt{\sum_{j=1}^{\tau} d_{ij}^2} \quad (3.14)$$

Όπου  $d_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|$  με  $i \neq j$ , και με συνήθη τιμή για το  $\tau$  να είναι:  $\tau = 2$  (Moody, Darken, 1989).

Τα βάρη,  $w_j$ , μπορούν να υπολογιστούν με μεθόδους ελαχίστων τετραγώνων, επειδή η έξοδος του νευρωνικού δικτύου συνάρτησης ακτινικής βάσης είναι γραμμική σε αυτά. Έτσι, για κάθε σημείο δεδομένο  $x_k$ , οι έξοδοι του νευρώνα υπολογίζονται από τη σχέση:

$$s_{kj} = g_j(x_k) \quad (3.15)$$

Και επομένως τοποθετούνται στον πίνακα  $\mathbf{A}=[s_{kj}]$ , όπου  $g_j(x_k)$  είναι η γκαουσιανή συνάρτηση. Εισάγοντας το διάνυσμα βάρους  $\mathbf{w} = [w_1, w_2, \dots, w_r]$  μπορεί να γραφεί η παρακάτω εξίσωση πίνακα για ολόκληρο το σύνολο δεδομένων:

$$\mathbf{\hat{y}} = \mathbf{A}\mathbf{w} \quad (3.16)$$

Τέλος τα ελάχιστα τετράγωνα υπολογίζουν τα βάρη  $\mathbf{w}$  που ελαχιστοποιούν το σφάλμα του δικτύου  $\mathbf{s} = \mathbf{\hat{y}} - \mathbf{y}$  που είναι:

$$\mathbf{w} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \mathbf{\hat{y}} \quad (3.17)$$

## 4.2. Δεύτερος Αλγόριθμος

Η δεύτερη μέθοδος προτάθηκε από τον Bezdek το 1995. Σε αυτή χρησιμοποιήθηκε ο αλγόριθμος Fuzzy c-Means για την εύρεση των κέντρων των συναρτήσεων βάσης, στην οποία με τη χρήση ασαφούς λογικής, υπολογιζόταν το ποσοστό της συμμετοχής του κάθε δεδομένου για όλες τις προκαθορισμένες συστάδες. Έπειτα από έναν ικανοποιητικό αριθμό επαναλήψεων το δεδομένο τοποθετούνταν στη συστάδα εκείνη στην οποία η τιμή της συνάρτησης συμμετοχής του είχε το μεγαλύτερο ποσοστό. Σε αυτή τη μέθοδο, τα κέντρα των συναρτήσεων ακτινικής βάσης, συνέπεσαν με τα κέντρα των συστάδων που υπολόγισε ο αλγόριθμος συσταδοποίησης Fuzzy c-Means. Τα συναπτικά βάρη βρέθηκαν και εδώ με τη μέθοδο των ελαχίστων τετραγώνων (least squares) όπως και στην πρώτη μέθοδο.

## 4.3. Τρίτος Αλγόριθμος

Ο τρίτος αλγόριθμος που χρησιμοποιήθηκε, προτάθηκε από τους Τσεκούρα και Νείρο το 2008. Και σε αυτή τη μέθοδο, τα κέντρα των συναρτήσεων βάσης βρέθηκαν με τον Fuzzy c-Means.

Η διασπορά των κόμβων των νευρωνικών δικτύων συνάρτησης ακτινικής βάσης υπολογίζεται ως εξής:

$$\sigma^j = \frac{2 * d_{max}^j}{3}, (1 \leq j \leq c) \quad (3.18)$$

το  $d_{max}$  δίνεται παρακάτω:

$$d_{max} = \left\{ \max_k \left\{ \|x_k - v_j\|^2 \right\} : x_k \in C_j, \text{έτσι ώστε } u_{jk} \geq 0.01 (1 \leq j \leq c) \right\} \quad (3.19)$$

Όπου  $C_j$  είναι η  $j$ -οστή συστάδα και  $u_{jk}$  ο βαθμός συμμετοχής του  $k$ -οστού διανυσματικού δεδομένου εκπαίδευσης στην  $j$ -οστή τελική ασαφή συστάδα.

Τέλος, τα βάρη υπολογίζονται με τη χρήση των ελάχιστων τετραγώνων και οι έξοδοι του νευρωνικού με τη γκαουσιανή συνάρτηση.

## 5. ΠΕΙΡΑΜΑΤΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο πειραματικό μέρος της παρούσας εργασίας τα αποτελέσματα που προέκυψαν συγκρίθηκαν μεταξύ τους με δύο τρόπους.

### 5.1. Πρώτο Πείραμα

Αρχικά, πραγματοποιήθηκε σύγκριση των αποτελεσμάτων μεταξύ των προσομοιώσεων όπου διαφοροποιούνταν οι τιμή του αριθμού των κοντινότερων γειτόνων,  $pn$ , του κέντρου των συστάδων για τον αλγόριθμο που χρησιμοποιήθηκε στην πρώτη μέθοδο, και η ασαφής παράμετρος,  $m$ , για τους αλγορίθμους της δεύτερης και της τρίτης μεθόδου.

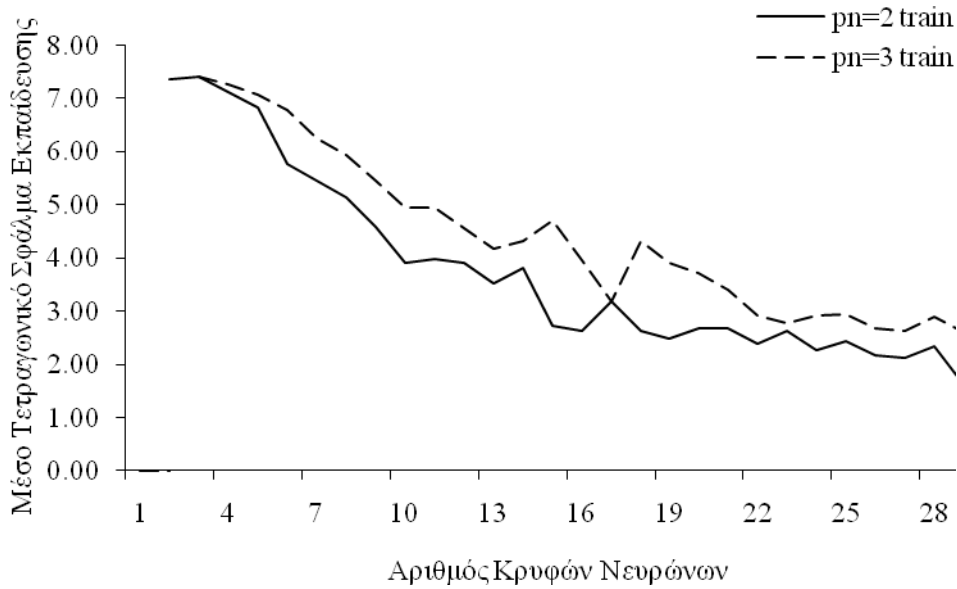
Παρακάτω παρουσιάζονται τα γραφήματα τόσο για την εκπαίδευση όσο και για τον έλεγχο του νευρωνικού δικτύου τριών διαφορετικών περιπτώσεων δεδομένων. Το πρώτο σύνολο δεδομένων είναι πραγματικού τύπου και είναι μετρήσεις που προέρχονται από ένα πείραμα που μελετούσε τη διαδικασία παλαίωσης οίνων. Ο αριθμός των δεδομένων ήταν 120 εκ των οποίων τα 80 χρησιμοποιήθηκαν για την εκπαίδευση του νευρωνικού δικτύου και τα υπόλοιπα 40 για την δοκιμή του νευρωνικού. Το δεύτερο σύνολο δεδομένων είναι μια προσέγγιση τριδιάστατης συνάρτησης  $y = (1 + x_1^{-2} + x_2^{-1.5} + x_3^{-1})^2$  με  $x_1, x_2, x_3 \in [1,5]$ , εκ των οποίων 120 από τα δεδομένα χρησιμοποιήθηκαν για εκπαίδευση και 80 για δοκιμή. Το τρίτο σύνολο δεδομένων είναι μια προσέγγιση μονοδιάστατης συνάρτησης,

$$y = 0.6 \sin(\pi x) + 0.3 \sin(3\pi x) + 0.1 \sin(5\pi x) \text{ όπου } x \in [-1,1]$$

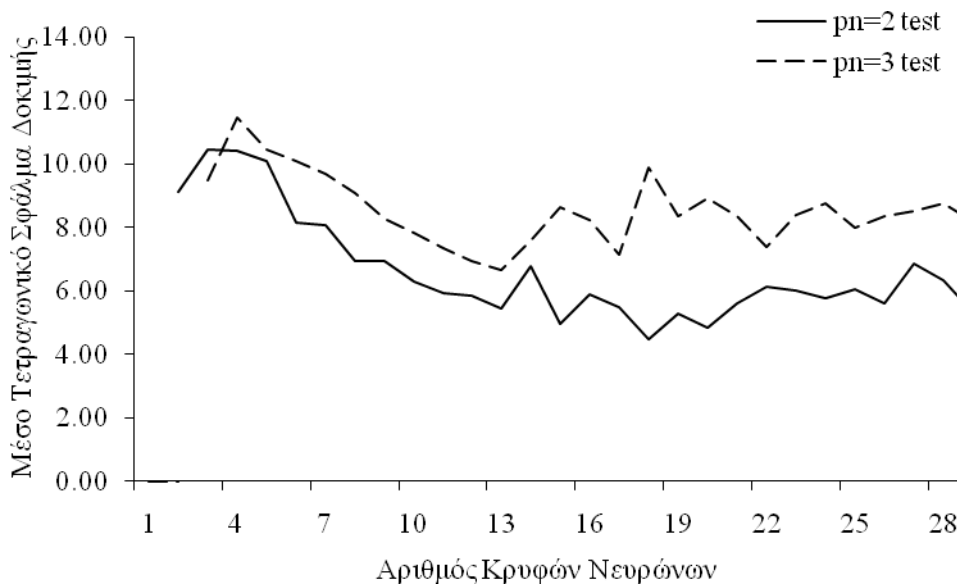
Πάλι, 120 δεδομένα χρησιμοποιήθηκαν για εκπαίδευση και 80 για δοκιμή του νευρωνικού.

Παρακάτω, φαίνονται τα γραφήματα των αποτελεσμάτων για το πρώτο σύνολο δεδομένων που προέκυψαν από τις προσομοιώσεις του πρώτου αλγόριθμου (Moody and Darken, 1989). Οι αρχικές τιμές των παραμέτρων που επιλέχθηκαν ήταν: αριθμός εισόδων = 3, αριθμός εξόδων = 1, και αριθμός κρυμμένων νευρώνων = 30. Για τον αριθμό κοντινότερων γειτόνων ( $pn$ ) επιλέχθηκαν δύο τιμές, 2 και 3, για να συγκριθούν τα αποτελέσματα που προέκυπταν. Πράγματι, όπως φαίνεται και στα γραφήματα, όταν ο αριθμός των κοντινότερων γειτόνων είναι δύο το νευρωνικό συγκλίνει παίρνοντας χαμηλότερες τιμές στο μέσο τετραγωνικό σφάλμα (Mean Square Error – MSE) τόσο της εκπαίδευσης όσο και της δοκιμής.





