Κεφάλαιο 4

Νευρωνικά δίκτυα στην ταξινόμηση προτύπων

4.1 Εισαγωγή

Στο κεφάλαιο αυτό περιγράφονται, και μελετώνται συστήματα ταξινόμησης προτύπων που περιέχουν διασυνδεδεμένες μη γραμμικές υπολογιστικές μονάδες. Τα συστήματα αυτά αποτελούν προσωμοιώσεις της φυσιολογίας των νευρικών κυττάρων των έμβιων όντων και ονομάζονται τεχνητά νευρωνικά δίκτυα.

Ενας από τους πρώτους ερευνητές που συνέλαβε την ιδέα ότι οι εγκεφαλικές λειτουργίες εκτελούνται από στοιχειώδεις υπολογιστικές μονάδες που ονομάζονται νευρώνες είναι ο Ramon y Cajal (1911). Πειράματα που έγιναν στην φυσιολογία των νευρικών κυττάρων έδειξαν ότι η λειτουργία τους μπορεί, σε απλοποιημένη μορφή, να προσωμοιωθεί με μη γραμμικούς τελεστές που δέχονται σήματα από έναν πεπερασμένο αριθμό εισόδων και διαθέτουν μονάχα μία έξοδο. Η εγκεφαλική ουσία των έμβιων όντων αποτελείται από εκατομμύρια διασυνδεδεμένων νευρώνων τα οποία αποτελούν ένα πολύπλοκο και ισχυρά μη γραμμικό νευρωνικό δίκτυο. Εχει αποδειχθεί ότι η γνώση του κόσμου συσσωρεύεται στην μνήμη του δικτύου. Η τοπολογία του διαφέρει σημαντικά στα έμβια όντα και είναι αυτή η οποία δίνει και τα διαφορετικά χαρακτηριστικά αντίληψης των ερεθισμάτων του περιβάλλοντος. Εχουν παρατηρηθεί ποσοτικές αλλά και ποιοτικές διαφοροποιήσεις στο είδος των συνδέσεων, τον μηχανισμό επεξεργασίας σημάτων και το πλήθος των εισόδων των νευρικών κυττάρων. Γιαυτό τον λόγο και τα νευρικά κύτταρα έχουν ομαδοποιηθεί σε εναν μικρό αριθμό κατηγοριών, ανάλογα με τους μηχανισμούς λειτουργίας των.

Η υπάρχουσα τεχνολογία δεν επιτρέπει την προσομοίωση των πολύπλοκων επεξεργασιών που εκτελούνται στα νευρικά κύτταρα. Εχει αποδειχθεί όμως ότι απλές προσομοιώσεις των νευρικών κυττάρων, που περιγράφουν τα βασικά τους γνωρίσματα, δίνουν εντυπωσιακά αποτελέσματα σε εφαρμογές ταξινόμησης προτύπων, προσεγγίσεων απόκρισης συστημάτων, προβλήματα αυτόματου ελέγχου, κ.α..

Τα χαρακτηριστικά γνωρίσματα των μηχανισμών των φυσικών νευρωνικών δικτύων είναι τα ακόλουθα:

1. *Μη γραμμικότητα.* Το βασικότερο γνώρισμα των φυσικών νευρωνικών δικτύων είναι ότι η έξοδος σε καμμία περίπτωση δεν μπορεί να θεωρηθεί ότι αποτελεί έναν γραμμικό συνδυασμό των εισόδων του.

2. *Εκπαίδευση από παραδείγματα.* Βασικό στοιχείο των νευρωνικών δικτύων είναι η ικανότητά

τους να εκπαιδεύονται αντλώντας γνώση και τροποποιώντας τα στοιχεία μνήμης του δικτύου. Προηγούμενες ενεργοποιήσεις του δικτύου μεταβάλλουν την συμπεριφορά του έτσι ώστε όταν ενεργοποιηθεί από ίδια ή ομοιάζοντα σήματα εισόδου να δίνουν με μεγαλύτερη ακρίβεια την επιθυμούμενη έξοδο.

3. *Προσαρμογή.* Τα φυσικά νευρωνικά δίκτυα έχουν την δυνατότητα να αλλάζουν τα δεδομένα εξόδου των και να προσαρμόζουν την συμπεριφορά τους όταν μεγάλης κλίμακας αλλαγές λαμβάνουν χώρα στην είσοδό τους.

   Αντιπροσωπευτικό παράδειγμα προσαρμογής αποτελεί η μεταβολή της συμπεριφοράς του δικτύου των νευρικών κυττάρων που βρίσκονται επάνω στον αμφιβλιστροειδή χιτώνα. Με αλλαγή του φωτισμού του περιβάλλοντος χώρου τα νευρικά κύτταρα αλλάζουν την απόκρισή τους έτσι ώστε να μεγιστοποιείται η διακριτική τους ικανότητα (η αντίληψη των αντικειμένων που βρίσκονται στον περιβάλλοντα χώρο). Η διαδικασία προσαρμογής μακροσκοπικά γίνεται αντιληπτή σε μεγάλες μεταβολές του φωτισμού που διαρκεί από δέκατα του δευτερολέπτου έως έναν μικρό αριθμό δευτερολέπτων. Οι περισσότεροι από εμάς θα έχουμε παρατηρήσει ότι ενώ βρισκόμαστε μέσα σε αυτοκίνητο που κινείται, όταν μπούμε μέσα σε σκοτεινό τούνελ αρχικά δεν διακρίνουμε κανένα αντικείμενο. Με την πάροδο του χρόνου παρατηρούμε ότι μπορούμε να διακρίνουμε τα λευκά αντικείμενα που βρίσκονται μέσα στο τούνελ όπως οι διαγραμίσεις επάνω στην άσφαλτο. Οταν ξαφνικά βγούμε πάλι στο φώς της ημέρας, για ένα μικρό χρονικό διάστημα μας τυφλώνει το φώς το οποίο πριν μπούμε στο τούνελ δεν μας ενοχλούσε καθόλου.

4. *Αντοχή σε διακοπές συνδέσεων και λειτουργιών νευρώνων.* Η συμπεριφορά του δικτύου δεν διαταράσσεται σημαντικά από τυχόν διακοπή συνδέσεων ή και αφαίρεση νευρώνων. Η ιδιότητα αυτή αποτελεί βασικό χαρακτηριστικό της συμπεριφοράς των νευρικών κυττάρων διότι είναι γνωστό ότι διακοπή της λειτουργίας μικρών τμημάτων του εγκεφάλου επηρεάζει συνήθως σε πολύ μικρό βαθμό τις υπόλοιπες εγκεφαλικές λειτουργίες.

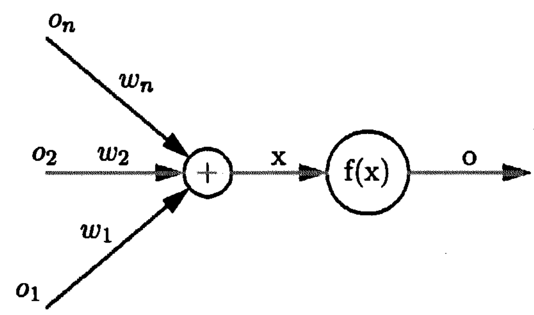
5. *Ομοιότητα λειτουργίας των νευρώνων.* Οι νευρώνες πρέπει να εκτελούν τις ίδιες βασικές λειτουργίες διότι τα νευρικά κύτταρα του ίδιου τύπου έχουν την ίδια φυσιολογία.

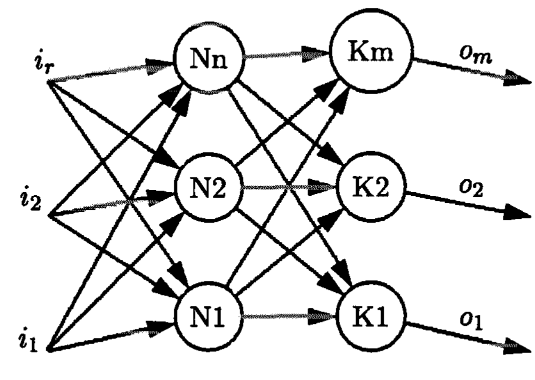
6. *Παράλληλη επεξεργασία δεδομένων.* Λαμβάνοντας υπόψη τον τρόπο συνδεσμολογίας των νευρικών κυττάρων, το τεράστιο πλήθος των πληροφοριών που επεξεργάζονται και την σχετικά μεγάλη καθυστέρηση απόκρισης του φυσικού κυττάρου (έχει υπολογιστεί ένας μέσος χρόνος απόκρισης της τάξης των 10^{-3} δευτερολέπτων, ενώ η απόκριση ενός ολοκληρωμένου κυκλώματος ανέρχεται σε 10^{-6} δευτερολέπτα), γίνεται φανερό ότι το φυσικό νευρωνικό δίκτυο επεξεργάζεται τα δεδομένα παράλληλα για να είναι σε θέση να επιτύχει απόκριση σε σχεδόν πραγματικό χρόνο και να επεξεργάζεται ταυτόχρονα έναν πολύ μεγάλο αριθμό πληροφοριών.

Η θεωρία των μη γραμμικών δικτύων που προσομοιάζουν την λειτουργία των νευρικών κυττάρων αναπτύχθηκε σημαντικά μέσα στην τελευταία τριακονταετία (1965-1995) προσφέροντας σημαντικές βελτιώσεις στα χαρακτηριστικά προσομοιώσεων όπου αυτή εφαρμόστηκε. Τα τεχνητά νευρωνικά δίκτυα αποτελούν μία γενική θεωρία μη γραμμικών συστημάτων και γιαυτό τον λόγο έχουν ένα ευρύ πεδίο εφαρμογών με σημαντικότερους τομείς την ταξινόμηση προτύπων, την ψηφιακή επεξεργασία σήματος, τον αυτόματο έλεγχο συστημάτων κ.ο.κ.

4.2 Ο νευρώνας

Κάθε νευρωνικό δίκτυο αποτελείται από διασυνδεδεμένες υπολογιστικές μονάδες που ονομάζονται νευρώνες. Κάθε νευρώνας μετασχηματίζει το διάνυσμα εισόδου του δίνοντας μία μονάχα έξοδο η οποία συνδέεται με εισόδους άλλων νευρώνων όπως φαίνεται και στο σχήμα 4.1.





Σχήμα 4.1: Νευρώνας και τυπική μορφή νευρωνικού δικτύου

Μελέτες που αναφέρονται στην φυσιολογία του νευρικού κυττάρου έδειξαν ότι η έξοδος αποτελεί έναν μη γραμμικό μετασχηματισμό των εισόδων του. Οι προσωμοιωτές που θα περιγραφούν αποτελούν προσέγγιση των μετασχηματισμών που πραγματοποιούνται στο εσωτερικό των φυσικών νευρικών κυττάρων.

Οι σημαντικότερες προσωμοιώσεις είναι υπολογιστικά απλές και αποδίδουν τους βασικούς μηχανισμούς ενεργοποίησης και εκπαίδευσης του φυσικού μοντέλου ικανοποιώντας τα χαρακτηριστικά του γνωρίσματα που περιγράφησαν στο προηγούμενο εδάφιο.

Οι περισσότερες προσωμοιώσεις που έχουν προταθεί βασίστηκαν σε μελέτες του τρόπου λειτουργίας των νευρικών κυττάρων στο φυσικό τους περιβάλλον. Απο την απλοποίηση των παρατηρήσεων της φυσιολογίας των νευρικών κυττάρων προέκυψαν τα διάφορα μαθηματικά μοντέλα. Το σημαντικότερο και το πλέον απλό μοντέλο περιγράφει την συμπεριφορά ενός νευρικού κυττάρου με δυό τελεστές, έναν γραμμικό και έναν μη γραμμικό που είναι συνδεδεμένοι σε σειρά, όπως φαίνεται και στο σχήμα 4.1.

Ο γραμμικός τελεστής παριστάνεται σαν το εσωτερικό γινόμενο του διανύσματος της εισόδου με το διάνυσμα της μνήμης του νευρώνα:

$$x = \mathbf{ow} = \sum_{i=1}^{N} o_i w_i \tag{4.1}$$

Ο μη γραμμικός τελεστής είναι μία μη γραμμική συνάρτηση της εξόδου του γραμμικού τελεστή:

$$o = f(x) \tag{4.2}$$

Η έξοδος του δικτύου συμβολίζεται με το ίδιο γράμμα με το οποίο συμβολίζεται και η είσοδος του νευρώνα διότι η έξοδος του νευρώνα είναι είσοδος σε άλλους νευρώνες. Η συνάρτηση του μη γραμμικού τελεστή ακολουθεί τα χαρακτηριστικά των σιγμοειδών (sigmoid) συναρτήσεων τα οποία είναι τα ακόλουθα:

1. *Είναι αύξουσα συνάρτηση*

$$\forall x_1, x_2 \in \mathbb{R}, x_1 > x_2 \quad \Rightarrow \quad f(x_1) \geq f(x_2) \tag{4.3}$$

2. *Εχει πεπερασμένα απειροστικά όρια*

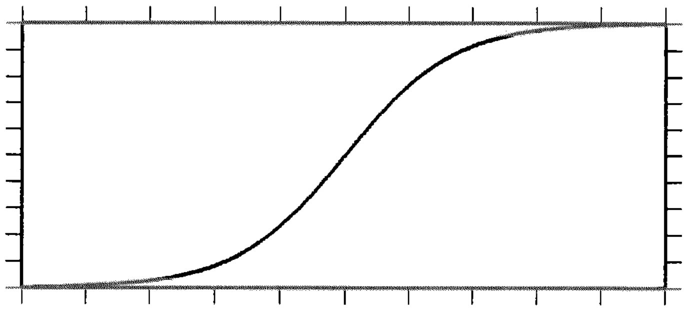
$$\lim_{x \to +\infty} f(x) = a, \qquad a \in \mathbb{R} - \{-\infty, +\infty\} \tag{4.4}$$

$$\lim_{x \to -\infty} f(x) = b, \qquad b \in \mathbb{R} - \{-\infty, +\infty\} \tag{4.5}$$

3. *Εχει πεδίο ορισμού το σύνολο των πραγματικών αριθμών και έχουν φραγμένο πεδίο τιμών.* Η ιδιότητα αυτή προκύπτει σαν αποτέλεσμα των δύο πρηγούμενων περιορισμών.

$$f : \mathbb{R} \to [a, b] \tag{4.6}$$

Μια συνάρτηση σιγμοειδών χαρακτηριστικών φαίνεται στο σχήμα 4.2.



Σχήμα 4.2: Τυπική μορφή σιγμοειδούς συνάρτησης

Οι πλέον διαδεδομένοι σε πρακτικές εφαρμογές γραμμικοί νευρωνικοί τελεστές είναι οι ακόλουθοι:

1. *Εκθετική σιγμοειδής.* Η συνάρτηση αυτή αποτελεί τον πλέον διαδεδομένο μη γραμμικό νευρωνικό τελεστή:

$$f(x) = \frac{1}{e^{-ax+b} + c}, \qquad a, b \in \mathbb{R}^+ \tag{4.7}$$

2. *Υπερβολική εφαπτομένη.*

$$o = \frac{e^{ax} - e^{-ax}}{e^{ax} + e^{-ax}}, \qquad a \in \mathbb{R}^+ \tag{4.8}$$

Είναι φανερό ότι ο υπολογισμός δύο εκθετικών όρων για την συνάρτηση της υπεβολικής εφαπτομένης και ενός εκθετικού όρου για την εκθετική σιγμοειδή συνάρτηση είναι μία υπολογιστικά χρονοβόρα διαδικασία, η οποία σε πολλές περιπτώσεις (ιδιαίτερα σε περιπτώσεις δικτύων που περιέχουν μικρό αριθμό κόμβων) υπερκαλύπτει τον χρόνο υπολογισμού του γραμμικού τελεστή του νευρώνα. Γιαυτό τον λόγο πολλές φορές καταφεύγουμε σε απλούστερες υπολογιστικά συναρτήσεις, όπως είναι οι ακόλουθες:

3. Τμηματικά γραμμική συνάρτηση. Η τμηματικά γραμμική συνάρτηση χρησιμοποιείται σπανιότερα στις εφαρμογές διότι δεν αποδίδει αξιόπιστα την μη γραμμική συμπεριφορά των νευρωνικών δικτύων.

$$f(x) = \begin{cases} \frac{1}{a}x & -a \leq x \leq a \\ -1 & x < -a \\ 1 & x > a \end{cases}, \quad a \in \mathbb{R}^+ \tag{4.9}$$

4. Συνάρτηση δύο κλάδων:

$$f(x) = \begin{cases} \frac{ax}{x+b}, & x \geq 0 \\ -\frac{ax}{x-b}, & x < 0 \end{cases}, \quad a, b \in \mathbb{R}^+ \tag{4.10}$$

5. Βηματική συνάρτηση:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{4.11}$$

6. Μη γραμμικός τελεστής του νευρώνα Adaline:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \tag{4.12}$$

Παράδειγμα 36 *Η συνάρτηση που δίνεται από την σχέση:*

$$f(x) = \frac{x}{x^2 + a}, \qquad a \in \mathbb{R}^+$$

μπορεί να χρησιμοποιηθεί σαν μη γραμμικός νευρωνικός τελεστής;

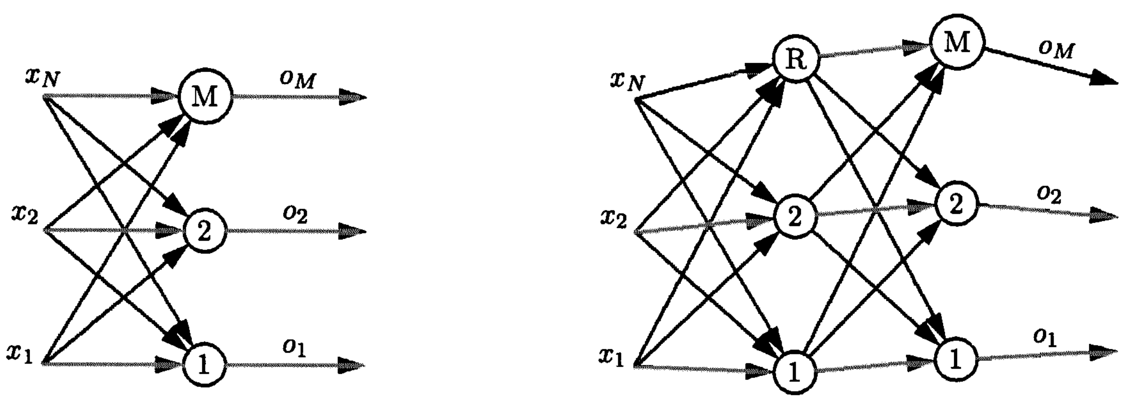
Η συνάρτηση $f(x)$ δεν είναι αύξουσα διότι αν υπολογίσουμε την πρώτη της παράγωγο:

$$f'(x) = \frac{-2x^2}{(a+x^2)^2} + \frac{1}{a+x^2} = \frac{a - x^2}{(a+x^2)^2}$$

βλέπουμε ότι ο αριθμητής είναι ένα πολυώνυμο δεύτερης τάξης το οποίο έχει πραγματικές ρίζες οπότε και η αντίστοιχη παράγωγος της συνάρτησης $f(x)$ παίρνει θετικές και αρνητικές τιμές. Η συνάρτηση $f(x)$ δεν είναι αύξουσα σε όλο το πεδίο ορισμού της και συνεπώς δεν μπορεί να είναι ένας μη γραμμικός νευρωνικός τελεστής.