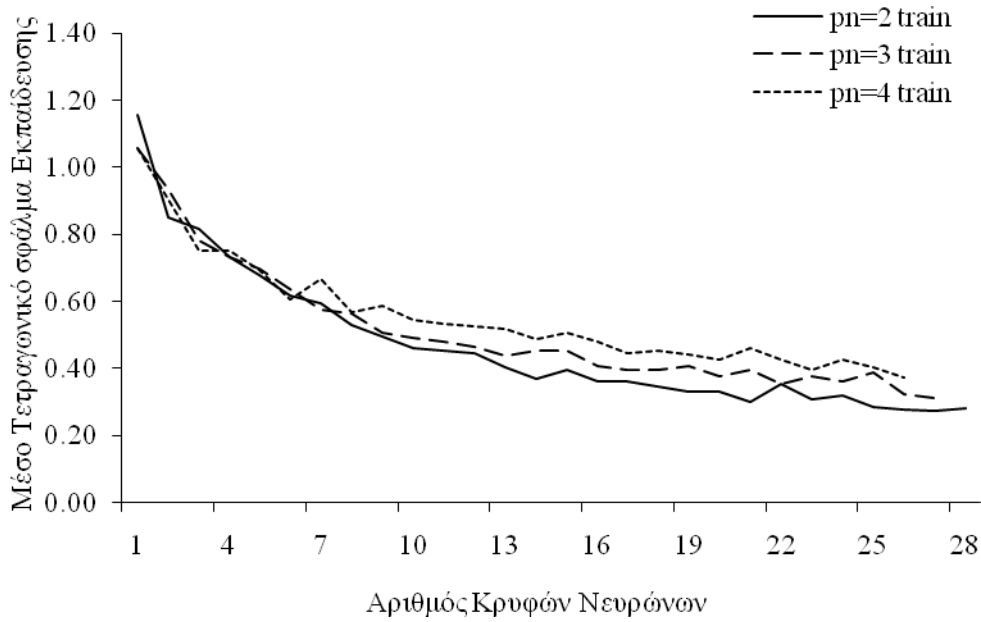
Εικόνα 5.1: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την πρώτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων



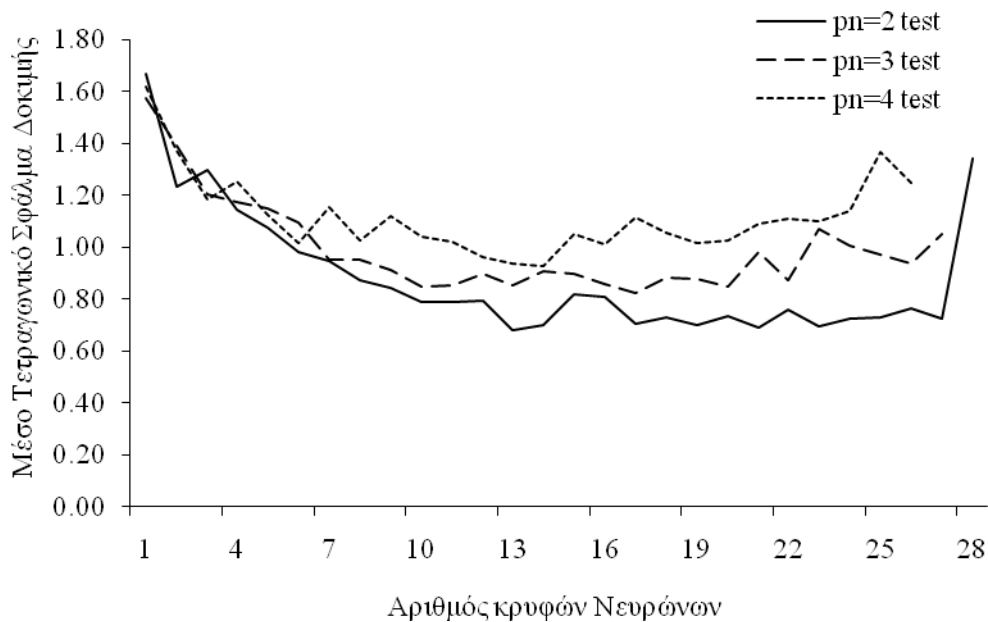
Εικόνα 5.2: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την πρώτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων

Για τα γραφήματα του δεύτερου συνόλου δεδομένων, των προσομοιώσεων του πρώτου αλγόριθμου (Moody and Darken, 1989), έχουν επιλεγεί οι ίδιες αρχικές τιμές για τις εκάστοτε παραμέτρους, και για τον αριθμό κοντινότερων γειτόνων ( $pn$ ), οι τιμές 2,3, και 4. Πάλι, και για αυτό το σύνολο δεδομένων, το μέσο τετραγωνικό σφάλμα της εκπαίδευσης και της δοκιμής συγκλίνει παίρνοντας τις χαμηλότερες τιμές του για  $pn=2$ , ακολουθούν οι τιμές του MSE για

$pn=3$ , και τέλος για  $pn=4$  παίρνει τις υψηλότερες τιμές του MSE για τις τιμές που δόθηκαν στο  $pn$  σε αυτό το πείραμα.

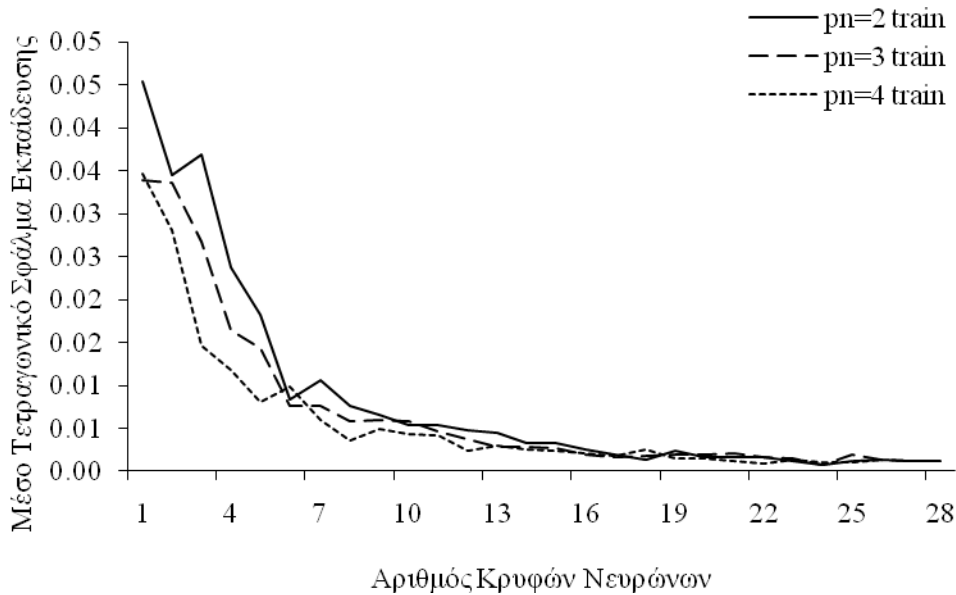


Εικόνα 5.3: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την πρώτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων

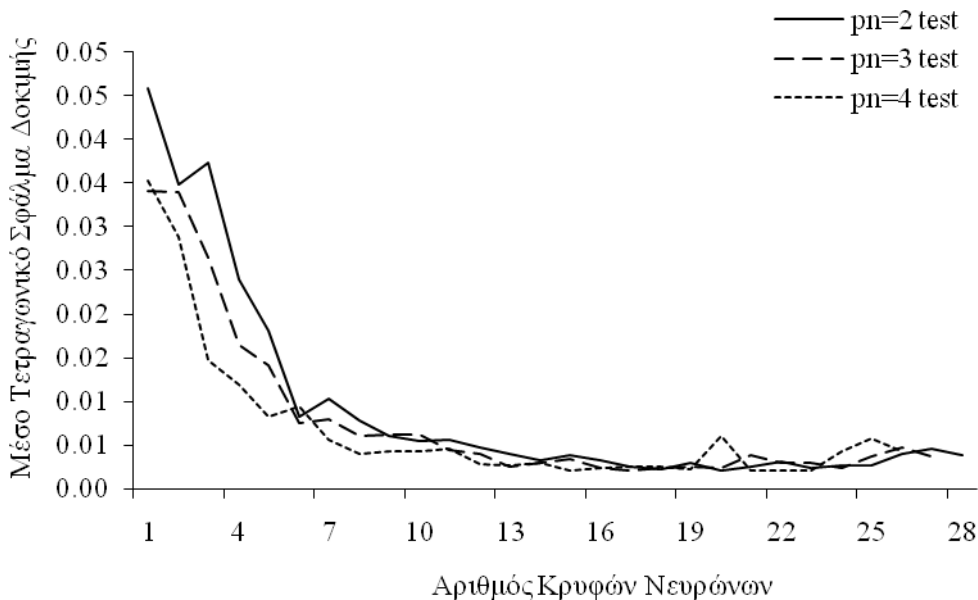


Εικόνα 5.4: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την πρώτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων

Στις προσομοιώσεις του τρίτου συνόλου δεδομένων, διατηρήθηκαν οι τιμές των αρχικών παραμέτρων στον αλγόριθμο, όπως και στις περιπτώσεις των προηγούμενων συνόλων δεδομένων, και δόθηκαν στο  $pn$  οι τιμές 2,3, και 4. Παρατηρείται ότι η σύγκλιση του μέσου τετραγωνικού σφάλματος και για τις τρεις τιμές του  $pn$  παίρνει τιμές που είναι κοντά μεταξύ τους και δεν είναι εμφανές για ποιο  $pn$  οι τιμές είναι χαμηλότερες.

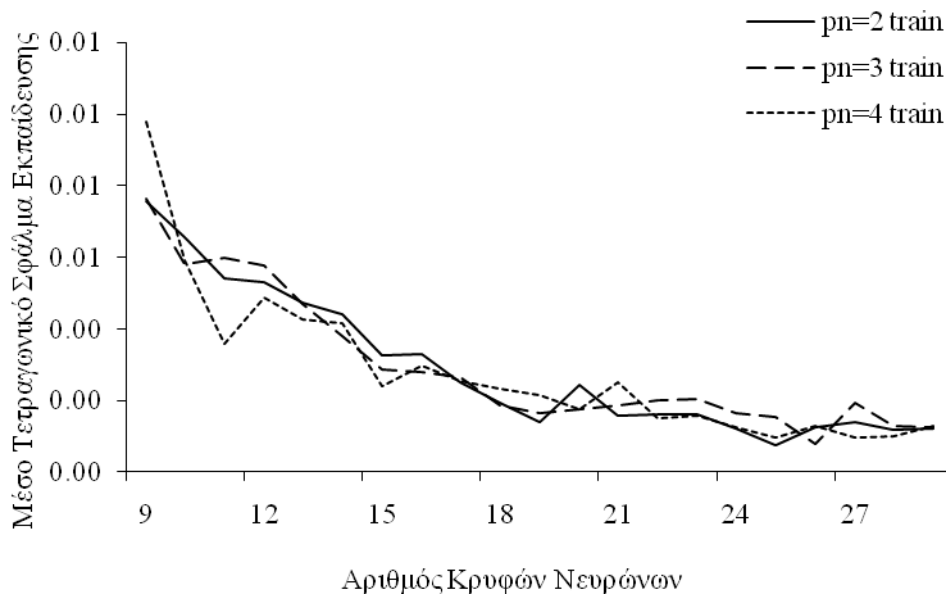


Εικόνα 5.5: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την πρώτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων



Εικόνα 5.6: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την πρώτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων

Όντως, αν μετατραπούν τα παραπάνω γραφήματα έτσι ώστε η σύγκλιση να φαίνεται από το 10<sup>0</sup> κρυφό κόμβο και έπειτα, συνεχίζει να μην είναι εμφανές για ποια τιμή προκύπτουν καλύτερα αποτελέσματα, αφού οι γραμμές των αποτελεσμάτων καθώς συγκλίνουν, τέμνονται σε κάποιους από τους κρυφούς κόμβους.