4.3 Δομή νευρωνικών δικτύων

Παρόλο που δεν υπάρχει περιορισμός για τον τρόπο με τον οποίο οργανώνονται οι συνάψεις των νευρώνων, υπάρχουν μερικές ομάδες δομών οι οποίες έχουν μελετηθεί εκτενέστερα και είναι αυτές οι οποίες χρησιμοποιούνται συχνότερα σε διάφορες εφαρμογές.



Σχήμα 4.3: Νευρωνικό δίκτυο ενός επιπέδου και Νευρωνικό δίκτυο πολλών επιπέδων

1. *Νευρωνικά δίκτυα ενός επιπέδου.* Αποτελεί την πλέον απλή περίπτωση οργάνωσης ενός νευρωνικού δικτύου. Οι είσοδοι κάθε νευρώνα συνδέονται με τις αντίστοιχες εισόδους του δικτύου και η έξοδος κάθε νευρώνα αποτελεί και έξοδο του δικτύου (σχήμα 4.3). Σε μερικές περιπτώσεις θεωρούμε ότι ο κάθε νευρώνας έχει και μία επιπλέον είσοδο η οποία συνδέεται με μία είσοδο σταθερής στάθμης.

   Τα δίκτυα ενός επιπέδου νευρώνων χρησιμοποιούνται συνήθως σε απλά προβλήματα διότι έχουν δύο σοβαρά μειονεκτήματα.

   (α) Αν υποθέσουμε ότι το σύστημα που θέλουμε να προσωμοιώσουμε έχει N εισόδους και M εξόδους, τότε στην καλύτερη των περιπτώσεων (όταν κάθε νευρώνας έχει συνάψεις που συνδέουν όλες τις εισόδους) το σύστημα προσωμοίωσης έχει $(N+1) \times M$ βαθμούς ελευθερίας, γεγονός που περιορίζει την ικανότητα του δικτύου να προσωμοιώνει πολύπλοκες διανυσματικές συναρτήσεις.

   Ο περιορισμός αυτός στην πράξη μπορεί να ξεπεραστεί επεκτείνοντας με τεχνητό τρόπο το διάνυσμα της εισόδου οπότε αυξάνουμε ταυτόχρονα και τους βαθμούς ελευθερίας του συστήματος προσωμοίωσης. Η επέκταση αυτή μπορεί να επιτευχθεί όταν συνδυάσουμε με γραμμικό ή μη γραμμικό τρόπο τις αρχικές εισόδους του συστήματος.

   (β) Ο δεύτερος περιορισμός έχει να κάνει με τις περιορισμένες δυνατότητες του δικτύου στην προσωμοίωση των μη γραμμικών χαρακτηριστικών της συνάρτησης που θέλουμε να προσομοιώσουμε. Συγκεκριμένα μπορούμε να δούμε ότι κάθε έξοδος μπορεί να εκφραστεί σαν συνάρτηση των δεδομένων εισόδου στην εξής μορφή:

   $$o_m = f(\sum_i x_i w_{mi}) \tag{4.13}$$

   Η έξοδος o_m μπορεί να προσωμοιώσει την πραγματική έξοδο του συστήματος μέσω μιας περιορισμένων δυνατοτήτων συνάρτησης.

   Κλασσικό πρόβλημα των περιορισμένων δυνατοτήτων των νευρωνικών δικτύων ενός επιπέδου στην ταξινόμηση προτύπων αποτελεί και η αδυναμία σωστής ταξινόμησης μη γραμμικά διαχωρίσιμων προτύπων.

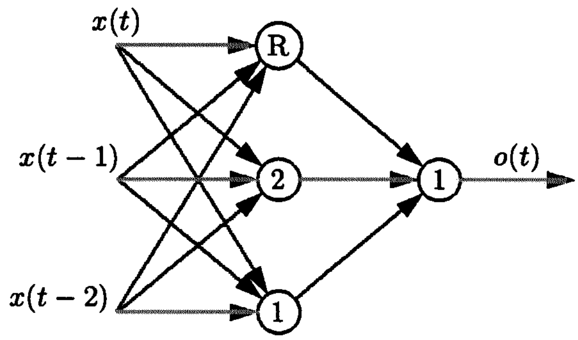
2. *Πολυεπίπεδα νευρωνικά δίκτυα.* Προσθέτονας ένα ή περισσότερα κρυφά επίπεδα μπορούμε να αυξήσουμε απεριόριστα τους βαθμούς ελευθερίας του νευρωνικού δικτύου. Συνήθης τακτική

είναι οι νευρώνες κάθε επιπέδου να συνδέονται με τις εξόδους των νευρώνων που βρίσκονται στο προηγούμενο επίπεδο (σχήμα 4.3). Οταν ο νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου η σύνδεση ονομάζεται *πλήρης* διαφορετικά ονομάζεται *μερική*.

Κατά την διάρκεια εκπαίδευσης του νευρωνικού δικτύου αν κάποιος συντελεστής βαρύτητας σύναψης πάρει την τιμή μηδέν, τότε το γεγονός αυτό ισοδυναμεί με την διακοπή σύνδεσης των νευρώνων διότι δεν μεταφέρεται πλέον πληροφορία από αυτή την σύναψη. Στην πράξη λοιπόν το δίκτυο μπορεί να εμφανίζει αρχιτεκτονική πλήρους σύνδεσης αλλά κατά την διάρκεια εκπαίδευσης κάποιες συνδέσεις μπορεί να αποκοπούν, γεγονός όχι σπάνιο σε πρακτικές εφαρμογές.

Στην ταξινόμηση προτύπων έχει αποδειχθεί ότι ένα δίκτυο που αποτελείται από νευρώνες perceptron δύο επιπέδων (ενός κρυφού και του επιπέδου εξόδου) μπορεί να ταξινομήσει σωστά γραμμικά και μη γραμμικά διαχωρίσιμες κατηγορίες προτύπων, ενώ ένα δίκτυο τριών επιπέδων νευρώνων μπορεί να ταξινιμήσει σωστά κατηγορίες προτύπων μη γραμμικά διαχωρίσιμες που επιπλέον βρίσκονται σε διαφορετικές νησίδες στο χώρο των μετήσεων.

Πολυεπίπεδα νευρωνικά δίκτυα έχουν χρησιμοποιηθεί με επιτυχία στην σχεδίαση πολύπλοκων συστημάτων ταξινόμησης προτύπων και καλύπτουν την πλειονότητα των εφαρμογών ταξινόμησης προτύπων με νευρωνικά δίκτυα. Αν στην είσοδο του δικτύου το οποίο έχει μόνο μία είσοδο, τοποθετηθούν δειγματοληπτημένα σήματα διακριτού χρόνου κατά αύξουσα χρονική σειρά τότε το δίκτυο λειτουργεί σαν ένα μη γραμμικό φίλτρο FIR (σχήμα 4.4).



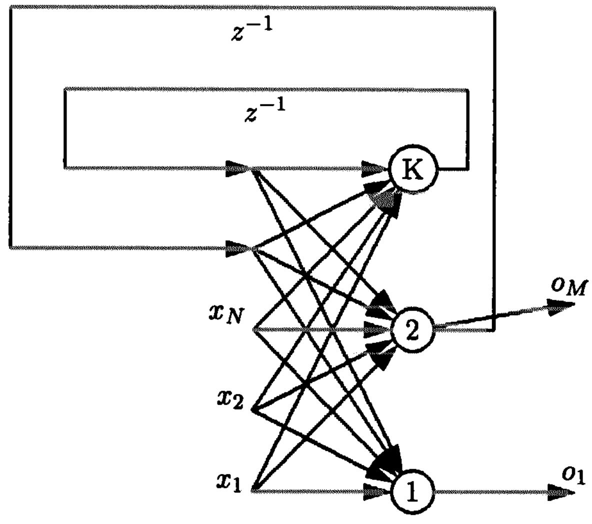
Σχήμα 4.4: Νευρωνικό δίκτυο πολλών επιπέδων σαν μη γραμμικό φίλτρο FIR

3. *Επανατροφοδοτούμενα νευρωνικά δίκτυα.* Οταν υπάρχει έστω και μία διαδρομή μέσω της οποίας, ξεκινώντας από έναν νευρώνα και μέσω των συνάψεων και κατά την φορά ενεργοποίησης του δικτύου μπορούμε να επανέλθουμε στον νευρώνα εκκίνησης, τότε το νευρωνικό δίκτυο θα λέγεται επανατροφοδοτούμενο (recurrent).

   Τυπικό παράδειγμα επανατροφοδοτούμενου νευρωνικού δικτύου δίνεται στο σχήμα 4.5.

   Η αρχιτεκτονική αυτή προσδίδει στο νευρωνικό δίκτυα κάποια ιδιαίτερα χαρακτηριστικά. Οι χρονικές καθυστερήσεις δίνουν την δυνατότητα στο δίκτυο να προσομοιώσει χρονικά μεταβαλλόμενα πρότυπα. Επειδή σε αυτό το κεφάλαιο, όπως και στα προηγούμενα, η μελέτη μεθόδων ταξινόμησης προτύπων γίνεται για στατικά πρότυπα γιαυτό τον λόγο και σε αυτό το κεφάλαιο δεν θα γίνει καμμία αναφορά στην επεξεργασία χρονικά μεταβαλλόμενων προτύπων με αυτά τα δίκτυα.

   Στο παρόν κεφάλαιο γίνεται ειδική αναφορά μόνο για το επανατροφοδοτούμενο δίκτυο Hopfield το οποίο χρησιμοποιείται για ταξινόμηση στατικών πρότυπων.



Σχήμα 4.5: Τυπική μορφή επανατροφοδοτούμενου νευρωνικού δικτύου

4.4 Εκπαίδευση νευρωνικών δικτύων

Οι εργασίες που πρέπει να πραγματοποιηθούν κατά την κατασκευή νευρωνικού δικτύου ταξινόμησης προτύπων είναι οι ακόλουθες:

1. Αρχικά πρέπει να οριστεί η τοπολογία του, δηλαδή ο αριθμός των νευρώνων που θα το απαρτίζουν και η οργάνωση των συνάψεων, ο τρόπος δηλαδή με τον οποίο η πληροφορία θα μεταβιβάζεται στους νευρώνες. Δυστυχώς δεν υπάρχει μία γενικά αποδεκτή μέθοδος η οποία να προσδιορίζει την αρχιτεκτονική του δικτύου σε σχέση με το πρόβλημα προσωμοίωσης που καλούμαστε να λύσουμε.

2. Υπολογίζεται οι συντελεστές βαρύτητας των συνάψεων οι οποίοι ονομάζονται και μνήμη του νευρωνικού δικτύου. Η διαδικασία εκπαίδευσης υπολογίζει συνήθως τους συντελεστές βαρύτητας από ένα σύνολο παραδειγμάτων. Αν είναι γνωστή και η επιθυμούμενη έξοδος τότε εφαρμόζουμε έναν αλγόριθμο κατευθυνόμενης εκπαίδευσης, διαφορετικά χρησιμοποιούμε αλγόριθμο αυτοεκπαίδευσης του δικτύου.

Σε αυτές τις σημειώσεις δίνεται ιδιαίτερη αναφορά στην κατηγορία εκείνων των προβλημάτων στα οποία κατά την εκπαίδευση η επιθυμούμενη έξοδος είναι γνωστή. Δόθηκε ιδιαίτερη βαρύτητα σε αυτή την κατηγορία προβλημάτων διότι αυτά συντάμε συνήθως σε πρακτικές εφαρμογές ταξινόμησης προτύπων και επιπλέον η κατεύθυνόμενη εκπαίδευση δίνει συνήθως πολύ καλύτερα αποτελέσματα από τις μεθόδους αυτοεκπαίδευσης.

4.4.1 Διόρθωση σφάλματος εξόδου

Η μέθοδος διόρθωσης του σφάλματος της εξόδου αποτελεί την πλέον διαδεδομένη τεχνική κατευθυνόμενης εκπαίδευσης. Είναι μία γενική μέθοδος υπολογισμού των παραμέτρων γραμμικών και μη γραμμικών συστημάτων.

Εστω σύστημα $\mathbf{s}(\mathbf{x}, \mathbf{w})$ όπου $\mathbf{x}$ είναι η είσοδος του συστήματος και $\mathbf{w}$ είναι ένα σύνολο παραμέτρων του συστήματος των οποίων θέλουμε να υπολογίσουμε την αριθμητική τους τιμή από παραδείγματα. Εστω ότι είναι διαθέσιμα M παραδείγματα. Τοποθετούμε τα παραδείγματα στην είσοδο του συστήματος και υπολογίζουμε την έξοδο για κάθε ένα από αυτά. Εστω ότι $\mathbf{x}_i, i = 1, M$ είναι οι είσοδοι και $\mathbf{t}_i, i = 1, M$ οι επιθυμούμενες έξοδοι. Το μέσο τετραγωνικό σφάλμα του συστήματος (LMS) είναι:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{t}_i - \mathbf{s}(\mathbf{x}_i, \mathbf{w}))^T(\mathbf{t}_i - \mathbf{s}(\mathbf{x}_i, \mathbf{w})) \tag{4.14}$$

Η μέθοδος διόρθωσης του σφάλματος της εξόδου προσπαθεί να υπολογίσει το διάνυσμα $\mathbf{w}$ το οποίο ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα. Η λύση δίνεται συνήθως από την επίλυση του συστήματος εξισώσεων:

$$\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial \mathbf{w}} = \mathbf{0} \tag{4.15}$$

Στις περιπτώσεις κατά τις οποίες δεν είναι δυνατή η εύρεση αναλυτικής λύσης χρησιμοποιούμε αριθμητικές μεθόδους επαναληπτικής ελάττωσης του σφάλματος. Με αυτές τις μεθόδους επαναπροσδιορίζουμε τους συντελεστές βαρύτητας των συνάψεων έτσι ώστε να πετύχουμε διαδοχικές μειώσεις του συνολικού σφάλματος. Οταν το σφάλμα εκφράζεται σαν το τετράγωνο της διαφοράς της πραγματικής από την επιθυμούμενη έξοδο του νευρωνικού δικτύου τότε ο μέθοδος καλείται αλγόριθμος LMS (least mean square).

4.4.2 Εκπαίδευση Hebbian

Ο νευροφυσιολόγος Hebb περιέγραψε την διαδικασία μάθησης των νευρικών κυττάρων στο βιβλίο του "The Organization of Behaviour" (1949) ως εξής: "Οταν ένα νευρικό κύτταρο A διεγείρει συστηματικά ένα νευρικό κύτταρο B τότε μία διαδικασία μεταβολισμού αλλάζει την συμπεριφορά του κυττάρου B έτσι ώστε η διέγερση του κυττάρου B να προχαλείται πλέον με ευκολότερο τρόπο από το κύτταρο A".

Είναι προφανές ότι η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας που περιέγραψε ο Hebb λαμβάνει υπόψη την σχέση ενεργοποίησης που έχουν αλληλοσυνδεόμενοι νευρώνες οπότε και δεν χρειάζεται να γνωρίζουμε την επιθυμούμενη έξοδο του δικτύου. Συνεπώς οι συνάψεις Hebb αποτελούν μία διαδικασία αυτο-εκπαίδευσης ενός νευρωνικού συστήματος.

Εικοσιπέντε χρόνια αργότερα η σύναψη Hebb ορίσθηκε σαν μία διαδικασία μεταβολής του συντελεστή βαρύτητας σύναψης εξαρτώμενη από την δραστηριότητα των δύο νευρώνων που συνδέονται στα άκρα της. Πιο συγκεκριμένα ορίσθηκαν δύο κανόνες που περιγράφουν την συμπεριφορά μιας σύναψης Hebb:

1. Οταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης Hebb είναι ενεργοποιημένοι ταυτόχρονα, τότε η συνδετική της ικανότητα αυξάνει. Η συνδετική ικανότητα της σύναψης η οποία ονομάζεται και κέρδος της σύναψης δεν είναι τίποτα άλλο από την αριθμητική τιμή του αντίστιχου συντελεστή βαρύτητας.

2. Οταν οι νευρώνες που βρίσκονται στα άκρα σύναψης Hebb ενεργοποιούνται ασύγχρονα, τότε η συνδετική ικανότητα της σύναψης μειώνεται.

Αν δώσουμε μία ποσοτική περιγραφή της διαδικασίας αυτοεκπαίδευσης μιας σύναψης Hebb, μπορούμε να πούμε ότι η μεταβολή του συντελεστή βαρύτητας w_{ij} που συνδέει τον νευρώνα i με τον νευρώνα j μπορεί να περιγραφεί από την ακόλουθη σχέση:

$$\Delta w_{ij} = H_\varepsilon(o_i, o_j) \tag{4.16}$$

όπου o_i, o_j είναι η στάθμη εξόδου των νευρώνων i (ο δέκτης της πληροφορίας που συνδέεται στην σύναψη) και j (πηγή της πληροφορίας) αντίστοιχα.

Η συνάρτηση $H(x, y)$ επιλέγεται συνήθως να έχει πεδίο τιμών το $\mathbb{R}^+$, γεγονός που περιορίζει τις δυνατότητες της σύναψης στον πρώτο κανόνα εκπαίδευσης. Η απλούστερη μορφή της συνάρτησης εκπαίδευσης που ικανοποιεί τους περιορισμούς που τέθηκαν είναι:

$$\Delta w_{ij} = \alpha o_i o_j \tag{4.17}$$

α είναι μικρός θετικός πραγματικός αριθμός που εκφράζει τον ρυθμό εκπαίδευσης της σύναψης.

Το σημαντικότερο μειονέκτημα των συνάψεων Hebb είναι το γεγονός του χορεσμού του συντελεστή βαρύτητας συνάψεων όταν επαναλαμβάνεται η διαδικασία επαναληπτικού επαναπροσδιορισμού λόγω διαδοχικών ενεργοποιήσεων των δύο νευρώνων που είναι συνδεδεμένοι στα άκρα της σύναψης. Για την αποφυγή τέτοιων καταστάσεων προτάθηκε πρόσφατα ένας επιπρόσθετος μηχανισμός που ελαττώνει την αυξητική τάση των συντελεστών βαρύτητας των συνάψεων ικανοποιώντας έτσι και τον δεύτερο κανόνα των συνάψεων Hebb. Ο όρος αυτός ο οποίος εξαρτάται από την αριθμητική τιμή του συντελεστή βαρύτητας της σύναψης και της κατάστασης του νευρώνα στον οποίο ανήκει η σύναψη. Το φυσικό νόημα αυτού του όρου είναι μία διαδικασίας αφαίρεσης της πληροφορίας που έχει συσσωρευτεί στην σύναψη του νευρώνα:

$$\Delta w_{ij} = H_\varepsilon(o_i, o_j) - H_\alpha(o_i, w_{ij}) \tag{4.18}$$

Η συνάρτηση $H_\alpha(o_i, w_{ij})$ έχει πεδίο τιμών το $\mathbb{R}^+$.

Η πιο απλοποιημένη έκφραση που μπορεί να περιγράψει την διπλή λειτουργία εκπαίδευσης μιας σύναψης Hebb είναι η ακόλουθη:

$$\Delta w_{ij} = \alpha o_i o_j - \beta o_i w_{ij} = \beta o_i(\gamma o_j - w_{ij}) \tag{4.19}$$

όπου $\gamma = \alpha/\beta$.

Η *στατιστική προσέγγιση των μηχανισμών εκπαίδευσης μιας σύναψης Hebb* μπορεί να περιγραφεί υπολογίζοντας την στατιστική συσχέτιση των εξόδων των νευρώνων που είναι συνδεδεμένοι στα άκρα της σύναψης. Αν οι νευρώνες ενεργοποιούνται και απενεργοποιούνται ταυτόχρονα, η συσχέτιση θα παρουσιάζει θετικές τιμές, διαφορετικά στην περίπτωση που ενεργοποιούνται ασύγχρονα, η συσχέτιση θα παρουσιάζει αρνητικές τιμές με συνέπεια την ελάττωση της αριθμητικής τιμής του συντελεστή βαρύτητας του νευρώνα:

$$\Delta w_{ij} = \alpha Conv(o_i, o_j) = \alpha E[(o_i - E[o_i])(o_j - E[o_j])] = \alpha E[o_i o_j] - \alpha E[o_i]E[o_j] \tag{4.20}$$

Η σύναψη που εκτελεί την αντίθετη της Hebbian διαδικασία εκπαίδευσης ονομάζεται anti-Hebbian σύναψη. Η λειτουργία μιας anti-Hebbian σύναψης μπορεί να περιγραφεί ως εξής:

1. Οταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης anti-Hebb είναι ενεργοποιημένοι ταυτόχρονα, τότε η συνδετική ικανότητα της σύναψης μειώνεται.

2. Οταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης anti-Hebb ενεργοποιούνται ασύγχρονα, τότε η συνδετική ικανότητα της σύναψης αυξάνεται.

Ορίζουμε σαν Non-Hebbian σύναψη κάθε διαδικασία εκπαίδευσης που δεν ούτε Hebbian ούτε anti-Hebbian.

Πρόγραμμα 6 *Στατιστική εκπαίδευση συνάψεων Hebbian.*

```
function [Classif] = HebbianOL(x,c,Repet,Alpha)
#
#  [Classif] = HebbianOL(x,c,Repet,Alpha)
#           Hebbian Learning method
#  Input
#      x: Pattern Vectors
#      c: Classes
#      HidNodes: Number of Hidden nodes
#      Repet: number of training repetitions
#      Alpha: Learning rate
#  Output
#      Classif: Classification rate using the C-method
#

NumOfClass = max(c) ;
NumOfPatterns = columns(x) ;
x = [ x ; ones(1,NumOfPatterns) ] ;
Dimens = rows(x) ;
OutWeights = rand(Dimens,NumOfClass) - 0.5 ;
Mi = sum(x')/NumOfPatterns ;

Classif = zeros( NumOfClass, NumOfClass ) ;
for j = 1:NumOfPatterns
Dist = 1.0 ./ (exp( OutWeights' * x(:,j) ) + 1 ) ;
Rec = ArgMax(Dist) ;
Classif(c(j),Rec) = Classif(c(j),Rec) + 1 ;
endfor
Classif

for i = 1:Repet
Out = 1.0 ./ (exp( OutWeights' * x ) + 1 ) ;
Mo = sum(Out')/NumOfPatterns ;
Moi = x * Out'/NumOfPatterns  ;
OutWeights = OutWeights + (Alpha * (Moi - Mi' * Mo) ) ;
#
#  C-Error
#
Rc = zeros(NumOfClass,1) ;
Classif = zeros( NumOfClass, NumOfClass ) ;
for j = 1:NumOfPatterns
Dist = 1.0 ./ (exp( OutWeights' * x(:,j) ) + 1 ) ;
Rec = ArgMax(Dist) ;
Classif(c(j),Rec) = Classif(c(j),Rec) + 1 ;
endfor
Classif
endfor
endfunction
```

Παράδειγμα 37 *Νευρωνικό δίκτυο διαθέτει ένα επίπεδο νευρώνων εξόδου που απαρτίζεται από δύο νευρώνες οι οποίοι συνδέονται μέσω συνάψεων Hebb με δύο εισόδους (σχήμα 4.6). Ο μή γραμμικός τελεστής των νευρώνων δίνεται από την σχέση:*

$$f(x) = \frac{1}{e^{2-0.2x} + 1}$$

Η σύνδεση είναι πλήρης και τα παραδείγματα εκπαίδευσης που είναι διαθέσιμα είναι τα ακόλουθα:

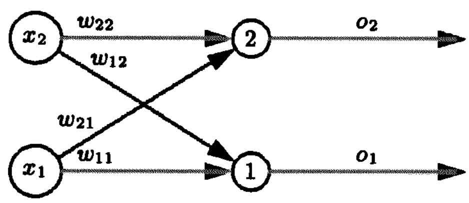
$$\{(1,4),(2,4),(1,5),(0,3),(5,2),(6,-1),(7,2),(5,1)\}$$

1. Υπολογίστε τις εξόδους του δικτύου όταν αυτά ενεργοποιηθούν από τα διαθέσιμα παραδείγματα τοποθετώντας σαν αρχικές τιμές $w_{11} = 0.3$, $w_{12} = 0.1$, $w_{21} = 0.1$, $w_{22} = 0.4$.

2. Υπολογίστε με την στατιστική μέθοδο εκπαίδευσης των συνάψεων Hebb τους συντελεστές βαρύτητας μετά από 40 διαδοχικές επαναλήψεις του αλγόριθμου, θέτοντας τον συντελεστή εκπαίδευσης στην τιμή $\alpha = 0.5$.

3. Δοκιμάστε την συμπεριφορά του δικτύου με διαφορετικές αρχικές τιμές για τους συντελεστές βαρύτητας των συνάψεων.

Στο σχήμα που ακολουθεί δίνεται η γραφική μορφή του δικτύου στο οποίο σημειώνονται και οι τέσσερεις συντελεστές βαρύτητας των αντίστοιχων συνάψεων.



Σχήμα 4.6: Νευρωνικό δίκτυο συνάψεων Hebb αποτελούμενο από δύο νευρώνες εξόδου

Στον πίνακα 4.1 δίνεται η αριθμητική τιμή των συντελεστών βαρύτητας των συνάψεων κατά την διαδοχική εφαρμογή του στατιστικού αλγόριθμου εκπαίδευσης συνάψεων Hebb. Για λόγους οικονομίας χώρου απεικονίζονται οι τιμές ανά δύο επαναλήψεις του αλγόριθμου.

Στον πίνακα 4.2 που ακολουθεί δίνεται η αριθμητική τιμή των εξόδων του δικτύου για τις οχτώ τιμές των διανυσμάτων εκπαίδευσης.

Παρατηρούμε ότι τα παραδείγματα μπορούν εύκολα να διαταχθούν σε δύο ομάδες. Η πρώτη ομάδα περιέχει τα τέσσερα πρώτα πρότυπα ενώ η δεύτερη ομάδα περιέχει τα υπόλοιπα τέσσερα πρότυπα. Τα πρότυπα αυτά είναι μάλιστα γραμμικά διαχωρισμένα απο την ευθεία που ενώνει τα σημεία (0,0) και (1,1).

Απο την μελέτη του πίνακα 4.2 μπορούμε εύκολα να διαπιστώσουμε ότι η αρχική επιλογή των συντελεστών βαρύτητας των νευρώνων επιτρέπει τον διαχωρισμό των προτύπων διότι όταν τοποθετούμε στην είσοδο πρότυπο της πρώτης ομάδας βρίσκεται σε κατάσταση μεγαλύτερης ενεργοποίησης ο δεύτερος νευρώνας. Το φαινόμενο αυτό οφείλεται στην αριθμητική τιμή της σύναψης που συνδέει την δεύτερη τιμή του διανύσματος εισόδου με τον δεύτερο νευρώνα (το w_{22}). Η αριθμητική τιμή του συντελεστή βαρύτητας της σύναψης αυξάνει διαδοχικά (απο 0.4 γίνεται 2.89) με συνέπεια την ολοένα και αυξανόμενη τιμή ενεργοποίησης του δεύτερου νευρώνα, όταν στην είσοδο τοποθετείται πρότυπο της πρώτης ομάδας. Το πρότυπο αυτό θέτει μεγαλύτερη αριθμητική τιμή στην πρώτη θέση της εισόδου. Το ίδιο γεγονός συμβαίνει και με την σύναψη που συνδέει την πρώτη είσοδο με τον πρώτο νευρώνα. Ο συντελεστής βαρύτητας αυξάνει από 0.3 σε 11.86. Σε αυτές τις δύο συνάψεις βλέπουμε στην πράξη την εφαρμογή του πρώτου κανόνα του Hebb: "Οταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης Hebb είναι ενεργοποιημένοι ταυτόχρονα, τότε η συνδετική της ικανότητα αυξάνει".

Στις άλλες δύο συνάψεις παρατηρούμε μία συνεχή μείωση των συντελεστών βαρύτητας. Ο μεν πρώτος συντελεστής w_{12} ελαττώνεται από 0.1 σε -3.51, ο δε δεύτερος συντελεστής w_{21} ελαττώνεται από 0.1 σε -7.16. Το φαινόμενο αυτό οφείλεται στο γεγονός ότι όταν το δίκτυο ενεργοποιείται από το διάνυσμα εισόδου η αριθμητική τιμή της εισόδου και του νευρώνα εξόδου που συνδέει η σύναψη έχουν τιμές αντίθετων χαρακτηριστικών, δηλαδή υψηλή τιμή εισόδου με χαμηλή τιμή εξόδου του νευρώνα και το αντίθετο. Συνεπώς σε αυτές τις δύο συνάψεις παρακολουθούμε την εφαρμογή του

Πίνακας 4.1: Αριθμητική τιμή των συντελεστών βαρύτητας κατά τον διαδοχική εφρμογή του στατιστικού αλγόριθμου εκπαίδευσης συνάψεων Hebb

Πριν	w_{11}	w_{12}	w_{21}	w_{22}
0	0.3	0.1	0.1	0.4
2	0.340	0.079	0.080	0.424
4	0.390	0.054	0.055	0.453
6	0.452	0.021	0.023	0.486
8	0.530	-0.018	-0.018	0.524
10	0.627	-0.069	-0.073	0.569
12	0.751	-0.131	-0.145	0.622
14	0.912	-0.207	-0.242	0.684
16	1.124	-0.299	-0.371	0.756
18	1.408	-0.410	-0.549	0.839
20	1.796	-0.542	-0.796	0.935
22	2.329	-0.697	-1.140	1.046
24	3.042	-0.878	-1.601	1.173
26	3.934	-1.087	-2.174	1.317
28	4.957	-1.326	-2.826	1.478
30	6.056	-1.597	-3.520	1.659
32	7.193	-1.902	-4.236	1.860
34	8.349	-2.244	-4.963	2.083
36	9.514	-2.624	-5.695	2.328
38	10.68	-3.046	-6.431	2.599
40	11.86	-3.514	-7.169	2.897

δεύτερου κανόνα του Hebb: "Οταν οι νευρώνες που βρίσκονται στα άκρα σύναψης Hebb ενεργοποιούνται ασύγχρονα, τότε η συνδετική ικανότητα της σύναψης μειώνεται".

Το φαινόμενο της ενεργοποίησης διαφορετικών νευρώνων, όταν πρότυπα διαφορετικών ομάδων τίθενται στην είσοδο του συστήματος είναι ένα επιθυμητό χαρακτηριστικό που είναι ιδιαίτερο χρήσιμο σε εφαρμογές ταξινόμησης προτύπων. Η ιδιότητα αυτή διατηρείται και μετά την εκπαίδευση Hebb η οποία μάλιστα αυξάνει την διαφορά των ενεργοποιήσεων.

Ας δούμε τι συμβαίνει όταν τοποθετήσουμε διαφορετικές αρχικές τιμές στους συντελεστές βαρύτητας των συνάψεων. Εκτελώντας τους ίδιους υπολογισμούς, με αρχικές τιμές των συντελεστών βαρύτητας των συνάψεων τις: $w_{11} = 0.1$, $w_{12} = 0.2$, $w_{21} = 0.2$, $w_{22} = 0.4$ έχουμε τα αποτελέσματα του πίνακα 4.3.

Τα αριθμητικά αποτελέσματα από την σκοπιά ενός συστήματος ταξινόμησης προτύπων δείχνουν ότι ενώ με την αρχική επιλογή των συντελεστών βαρύτητας το σύστημα διατηρεί πάντα περισσότερο ενεργοποιημένο τον δεύτερο νευρώνα (το νευρωνικό σύστημα ταξινόμησης προτύπων έχει ελάχιστο σφάλμα επιτυχούς ταξινόμησης 50%), μετά την

Πίνακας 4.2: Εξοδος του νευρωνικού δικτύου πριν και μετά την εκπαίδευση με είσοδο τα διανύσματα εκπαίδευσης

Αρχικά	1	2	3	4	5	6	7	8
o_1	0.134	0.141	0.137	0.125	0.159	0.159	0.176	0.157
o_2	0.159	0.162	0.170	0.146	0.149	0.123	0.154	0.13
Εκπαίδευση	1	2	3	4	5	6	7	8
o_1	0.004	0.047	0.001	0.001	0.999	0.999	0.999	0.999
o_2	0.404	0.251	0.548	0.434	0.012	0.001	0.003	0.007

Πίνακας 4.3: Υπολογισμός συντελεστών βαρύτητας με τον στατιστικό αλγόριθμο εκπαίδευσης συνάψεων Hebb

Επανάληψη	w_{11}	w_{12}	w_{21}	w_{22}
0	0.1	0.2	0.2	0.4
2	0.0983599	0.196429	0.207349	0.416376
4	0.0958321	0.190634	0.215433	0.434603
6	0.0922128	0.182031	0.224386	0.455061
8	0.087255	0.169896	0.234371	0.478215
10	0.0806597	0.153332	0.245583	0.504634
12	0.0720661	0.131245	0.258254	0.535007
14	0.0610398	0.102298	0.272665	0.570166
16	0.0470603	0.0648842	0.289149	0.611101
18	0.0295068	0.0170937	0.308101	0.658978
20	0.00764311	-0.0433071	0.329986	0.715142
22	-0.0193973	-0.118875	0.355351	0.781115
24	-0.052625	-0.212459	0.384826	0.85857
26	-0.0932057	-0.327138	0.419136	0.949283
28	-0.142466	-0.466112	0.459098	1.05507
30	-0.201892	-0.632569	0.505617	1.17772
32	-0.273114	-0.829551	0.55968	1.3189
34	-0.357887	-1.05986	0.622332	1.48014
36	-0.458045	-1.32603	0.694653	1.66283
38	-0.575454	-1.63045	0.777724	1.86832
40	-0.711962	-1.97548	0.872598	2.09797

εκπαίδευση των συνάψεων Hebb με την στατιστική μέθοδο, που επαναλαμβάνεται 40 φορές, βλέπουμε ότι το δίκτυο αυτοοργανώνεται. Η οργάνωση του δικτύου είναι τέτοια ώστε όταν τίθεται πρότυπο της πρώτης ομάδας στην είσοδο ενεργοποιείται περισσότερο ο δεύτερος νευρώνας ενώ όταν στην είσοδο τίθεται διάνυσμα της δεύτερης ομάδας προτύπων ενεργοποιείται περισσότερο ο πρώτος νευρώνας. Το ελάχιστο σφάλμα του συστήματος ταξινόμησης πρότυπων πέφτει στο 0%.

Με αυτό το παράδειγμα βλέπουμε ότι οι συνάψεις Hebb είναι δυνατόν να διαχωρίσουν με αυτόματο τρόπο τα πρότυπα εκπαίδευσης.

Πίνακας 4.4: Εξοδοι του δικτύου πριν και μετά την εκπαίδευση

Αρχικά	1	2	3	4	5	6	7	8
o_1	0.139	0.141	0.144	0.132	0.139	0.127	0.144	0.134
o_2	0.162	0.167	0.173	0.146	0.162	0.137	0.173	0.151
Εκπαίδευση	1	2	3	4	5	6	7	8
o_1	0.190	0.169	0.219	0.185	0.086	0.046	0.066	0.073
o_2	0.328	0.247	0.426	0.322	0.041	0.008	0.019	0.027

4.4.3 Ανταγωνιστική Εκπαίδευση

Η δεύτερη μέθοδος αυτοεκπαίδευσης που θα εξετάσουμε είναι η ανταγωνιστική εκπαίδευση που εφαρμόζεται σε **ανεξάρτητες ομάδες** νευρώνων του δικτύου.

Κατά την εκπαίδευση συνάψεων Hebb είδαμε ότι η συνάρτηση επαναπροσδιορισμού του συντελεστή βαρύτητας της σύναψης εξαρτάται μόνο από την έξοδο των νευρώνων που η σύναψη συνδέει. Με αυτή την μέθοδο επαναπροσδιορίζονται οι συντελεστές βαρύτητας των συνάψεων κάθε φορά που ενεργοποιείται η διαδικασία εκπαίδευσης. Το μεγάλο μειονέκτημα των συνάψεων Hebb είναι το φαινόμενο του κορεσμού των νευρώνων το οποίο εμφανίζεται όταν η διαδικασία εκπαίδευσης επαναληφθεί πολλές φορές και ο συντελεστής εκπαίδευσης έχει μεγάλη αριθμητική τιμή. Σε αυτή την περίπτωση, όπως μπορούμε να παρατηρήσουμε και από τα αριθμητικά αποτελέσματα του προηγούμενου παραδείγματος, οι συντελεστές βαρύτητας τείνουν να πάρουν πολύ μεγάλες ή πολύ μικρές αριθμητικές τιμές.

Σε αντίθεση με την λειτουργία εκπαίδευσης των συνάψεων Hebb, η ανταγωνιστική εκπαίδευση επαναπροσδιορίζει τους συντελεστές βαρύτητας των συνάψεων **ενός μόνο νευρώνα** από ένα σύνολο ανεξάρτητων ομάδων νευρώνων. Ανεξάρτητη ομάδα νευρώνων ονομάζεται κάθε υποσύνολο νευρώνων του δικτύου που περιέχει όλους τους νευρώνες οι συνάψεις των οποίων δέχονται σήματα από τους ίδιους ακριβώς νευρώνες.

Πιο συγκεκριμένα, αν θεωρήσουμε δύο διαδοχικές ομάδες νευρώνων την Α και την Β, όπου οι έξοδοι της Α συνδέονται μέσω συνάψεων με τους νευρώνες της Β τότε η διαδικασία ανταγωνιστικής εκπαίδευσης έχει ως εξής:

1. *Αρχικές τιμές.* Τοποθετούμε τυχαίους θετικούς πραγματικούς αριθμούς στους συντελεστές βαρύτητας των συνάψεων και κανονικοποιούμε τις τιμές τους για κάθε ένα νευρώνα ξεχωριστά, έτσι ώστε:

$$\sum_i w_{ji} = 1 \tag{4.21}$$

Εναλλακτικά έχει χρησιμοποιηθεί με επιτυχία και η κανονικοποίηση του Ευκλείδιου χώρου:

$$\sum_i {w_{ji}}^2 = 1 \tag{4.22}$$

2. Επιλέγουμε τυχαία ένα πρότυπο εκπαίδευσης και με αυτό ενεργοποιούμε το δίκτυο.