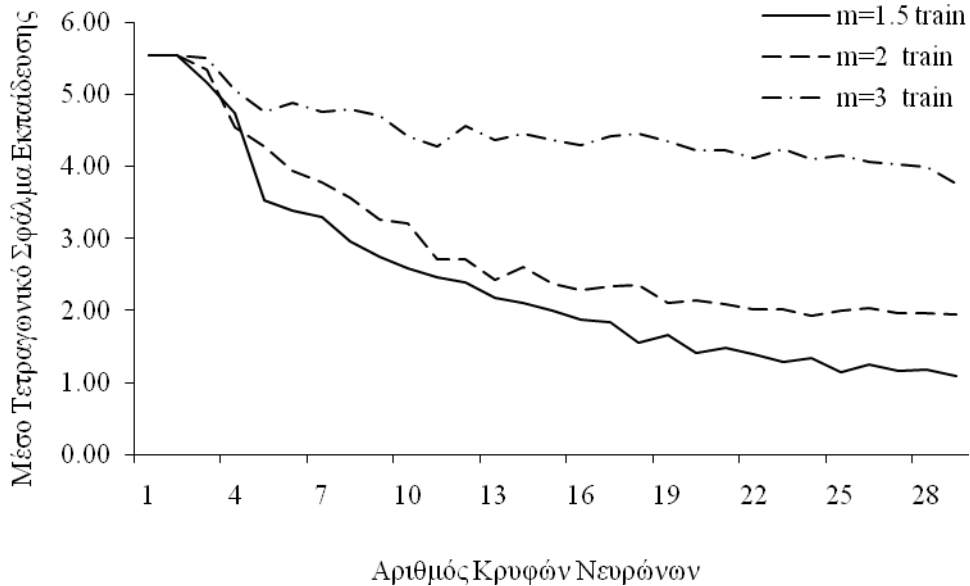


Εικόνα 5.7: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την πρώτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων από το 10<sup>0</sup> κρυφό νευρώνα

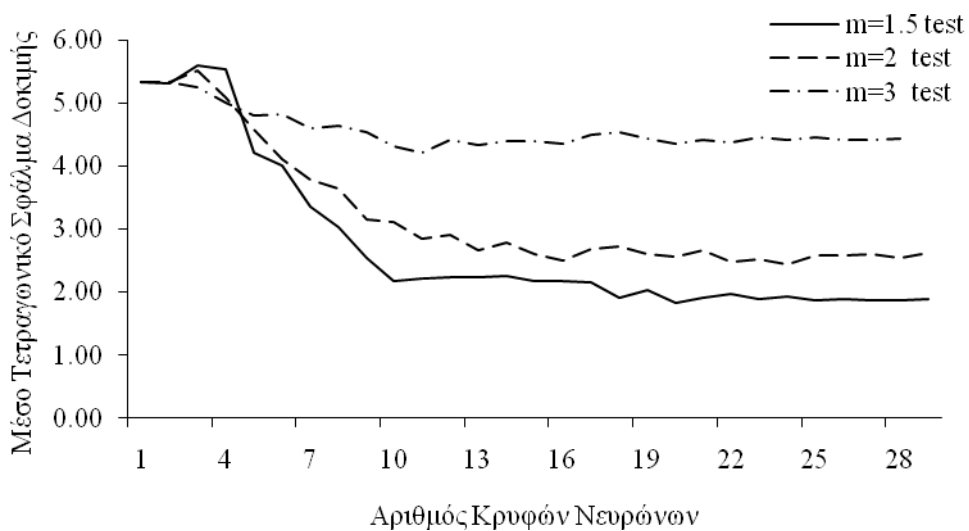


Εικόνα 5.8: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την πρώτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορους αριθμούς των κοντινότερων γειτόνων από το 10<sup>0</sup> κρυφό νευρώνα

Στο δεύτερο αλγόριθμο που χρησιμοποιήθηκε (Bezdek, 1995) οι παράμετροι που διατηρήθηκαν σταθερές είναι ο αριθμός των εξόδων με τιμή 1 και το πλήθος των κρυφών νευρώνων με τιμή 30, και άλλαξε η παράμετρος ασάφειας στην οποία δόθηκαν οι τιμές 1.5, 2, και 3. Τα αποτελέσματα που προέκυψαν για το μέσο τετραγωνικό σφάλμα της εκπαίδευσης και της δοκιμής καθώς συγκλίνουν οι γραμμές των γραφημάτων, έχουν χαμηλότερες τιμές για  $m=1.5$ , ακολουθούν οι τιμές για  $m=2$ , και τέλος οι τιμές του MSE για  $m=3$ . Έτσι για το πρώτο σύνολο δεδομένων το μέσο τετραγωνικό σφάλμα, φαίνεται παρακάτω.

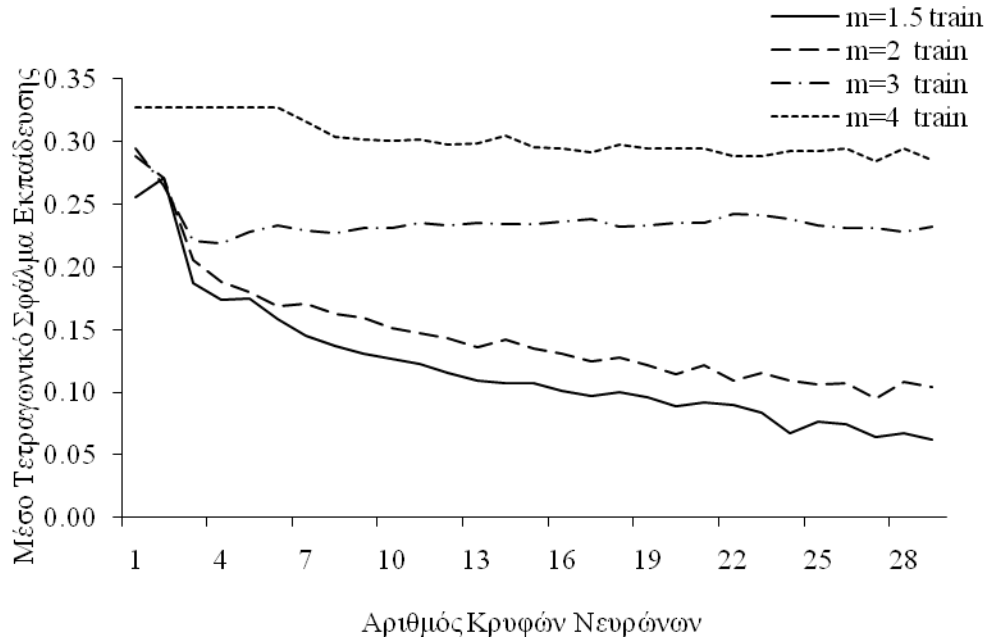


Εικόνα 5.9: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την δεύτερη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

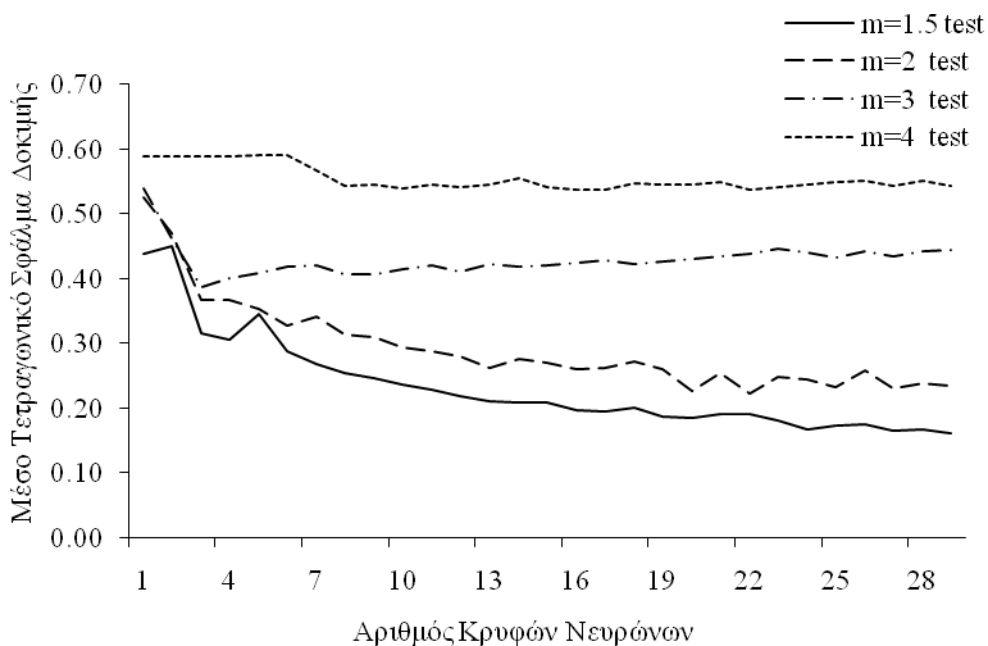


Εικόνα 5.10: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την δεύτερη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

Στο δεύτερο σύνολο δεδομένων δόθηκαν στην παράμετρο ασάφειας οι τιμές 1.5, 2, 3, και 4. Πάλι οι υπόλοιπες παράμετροι διατήρησαν τις ίδιες τιμές που είχαν και στο προηγούμενο σύνολο δεδομένων. Το μέσο τετραγωνικό σφάλμα της εκπαίδευσης και της δοκιμής παίρνει χαμηλότερες τιμές καθώς θέτονται μικρότερες τιμές στην παράμετρο  $m$ .

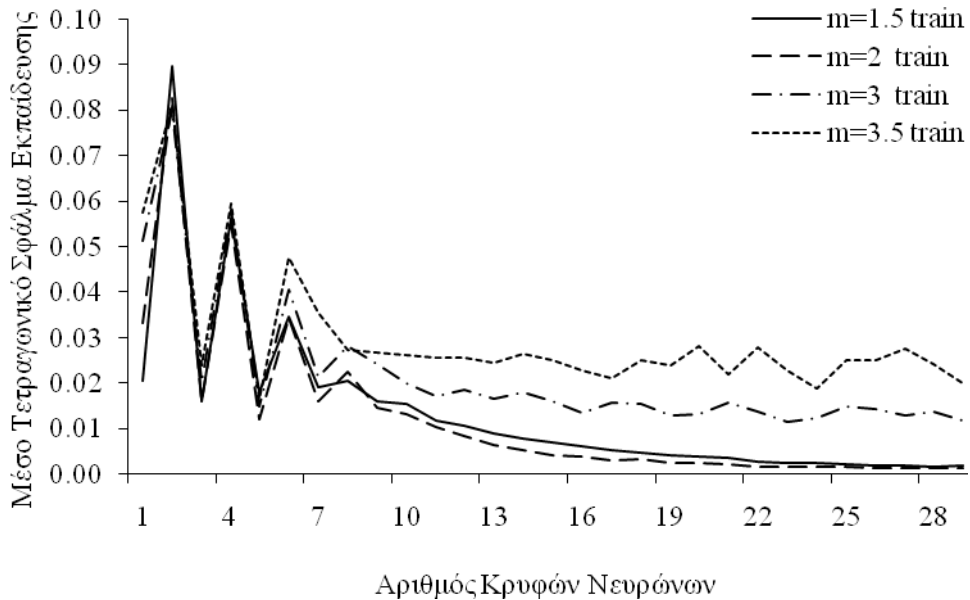


Εικόνα 5.11: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την δεύτερη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

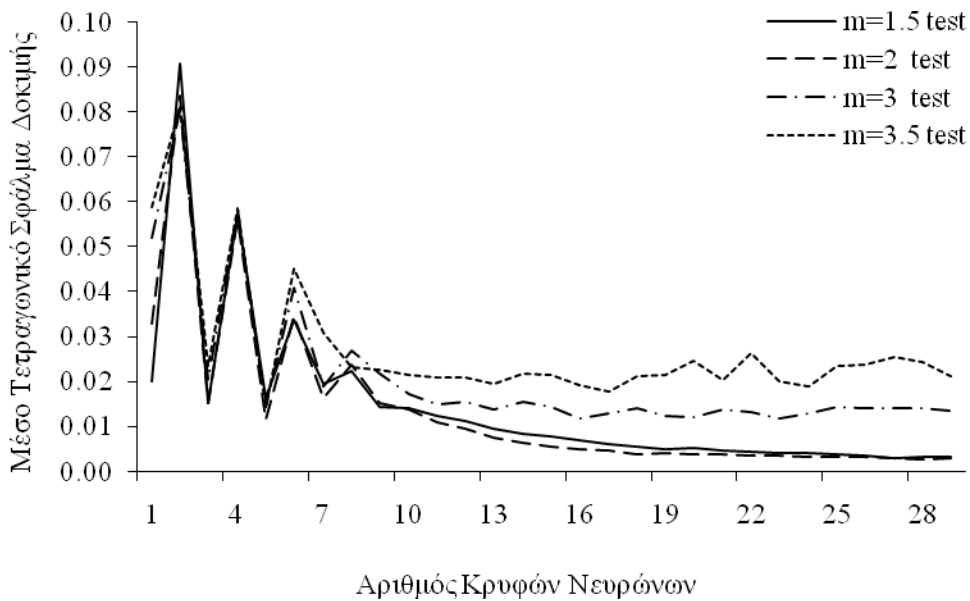


Εικόνα 5.12: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την δεύτερη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

Στο τρίτο σύνολο δεδομένων δόθηκαν στην παράμετρο ασάφειας οι τιμές 1.5, 2, 3, και 3.5. Σε αυτή τη προσομοίωση εμφανίζεται μια διαφοροποίηση σε σχέση με τα γραφήματα των αποτελεσμάτων από τα προηγούμενα πειράματα. Ενώ το αναμενόμενο θα ήταν να είναι χαμηλότερες οι τιμές του MSE για  $m=1.5$ , τελικά είναι για  $m=2$  και για τα δύο γραφήματα, εκπαίδευσης και δοκιμής.