3. Βρίσκουμε τον νευρώνα j της ομάδας Β ο οποίος βρίσκεται στην υψηλότερη στάθμη, έχει δηλαδή την υψηλότερη τιμή εξόδου. Τον νευρώνα αυτό ονομάζουμε νικητή (winning neuron).

4. Επαναπροσδιορίζουμε όλες τις συνάψεις του νικητή νευρώνα σύμφωνα με την σχέση:

$$\Delta w_{ji} = \alpha(o_i - w_{ji}) \tag{4.23}$$

όπου α είναι μικρός θετικός πραγματικός αριθμός που καθορίζει τον ρυθμό εκπαίδευσης του δικτύου και o_i είναι η έξοδος του νευρώνα i του προηγούμενου επιπέδου (Α) το οποίο συνδέεται μέσω της σύναψης που έχει συντελεστή βαρύτητας w_{ji} με τον νευρώνα j.

5. Οι συντελεστές βαρύτητας των συνάψεων του νικητή νευρώνα κανονικοποιούνται.

Πρόγραμμα 7 *Ανταγωνιστική εκπαίδευση νευρωνικού δικτύου.*

```
function [Classif] = CompetOLLin(x,c,Repet,Alpha)
#
#  [Classif] = CompetOLLin(x,c,Repet,Alpha)
#          Competitive Learning method (Sum of weights equal one)
#  Input
#      x: Pattern Vectors
#      c: Classes
#      HidNodes: Number of Hidden nodes
#      Repet: number of training repetitions
#      Alpha: Learning rate
#  Output
#      Classif: Classification rate using the C-method
#

NumOfClass = max(c) ;
NumOfPatterns = columns(x) ;
x = [ x ; ones(1,NumOfPatterns) ] ;
Dimens = rows(x) ;
OutWeights = rand(Dimens,NumOfClass) - 0.5 ;
Norm = sum( OutWeights ) ;
for i=1:NumOfClass
OutWeights(:,i) = OutWeights(:,i) / Norm(i) ;
endfor

#
#  C-Error
#
Classif = zeros( NumOfClass, NumOfClass ) ;
for j = 1:NumOfPatterns
Dist = 1.0 ./ (exp( OutWeights' * x(:,j) ) + 1 ) ;
Rec = ArgMax(Dist) ;
Classif(c(j),Rec) = Classif(c(j),Rec) + 1 ;
endfor
Classif
fflush(stdout) ;

for i = 1:Repet
Pat = floor( rand() * NumOfPatterns ) + 1
Out = 1.0 ./ (exp( OutWeights' * x(:,Pat) ) + 1 ) ;
Win = ArgMax(Out) ;
OutWeights(:,Win) = OutWeights(:,Win) + Alpha * (x(:,Pat) - OutWeights(:,Win) ) ;
Norm = sum( OutWeights(:,Win) ) ;
OutWeights(:,Win) = OutWeights(:,Win) / Norm ;
#
#  C-Error
#
Classif = zeros( NumOfClass, NumOfClass ) ;
for j = 1:NumOfPatterns
Dist = 1.0 ./ (exp( OutWeights' * x(:,j) ) + 1 ) ;
Rec = ArgMax(Dist) ;
Classif(c(j),Rec) = Classif(c(j),Rec) + 1 ;
endfor
Classif
fflush(stdout) ;
endfor
endfunction


function [Res] = ArgMax(x)
```

```
#
# Res = ArgMax(x)
# Return the POSITION of the maximum value
#
Num = length(x) ;
Res = 1 ;
MaxNum = x(1) ;
for i = 1:Num
    if (x(i) > MaxNum)
       MaxNum = x(i) ;
       Res = i ;
    endif
endfor
endfunction
```

Παράδειγμα 38 *Για το νευρωνικό δίκτυο του προηγούμενου παραδείγματος υπολογίστε τους συντελεστές βαρύτητας των συνάψεων με την μέθοδο της ανταγωνιστικής εκπαίδευσης. Συγκρίνατε τα αποτελέσματα της εκπαίδευσης με τα αντίστοιχα αποτελέσματα με την εκπαίδευση των συνάψεων Hebb . Ποιά είναι τα συμπεράσματά σας;*

Αρχικά αναδιατάσσουμε τα παραδείγματα έτσι ώστε η εκπαίδευση να εκτελείται κάθε φορά χρησιμοποιώντας πρότυπα διαφορετικών κατηγοριών. Σε πρακτικά συστήματα αυτό που κάνουμε είναι να αναδιατάσσουμε με τυχαίο τρόπο τα πρότυπα εκπαίδευσης.

$$\{(1,4),(5,2),(2,4),(6,-1),(1,5),(7,2),(0,3),(5,1)\}$$

Θεωρούμε ότι το δίκτυο που θα κατασκευάσουμε έχει νευρώνες με μη γραμμικό τελεστή την σιγμοειδή συνάρτηση:

$$f(x) = \frac{1}{e^{2-0.2x}+1}$$

Τοποθετούμε σαν αρχικές τιμές για τους συντελεστές βαρύτητας όλων των συνάψεων την τιμή 0.5. Για όλα τα παραδείγματα εκτελούμε τον αλγόριθμο της ανταγωνιστικής εκπαίδευσης οπότε σε κάθε κύκλο εκπαίδευσης έχουμε τους συντελεστές βαρύτητας που δίνονται στον πίνακα που ακολουθεί:

Πίνακας 4.5: Υπολογισμός συντελεστών βαρύτητας με τον ανταγωνιστικό αλγόριθμο εκπαίδευσης

Επανάληψη	w_{11}	w_{12}	w_{21}	w_{22}
0	0.5	0.5	0.5	0.5
1	0.148571	0.851429	0.82893	0.17107
2	0.119053	0.880947	0.838788	0.161212
3	0.117854	0.882146	0.838984	0.161016
4	0.117805	0.882195	0.838988	0.161012
5	0.117803	0.882197	0.838988	0.161012

Οπως και στην περίπτωση της εκπαίδευσης των συνάψεων Hebb έτσι και στην περίπτωση της ανταγωνιστικής εκπαίδευσης το δίκτυο φαίνεται να αυτοοργανώνεται και μάλιστα σε μικρότερο αριθμό επαναλήψεων. Οι συντελεστές βαρύτητας των συνάψεων φαίνεται να συγκλίνουν στις τελικές τους τιμές πολύ γρήγορα.

Το σύστημα που κατασκευάσαμε διαχωρίζει τα πρότυπα των κατηγοριών σε δύο ομάδες οι οποίες ενεργοποιούν κάθε φορά διαφορετικό νευρώνα εξόδου.

Αν επαναλάβουμε την διαδικασία εκπαίδευσης με διαφορετικές αρχικές τιμές, τις $w_{11} = 0.1$, $w_{12} = 0.9$, $w_{21} = 0.9$, $w_{22} = 0.1$ και τον ίδιο συντελεστή εκπαίδευσης θα έχουμε, μετά την σύγκλιση του ανταγωνιστικού αλγόριθμου εκπαίδευσης, τις ακόλουθες τιμές εξόδου των νευρώνων για τα πρότυπα εκπαίδευσης:

Πίνακας 4.6: Εξοδοι του δικτύου πριν και μετά την εκπαίδευση

Αρχικά	1	2	3	4	5	6	7	8
o_1	0.182	0.214	0.197	0.182	0.197	0.249	0.154	0.197
o_2	0.182	0.214	0.197	0.182	0.197	0.249	0.154	0.197
Εκπαίδευση	1	2	3	4	5	6	7	8
o_1	0.213	0.175	0.217	0.116	0.242	0.182	0.182	0.152
o_2	0.155	0.258	0.179	0.274	0.159	0.331	0.129	0.252

Αρχικά	1	2	3	4	5	6	7	8
o_1	0.213	0.175	0.217	0.116	0.242	0.182	0.182	0.152
o_2	0.155	0.258	0.179	0.274	0.159	0.331	0.129	0.252

Η επιλογή του ρυθμού εκπαίδευσης δεν επηρεάζει, για το συγκεκριμένο πρόβλημα, την αξιοπιστία του συστήματος ταξινόμησης. Αυτό που φαίνεται να επηρεάζεται είναι η ταχύτητα σύγκλισης του αλγόριθμου εκπαίδευσης, όπως φαίνεται και στον επόμενο πίνακα ο οποίος μας δείχνει την αριθμητική τιμή των συντελεστών βαρύτητας των συνάψεων μετά από κάθε ολοκληρωμένο κύκλο εκμάθησης του δικτύου με όλα τα παραδείγματα.

Στο συγκεκριμένο παράδειγμα η ταχύτητα σύγκλισης μεγιστοποιείται όταν επιλέξουμε τον ρυθμό εκπαίδευσης στην τιμή 0.5.

Εκπαίδευση με 0.01	w_{11}	w_{12}	w_{21}	w_{22}
0	0.1	0.9	0.9	0.1
1	0.188467	0.125648	0.151871	0.253506
2	0.117106	0.882894	0.888963	0.111037
3	0.131156	0.868844	0.880482	0.119518
4	0.142697	0.857303	0.873965	0.126035
5	0.152175	0.847825	0.868958	0.131042
6	0.15996	0.84004	0.86511	0.13489
7	0.166355	0.833645	0.862153	0.137847
8	0.171607	0.828393	0.859881	0.140119
9	0.17592	0.82408	0.858135	0.141865
10	0.179463	0.820537	0.856794	0.143206
11	0.182374	0.817626	0.855763	0.144237
Εκπαίδευση με 0.1	w_{11}	w_{12}	w_{21}	w_{22}
1	0.188467	0.125648	0.151871	0.253506
2	0.146643	0.853357	0.85314	0.14686
3	0.154739	0.845261	0.848056	0.151944
4	0.156144	0.843856	0.847505	0.152495
5	0.156388	0.843612	0.847445	0.152555
Εκπαίδευση με 0.2	w_{11}	w_{12}	w_{21}	w_{22}
1	0.188467	0.125648	0.151871	0.253506
2	0.117079	0.882921	0.840202	0.159798
3	0.117773	0.882227	0.839013	0.160987
4	0.117802	0.882198	0.838989	0.161011
Εκπαίδευση με 0.5	w_{11}	w_{12}	w_{21}	w_{22}
1	0.188467	0.125648	0.151871	0.253506
2	0.0468537	0.953146	0.830327	0.169673

Στα επόμενα εδάφια θα περιγράψουμε αναλυτικά την μεθοδολογία που ακολουθούμε κατά την σχεδίαση νευρωνικών συστημάτων ταξινόμησης προτύπων με κατευθυνόμενη εκπαίδευση.

4.5 Το γραμμικό φίλτρο Wiener

Εστω ότι διαθέτουμε ένα δίκτυο νευρώνων ενός επιπέδου το οποίο δεν περιέχει μη-γραμμικούς τελεστές. Οι έξοδοι του δικτύου αποτελούν έναν γραμμικό συνδυασμό των εισόδων του. Κάθε έξοδος του δικτύου δίνεται από την ακόλουθη σχέση:

$$o = \sum_{i=1}^{N} x_i w_i \tag{4.24}$$

όπου N είναι ο αριθμός των εισόδων του δικτύου και o η έξοδος ενός νευρώνα.

Εστω d η επιθυμούμενη έξοδος για το πρότυπο της εισόδου: $\mathbf{x} = (x_1, x_2, ..., x_N)^T$.

Το σφάλμα της εξόδου ισούται με την διαφορά της επιθυμούμενης από την πραγματική έξοδο:

$$e = d - o \tag{4.25}$$

Οταν έχουμε στην διάθεσή μας έναν αριθμό παραδειγμάτων τότε ορίζουμε σαν συνολικό σφάλμα το μισό του στατιστικά αναμενόμενου τετραγώνου του σφάλματος:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = \frac{1}{2}E[e^2] \tag{4.26}$$

Οταν υπολογιστούν οι συντελεστές βαρύτητας του γραμμικού δικτύου που ελαχιστοποιεί το στατιστικά αναμενόμενο σφάλμα τότε θα λέμε ότι έχουμε κατασκευάσει ένα φίλτρο Wiener. Ο αλγόριθμος που υπολογίζει του συντελεστές του φίλτρου Wiener ανήκει στην ομάδα των αλγορίθμων LMS.

Οι συντελεστές βαρύτητας των συνάψεων μπορούν να υπολογιστούν είτε με επαναληπτικό είτε με αναλυτικό τρόπο. Εμείς θα περιγράψουμε την αναλυτική μέθοδο η οποία δίνει την βέλτιστη λύση και έχει την μικρότερη υπολογιστική πολυπλοκότητα.

Αντικαθιστώντας τις αναλυτικές εκφράσεις στην συνάρτηση του στατιστικά αναμενόμενου σφάλματος έχουμε:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = \frac{1}{2}E[d^2] - E[\sum_{i=1}^{N} x_i w_i d] + \frac{1}{2}E[\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j x_i x_j] \tag{4.27}$$

Οι συντελεστές βαρύτητας των συνάψεων είναι σταθεροί αριθμοί, συνεπώς οι αναμενόμενες τιμές απλοποιούνται σε:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = \frac{1}{2}E[d^2] - \sum_{i=1}^{N} w_i E[x_i d] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j E[x_i x_j] \tag{4.28}$$

Η εύρεση των τιμών των συντελεστών βαρύτητας που ελαχιστοποιούν το στατιστικά αναμενόμενο σφάλμα υπολογίζονται με την μέθοδο του ελάχιστου μέσου τετραγωνικού σφάλματος (LMS), προκύπτουν δηλαδή από την λύση του ακόλουθου συστήματος εξισώσεων:

$$\frac{\partial}{\partial w_i}\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = 0 \tag{4.29}$$

Αντικαθιστώντας την αναλυτική έκφραση του σφάλματος έχουμε το τελικό γραμμικό σύστημα εξισώσεων, το οποίο συναντάται στην βιβλιογραφία και σαν σύστημα εξισώσεων των Wiener-Hopf:

$$\sum_{j=1}^{N} w_j E[x_i x_j] = E[x_i d] \tag{4.30}$$

Ο όρος $E[x_i d]$ εκφράζει την ετεροσυσχέτιση της εισόδου με την επιθυμούμενη έξοδο, ενώ ο όρος $E[x_i x_j]$ είναι η αυτοσυσχέτιση της εισόδου.

Πρόγραμμα 8 *Υπολογισμός των συντελεστών του γραμμικού φίλτρου Wiener και υπολογισμός του ελάχιστου και μέγιστου σφάλματος του αντίστιχου γραμμικού συστήματος ταξινόμησης προτύπων δύο κατηγοριών.*

```
function [Rc,Ru,Rep] = Wiener(x1,x2)
#
#  [Rc,Ru,Rep] = Wiener(x1,x2)
#
#  Input
#      x1: Pattern Vectors for the first class
#      x2: Pattern Vectors for the second class
#  Output
#      Rc: Correct classification rate using the C-method
#      Ru: Correct classification rate using the U-method
#      Rep: Pattern vectors on each class
#
NumOfP1 = columns(x1) ;
NumOfP2 = columns(x2) ;
Rep = [NumOfP1,NumOfP2] ;
TotPat = sum(Rep) ;
Rc = zeros(2,1) ;
Ru = zeros(2,1) ;
if ( rows(x1) != rows(x2) )
printf( "Error in vectors x1, x2\n" ) ;
endif
x1 = [x1;ones(1,NumOfP1)] ;
x2 = [x2;ones(1,NumOfP2)] ;

#
#  C-Error
#
d1 = ones(NumOfP1,1) ;
d2 = zeros(NumOfP2,1) ;
d = [d1;d2] ;
x = [x1,x2] ;
A = x * x' ;
B = x * d ;
Weights = A \ B ;
for i=1:NumOfP1
if ( Weights' * x1(:,i) >= 0.5 )
Rc(1) = Rc(1) + 1 ;
endif
endfor
for i=1:NumOfP2
if ( Weights' * x2(:,i) < 0.5 )
Rc(2) = Rc(2) + 1 ;
endif
endfor
```

```
#
# U-Error
#
d1 = ones(NumOfP1-1,1) ;
d2 = zeros(NumOfP2,1) ;
d = [d1;d2] ;
for k = 1:NumOfP1
x = [x1(:,1:k-1),x1(:,k+1:NumOfP1),x2] ;
A = x * x' ;
B = x * d ;
Weights = A \ B ;
if ( Weights' * x1(:,k) >= 0.5 )
Ru(1) = Ru(1) + 1 ;
endif
endfor
d1 = ones(NumOfP1,1) ;
d2 = zeros(NumOfP2-1,1) ;
d = [d1;d2] ;
for k=1:NumOfP2
x = [x1,x2(:,1:k-1),x2(:,k+1:NumOfP2)] ;
A = x * x' ;
B = x * d ;
Weights = A \ B ;
if ( Weights' * x2(:,k) < 0.5 )
Ru(2) = Ru(2) + 1 ;
endif
endfor

endfunction
```

Παράδειγμα 39 *Θέλουμε να κατασκευάσουμε γραμμικό σύστημα ταξινόμησης προτύπων δύο κατηγοριών απο τα ακόλουθα παραδείγματα εκπαίδευσης:*

$$\Omega_1 = \{(1,1),(2,2),(1,0),(0,1)\}$$

$$\Omega_2 = \{(-10,-10),(-9,-9),(-8,-9),(-9,-10)\}$$

1. Χρησιμοποιώντας το σύστημα εξισώσεων των Wiener-Hopf υπολογίστε τους συντελεστές βαρύτητας των συνάψεων.

2. Ποιό είναι το ελάχιστο σφάλμα του συστήματος ταξινόμησης που κατασκευάσατε;

Θα κατασκευάσουμε ένα γραμμικό σύστημα ταξινόμησης προτύπων το οποίο θα διαθέτει δύο εξόδους. Οι έξοδοι θα προσομοιώνουν τις συναρτήσεις διάκρισης των αντίστοιχων κατηγοριών. Ορίζουμε επίσης ότι όταν στην είσοδο τίθεται πρότυπο της πρώτης κατηγορίας η πρώτη έξοδος να δίνει τιμή 1.0 και η δεύτερη 0, ενώ όταν στην είσοδο τίθεται πρότυπο της δεύτερης κατηγορίας τότε οι επιθυμούμενες τιμές πρέπει να αντιστρέφονται.

Με βάση τους περιορισμούς που θέσαμε, τα παραδείγματα της κατευθυνόμενης εκπαίδευσης γίνονται ως εξής:

$$\Omega = \{((1,1),(0,1)),((2,2),(0,1)),((1,0),(0,1)),((0,1),(0,1)),$$

$$((-10,-10),(1,0)),((-9,-9),(1,0)),((-8,-9),(1,0)),((-9,-10),(1,0))\}$$

Οι όροι των συντελεστών της γραμμικής εξίσωσης γιά την πρώτη κατηγορία, βάσει των αριθμητικών δεδομένων των παραδειγμάτων, υπολογίζονται:

$$E[x_i x_j] = \begin{pmatrix} 41.5 & -43.5 \\ -43.5 & 46.0 \end{pmatrix}$$

Για τον πρώτο νευρώνα έχουμε $E[x_i d] = (-36.0, 38.0)^T$. Η επίλυση της γραμμικής εξίσωσης δίνει τους εξής συντελεστές βαρύτητας: $(-0.179104, 0.656716)^T$.

Επαναλαμβάνοντας τους ίδιους υπολογισμούς για τον δεύτερο νευρώνα έχουμε $E[x_i d] = (-4.0, 4.0)^T$ και επιλύοντας το σύστημα των εξισώσεων έχουμε τους συντελεστές βαρύτητας $(0.597015, 0.477612)^T$.

Στον πίνακα που ακολουθεί δίνουμε τις εξόδους των νευρώνων για κάθε ένα από τα παραδείγματα εκπαίδευσης. Παρατηρούμε ότι αν ορίσουμε σαν κριτήριο ταξινόμησης τον νευρώνα που περιέχει την μεγαλύτερη αριθμητική τιμή, τότε το σύστημα που κατασκευάσαμε έχει ελάχιστο σφάλμα 0%.

Πίνακας 4.7: Είσοδος και έξοδος γραμμικής προσωμοίωσης συναρτήσεων διάκρισης

Πρότυπο	Εξοδος της ω_1	Εξοδος της ω_2	Ταξινόμηση
1,-1	-0.835821	0.119403	ω_1
2,-1	-1.67164	0.238806	ω_1
1,0	-0.179104	0.597015	ω_1
0,-1	-0.656716	-0.477612	ω_1
-10,10	8.35821	-1.19403	ω_2
-9,9	7.52239	-1.07463	ω_2
-8,9	7.34328	-0.477612	ω_2
-9,10	8.1791	-0.597015	ω_2

4.6 Πολυεπίπεδο δίκτυο perceptron

Το πλέον διαδεδομένο νευρωνικό δίκτυο συνεχών τιμών εισόδου και εξόδου είναι το πολυεπίπεδο perceptron. Μια τυπική μορφή του δικτύου φαίνεται στο σχήμα 4.3. Αποτελείται απο νευρώνες του τύπου perceptron οι οποίοι διαθέτουν έναν γραμμικό και έναν μη-γραμμικό τελεστή συνδεδεμένους σε σειρά:

$$y = \sum_{j=1}^{N} w_i x_i \tag{4.31}$$

$$o = f(y) \tag{4.32}$$

Το δίκτυο αποτελείται από περισσότερες των δύο ομάδων νευρώνων perceptron που ονομάζονται και επίπεδα του δικτύου. Ενα ή περισσότερα κρυφά επίπεδα επεξεργάζονται τα δεδομένα εισόδου και το επίπεδο εξόδου περιέχει τους νευρώνες, η έξοδος των οποίων είναι ταυτόχρονα και έξοδος του δικτύου.

Η συνήθης σύνδεση των νευρώνων perceptron ενός επιπέδου περιλαμβάνει συνάψεις που συνδέουν κάθε νευρώνα με όλους τους νευρώνες του προηγούμενου επιπέδου. Η σύνδεση αυτή ονομάζεται πλήρης (full connected) και χρησιμοποιείται σε περιπτώσεις στις οποίες δεν μπορούμε να ορίσουμε εκ των προτέρων την ακριβή τοπολογία του δικτύου. Η γενική αυτή τοπολογία καλύπτει και την πλειονότητα των εφαρμογών.

Η ευρεία διάδοση του πολυεπίπεδου δικτύου perceptron οφείλεται σε δύο κυρίως λόγους:

1. Η ισχυρή πολυπλοκότητα και μη γραμμικότητα του δικτύου μας παρέχει την δυνατότητα να προσομοιώσουμε με ικανοποιητική ακρίβεια συνεχείς μη γραμμικές διανυσματικές συναρτήσεις.

2. Ο αλγόριθμος εκπαίδευσης LMS δίνει την δυνατότητα υπολογισμού των συντελεστών βαρύτητας των συνάψεων από παραδείγματα. Η δυνατότητα αυτή μας επιτρέπει να χρησιμοποιήσουμε το δίκτυο σε πρακτικές εφαρμογές.

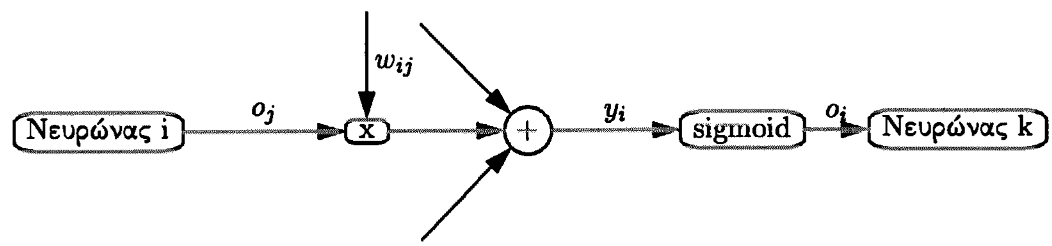
Πράγματι από το 1985, όταν ο Werdos στην διδακτορική του διατριβή παρουσίασε την αρχική μορφή ενός αλγόριθμου εκπαίδευσης που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα της πραγματικής από την επιθυμούμενη έξοδο του δικτύου και που αργότερα πήρε το όνομα οπισθοδρομική διάδοση του σφάλματος (backpropagation of error), μεγάλος αριθμός εργασιών που αφορούσαν το πολυεπίπεδο δικτύο perceptron παρουσιάστηκαν σε εφαρμογές ταξινόμησης προτύπων, αυτόματο έλεγχο, προσωμοίωση συναρτήσεων, κ.ο.κ.

4.6.1 Ο αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος

Οπως είδαμε και στο προηγούμενο εδάφιο κατά την σχεδίαση ενός γραμμικού φίλτρου Wiener η μέθοδος της κατευθυνόμενης εκπαίδευσης LMS έχει σαν σκοπό την εύρεση των συντελεστών βαρύτητας των συνάψεων που ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα της πραγματικής από την επιθυμούμενη έξοδο του δικτύου.

Οι ίδιες αρχές ακολουθούνται και στον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος με την διαφορά ότι στο πολυεπίπεδο δίκτυο perceptron δεν μπορούμε να πετύχουμε αναλυτική λύση. Γιαυτό τον λόγο καταφεύγουμε στην μέθοδο της διαδοχικής προσέγγισης των συντελεστών βαρύτητας των συνάψεων κατά την διεύθυνση της αρνητικής τιμής της πρώτης παραγώγου του σφάλματος (steepest descent - βαθύτατη κάθοδος). Η μέθοδος αυτή που είναι η πλέον διαδεδομένη τεχνική προσδιορισμού της μνήμης πολύπλοκων νευρωνικών ονομάζεται αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος.

Ο αλγόριθμος οφείλει την ονομασία του στο γεγονός ότι ενώ κατά την ενεργοποίηση του νευρωνικού δικτύου οι υπολογισμοί ξεκινούν από τους νευρώνες εισόδου και κατευθύνονται διαδοχικά προς τους νευρώνες του εξόδου, υπολογίζοντας την τιμή της εξόδου των νευρώνων στα κρυφά επίπεδα, η εκπαίδευση εκτελεί την αντίστροφη διαδικασία. Μετά την ενεργοποίηση του δικτύου αρχίζει η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων αρχικά στις συνάψεις που βρίσκονται στους νευρώνες εξόδου και στην συνέχεια ακολουθεί ο επαναπροσδιορισμός των συντελεστών βαρύτητας προς τους νευρώνες εισόδου.



Σχήμα 4.7: Διάταξη νευρώνων στον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος

Στην γενική περίπτωση ένα πολυεπίπεδο δίκτυο perceptron αποτελείται από έναν αριθμό κρυφών επιπέδων και το επίπεδο εξόδου που περιέχει M νευρώνες. Στην κατευθυνόμενη εκπαίδευση τα παραδείγματα αποτελούνται από ζεύγη διανυσμάτων εισόδου-εξόδου. Αν η είσοδος του δικτύου αποτελείται από N εισόδους τότε τα παραδείγματα είναι διατεταγμένες δυάδες διανυσμάτων $(\mathbf{a}, \mathbf{b})$, $\mathbf{a} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^M$.

Το σφάλμα εκτίμησης για τυχαία είσοδο **a** είναι το τετράγωνο των διαφορών της πραγματικής εξόδου του δικτύου **o** με το αντίστοιχο διάνυσμα των αναμενόμενων τιμών **b**:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = \frac{1}{2}(\mathbf{b}-\mathbf{o})^T(\mathbf{b}-\mathbf{o}) = \frac{1}{2}\sum_{i=1}^{M}(b_i - o_i)^2 \tag{4.33}$$

Δεδομένου ότι συνήθως έχουμε στην διάθεσή μας Q παραδείγματα, το αναμενόμενο σφάλμα υπολογίζεται από την σχέση:

$$E[\Sigma\varphi\acute{\alpha}\lambda\mu\alpha] = \frac{1}{2}E[\sum_{i=1}^{M}(b_i - o_i)^2] = \frac{1}{2}\sum_{i=1}^{M}E[(b_i - o_i)^2] = \frac{1}{2Q}\sum_{i=1}^{M}\sum_{j=1}^{Q}(b_{ij} - o_{ij})^2 \tag{4.34}$$

Είναι γνωστό ότι οι προσπάθειες των ερευνητών για την αναζήτηση μεθόδου αναλυτικού υπολογισμού των συντελεστών βαρύτητας των συνάψεων, μιάς λύσης παρόμοιας με αυτή των φίλτρων Wiener, έχουν καταλήξει σε αποτυχία. Γιαυτό τον λόγο καταφεύγουμε στην μέθοδο της διαδοχικής προσέγγισης των συντελεστών βαρύτητας των συνάψεων κατά την διεύθυνση της αρνητικής τιμής της πρώτης παραγώγου του σφάλματος. Η μέθοδος αυτή επαναπροσδιορίζει την αριθμητική τιμή του συντελεστή βαρύτητας των συνάψεων σύμφωνα με την ακόλουθη σχέση:

$$\Delta w_{ij} = -n\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial w_{ij}} \tag{4.35}$$

όπου n είναι μικρός θετικός πραγματικός αριθμός που ονομάζεται ρυθμός εκπαίδευσης του αλγόριθμου.

Η εξίσωση μπορεί να εκφραστεί σαν συνάρτηση της εξόδου και της εσωτερικής κατάστασης του νευρώνα:

$$\Delta w_{ij} = -n\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial y_i}\frac{\partial y_i}{\partial w_{ij}} \tag{4.36}$$

Αν ορίσουμε σαν δ_i την αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα y_i (εσωτερική κατάσταση του νευρώνα είναι η έξοδος του γραμμικού του τελεστή), τότε:

$$\delta_i = -\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial y_i} = -\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial o_i}\frac{\partial o_i}{\partial y_i} \tag{4.37}$$

και

$$\Delta w_{ij} = n\delta_i\frac{\partial y_i}{\partial w_{ij}} = n\delta_i\frac{\partial}{\partial w_{ij}}\sum_k w_{ik}o_k = n\delta_i o_j \tag{4.38}$$

Για να πετύχουμε μία αναλυτική έκφραση της συνάρτησης επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων χωρίζουμε τους νευρώνες σε δύο κατηγορίες.

1. **Νευρώνες εξόδου.**

   Η αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα (δ_i) γίνεται:

$$\delta_i = -\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial o_i}\frac{\partial o_i}{\partial y_i} = -(-(b_i - o_i))\frac{\partial}{\partial y_i}f(y_i) \tag{4.39}$$

Ο τελευταίος πολλαπλασιαστικός όρος δεν είναι τίποτα άλλο από την πρώτη παράγωγο του μη γραμμικού τελεστή του νευρώνα. Η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους νευρώνες εξόδου γίνεται:

$$\Delta w_{ij} = n\delta_i o_j = n(b_i - o_i)o_j\frac{\partial}{\partial y_i}f(y_i) \tag{4.40}$$

2. **Κρυφοί νευρώνες.** Η περίπτωση των κρυφών νευρώνων είναι ελαφρώς πολυπλοκότερη. Αν υποθέσουμε ότι στο αμέσως υψηλότερο επίπεδο βρίσκονται K νευρώνες, τότε η αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα (δ_i) γίνεται:

$$\delta_i = -\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial o_i}\frac{\partial o_i}{\partial y_i} = -\frac{\partial o_i}{\partial y_i}\sum_{k=1}^{K}\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial y_k}\frac{\partial y_k}{\partial o_i} = \frac{\partial o_i}{\partial y_i}\sum_{k=1}^{K}(-\frac{\partial \Sigma\varphi\acute{\alpha}\lambda\mu\alpha}{\partial y_k})\frac{\partial}{\partial o_i}\sum_{m=1}^{I} w_{km}o_m \tag{4.41}$$

όπου I είναι ο αριθμός των νευρώνων του επιπέδου που μελετάμε. Μετά την απλοποίηση των παραπάνω εκφράσεων έχουμε:

$$\delta_i = \frac{\partial o_i}{\partial y_i}\sum_{k=1}^{K}\delta_k w_{ki} \tag{4.42}$$

Βλέπουμε λοιπόν ότι ο όρος δ του νευρώνα i εξαρτάται από τους αντίστοιχους όρους δ των νευρώνων που βρίσκονται στο αμέσως υψηλότερο επίπεδο και τους συντελεστές βαρύτητας των συνάψεων που συνδέουν την έξοδο του νευρώνα. Προκύπτει μάλιστα σαν το εσωτερικό γινόμενο των αντίστοιχων διανυσμάτων τους.

Η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους κρυφούς νευρώνες γίνεται:

$$\Delta w_{ij} = n\delta_i o_j = n o_j\frac{\partial o_i}{\partial y_i}\sum_{k=1}^{K}\delta_k w_{ki} \tag{4.43}$$

Επειδή ο υπολογισμός των συντελεστών δ είναι αναδρομικός, από τους νευρώνες υψηλότερων επιπέδων προς χαμηλότερα επίπεδα γιαυτό και η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων στον πολυεπίπεδο δίκτυο perceptron ακολουθεί τα εξής βήματα:

1. *Τοποθέτηση αρχικών τιμών.* Ορίζουμε τον ρυθμό εκπαίδευσης n και τοποθετούμε τυχαίες αρχικές τιμές στους συντελεστές βαρύτητας των συνάψεων.

2. *Ενεργοποίηση του δικτύου.* Τοποθετούμε στην είσοδο του δικτύου τυχαίο πρότυπο από τα παραδείγματα και στην συνέχεια υπολογίζουμε από χαμηλότερο προς υψηλότερο επίπεδο την έξοδο των νευρώνων κάθε επιπέδου.

3. *Υπολογισμός των συντελεστών δ.* Αντιστέφοντας την φορά των υπολογισμών υπολογίζουμε τους συντελεστές δ από υψηλότερα σε χαμηλότερα επίπεδα.

4. *Επαναπροσδιορισμός των συντελεστών βαρύτητας των συνάψεων.* Εχοντας υπολογίσει τους συντελεστές δ και την αριθμητική τιμή των εξόδων των κόμβων μπορούμε να επαναπροσδιορίσουμε τους συντελεστές βαρύτητας των συνάψεων όλων των νευρώνων του δικτύου.

5. *Ελεγχος σύγκλισης.* Υπάρχουν διάφορα κριτήρια ελέγχου της σύγκλισης του αλγόριθμου. Οι συνήθεις έλεγχοι που γίνονται είναι οι εξής:

   (α) Υπολογίζεται η συνολική απόλυτη μεταβολή των συντελεστών βαρύτητας σε σχέση με έναν πολύ μικρό θετικό αριθμό, το κριτήριο σύγκλισης.

   (β) Υπολογίζεται ο μέσος ρυθμός μεταβολής του σφάλματος στα παραδείγματα. Αν η μεταβολή αυτή είναι πολύ μικρή τότε ο αλγόριθμος τερματίζει.

   (γ) Αν ο μέσος ρυθμός μεταβολής των συντελεστών δ για όλα τα παραδείγματα είναι πολύ μικρός, αποδεικνύεται ότι η αντίστοιχη μεταβολή των συντελεστών βαρύτητας των συνάψεων είναι επίσης πολύ μικρή.

6. *Επανάληψη του αλγόριθμου.* Τοποθετούμε νέο παράδειγμα στην είσοδο του δικτύου και επαναλαμβάνουμε την διαδικασία από το δεύτερο βήμα.