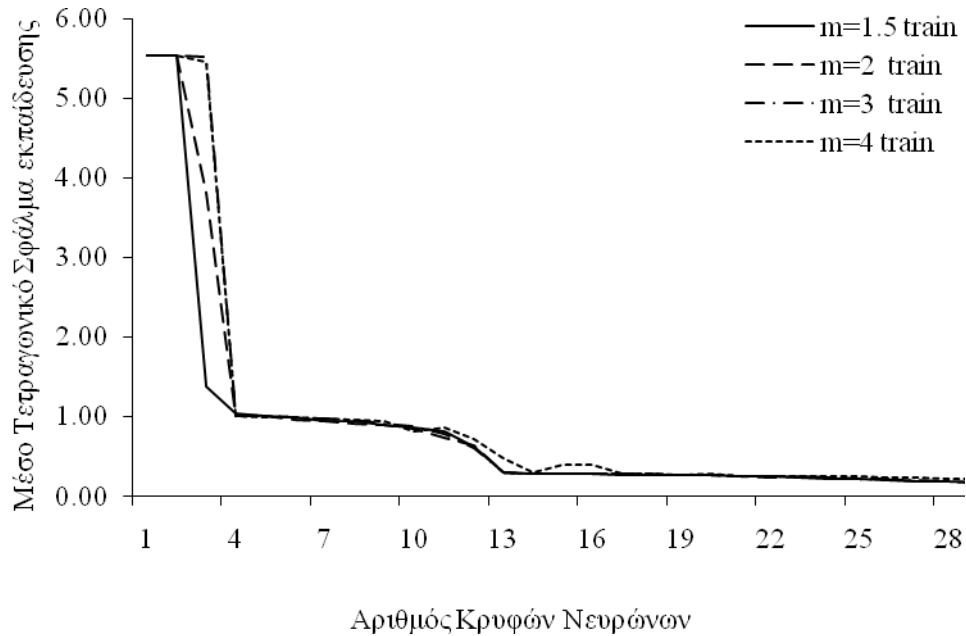


Εικόνα 5.13: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την δεύτερη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

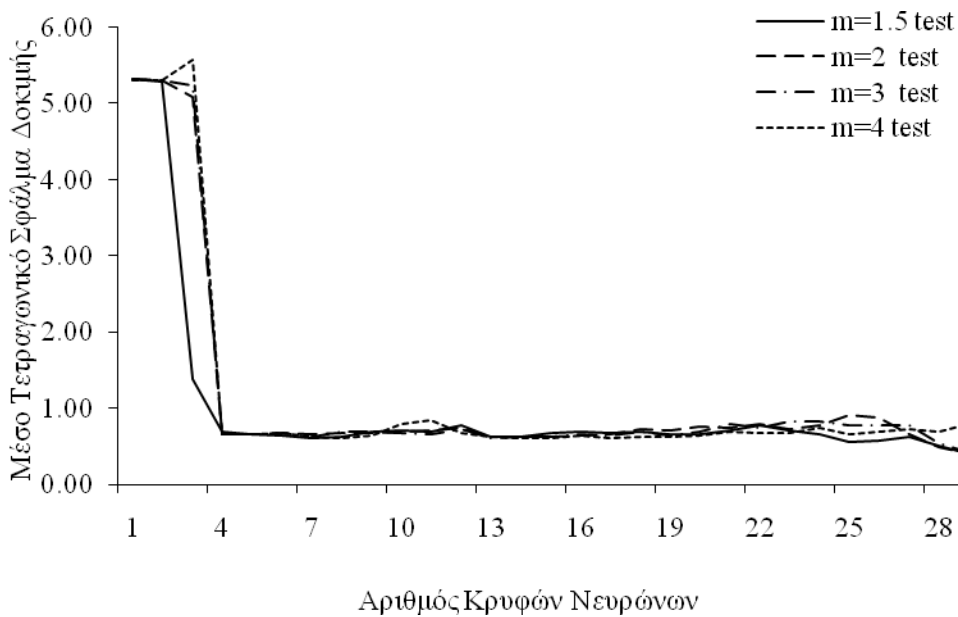


Εικόνα 5.14: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την δεύτερη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

Στον τρίτος αλγόριθμο που χρησιμοποιήθηκε, (Tsekouras and Neiros, 2008), ο αριθμός των εισόδων ισούται με 3, των εξόδων με 1, και το σύνολο των κρυφών νευρώνων είναι ίσο με 30. Ο βαθμός μάθησης  $\eta$  είναι ίσος με 0.0001. Σε αυτές τις προσομοιώσεις η μεταβλητή που άλλαζε ήταν η παράμετρος ασάφειας παίρνοντας τις τιμές 1.5, 2, 3, και 4.



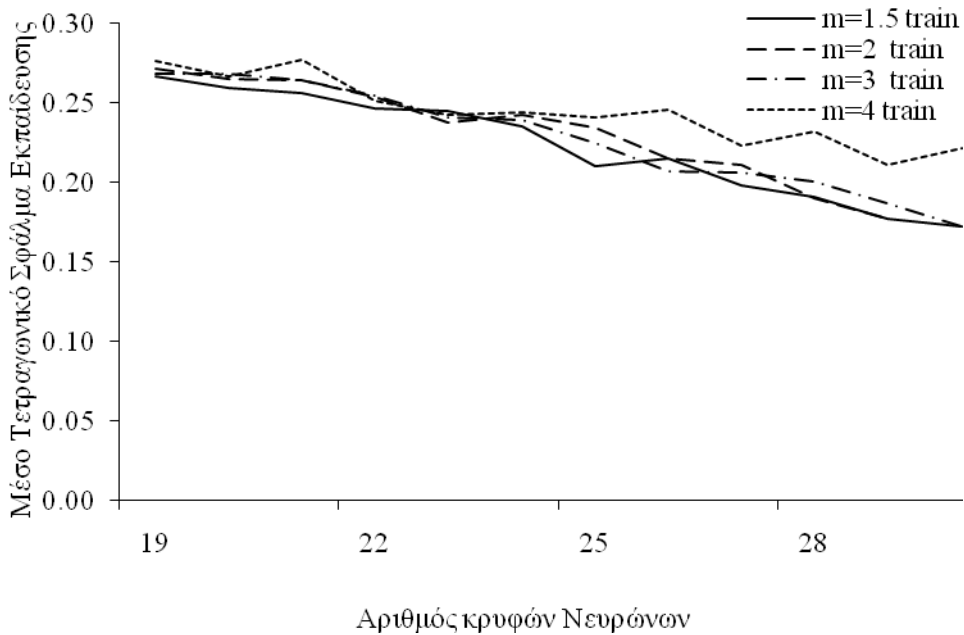
Εικόνα 5.15: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την τρίτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας



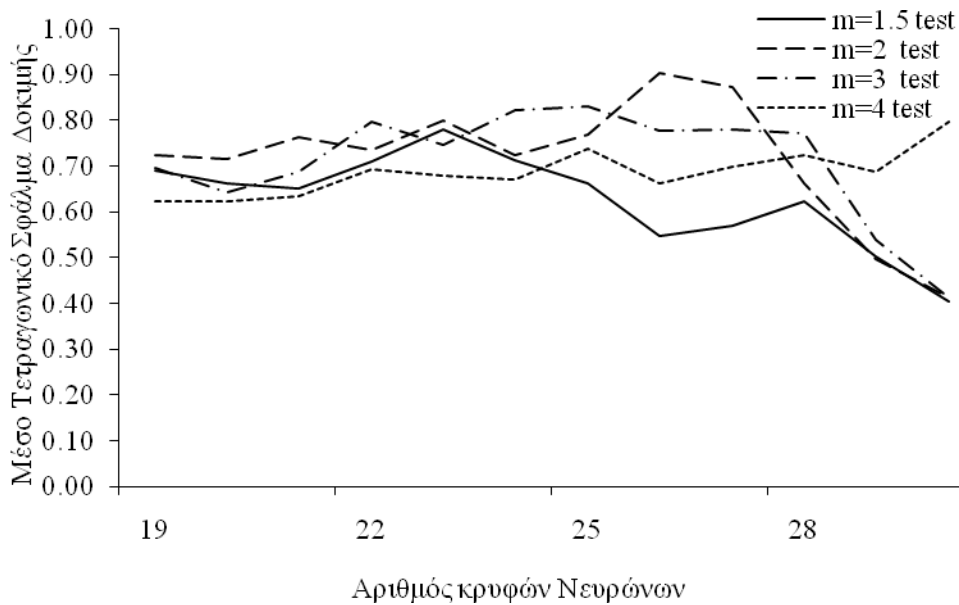
Εικόνα 5.16: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας



Όπως φαίνεται στα γραφήματα επιτυγχάνεται γρήγορη σύγκλιση του μέσου τετραγωνικού σφάλματος, ήδη από τον τέταρτο κρυφό κόμβο, και οι τιμές που παίρνει το MSE για τις διάφορες τιμές του  $m$  δεν έχει τεράστιες αποκλίσεις μεταξύ τους. Μετατρέποντας τα γραφήματα έτσι ώστε να φαίνεται το MSE από το 19<sup>ο</sup> κρυφό νευρώνα και έπειτα φαίνεται ότι οι γραμμές πλέκονται μεταξύ τους και δεν υπάρχει τιμή για την παράμετρο ασάφειας για την οποία το MSE να παίρνει τις χαμηλότερες τιμές ή τις υψηλότερες για την εκπαίδευση και τη δοκιμή.

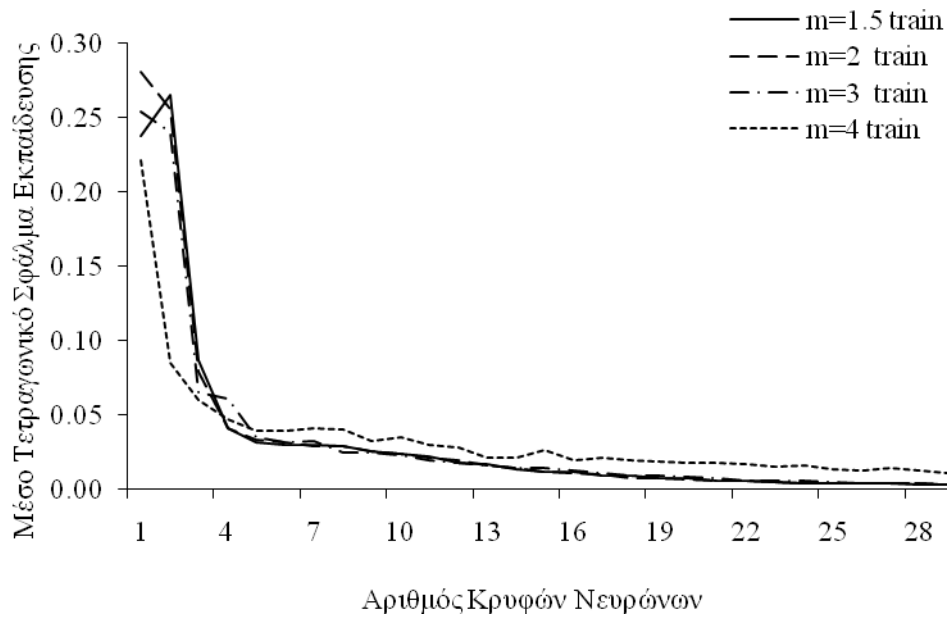


Εικόνα 5.17: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την τρίτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 19<sup>ο</sup> κρυφό νευρώνα