Πρόγραμμα 9 *Υπολογίζουμε τους συντελεστές βαρύτητας νευρωνικού δικτύου με νευρώνες τύπου perceptron και ένα κρυφό επίπεδο. Κάθε νευρώνας διαθέτει σύναψη η οποία συνδέεται με σταθερή είσοδο 1.*

```
function [Recs,Sums] = MlpEbp1LBias(x,c,HidNodes,Lr,Tr,Step)
#
#  [Recs,Sums] = MlpEbp1LBias(x,c,HidNodes,Lr,Tr,Step)
#      Back-propagation of error. One hidden layer with bias inputs
#  Input
#      x: Pattern Vectors
#      c: Classes
#      HidNodes: Number of Hidden nodes
#      Lr: Learning rate
#      Tr: Training epoches
#      Step: Classification accuracy and Estimation Error
#            after "Step" training cycles
#  Output
#      Recs: Correct classification rate using the C-method
#      Sums: Error Function

NumOfClass = max(c) ;
NumOfPatterns = columns(x) ;
x = [ x ; 0.5 * ones( 1, NumOfPatterns ) ] ;
Dimens = rows(x) ;
Rec = zeros( 1, Tr/Step+1) ;
Sums = zeros( 1, Tr/Step+1) ;
HidWeights = rand(Dimens,HidNodes+1) - 0.5 ;
OutWeights = rand( HidNodes+1, NumOfClass ) - 0.5 ;
HidOut = 0.5 * ones( 1, HidNodes+1 ) ;
HidDelta = zeros( 1, HidNodes+1) ;
Out = zeros( 1, NumOfClass ) ;
OutDelta = zeros( 1, NumOfClass ) ;
mr = 0 ;
for m = 0:Tr
Nxv = floor(rand() * NumOfPatterns ) + 1 ;
Xinp = x(:,Nxv)' ;
for j = 1:HidNodes
```

```
HidOut(1,j) = 1 / ( exp(-(Xinp * HidWeights(:,j))) + 1 ) ;
endfor
for j = 1:NumOfClass
Out(1,j) = 1 / ( exp(-(HidOut * OutWeights(:,j)) ) + 1 ) ;
endfor
Desired = zeros( 1, NumOfClass ) + 0.1 ;
Desired(1,c(Nxv)) = 0.9 ;
OutDelta = ( Desired - Out ) .* Out .* ( 1 - Out ) ;
HidDelta = HidOut .* ( 1 - HidOut ) .* ( OutWeights * OutDelta' )' ;
OutWeights = OutWeights + Lr * HidOut' * OutDelta ;
HidWeights = HidWeights + Lr * Xinp' * HidDelta ;
if( rem( m, Step ) == 0 )
mr++ ;
Rc = zeros(NumOfClass,1) ;
Sum = 0 ;
for i = 1:NumOfPatterns
k = c(i) ;
Xinp = x(:,i)' ;
for j = 1:HidNodes
HidOut(1,j) = 1 / ( exp(-(Xinp * HidWeights(:,j)) ) + 1 ) ;
endfor
for j = 1:NumOfClass
Out(1,j) = 1 / ( exp(-(HidOut * OutWeights(:,j)) ) + 1 ) ;
endfor
for j = 1:NumOfClass
if ( k == j )
Sum = Sum + ( 0.9 - Out(1,j) )^2 ;
else
Sum = Sum + ( Out(1,j) - 0.1 )^2 ;
endif
endfor
Rec = ArgMax(Out) ;
if (Rec == k )
Rc(Rec) = Rc(Rec) + 1 ;
endif
endfor
Sums(1,mr) = Sum ;
TotRec = sum(Rc') ;
Recs(1,mr) = 100.0 * TotRec / NumOfPatterns ;
printf( "%5d %15.4f %8.6f\n", m, Sum, Recs(1,mr)) ;
fflush( stdout ) ;
if ( TotRec == NumOfPatterns )
return ;
endif
endif
endfor
endfunction
```

Παράδειγμα 40 *Στην περίπτωση που χρησιμοποιούμε σαν μη γραμμικό τελεστή των νευρώνων την σιγμοειδή συνάρτηση υπολογίστε την αναλυτική έκφραση της συνάρτησης επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων σε ένα πολυεπίπεδο δίκτυο perceptron.*

Η πρώτη παράγωγος του μη γραμμικού τελεστή του νευρώνα για την σιγμοειδή συνάρτηση είναι:

$$\frac{\partial}{\partial y} f(y) = f(y)(1 - f(y)) = o(1 - o)$$

Αντικαθιστώντας την σχέση στην συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους νευρώνες της εξόδου έχουμε:

$$\delta_i = (b_i - o_i)o_i(1 - o_i)$$

$$\Delta w_{ij} = n o_j \delta_i = n o_j (b_i - o_i)o_i(1 - o_i)$$

Αντίστοιχα, η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους κρυφούς νευρώνες γίνεται:

$$\delta_i = o_i(1 - o_i) \sum_{k=1}^{K} \delta_k w_{ki}$$

$$\Delta w_{ij} = n o_j \delta_i = n o_j o_i(1 - o_i) \sum_{k=1}^{K} \delta_k w_{ki}$$

Παράδειγμα 41 *Αν ο μη γραμμικός τελεστής των νευρώνων είναι η συνάρτηση της υπερβολικής εφαπτομένης υπολογίστε την αναλυτική έκφραση της συνάρτησης επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων ενός πολυεπίπεδου δικτύου perceptron.*

Η πρώτη παράγωγος της υπερβολικής εφαπτομένης είναι:

$$\frac{\partial}{\partial y} f(y) = (f(y) + 1)(1 - f(y)) = (o + 1)(1 - o)$$

Αντικαθιστώντας την παραπάνω σχέση στην συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους νευρώνες της εξόδου έχουμε:

$$\delta_i = (b_i - o_i)(o_i + 1)(1 - o_i)$$

$$\Delta w_{ij} = n o_j \delta_i = n o_j (b_i - o_i)(o_i + 1)(1 - o_i)$$

Η συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων για τους κρυφούς νευρώνες γίνεται:

$$\delta_i = (o_i + 1)(1 - o_i) \sum_{k=1}^{K} \delta_k w_{ki}$$

$$\Delta w_{ij} = n o_j \delta_i = n o_j (o_i + 1)(1 - o_i) \sum_{k=1}^{K} \delta_k w_{ki}$$

4.6.2 Βελτιώσεις του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος

Ο αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος εξασφαλίζει την σύγκλιση σε ένα τοπικά ελάχιστο σημείο με την έννοια ότι το σημείο αυτό, που ορίζει ένα σύνολο τιμών για τους συντελεστές σύναψης του δικτύου, εξασφαλίζει την ελάχιστη τιμή του σφάλματος της επιθυμούμενης από την πραγματική έξοδο του δικτύου για **οποιαδήποτε μικρή μεταβολή** κάθε συντελεστή βαρύτητας του δικτύου. Το γεγονός αυτό δεν μας εξασφαλίζει βέβαια και το συνολικά ελάχιστο το οποίο ορίζεται σαν το σύνολο τιμών των συντελεστών βαρύτητας των συνάψεων για το οποίο **οποιαδήποτε** μεταβολή των συντελεστών βαρύτητας δίνει μεγαλύτερο σφάλμα.

Μόνο στην περίπτωση κατά την οποία μπορούμε να εξασφαλίσουμε την ύπαρξη ενός μόνο τοπικά ελάχιστου σημείου (το οποίο θα είναι τότε και συνολικά ελάχιστο) θα είχαμε σύγκλιση του αλγόριθμου στην επιθυμούμενη λύση. Η συνθήκη αυτή δεν μπορεί να εξασφαλιστεί στην περίπτωση του πολυεπίπεδου δικτύου τύπου perceptron.

Είναι επίσης σημαντικό να επισημάνουμε ότι λανθασμένη επιλογή του ρυθμού εκπαίδευσης έχει σαν αποτέλεσμα δύο αντιθετικά συμβάντα. Αν ο συντελεστής εκπαίδευσης έχει μεγάλη αριθμητική τιμή τότε υπάρχει περίπτωση να μην επιτύχουμε σύγκλιση του αλγόριθμου σε κάποιο τοπικά ελάχιστο σημείο του σφάλματος αλλά να παρατηρήσουμε κατά την διάρκεια της εκπαίδευσης ταλαντώσεις γύρω από ένα ή περισσότερα τοπικά ελάχιστα. Οι ταλαντώσεις αυτές είναι μεγαλύτερες όσο μεγαλύτερη είναι και η αριθμητική τιμή του ρυθμού εκπαίδευσης. Αντίθετα, αν επιλεγεί μικρός συντελεστής εκπαίδευσης τότε η σύγκλιση πιθανόν να απαιτεί πολύ μεγάλο αριθμό επαναλήψεων. Σε αυτή την περίπτωση ο αλγόριθμος σε μερικές περιοχές μεταβάλλει με πολύ χαμηλούς ρυθμούς τους συντελεστές βαρύτητας των συνάψεων, με άμεση συνέπεια την αύξηση των βημάτων του αλγόριθμου μέχρι την τελική σύγκλιση. Ταυτόχρονα αυξάνεται και η πιθανότητα σύγκλισης του αλγόριθμου σε ένα τοπικά ελάχιστο σημείο το οποίο δεν δίνει μικρό σφάλμα.

Σε αυτό το εδάφιο δίνουμε μερικές χρήσιμες οδηγίες για το πως θα βελτιώσουμε την απόδοση του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος.

1. Τα παραδείγματα πρέπει να επιλέγονται με τυχαίο τρόπο. Μια καλή στρατηγική είναι να αναδιατάξουμε με τυχαίο τρόπο τα παραδείγματα και κατόπιν να εφαρμόσουμε μία κυκλική διαδοχή χρήσης των παραδειγμάτων στον αλγόριθμο.

2. Οι υπολογισμοί του αλγόριθμου αναφέρονται στο σφάλμα που προκύπτει από την τοποθέτηση ενός μόνο παραδείγματος στην είσοδο του δικτύου. Η συνάρτηση σφάλματος που ελαχιστοποιείται αναφέρεται στο στιγμιαίο σφάλμα που ορίστηκε σαν την διαφορά της επιθυμούμενης από την πραγματική έξοδο για ένα μόνο παράδειγμα. Αν ελαχιστοποιήσουμε την συνάρτηση σφάλματος που ορίζεται σαν την μέση τιμή του στιγμιαίου σφάλματος στα διαθέσιμα παραδείγματα εκπαίδευσης, τότε προκύπτουν οι ίδιες αναδρομικές εξισώσεις και όλοι οι υπολογισμοί πραγματοποιούνται από τις αντίστοιχες αναμενόμενες τιμές, εφόσον τοποθετηθούν όλα τα παραδείγματα στο δίκτυο.

   Αυτή η παραλλαγή του αλγόριθμου αποφεύγει σε σημαντικό βαθμό τα τοπικά ελάχιστα, αλλά απαιτεί χρονοβόρους υπολογισμούς διότι για κάθε ένα βήμα επαναπροσδιορισμού των συντελεστών βαρύτητας απαιτείται η τοποθέτηση όλων των διαθέσιμων παραδειγμάτων στο δίκτυο. Σε περιπτώσεις κατά τις οποίες το νευρωνικό δίκτυο περιέχει πολλά επίπεδα, έχουμε μεγάλο αριθμό νευρώνων και διαθέτουμε πολλά παραδείγματα, η παραλλαγή του αλγόριθμου που προτείνουμε είναι στην πράξη ανεφάρμοστη.

   Συνοπτικά, η στατιστική εκδοχή του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος έχει ως εξής:

   (α) Τοποθετώ αρχικές τιμές για τον ρυθμό εκπαίδευσης και τους συντελεστές βαρύτητας των νευρώνων.

   (β) Τοποθετούμε ένα παράδειγμα στην είσοδο του δικτύου και υπολογίζουμε την πραγματική του έξοδο για την τρέχουσα εκτίμηση των συντελεστών βαρύτητας των νευρώνων

   (γ) Κατά την αντίστροφη φορά υπολογίζουμε τους συντελεστές δ.

   (δ) Επιλέγουμε το επόμενο παράδειγμα και επαναλαμβάνουμε το αλγόριθμο από το δεύτερο βήμα.

   (ε) Οταν τοποθετήσουμε όλα τα παραδείγματα υπολογίζουμε την μέση τιμή των συντελεστών δ κάθε νευρώνα. Με την βοήθεια της μέσης τιμής δ επαναπροσδιορίζουμε τους συντελεστές βαρύτητας των νευρώνων.

   (στ) Ελέγχουμε την συνθήκη σύγκλισης. Αν δεν επιτευχθεί σύγκλιση ο αλγόριθμος επαναλαμβάνεται από την αρχή.

3. Πειράματα έχουν δείξει ότι ο αριθμός των νευρώνων που απαρτίζουν κάθε επίπεδο του νευρωνικού δικτύου πρέπει να μειώνεται καθώς προχωράμε από την είσοδο προς την έξοδό του. Η αρχιτεκτονική αυτή τείνει να μειώσει τον αριθμό των τοπικά ελάχιστων του δικτύου.

4. Καλή επιλογή των αρχικών τιμών των συντελεστών βαρύτητας των νευρώνων είναι μικροί πραγματικοί αριθμοί στην περιοχή του μηδενός. Οι αριθμοί αυτοί δεν πρέπει να προσεγγίζουν την τιμή του μηδενός.

5. Η αριθμητική τιμή του ρυθμού εκπαίδευσης (n) επιλέγεται συνήθως εμπειρικά. Συνήθως οι τιμές που εξασφαλίζουν καλή ταχύτητα σύγκλισης και αποφεύγουν τις ταλαντώσεις γύρω από τοπικά βέλτιστα είναι θετικοί πραγματικοί κοντά στην μονάδα.

6. Συνήθως ο παράγοντας δ για τους νευρώνες που περιέχονται σε χαμηλότερα επίπεδα λαμβάνει μικρότερες αριθμητικές τιμές. Συνεπώς καλό θα ήταν ο ρυθμός εκπαίδευσης να έχει μεγαλύτερη αριθμητική τιμή για τους νευρώνες των χαμηλότερων επιπέδων.

7. Η σύγκλιση του αλγόριθμου επιβραδύνεται σημαντικά με την αύξηση του αριθμού των επιπέδων του δικτύου. Γιαυτό τον λόγο πρέπει να διατηρούμε τον χαμηλότερο αριθμό επιπέδων στο δίκτυο για το πρόβλημα που καλούμαστε να λύσουμε. Αν είναι απαραίτητο μπορούμε να αυξήσουμε τον αριθμό των νευρώνων στα ήδη υπάρχοντα επίπεδα. Αν το σφάλμα δεν μειώνεται κατά την επιθημητή τιμή τότε πρέπει να δοκιμάσουμε να προσθέσουμε ένα επίπεδο νευρώνων.

8. Οι επιθυμούμενες τιμές των παραδειγμάτων δεν θα πρέπει να συμπίπτουν με τα απειροστά όρια της σιγμοειδούς συνάρτησης, αλλά πρέπει να βρίσκονται μέσα στο πεδίο τιμών της.

9. Καλό θα ήταν να επιλέξουμε έναν μη γραμμικό τελεστή που να ικανοποιεί τη συνθήκη:

$$f(-x) = -f(x) \tag{4.44}$$

   Εχει παρατηρηθεί ότι στις περισσότερες των περιπτώσεων ο αλγόριθμος "μαθαίνει" γρηγορότερα όταν ο μη γραμμικός τελεστής ικανοποιεί την παραπάνω σχέση. Τυπικό παραδειγμα μη γραμμικού νευρωνικού τελεστή που ικανοποιεί αυτή την συνθήκη είναι η συνάρτηση της υπερβολικής εφαπτομένης.

10. Αυξημένες δυνατότητες αποφυγής τοπικά ελάχιστων σημείων παρουσιάζονται αν σε κάθε νευρώνα προστεθεί μία σύναψη η οποία έχει σταθερή είσοδο (συνήθως επιλέγονται οι τιμές 1 ή -1). Η επέκταση του διανύσματος εισόδου του νευρώνα έχει σαν συνέπεια την αύξηση των βαθμών ελευθερίας του δικτύου και πιστεύεται ότι αυτή είναι η κύρια αιτία ελάττωσης της πιθανότητας σύγκλισης σε τοπικά ελάχιστο σημείο της συνάρτησης του σφάλματος.

11. Μια έξυπνη τεχνική αύξησης της ταχύτητας εκπαίδευσης του δικτύου συνίσταται στην ανεξάρτητη (ανά νευρώνα) αύξηση του συντελεστή εκπαίδευσης όταν η πρώτη παράγωγος του σφάλματος ως προς τους συντελεστές βαρύτητας των συνάψεων του νευρώνα παραμένουν ομόσημες για έναν αριθμό διαδοχικών επαναλήψεων του αλγόριθμου με διαφορετικά παραδείγματα. Αντίθετα όταν έχουμε φαινόμενα αλλαγής προσήμου (πιθανή ύπαρξη φαινομένων ταλάντωσης γύρω από κάποιο τοπικά ελάχιστο) ο συντελεστής εκπαίδευσης πρέπει να ελαττώνεται.

12. Σε πολλές εφαρμογές έχει παρατηρηθεί βελτίωση των χαρακτηριστικών σύγκλισης του αλγόριθμου όταν στην συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων προστεθεί και ένας σταθμισμένος όρος που περιέχει την μεταβολή του συντελεστή βαρύτητας που προήλθε από την εφαρμογή του αλγόριθμου στο προηγούμενο παράδειγμα:

$$\Delta w_{ij}(t) = n\delta_i o_j + m\Delta w_{ij}(t-1) \tag{4.45}$$

Αυτή η παραλλαγή του αλγόριθμου της οπισθοδρομικής διάδοσης του σφάλματος ονομάζεται momentum-term.

Ο αριθμός των εργασιών που μελετούν τα θεωρητικά προβλήματα και την συμπεριφορά του αλγόριθμου της οπισθοδρομικής διάδοσης του σφάλματος είναι πολύ μαγάλος στην διεθνή βιβλιογραφία, γεγονός που επιβεβαιώνει την σημαντικότητα των εφαρμογών του πολυεπίπεδου δικτύου perceptron.

Παράδειγμα 42 *Κατασκευάστε σύστημα ταξινόμησης προτύπων δύο κατηγοριών με ένα πολυεπίπεδο δίκτυο perceptron το οποίο έχει τρείς νευρώνες στο πρώτο κρυφό επίπεδο και μονάχα έναν νευρώνα εξόδου. Δύο παραδείγματα από κάθε κατηγορία είναι διαθέσιμα, τα $\Omega_1 = \{(0,0),(1,1)\}$ και $\Omega_2 = \{(0,1),(1,0)\}$. Θεωρείστε ότι όλοι οι νευρώνες έχουν σαν μη γραμμικό νευρωνικό τελεστή την σιγμοειδή συνάρτηση.*

Επειδή θέλουμε σε μία έξοδο να αναγνωρίζουμε δύο κατηγορίες αντικειμένων πρέπει να κατασκευάσουμε έναν κανόνα με τον οποίο θα πραγματοποιούμε την ταξινόμηση των προτύπων ανάλογα με την τιμή που θα λαμβάνουμε στην έξοδο του δικτύου.

Εστω ότι θέλουμε τα πρότυπα της μιας κατηγορίας να λαμβάνουν την τιμή 0 στην έξοδο του δικτύου ενώ τα πρότυπα της δεύτερης κατηγορίας να λαμβάνουν την τιμή 1.

Συνεπώς τα παραδείγματα της κατευθυνόμενης εκπαίδευσης παίρνουν την μορφή μιας πύλης XOR:

$$\{((0,0),0),((1,1),0),((0,1),1),((1,0),1)\}$$

Επειδή πρέπει να αποφύγουμε να θέσουμε σαν επιθυμούμενες τιμές εξόδου τα όρια της σιγμοειδούς συνάρτησης, οι επιθυμούμενες τιμές των παραδειγμάτων τροποποιούνται ως εξής:

$$\{((0,0),0.1),((1,1),0.1),((0,1),0.9),((1,0),0.9)\}$$

Η ταξινόμηση επιτυγχάνεται με την τοποθέτηση στην είσοδο του άγνωστου πρότυπου, την ενεργοποίηση του δικτύου και την διαδικασία ταξινόμησης η οποία εξετάζει την αριθμητική τιμή της εξόδου του δικτύου:

$$|o-0.1| < |o-0.9| \Rightarrow x \in \Omega_1 \quad \Leftrightarrow \quad o < \frac{1}{2} \Rightarrow x \in \Omega_1$$

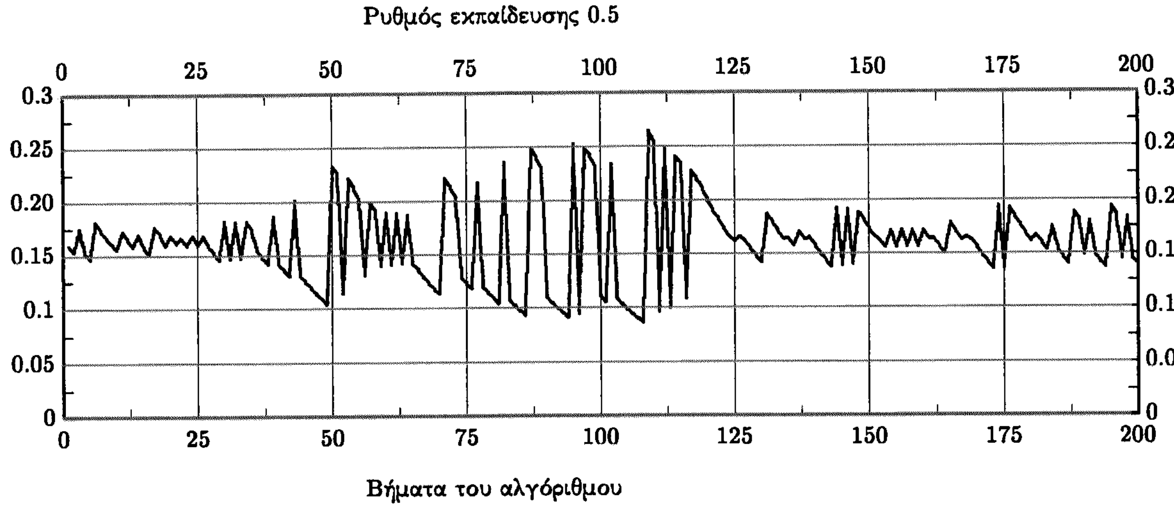
Λαμβάνοντας υπόψιν τα σχόλια που κάναμε για την συμπεριφορά του αλγόριθμου της οπισθοδρομικής διάδοσης του σφάλματος, επιλέγουμε τυχαίους μικρούς πραγματικούς αριθμούς και τους θέτουμε σαν αρχικές τιμές στους συντελεστές βαρύτητας των συνάψεων που αποτελούν συνολικά γιά όλο το δίκτυο $3 \times 2 + 3 \times 1 = 9$ μεταβλητές. Επιλέγουμε σαν συντελεστή εκπαίδευσης τον αριθμό 0.5 και διαλέγοντας τυχαία τα παραδείγματα εφαρμόζουμε τον αλγόριθμο εκπαίδευσης επαναπροσδιορίζοντας τους συντελεστές βαρύτητας κάθε φορά που ενεργοποιούμε το δίκτυο.