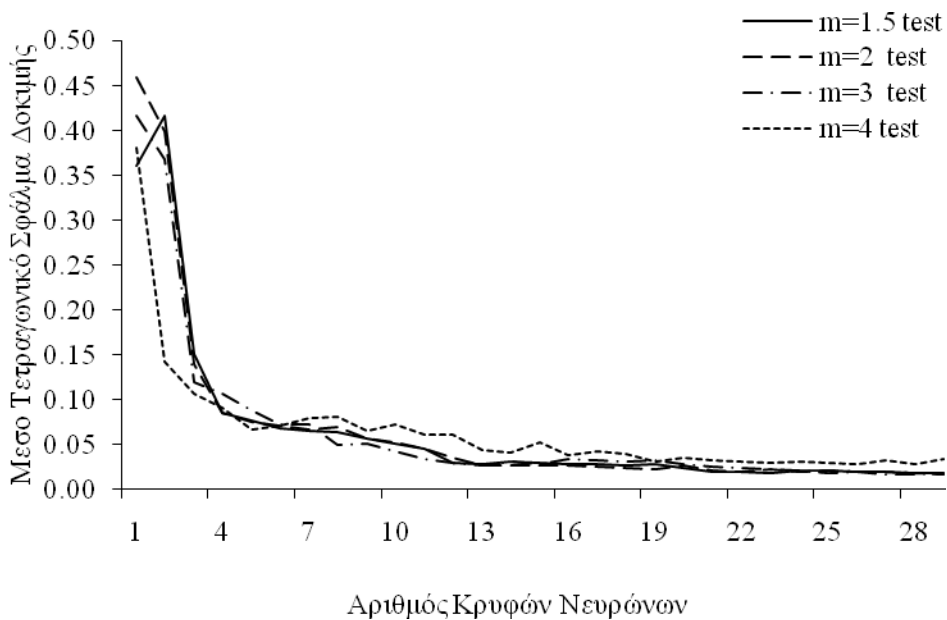


Εικόνα 5.18: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το πρώτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 19<sup>ο</sup> κρυφό νευρώνα

Στο δεύτερο σύνολο δεδομένων, οι παράμετροι του αλγόριθμου πήραν πάλι τις ίδιες αρχικές τιμές, και στην παράμετρο ασάφειας δόθηκαν οι τιμές 1.5, 2, 3, και 4. Επειδή για άλλη μία φορά δεν είναι εμφανές για ποιο  $m$  επιτυγχάνονται οι χαμηλότερες τιμές του MSE, παρακάτω τροποποιούνται τα γραφήματα από το 15<sup>ο</sup> κρυφό νευρώνα μέχρι τέλους.

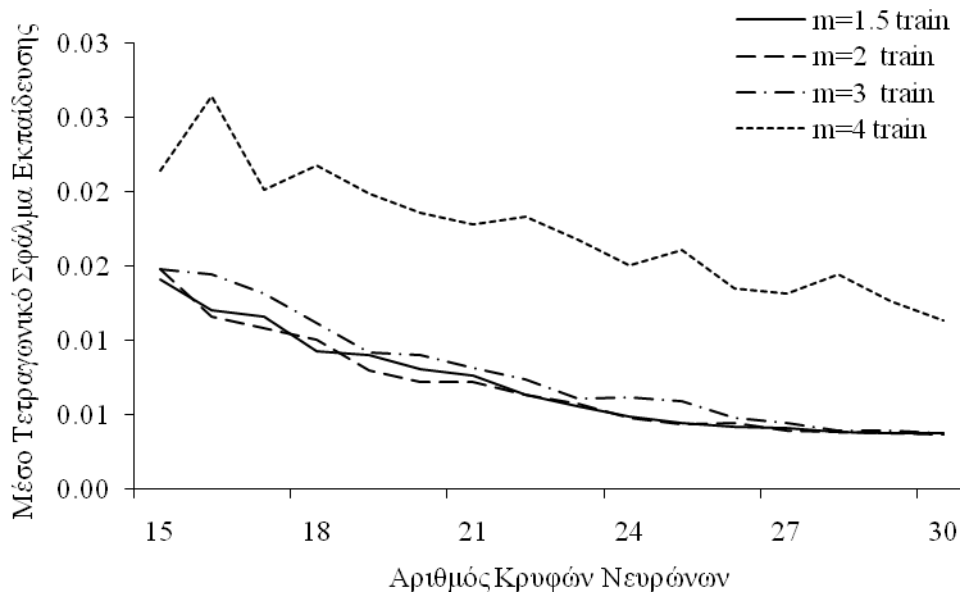


Εικόνα 5.19: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την τρίτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

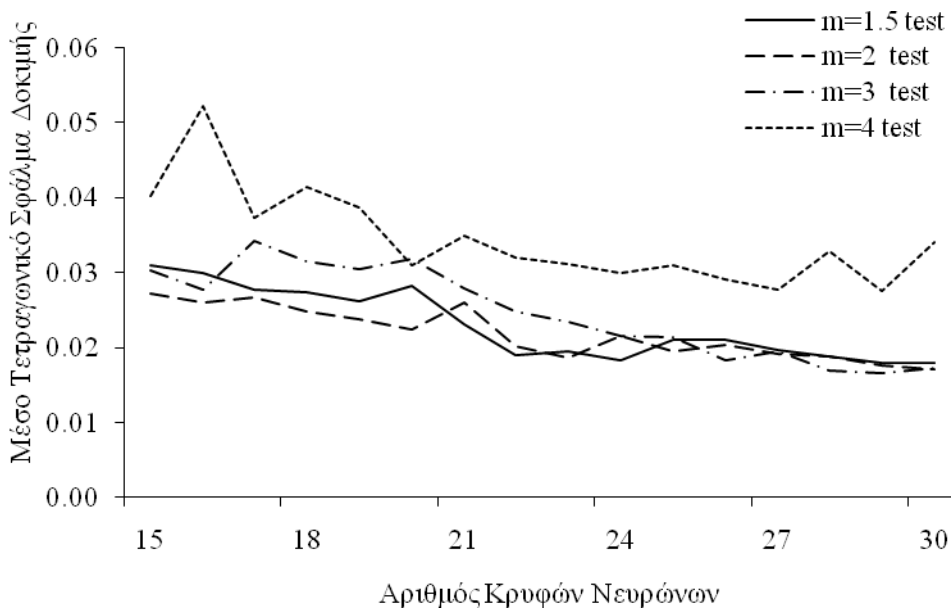


Εικόνα 5.20: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας

Όντως, όταν η παράμετρος ασάφειας παίρνει τις τιμές 1.5, 2, και 3, οι διαφορές των τιμών του MSE είναι μικρές μεταξύ τους. Όμως για  $m=4$ , το MSE παίρνει τις υψηλότερες τιμές σε σύγκριση με τις άλλες τρεις.

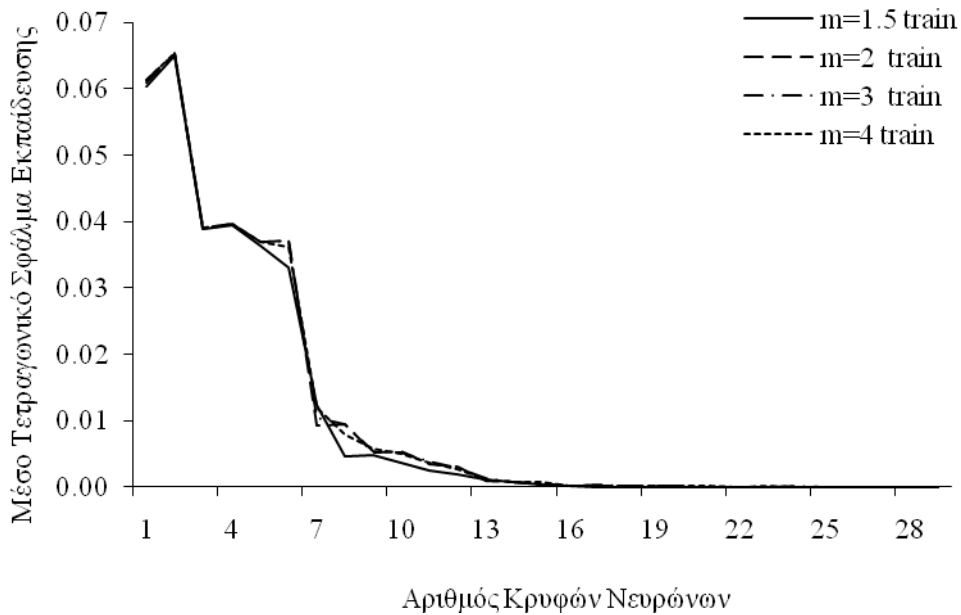


Εικόνα 5.21: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την τρίτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

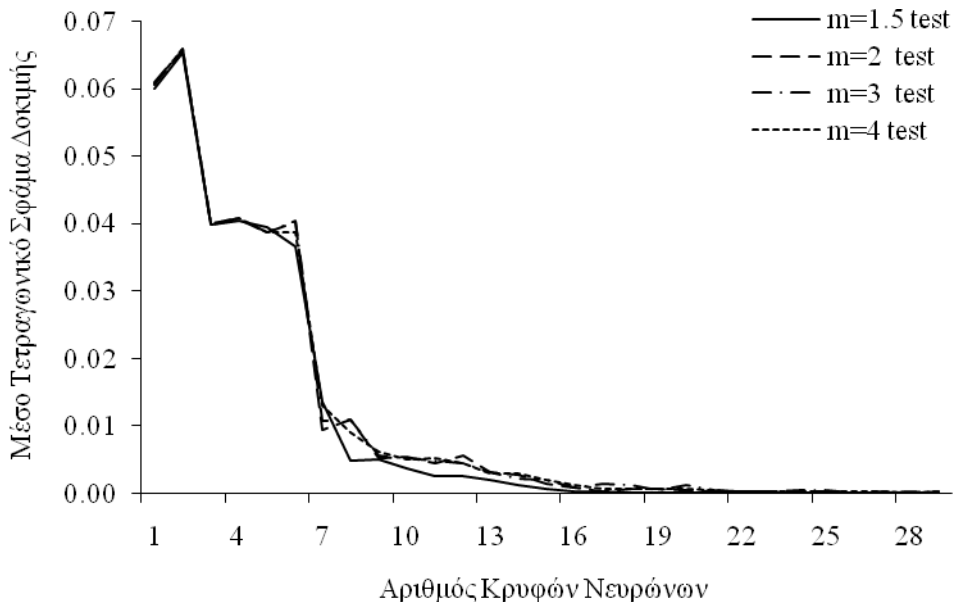


Εικόνα 5.22: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το δεύτερο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

Στο τρίτο σύνολο δεδομένων, οι παράμετροι του αλγόριθμου πήραν πάλι τις ίδιες αρχικές τιμές, και στην παράμετρο ασάφειας δόθηκαν οι τιμές 1.5, 2, 3, και 4. Επειδή για άλλη μία φορά δεν είναι εμφανές για ποιο  $m$  επιτυγχάνονται οι χαμηλότερες τιμές του MSE, παρακάτω τροποποιούνται τα γραφήματα από το 15<sup>ο</sup> κρυφό νευρώνα μέχρι τέλους.

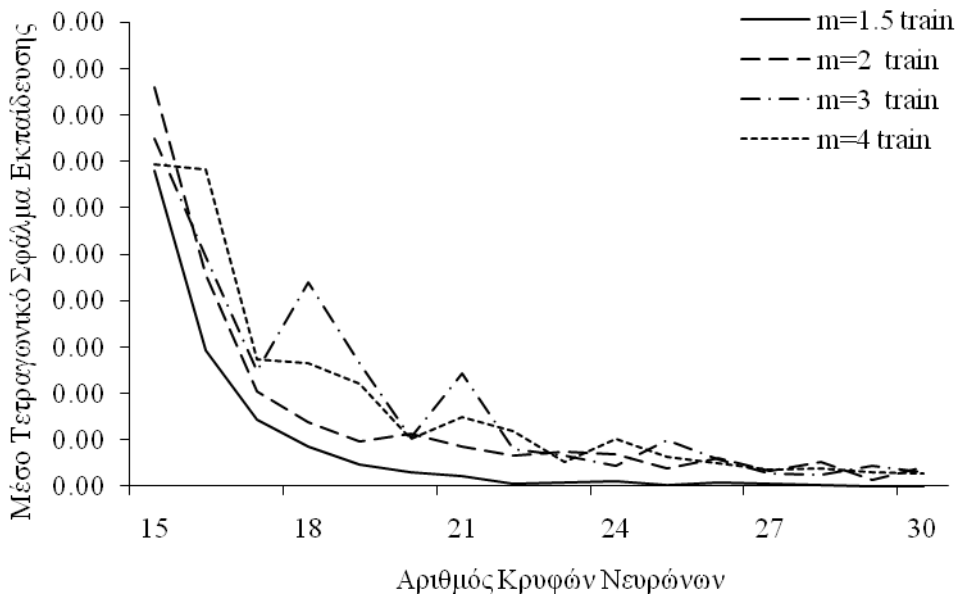


Εικόνα 5.23: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

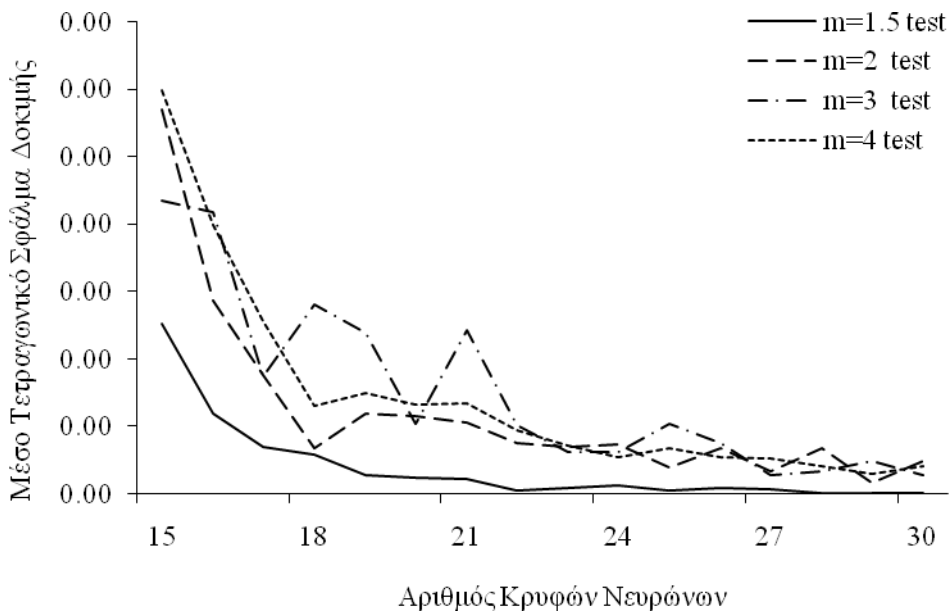


Εικόνα 5.24: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

Όντως, όταν η παράμετρος ασάφειας παίρνει τις τιμές 2,3, και 4, οι διαφορές των τιμών του MSE είναι μικρές μεταξύ τους. Όμως για  $m=1.5$  φαίνεται ότι το MSE παίρνει τις χαμηλότερες τιμές σε σύγκριση με τις άλλες τρεις.



Εικόνα 5.25: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για την τρίτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

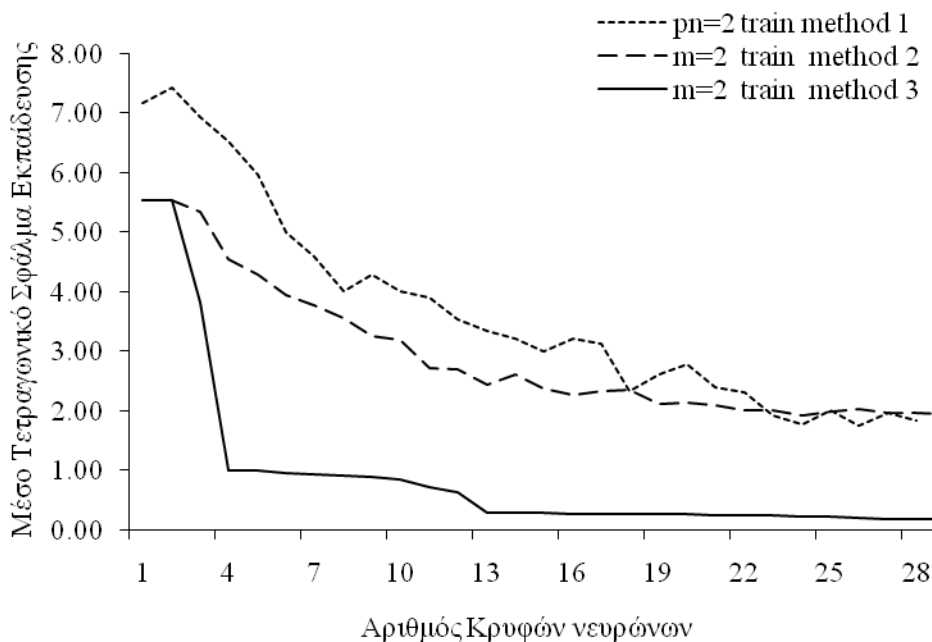


Εικόνα 5.26: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για την τρίτη μέθοδο με το τρίτο σύνολο δεδομένων για διάφορες παραμέτρους ασάφειας από το 15<sup>ο</sup> κρυφό νευρώνα

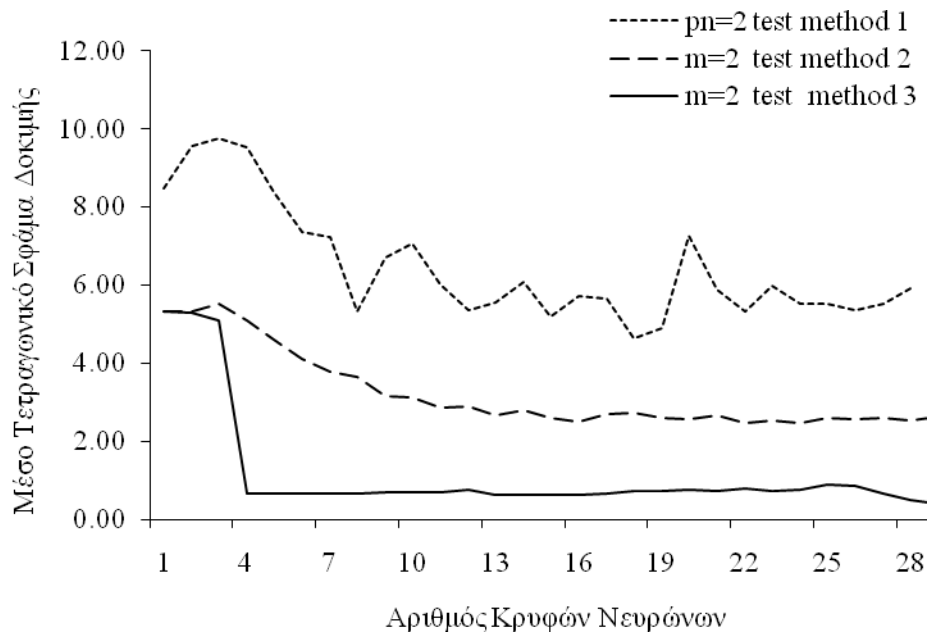
## 5.2. Δεύτερο Πείραμα

Η δεύτερη σύγκριση πραγματοποιήθηκε εξετάζοντας τον τρόπο με τον οποίο το κάθε σύνολο δεδομένων συμπεριφερόταν και στις τρεις πειραματικές μεθόδους. Από τις προσομοιώσεις που είχαν εκτελεστεί επιλέχθηκαν τα αποτελέσματα για τα οποία ο αριθμός των κοντινότερων γειτόνων και η παράμετρος ασάφειας ισούταν με δύο. Η αιτία για την οποία επιλέχθηκαν οι συγκεκριμένες περιπτώσεις, είναι διότι και για τις δύο παραμέτρους, συνήθως προτιμάται ο αριθμός 2, καθώς εκφέρει καλά αποτελέσματα, χωρίς να βρίσκει τοπικό ελάχιστο ή να αδυνατεί να συσταδοποιήσει τα δεδομένα στα κατάλληλα κέντρα.

Παρακάτω παρουσιάζονται τα διαγράμματα για τις τρεις πειραματικές μεθόδους με σύνολο δεδομένων τις μετρήσεις για το Wine, που είναι τα δεδομένα πραγματικού τύπου. Όπως φαίνεται στα γραφήματα, βέλτιστα αποτελέσματα στο μέσο τετραγωνικό σφάλμα της εκπαίδευσης και της δοκιμής, δίνει ο αλγόριθμος των Τσεκούρα και Νείρου, ακολουθεί η μέθοδος του Bezdek, και τέλος των Moody και Darken.

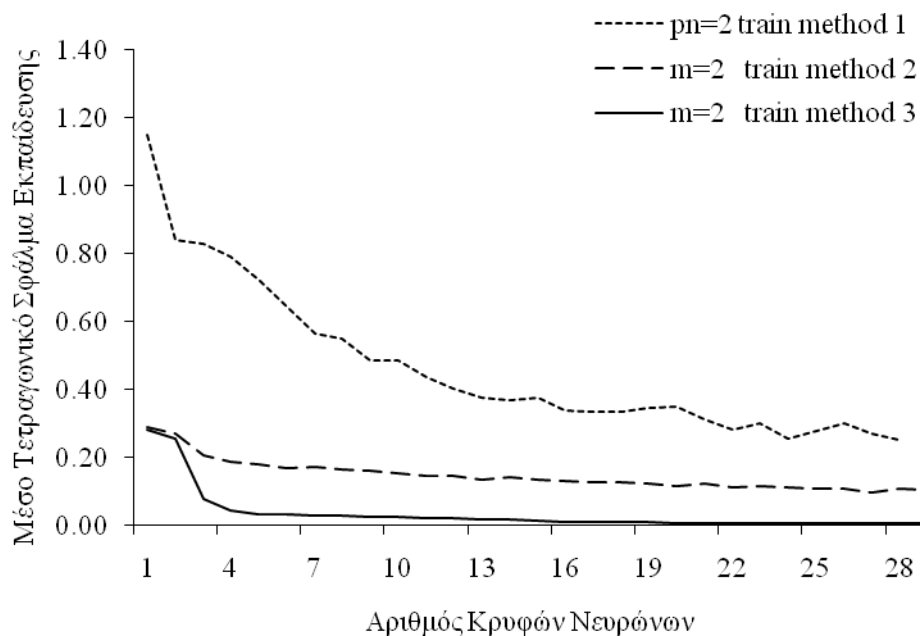


Εικόνα 5.27: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για το πρώτο σύνολο δεδομένων και των τριών νευρωνικών δικτύων

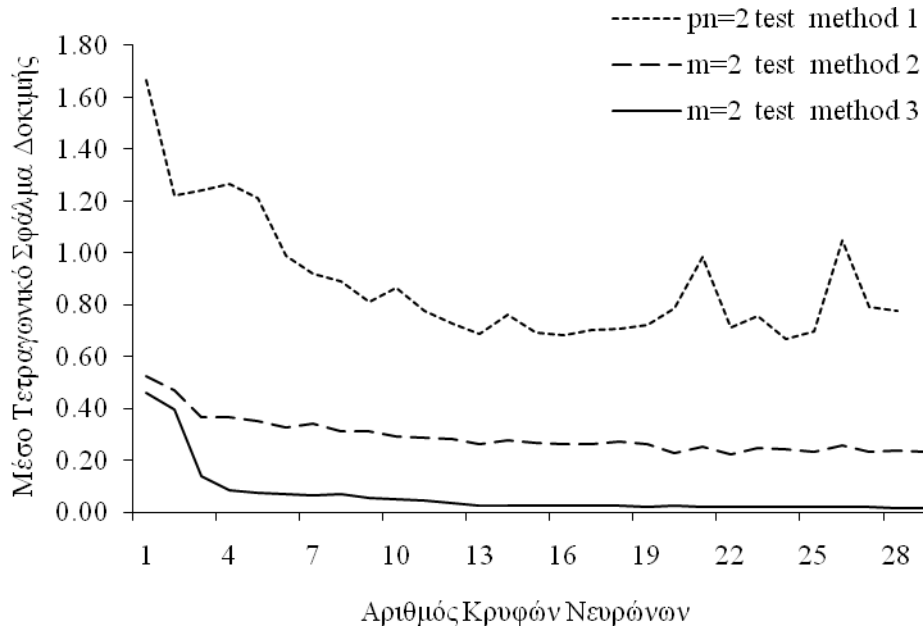


Εικόνα 5.28: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για το πρώτο σύνολο δεδομένων και των τριών νευρωνικών δικτύων

Στο δεύτερο σύνολο δεδομένων, δηλαδή ο πίνακας με τον τρισδιάστατο αριθμητικό πίνακα, τα αποτελέσματα του μέσου τετραγωνικού σφάλματος συμπεριφέρονται όπως και για το σύνολο των δεδομένων του Wine. Δηλαδή καλύτερα αποτελέσματα προκύπτουν από τον αλγόριθμο των Τσεκούρα και Νείρο, ακολουθεί η μέθοδος του Bezdek, και τέλος των Moody και Darken.