Παρατηρούμε ότι ο αλγόριθμος δεν δείχνει να συγκλίνει παρόλο που τα βήματα επανάληψης του αλγόριθμου αυξάνουν σημαντικά. Το σφάλμα δεν μειώνεται στα επιθημητά επίπεδα και το δίκτυο δεν καταφέρνει να ταξινομήσει σωστά όλα τα παραδείγματα εκπαίδευσης. Η μέγιστη αξιοπιστία του συστήματος ταξινόμησης που φτιάξαμε δεν ξεπερνά το 75%.

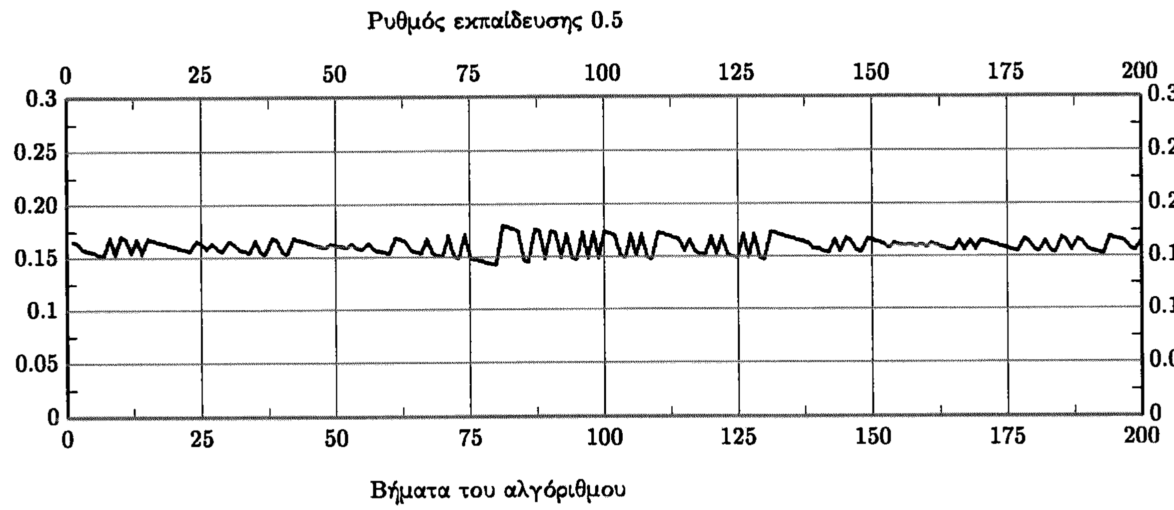
Παρατηρούμε επίσης ότι ο αλγόριθμος δεν συγκλίνει ακόμα και αν αυξήσουμε τα βήματα του αλγόριθμου. Το επόμενο βήμα μας πρέπει να είναι η ελάττωση του ρυθμού εκπαίδευσης για να δούμε αν η τυχόν μεγάλη αριθμητική τιμή του ρυθμού εκπαίδευσης είναι και η αιτία της μη σύγκλισης του αλγόριθμου.

Στο σχήμα 4.9 βλέπουμε το μέσο σφάλμα του αλγόριθμου εκπαίδευσης με ρυθμό εκπαίδευσης 0.1. Τα αποτελέσματα που παίρνουμε είναι παρόμοια με τα προηγούμενα διότι το μέσο σφάλμα παραμένει στα ίδια επίπεδα και η επιτυχία της ταξινόμησης δεν υπερβαίνει πάλι το 75%.

Εφόσον με την αλλαγή του ρυθμού εκπαίδευσης δεν βελτιώθηκε η ποιότητα εκπαίδευσης του συστήματος ταξινόμησης καταλαβαίνουμε ότι πρέπει να αυξήσουμε τους βαθμούς ελευθερίας του δικτύου προσθέτοντας συνάψεις. Υπάρχουν τρείς διαφορετικές τεχνικές με τις οποίες μπορούμε να επιτύχουμε αύξηση του αριθμού των συνάψεων. Η πρώτη τεχνική αυξάνει τον αριθμό των συνάψεων κατά ποσότητα ίση με τον αριθμό των νευρώνων του δικτύου προσθέτοντας σε κάθε νευρώνα μία σύναψη που συνδέεται με σταθερή είσοδο που έχει αριθμητική τιμή 1. Με την δεύτερη τεχνική



Σχήμα 4.8: Μέσο σφάλμα του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος για πρόβλημα κατασκευής συστήματος ταξινόμησης δύο κατηγοριών και ρυθμό εκπαίδευσης 0.5.



Σχήμα 4.9: Μέσο σφάλμα του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος με συντελεστή εκπαίδευσης 0.1

αυξάνουμε τον αριθμό των νευρώνων στα διάφορα επίπεδα. Η τελευταία τεχνική έχει ήδη χρησιμοποιηθεί στο κεφάλαιο στο οποίο περιγράφονται τα δομικά συστήματα ταξινόμησης και αναφέρεται στην επέκταση του διανύσματος εισόδου με συναρτήσεις που περιέχουν μη γραμμικούς συνδυασμούς των συνιστωσών του διανύσματος εισόδου.

Σε αυτή την άσκηση θα αυξήσουμε κατά έναν τους νευρώνες εξόδου τροποποιώντας αναγκαστικά και την τεχνική ταξινόμησης. Η ταξινόμηση πραγματοποιείται ελέγχοντας ποιός από τους δύο νευρώνες βρίσκεται στην υψηλότερη στάθμη ενεργοποίησης:

$$o_1 > o_2 \Rightarrow x \in \Omega_1$$

Το κριτήριο ταξινόμησης που θέσαμε είναι ίδιο με αυτό που χρησιμοποιούν οι συναρτήσεις διάκρισης δομικών συστημάτων. Συνεπώς μπορούμε να ισχυριστούμε ότι η δομή του δικτύου ισοδυναμεί με ένα δομικό σύστημα ταξινόμησης προτύπων που περιέχει μη γραμμικές συναρτήσεις διάκρισης.

Τα παραδείγματα εκπαίδευσης μετατρέπονται έτσι ώστε η επιθυμούμενη έξοδος να είναι 0.9 για τον νευρώνα της κατηγορίας στην οποία ανήκει το πρότυπο και 0.1 για τον νευρώνα της άλλης κατηγορίας:

$$\{((0,0),(0.9,0.1)),((1,1),(0.9,0.1)),((0,1),(0.1,0.9)),((1,0),(0.1,0.9))\}$$

Στο νέο δίκτυο ο αριθμός των συνάψεων αυξάνει και γίνεται $3 \times 2 + 3 \times 2 = 12$. Δοκιμάζοντας την συμπεριφορά του αλγόριθμου εκπαίδευσης προσθέτοντας ταυτόχρονα και τον όρο momentum με ρυθμούς εκπαίδευσης $n = 2$ και $m = 0.5$ βλέπουμε ότι η σύγκλιση του αλγόριθμου επιτυγχάνεται σε πολύ μικρό αριθμό επαναλήψεων. Σε 70 περίπου επαναλήψεις του αλγόριθμου επιτυγχάνουμε αξιοπιστία ταξινόμησης 100%.

Οι συντελεστές βαρύτητας των συνάψεων για τους τρείς νευρώνες του κρυφού επιπέδου μετά από 200 επαναλήψεις του αλγόριθμου έχουν λάβει τις ακόλουθες τιμές:

$$\{(-3.079, 0.484604), (-3.25359, 0.570598), (2.88307, -0.465273)\}$$

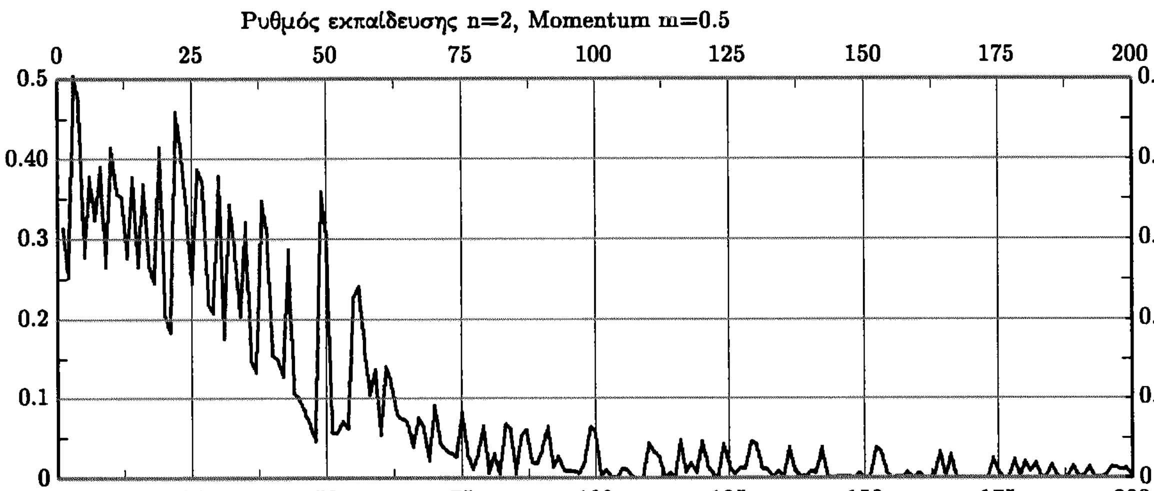
Οι δε συντελεστές βαρύτητας των συνάψεων για τους δύο νευρώνες της εξόδου έχουν τις τιμές:

$$\{(3.0407, 3.58521, -2.36929), (-3.24589, -3.41236, 2.38413)\}$$

Αν τοποθετήσουμε τα διανύσματα εισόδου των παραδειγμάτων στο δίκτυο και το ενεργοποιήσουμε θα λάβουμε στους νευρώνες εξόδου τις τιμές που δίνονται στον πίνακα που ακολουθεί.

Πίνακας 4.8: Είσοδος, έξοδος, επιθυμούμενη έξοδος και κριτήριο ταξινόμησης

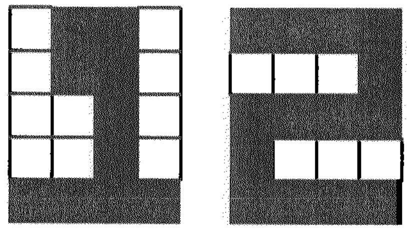
Πρότυπο	Εξοδος δικτύου	Επιθυμούμενη Εξοδος	Ταξινόμηση
$(0,0) \in \omega_1$	0.825004, 0.17411	0.9, 0.1	ω_1
$(1,1) \in \omega_1$	0.922083, 0.077366	0.9, 0.1	ω_1
$(0,1) \in \omega_2$	0.139699, 0.861543	0.1, 0.9	ω_2
$(1,0) \in \omega_2$	0.172902, 0.828438	0.1, 0.9	ω_2



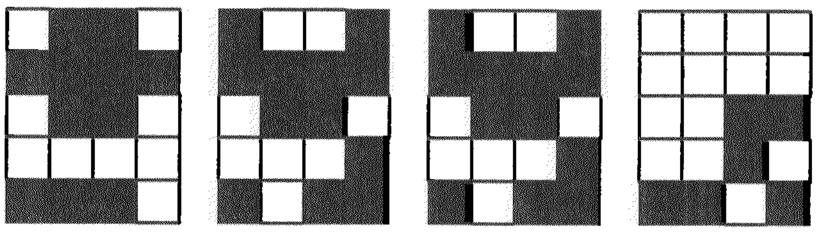
Σχήμα 4.10: Μέσο σφάλμα του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος με τον προσθετικό όρο momentum.

Παράδειγμα 43 *Νευρωνικό σύστημα οπτικής ταξινόμησης χαρακτήρων ταξινομεί τα δύο πρώτα ψηφία (1 και 2). Κάθε εικόνα αποτελείται από διδιάστατα λευκά ή μαύρα σημεία διαστάσεων* 5×4. *Εχουμε στην διάθεσή μας ένα παράδειγμα από κάθε ψηφίο, όπως αυτά που απεικονίζονται στο σχήμα που ακολουθεί:*

Ποιά είναι η αριθμητική τιμή των συντελεστών βαρύτητας των συνάψεων οι οποίοι μπορούν να ταξινομούν σωστά τα εκπαιδευόμενα ψηφία, δεδομένου ότι το δίκτυο αποτελείται από ένα κρυφό επίπεδο το οποίο έχει τέσσερις νευρώνες και το επίπεδο εξόδου το οποίο έχει δύο νευρώνες;



Σχήμα 4.11: Οπτική απεικόνιση παραδειγμάτων των ψηφίων 1 και 2



Σχήμα 4.12: Οπτική απεικόνιση άγνωστων ψηφίων

Στην συνέχεια ταξινομήστε τα σχήματα που ακολουθούν.

Αρχικά σχεδιάζουμε την δομή του πολυεπίπεδου δικτύου perceptron. Λαμβάνοντας υπόψιν τις οδηγίες που δόθηκαν κατασκευάζουμε ένα δίκτυο που έχει στο κρυφό επίπεδο τέσσερις νευρώνες και στο επίπεδο εξόδου δύο νευρώνες. Το διάνυσμα εισόδου αποτελείται από 20 διαδικές τιμές, όσες και τα σημεία κάθε πλαισίου. Με βάσει αυτά τα στοιχεία, ο αριθμός των συνάψεων του δικτύου είναι $20 \times 4 + 4 \times 2 = 88$.

Τα παραδείγματα εκπαίδευσης είναι δύο, ένα για κάθε μία κατηγορία. Οι συνιστώσες του διανύσματος εισόδου έχουν την τιμή 0 όταν το σημείο είναι λευκό και 1 όταν το σημείο που περιγράφεται είναι μαύρο. Τα σημεία της εικόνας σαρώνονται από επάνω προς τα κάτω και από αριστερά προς τα δεξιά, ως εξής:

$$\Omega_1 = \{((0,1,1,0,0,1,1,0,0,0,1,0,0,0,1,0,1,1,1,1),(0.9,0.1))\}$$

$$\Omega_2 = \{((1,1,1,1,0,0,0,1,1,1,1,1,1,0,0,0,1,1,1,1),(0.1,0.9))\}$$

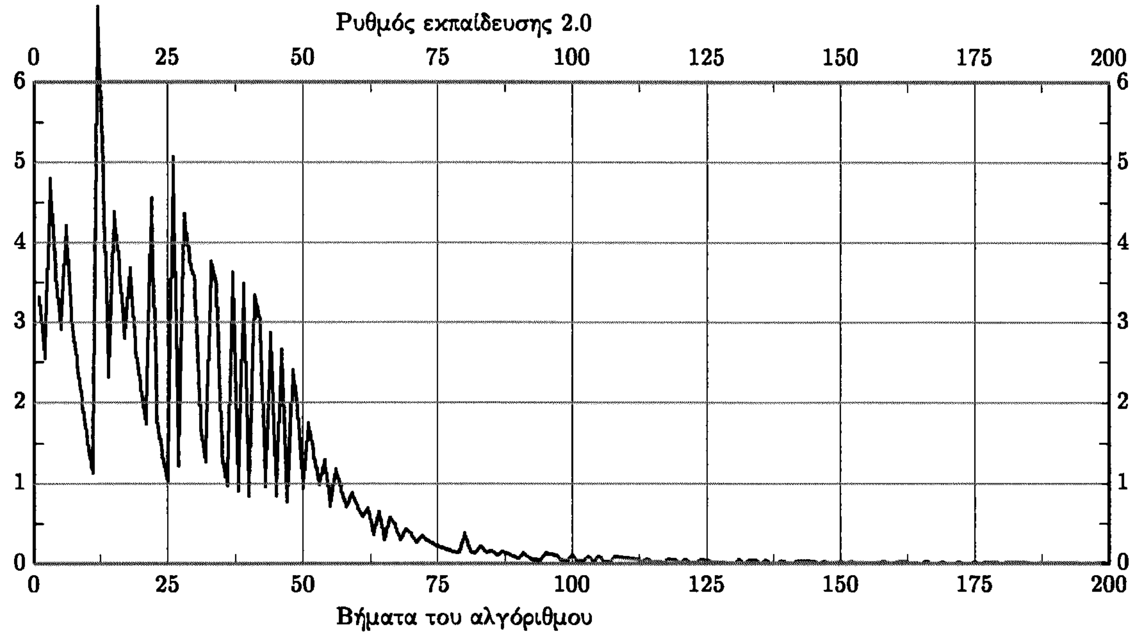
Τοποθετώντας τυχαίες αρχικές τιμές στους συντελεστές βαρύτητας εφαρμόζουμε τον αλγόριθμο οπισθοδρομικής διάδοσης του σφάλματος με ρυθμό εκπαίδευσης 2.0, οπότε επιτυγχάνουμε μία ταχύτατη σύγκλιση όπως φαίνεται στο σχήμα 4.13.

Μετά την διαδικασία εκπαίδευσης ταξινομούμε τα άγνωστα πρότυπα με την μέθοδο αναζήτησης του νευρώνα που έχει την πλέον ενεργοποιημένη έξοδο.

Τα τέσσερα πρότυπα κατά αύξουσα σειρά απεικόνισης θα ταξινομηθούν στις κατηγορίες του πίνακα 4.9:

4.7 Δίκτυα ακτινικών συναρτήσεων

Υπάρχει μία ιδιαίτερη ομάδα νευρωνικών δικτύων τα οποία ονομάζονται δίκτυα ακτινικών συναρτήσεων (Radial Basis Function networks - RBF networks). Τα δίκτυα αυτά αποτελούνται από ένα κρυφό επίπεδο και ένα επίπεδο νευρώνων εξόδου. Η ονομασία τους οφείλεται στο γεγονός της αντικατάστασης του γραμμικού τελεστή που υπάρχει στους νευρώνες του κρυφού επιπέδου. Ο μη γραμμικός τελεστής είναι μία συνάρτηση που περιέχει έκφραση της γενικευμένης Ευκλείδιας απόστασης του



Σχήμα 4.13: Μέσο σφάλμα του αλγόριθμου οπισθοδρομικής διάδοσης του σφάλματος για τον υπολογισμό των συντελεστών βαρύτητας των συνάψεων συστήματος οπτικής ταξινόμησης των ψηφίων 1 και 2.