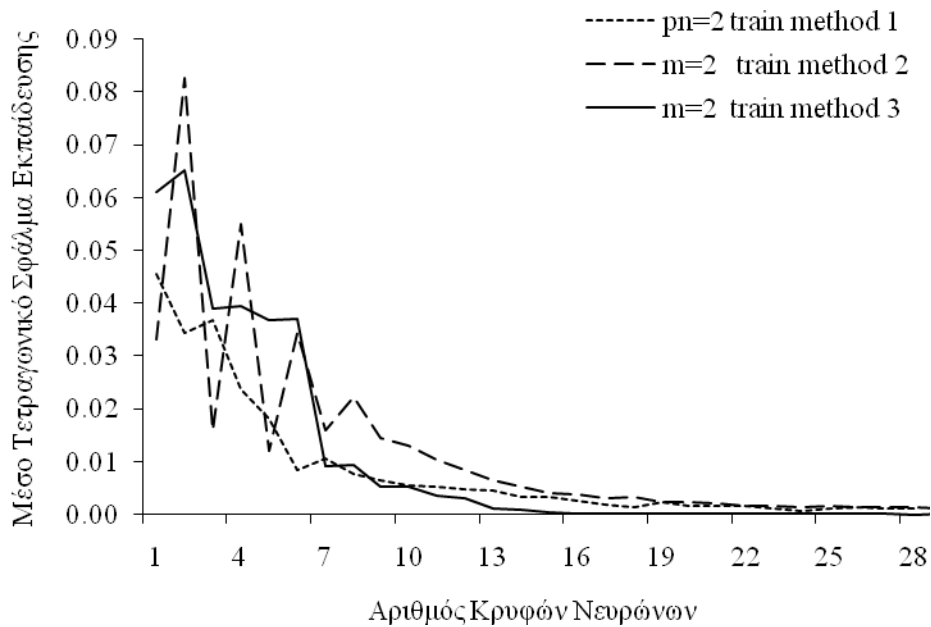


Εικόνα 5.29: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για το δεύτερο σύνολο δεδομένων και των τριών νευρωνικών δικτύων



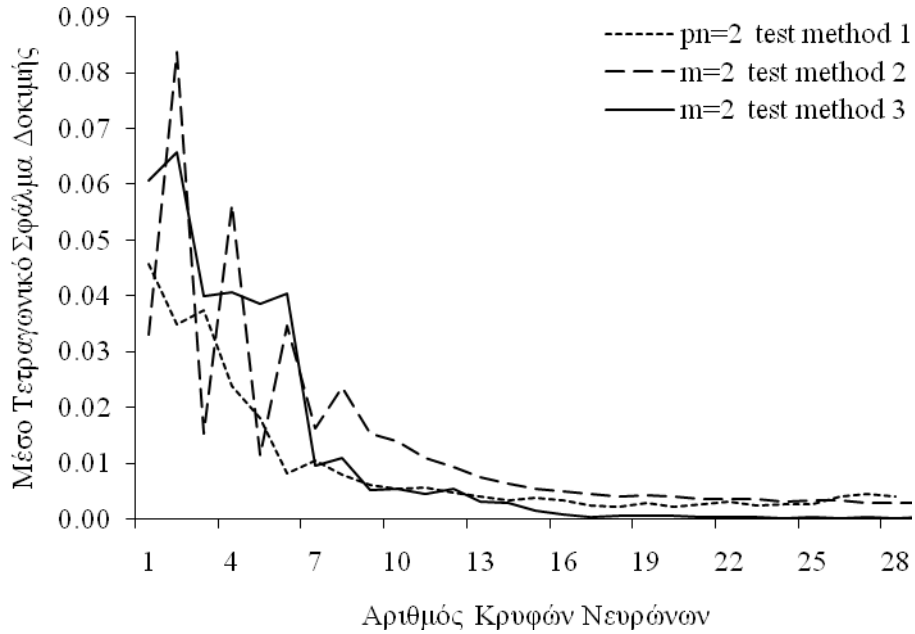
Εικόνα 5.30: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για το δεύτερο σύνολο δεδομένων και των τριών νευρωνικών δικτύων

Στο τρίτο σύνολο δεδομένων, οι γραμμές των γραφημάτων πλέκονται μεταξύ τους και δεν φαίνεται που παρουσιάζονται τα καλύτερα αποτελέσματα.



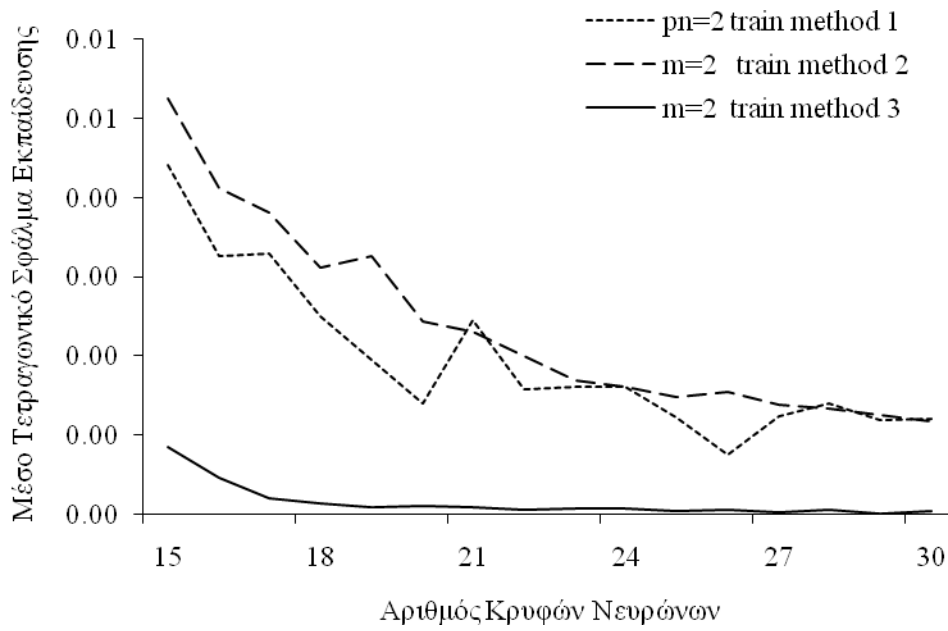
Εικόνα 5.31: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για το τρίτο σύνολο δεδομένων και των τριών νευρωνικών δικτύων



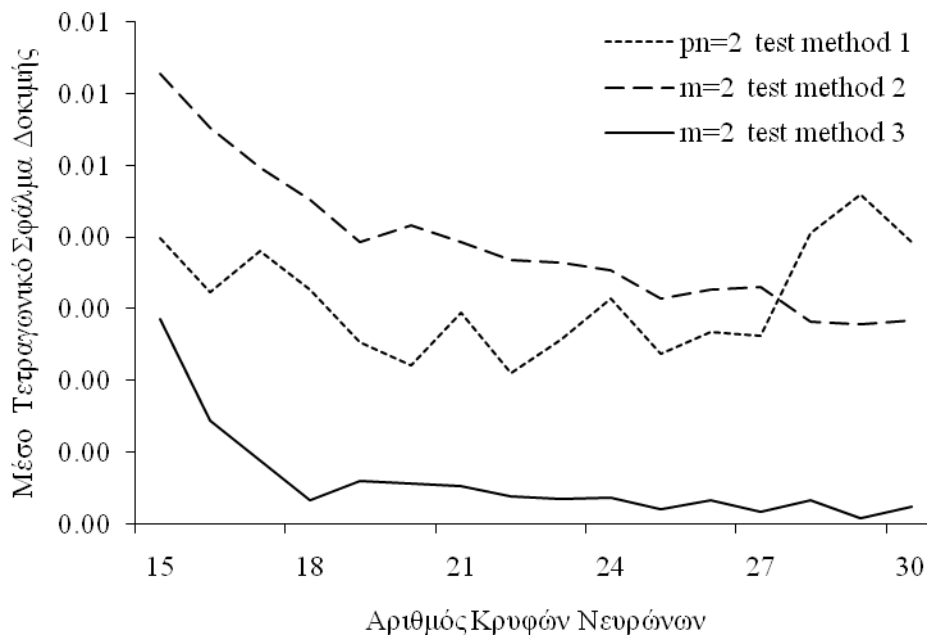


Εικόνα 5.32: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για το τρίτο σύνολο δεδομένων και των τριών νευρωνικών δικτύων

Γι αυτό, τα γραφήματα τροποποιήθηκαν έτσι ώστε το μέσο τετραγωνικό σφάλμα να φαίνεται από το δέκατο πέμπτο κρυφό νευρώνα και έπειτα. Όντως, εκτός από το γεγονός ότι τα αποτελέσματα της μεθόδου του Bezdek, τέμνονται σε ορισμένους κρυφούς νευρώνες με αυτά της μεθόδου των Moody και Darken, η μέθοδος των Τσεκούρα και Νείρου, συνεχίζει να παράγει τα βέλτιστα αποτελέσματα του μέσου τετραγωνικού σφάλματος τόσο στην εκπαίδευση, όσο και στη δοκιμή.



Εικόνα 5.33: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Εκπαίδευσης για το τρίτο σύνολο δεδομένων και των τριών νευρωνικών από το 15<sup>ο</sup> νευρώνα



Εικόνα 5.34: Γράφημα του Μέσου Τετραγωνικού Σφάλματος Δοκιμής για το τρίτο σύνολο δεδομένων και των τριών νευρωνικών δικτύων από το 15<sup>ο</sup> νευρώνα

## ΣΥΜΠΕΡΑΣΜΑΤΑ

Σε αυτή την εργασία πραγματοποιήθηκε η σύγκριση των τιμών του μέσου τετραγωνικού σφάλματος που προέκυπτε καθώς άλλαζαν οι αρχικές τιμές της παραμέτρου που επηρέαζε τη συσταδοποίηση τριών αλγόριθμων εκπαίδευσης νευρωνικών δικτύων, στους οποίους δόθηκαν σαν είσοδο τρία διαφορετικά σύνολα δεδομένων. Επιπλέον, συγκρίθηκαν και οι μέθοδοι μεταξύ τους για να διερευνηθεί ποια από τις τρεις πραγματοποιούσε καλύτερη μάθηση. Σκοπός ήταν να εξεταστούν οι επιπτώσεις που ασκούν οι αλλαγές στις τιμές ορισμένων παραμέτρων στη διασπορά των συναρτήσεων βάσης και στο σχεδιασμό των αντίστοιχων νευρωνικών δικτύων.

Αφού έγιναν οι απαραίτητες προσομοιώσεις, τα αποτελέσματα εισήχθησαν σε γραφήματα που διευκόλυναν τη μεταξύ τους σύγκριση. Έτσι, οι συγκρίσεις που πραγματοποιήθηκαν αφορούσαν τα αποτελέσματα του μέσου τετραγωνικού σφάλματος για την εκπαίδευση και τη δοκιμή, για τις διάφορες τιμές του αριθμού των κοντινότερων γειτόνων του κέντρου της πρώτης συνάρτησης ακτινικής βάσης, και τις διάφορες τιμές της παραμέτρου ασάφειας για τη δεύτερη και τρίτη συνάρτηση.