Πίνακας 4.9: Είσοδος, έξοδος και ταξινόμηση των παραμορφωμένων οπτικά ψηφίων

Πρότυπο	Εξοδος της ω_1	Εξοδος της ω_2	Ταξινόμηση
1	0.890299	0.109883	ω_1
2	0.117867	0.882141	ω_2
3	0.147734	0.853502	ω_2
4	0.557962	0.445515	ω_1

διανύσματος εισόδου με το αντίστοιχο διάνυσμα των συντελεστών βαρύτητας των συνάψεων του νευρώνα. Οι νευρώνες εξόδου δεν διαθέτουν μη γραμμικό τελεστή και γιαυτό το λόγο εκτελούν μονάχα ένα γραμμικό μετασχηματισμό των εξόδων των νευρώνων του κρυφού επιπέδου.

Η συνάρτηση μεταφοράς αυτού του δικτύου ισοδυναμεί με τις γενικευμένες αντινικές συναρτήσεις απόφασης που έχουν ήδη περιγραφεί στο κεφάλαιο 2. Γιαυτό το λόγο δίνεται σε συντομία η διαδικασία εκπαίδευσης από παραδείγματα που πραγματοποιείται με αναλυτικές εξισώσεις στο επίπεδο εξόδου.

Η συνάρτηση που δίνει την σχέση μιας εξόδου του δικτύου με την είσοδο και τους συντελεστές βαρύτητας των συνάψεων είναι η ακόλουθη:

$$o = \sum_{i=1}^{N} w_i G(\mathbf{x}, \mathbf{q}_i) \tag{4.46}$$

οπου $\mathbf{x}$ είναι η είσοδος του δικτύου, $\mathbf{q}_i$ είναι ένα από N σταθερά σημεία στο χώρο των προτύπων, και $w_i, i = 1, N$ είναι οι συντελεστές βαρύτητας ενός νευρώνα εξόδου.

Το πρόβλημα της κατευθυνόμενης εκπαίδευσης των δικτύων ακτινικών συναρτήσεων μπορεί να εκφραστεί ως εξής: Δοσμένου ενός αριθμού παραδειγμάτων εκπαίδευσης βρείτε τους συντελεστές βαρύτητας των συνάψεων που ελαχιστοποιούν το μέσο σφάλμα της επιθυμούμενης από την πραγματική έξοδο του δικτύου.

Η εύρεση των συντελεστών βαρύτητας κάθε νευρώνα αποτελεί μια ανεξάρτητη λύση. Αν ορίσουμε σαν b την επιθυμούμενη έξοδο κάθε νευρώνα, το μέσο σφάλμα της προσομοίωσης για Μ παραδείγματα είναι:

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = (\mathbf{b} - \mathbf{G}\mathbf{w})^T(\mathbf{b} - \mathbf{G}\mathbf{w}) = |\mathbf{b} - \mathbf{G}\mathbf{w}|^2 = \sum_{j=1}^{M} (b_j - \sum_{i=1}^{N} w_i G(\mathbf{x}, \mathbf{q}_i))^2 \qquad (4.47)$$

όπου $\mathbf{b} = (b_1, b_2, ..., b_M)^T$ το διάνυσμα των επιθυμούμενων εξόδων, και $\mathbf{w} = (w_1, w_2, ..., w_N)^T$ οι συντελεστές βαρύτητας του νευρώνα. Το $\mathbf{G}$ περιγράφει ένα σύνολο σταθερών αριθμών:

$$\mathbf{G} = \begin{pmatrix} G(\mathbf{x}_1, \mathbf{q}_1) & G(\mathbf{x}_1, \mathbf{q}_2) & ... & G(\mathbf{x}_1, \mathbf{q}_N) \\ G(\mathbf{x}_2, \mathbf{q}_1) & G(\mathbf{x}_2, \mathbf{q}_2) & ... & G(\mathbf{x}_2, \mathbf{q}_N) \\ ... & ... & ... & ... \\ G(\mathbf{x}_M, \mathbf{q}_1) & G(\mathbf{x}_M, \mathbf{q}_2) & ... & G(\mathbf{x}_M, \mathbf{q}_N) \end{pmatrix} \qquad (4.48)$$

Η LMS λύση δίνεται από την εξίσωση:

$$\frac{\partial}{\partial \mathbf{w}} \Sigma\varphi\acute{\alpha}\lambda\mu\alpha = 0 \quad \Rightarrow \quad \mathbf{G}^T\mathbf{G}\mathbf{w} = \mathbf{G}^T\mathbf{b} \quad \Rightarrow \quad \mathbf{w} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{b} \quad \Rightarrow \quad \mathbf{w} = \mathbf{G}^{+}\mathbf{b} \qquad (4.49)$$

Ο πίνακας $\mathbf{G}^+$ ονομάζεται ψευδοαντίστροφος πίνακας της $\mathbf{G}$ διότι η διαδικασία υπολογισμού της έχει ομοιάζοντα χαρακτηριστικά με τον τρόπο υπολογισμού του αντίστροφου ενος τετραγωνικού πίνακα.

Από την λύση που περιγράψαμε βλέπουμε ότι στην περίπτωση των δικτύων ακτινικών συναρτήσεων μπορούμε να βρούμε με αναλυτικό τρόπο τους συντελεστές βαρύτητας των συνάψεων. Το γεγονός αυτό αποτελεί σαφές πλεονέκτημα απέναντι στις μεθόδους εκπαίδευσης του πολυεπίπεδου perceptron. Επίσης έχει αποδειχθεί ότι τα δίκτυα ακτινικών συναρτήσεων μπορούν να προσωμοιώσουν με την επιθυμητή ακρίβεια οποιαδήποτε συνεχή διανυσματική συνάρτηση.

Σε προβλήματα κατασκευής συστημάτων ταξινόμησης προτύπων τα δίκτυα ακτινικών συναρτήσεων χρησιμοποιούνται πλέον συχνότερα από το πολυεπίπεδο perceptron διότι ο αλγόριθμος εκπαίδευσης έχει πολύ μικρή υπολογιστική πολυπλοκότητα.

Πρόγραμμα 10 *Υπολογισμός των συντελεστών βαρύτητας νευρωνικού δικτύου RBF.*

```
function [Rc,Rep] = RbfRnd(x,c,HidNodes)
#
#  [Rc,Rep] = RbfRnd(x,c)
#          Random Weights in the first Layer
#  Input
#      x: Pattern Vectors for the first class
#      c: Classes
#      HidNodes: Number of Hidden nodes
#  Output
#      Rc: Correct classification rate using the C-method
#      Rep: Pattern vectors on each class
#
```

```
NumOfClass = max(c) ;
NumOfPatterns = columns(x) ;
Dimens = rows(x) ;
MaxVal = max(x') ;
MinVal = min(x') ;
VarVal = (MaxVal - MinVal)' ;
MinVal = MinVal' ;
HidWeights = rand(Dimens,HidNodes) ;
for i = 1:HidNodes
HidWeights(:,i) = HidWeights(:,i) .* VarVal .+ MinVal ;
endfor
Rep = zeros(NumOfClass,1) ;
G = zeros(NumOfPatterns,HidNodes) ;

#
#  C-Error
#
for i = 1:HidNodes
for j = 1:NumOfPatterns
G(j,i) = 1 / ( sum(abs( HidWeights(:,i) - x(:,j))) + 1 ) ;
endfor
endfor
A = inv(G' * G) * G' ;
OutWeights = zeros( HidNodes, 0 ) ;
for i = 1:NumOfClass
for j = 1:NumOfPatterns
if ( c(j) == i )
b(j) = 1 ;
else
b(j) = 0 ;
endif
endfor
OutWeights = [ OutWeights, A * b ] ;
endfor

Rc = zeros(NumOfClass,1) ;
for j = 1:NumOfPatterns
k = c(j) ;
Rep(k,1) = Rep(k,1) + 1 ;
for i = 1:HidNodes
Ho(i) = 1 / ( sum(abs( HidWeights(:,i) - x(:,j))) + 1 ) ;
endfor
Dist = Ho' * OutWeights ;
Rec = ArgMax(Dist) ;
if (Rec == k )
Rc(Rec) = Rc(Rec) + 1 ;
endif
endfor
endfunction
```

Παράδειγμα 44 *Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης διδιάστατων προτύπων τριών κατηγοριών έχοντας στην διάθεσή μας τα ακόλουθα παραδείγματα:*

$$\Omega_1 = \{(0.5, 0), (1.5, 0), (0, 0.5), (1, 1)\}$$

$$\Omega_2 = \{(0, 0), (0, -1), (-1, 1), (-1, 0)\}$$

$$\Omega_3 = \{(1,0),(1,-0.5),(1,-1),(1.5,-0.5)\}$$

Υπολογίστε τους συντελεστές βαρύτητας των συνάψεων ενός δικτύου ακτινικών συναρτήσεων και εκτιμήστε το ελάχιστο σφάλμα του συστήματος ταξινόμησης που κατασκευάσατε.

Οπως έχουμε δεί και στην θεωρία του κεφαλαίου 2 η έξοδος κάθε νευρώνα αποτελεί μία προσομοίωση της συνάρτησης διάκρισης για κάθε μία από τις τρείς κατηγορίες προτύπων.

Αρχικά θα πρέπει να επιλέξουμε εναν αριθμό σημείων στο χώρο των προτύπων και να ορίσουμε την μορφή της συνάρτησης $G(.)$. Σαφείς κανόνες για το πως θα πρέπει να ορίσουμε αυτά τα μεγέθη δεν υπάρχουν. Γενικά θεωρείται καλή επιλογή τα σημεία αυτά να είναι κοντά στις τιμές των προτύπων των παραδειγμάτων και ο αριθμός αυτών των σημείων πρέπει να αυξάνει όταν υπάρχει ισχυρή επικάλυψη των προτύπων των κατηγοριών στο χώρο των μετρήσεων.

Ορίζουμε αυθαίρετα την συνάρτηση $G(.)$:

$$G(\mathbf{x},\mathbf{q}) = \frac{1}{|\mathbf{x}-\mathbf{q}|+1}$$

και τα ακόλουθα πέντε σημεία στον χώρο των μετρήσεων:

$$Q = \{(0,0),(-1,0),(1,-1),(0.5,-0.5),(1,0.5)\}$$

Οι συντελεστές βαρύτητας των συνάψεων κάθε νευρώνα προκύπτουν σαν την μοναδική λύση ενός γραμμικού συστήματος, από τα πρότυπα των παραδειγμάτων εκπαίδευσης και τις αντίστοιχες επιθυμούμενες τιμές εξόδου που είναι 1 για τα πρότυπα που ανήκουν στην κατηγορία και 0 για τα πρότυπα που ανήκουν σε άλλες κατηγορίες.

$$\mathbf{G}^T = \begin{pmatrix} 0.8 & 0.307 & 0.8 & 0.333 & 1.0 & 0.5 & 0.333 & 0.5 & 0.5 & 0.444 & 0.333 & 0.285 \\ 0.307 & 0.137 & 0.444 & 0.166 & 0.5 & 0.333 & 0.5 & 1.0 & 0.2 & 0.190 & 0.166 & 0.133 \\ 0.444 & 0.444 & 0.235 & 0.2 & 0.333 & 0.5 & 0.111 & 0.166 & 0.5 & 0.8 & 1.0 & 0.666 \\ 0.8 & 0.444 & 0.444 & 0.285 & 0.666 & 0.666 & 0.181 & 0.285 & 0.666 & 0.8 & 0.666 & 0.5 \\ 0.666 & 0.666 & 0.5 & 0.8 & 0.444 & 0.235 & 0.19 & 0.19 & 0.8 & 0.5 & 0.307 & 0.444 \end{pmatrix}$$

Ο ψευδοαντίστροφος πίνακας $\mathbf{G}^+$ είναι:

$$\mathbf{G}^+ = \begin{pmatrix} -0.27 & -0.24 & 1.43 & -0.01 & 1.29 & -0.77 & -0.06 & -0.75 & -0.74 & -0.67 & 0.71 & 0.27 \\ -0.15 & 0.07 & -0.36 & 0.05 & -0.45 & 0.20 & 0.43 & 1.13 & 0.1 & 0.1 & -0.16 & -0.05 \\ -0.93 & 0.11 & 0.76 & 0.13 & 0.24 & -0.86 & 0.12 & -0.16 & -0.67 & -0.31 & 1.54 & 0.83 \\ 1.38 & -0.21 & -1.67 & -0.74 & -0.68 & 1.8 & -0.29 & 0.29 & 1.02 & 1.14 & -1.51 & -0.91 \\ -0.07 & 0.5 & 0.05 & 0.84 & -0.34 & -0.44 & 0.08 & 0.03 & 0.4 & -0.14 & -0.23 & 0.13 \end{pmatrix}$$

Για τον πρώτο νευρώνα τα παραδείγματα εκπαίδευσης είναι τα ακόλουθα:

$$\{((0.5,0),1),((1.5,0),1),((0,0.5),1),((1,1),1),$$

$$((0,0),0),((0,-1),0),((-1,1),0),((-1,0),0),((1,0),0),((1,-0.5),0),((1,-1),0),((1.5,-0.5),0)\}$$

οπότε οι επιθυμούμενες έξοδοι είναι:

$$\mathbf{b}_1 = (1,1,1,1,0,0,0,0,0,0,0,0)^T$$

Οι συντελεστές βαρύτητας των συνάψεων υπολογίζονται από την σχέση:

$$\mathbf{w}_1 = \mathbf{G}^+\mathbf{b}_1 = (0.894205,-0.399421,0.0775641,-1.24777,1.33348)^T$$

Για τον δεύτερο νευρώνα τα παραδείγματα εκπαίδευσης είναι τα ακόλουθα:

$$\{((0.5,0),0),((1.5,0),0),((0,0.5),0),((1,1),0),((0,0),1),((0,-1),1),((-1,1),1),((-1,0),1),$$

$$((1,0),0),((1,-0.5),0),((1,-1),0),((1.5,-0.5),0)\}$$

οπότε και οι επιθυμούμενες έξοδοι είναι οι εξής:

$$\mathbf{b}_2 = (0,0,0,0,1,1,1,1,0,0,0,0)^T$$

Ο ψευδοαντίστροφος πίνακας παραμένει ο ίδιος. Οι συντελεστές βαρύτητας των συνάψεων για τον δεύτερο νευρώνα υπολογίζονται απο την σχέση:

$$\mathbf{w}_2 = \mathbf{G}^{+}\mathbf{b}_2 = (-0.296934, 1.31657, -0.650708, 1.1692, -0.670198)^T$$

Για τον τρίτο νευρώνα οι επιθυμούμενες έξοδοι είναι:

$$\mathbf{b}_3 = (0,0,0,0,0,0,0,0,1,1,1,1)^T$$

Οι συντελεστές βαρύτητας των συνάψεων για τον τρίτο νευρώνα υπολογίζονται απο την σχέση:

$$\mathbf{w}_3 = \mathbf{G}^{+}\mathbf{b}_3 = (-0.429196, -0.0230214, 1.40181, -0.256063, 0.16657)^T$$

Το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων που κατασκευάσαμε υπολογίζεται, με την τοποθέτηση στην είσοδο του δικτύου των προτύπων των παραδειγμάτων, την ενεργοποίηση του δικτύου και την ταξινόμηση κάθε ενός από αυτά στην κατηγορία εκείνη που έχει τον πλέον ενεργοποιημένο νευρώνα εξόδου. Στον πίνακα που ακολουθεί δίνονται οι έξοδοι του δικτύου RBF και η διαδικασία ταξινόμησης των προτύπων.

Παρατηρούμε ότι μόνο ένα πρότυπο που ανήκει στην τρίτη κατηγορία ταξινομήθηκε λανθασμένα, σαν πρότυπο που ανήκει στην πρώτη κατηγορία. Συνεπώς το ελάχιστο σφάλμα του συστήματος ταξινόμησης είναι $1/12 \times 100 = 8.33\%$.

Πίνακας 4.10: Εξοδος δικτύου RBF και ταξινόμηση των προτύπων εκπαίδευσης

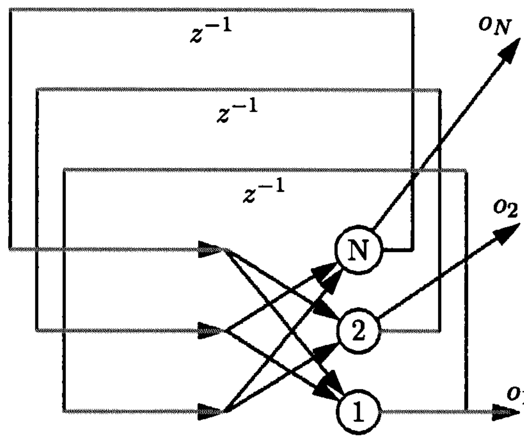
Πρότυπο	1	2	3	4	5	6
Εξοδος 1	0.5177	0.5889	0.6682	0.9572	0.4811	-0.165
Εξοδος 2	0.3669	-0.1261	0.379	-0.2112	0.626	0.5868
Εξοδος 3	0.1787	0.485	-0.0542	0.1935	-0.0701	0.3471
Ταξιν/ση	ω_1	ω_1	ω_1	ω_1	ω_2	ω_2
Κατηγορία	ω_1	ω_1	ω_1	ω_1	ω_2	ω_2
Πρότυπο	7	8	9	10	11	12
Εξοδος 1	0.1341	-0.0418	0.6409	0.0519	-0.1124	0.2227
Εξοδος 2	0.5719	1.266	0.0328	0.1985	0.0429	-0.0563
Εξοδος 3	-0.0136	-0.0454	0.4442	0.8047	1.1354	0.7548
Ταξιν/ση	ω_2	ω_2	ω_1	ω_3	ω_3	ω_3
Κατηγορία	ω_2	ω_2	ω_3	ω_3	ω_3	ω_3

4.8 Το δίκτυο Hopfield

Η δομή του δικτύου Hopfield φαίνεται στο σχήμα 4.14. Αποτελείται από ένα μόνο επίπεδο νευρώνων το οποίο είναι ταυτόχρονα είσοδος και έξοδος για το δίκτυο. Η έξοδος κάθε νευρώνα συνδέεται με συνάψεις με όλους τους άλλους νευρώνες και δεν διαθέτει σύναψη που να συνδέει την έξοδο με την είσοδο του ίδιου νευρώνα.

Οι N νευρώνες διαθέτουν δύο τελεστές που συνδέονται σε σειρά και περιγράφονται απο τις ακόλουθες αλγεβρικές εκφράσεις:

$$y_i = \sum_{j=1}^{N} w_{ij} o_j \tag{4.50}$$



Σχήμα 4.14: Τυπική μορφή δικτύου Hopfield

$