Πράγματι, όπως προέκυψε, για την πρώτη μέθοδο όπου χρησιμοποιήθηκε ο αλγόριθμος c-Means για συσταδοποίηση, όσον αφορά το πρώτο και το δεύτερο σύνολο δεδομένων, τα αποτελέσματα του μέσου τετραγωνικού σφάλματος παίρνουν μικρότερες και άρα καλύτερες τιμές όταν οι κοντινότεροι γείτονες είναι δύο, αντί όταν είναι τρεις ή τέσσερις. Επιβεβαιώνεται με αυτό τον τρόπο η θεωρία που συνηθίζει να δίνει την τιμή δύο σε αυτή την παράμετρο ως βέλτιστη. Παρέκκλιση, από αυτή τη συμπεριφορά, παρουσιάζει ο αλγόριθμος για το τρίτο σύνολο δεδομένων, στο οποίο δεν διακρίνεται για ποια τιμή της παραμέτρου απορρέουν οι χαμηλότερες τιμές του MSE, καθώς τα αποτελέσματα του MSE παίρνουν τιμές που είναι κοντά η μία στην άλλη με αποτέλεσμα οι γραμμές να πλέκονται μεταξύ τους.

Στη δεύτερη και την τρίτη μέθοδο, δόθηκαν διάφορες τιμές, στην παράμετρο ασάφειας. Θεωρητικά, αυτό που ισχύει για αυτή την παράμετρο είναι ότι όταν οι τιμές της τείνουν προς το ένα, τότε οι συστάδες τείνουν να είναι μη ασαφής και τα όριά τους να είναι αυστηρώς ορισμένα. Αντίθετα, όταν παίρνει μεγάλες τιμές ή τείνει προς το άπειρο, τότε τα στοιχεία ανήκουν σε όλες τις υπό μελέτη συστάδες με τον ίδιο βαθμό συμμετοχής. Βέλτιστη τιμή και για αυτή την παράμετρο θεωρείται το δύο (Βαζιργιάννης, Χαλκίδη, 2005). Τα αποτελέσματα του μέσου τετραγωνικού σφάλματος για το πρώτο και το δεύτερο σύνολο δεδομένων είναι καλύτερα για την τιμή του  $m=1.5$ , και ακολουθούν διαδοχικά τα αποτελέσματα για τις τιμές 2,3 και 3.5. Για το τρίτο σύνολο δεδομένων, όμως, προκύπτουν καλύτερα αποτελέσματα για  $m=2$  και ακολουθούν τα αποτελέσματα για  $m=1.5,3$  και 4. Τα αποτελέσματα της τρίτης μεθόδου, για όλες τις τιμές του

m, οι γραμμές του μέσου τετραγωνικού σφάλματος πλέκονται μεταξύ τους. Διαφοροποίηση παρατηρείται στα αποτελέσματα του δεύτερου και του τρίτου συνόλου δεδομένων, όπου στο δεύτερο για  $m=4$ , το μέσο τετραγωνικό σφάλμα είναι εμφανώς μεγαλύτερο, ενώ στο τρίτο σύνολο για  $m=1.5$ , είναι εμφανώς μικρότερο.

Στη σύγκριση μεταξύ των νευρωνικών δικτύων, όπου άλλαζε η μέθοδος και όχι το σύνολο δεδομένων, ο αλγόριθμος που προτάθηκε από τους Τσεκούρα και Νείρο, απέδιδε τα καλύτερα αποτελέσματα σε σύγκριση με τους αλγόριθμους των νευρωνικών που εξετάστηκαν. Έπειτα ακολουθούσε η μέθοδος του Bezdek, και τέλος των Moody και Darken. Η συμπεριφορά αυτή των νευρωνικών ήταν αναμενόμενη, καθώς οι Τσεκούρας και Νείρος, και ο Bezdek χρησιμοποιούν ασαφή λογική και τον Fuzzy c-Means, για να συσταδοποιήσουν τα δεδομένα, ενώ οι Moody και Darken συσταδοποιούν με τον c-Means, του οποίου η συνάρτηση συμμετοχής βασίζεται στη δυαδική λογική, δηλαδή το δεδομένο είτε ανήκει στο σύνολο είτε όχι. Επιπλέον ο αλγόριθμος των Τσεκούρα και Νείρο, παράγει καλύτερα αποτελέσματα και από τους τρεις, διότι εκτός από τον αλγόριθμο Fuzzy c-Means, χρησιμοποιεί και τη σχέση (3.18) για να υπολογίσει τη διασπορά της γκαουσιανής και στη συνέχεια τη μέθοδο ελαχίστων τετραγώνων για τον υπολογισμό των βαρών διασύνδεσης των νευρώνων, σε αντίθεση με τον Bezdek, που χρησιμοποιεί για τον ίδιο σκοπό τη συνάρτηση συμμετοχής.

Ολοκληρώνοντας το πείραμα και αφού πραγματοποιήθηκαν οι συγκρίσεις των αποτελεσμάτων προκύπτει το εξής συμπέρασμα.

Το μέσο τετραγωνικό σφάλμα, είναι το τετράγωνο της διαφοράς της επιθυμητής έξοδου του νευρωνικού από την πραγματική του έξοδο. Όσο μικρότερη είναι η τιμή του μέσου τετραγωνικού σφάλματος, τόσο περισσότερο το νευρωνικό προσεγγίζει την επιθυμητή έξοδο του νευρωνικού δικτύου και επιτυγχάνεται καλύτερη μάθηση. Το ιδανικό θα ήταν η πραγματική με την επιθυμητή έξοδο του νευρωνικού να συμπίπτουν, αλλά πρακτικά είναι αδύνατο, ιδιαίτερα όταν το σύνολο των δεδομένων είναι μεγάλο και δεν γίνεται ο αριθμός των κέντρων των συστάδων να είναι ίσος με το πλήθος των δεδομένων, λόγω υπολογιστικού κόστους.

Η έξοδος του νευρωνικού επηρεάζεται από τη διασπορά της γκαουσιανής συνάρτησης. Όσο μικρότερο είναι το μέσο τετραγωνικό σφάλμα τόσο καλύτερη είναι η μέθοδος που χρησιμοποιήθηκε για τον υπολογισμό της διασποράς και όσο καλύτερη η διασπορά, τόσο πιο καλά είναι κατανομημένα τα δεδομένα γύρω από τα κέντρα των συστάδων.

Μεταβάλλοντας τις τιμές των κοντινότερων γειτόνων και της παραμέτρου ασάφειας, εξετάστηκε πως επηρεαζόταν η διασπορά των συναρτήσεων βάσης που χρησιμοποιήθηκαν στο πείραμα, με ιδιαίτερη βάση στη γκαουσιανή συνάρτηση. Προέκυψε ότι όσο μικρότερες ήταν οι τιμές αυτών των παραμέτρων τόσο καλύτερες ήταν οι τιμές των αποτελεσμάτων του μέσου

τετραγωνικού σφάλματος. Επομένως, στο σχεδιασμό νευρωνικών δικτύων ακτινικής βάσης, είναι σημαντικό σε αυτές τις αρχικές παραμέτρους να δίνονται οι κατάλληλες τιμές, έτσι ώστε να βελτιστοποιείται η μέθοδος με την οποία υπολογίζεται η διασπορά της συνάρτησης και να ελαχιστοποιείται το μέσο τετραγωνικό σφάλμα. Αν εξασφαλιστούν αυτά τότε το νευρωνικό δίκτυο επιτυγχάνει τη βέλτιστη μάθηση.

Τέλος, έχοντας εξετάσει όλες τις συγκρίσεις, προκύπτει ότι τόσο στο θεωρητικό κομμάτι όσο και στο πειραματικό, ότι ο αλγόριθμος που προτάθηκε από τους Τσεκούρα και Νείρο για τις τιμές 1.5 και 2 της παραμέτρου ασάφειας, χρησιμοποιεί την καλύτερη μέθοδο εύρεσης της διασποράς της συνάρτησης βάσης και φέρει τα καλύτερα αποτελέσματα στο μέσο τετραγωνικό σφάλμα τόσο στην εκπαίδευση των δεδομένων όσο και στη δοκιμή τους.

Μελλοντικά, τα αποτελέσματα αυτής της εργασίας, θα μπορούσαν να χρησιμοποιηθούν σαν βάση για να βρεθούν νέες μέθοδοι για την εύρεση της διασποράς, ή σαν μέτρο σύγκρισης με άλλες μεθόδους υπολογισμού της διασποράς που μπορεί να προκύψουν οι οποίες θα εξασφαλίζουν στο νευρωνικό δίκτυο καλύτερα αποτελέσματα μέσου τετραγωνικού σφάλματος και ταυτόχρονα επίτευξη καλύτερης μάθησης.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- Βαζιργιάννης, Μ., Χαλκίδη, Μ., 2005, *Εξόρυξη από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό*, Τυπωθήτω – Γιώργος Δαρδάνος, Αθήνα
- Βλαχάβας, Ι., Βασιλειάδης, Ν., Κόκκορας, Φ., Σακελλαρίου, Η., 2006, *Τεχνητή Νοημοσύνη*, Γκιούρδας, Αθήνα
- Διαμαντάρας Κ., 2007, *Τεχνητά Νευρωνικά Δίκτυα*, Κλειδάριθμος, Αθήνα
- Λύκας, Α., 2008, *Τεχνητά Νευρωνικά Δίκτυα – Εφαρμογές*, ΕΑΠ
- Bezdek, J., C., 1973, *Fuzzy Mathematics in Pattern Classification*, PhD Dissertation, Cornell University, Ithaca, NY
- Bezdek, J., C., 1974, “*Cluster Validity with the Fuzzy Sets*”, Journal of Cybernetics 3, p. 58-73
- Bezdek, J., C., 1980, “*A Convergence Theorem for the fuzzy ISODATA Clustering Algorithms*”, IEEE Trans. Pattern Anal. Machine Intelligence. PAMI-2, p.1-8
- Bezdek, J., C., 1987, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, USA
- Bezdek, J., C., Pal, D., K., 1992, *Fuzzy Models for Pattern Recognition, Methods that Search for Patterns in Data*, IEEE Press, New York, p.539
- Bezdek, J., C., Pal, D., K., 1995, *On clustering validity for the fuzzy c-means model*, IEEE Transactions on Fuzzy Systems 3, p. 370-379.
- Dunham, M., 2005, *Data Mining: εισαγωγικά και προηγμένα θέματα εξόρυξης γνώσης από δεδομένα*, Εκδόσεις Νέων Τεχνολογιών, Αθήνα
- Haykin, S., 1999, *Neural Networks - A Comprehensive Foundation*, Pearson Education, Delhi
- Klir, G., J., 1995, “*Principles of uncertainty: What are they? Why do we need them?*”, Sets and systems 74, p.15-31
- Kohonen, T., 1984, *Self-Organization and Associative Memory*, Springer Verlag, Berlin
- Moody, J., Darken, C., J., 1989, “*Fast Learning in Networks of locally-Tuned processing Units*”, Neural Computation 1, p.281-294
- Pedrycz, W., 1984, “*Construction of Fuzzy Relational Models. In: Trappl, R., Editor*”, Cybernetics and Systems Research Vol.2, North Holland, Amsterdam, p.545-549
- Pedrycz, W., 1993, *Fuzzy Control and Fuzzy Systems*, Research Studies Press/ Wiley, Taunton/ New York
- Tikhonov, A., Arsenin, V., 1977, *Solution of ill-posed Problems*, Washington, DC

Tsekouras, G., Neiros, A., 2008, "A Fuzzy clustering algorithm to estimate the parameters of Radial Basis Functions Neural Networks and its Application to System Modelling" *SETN 2008, LNAI 5138*, p.194-204

Zadeh, L., A., 1965, "Fuzzy Sets, Information and Control", p.338-353

Zurada, J., M., 1992, *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul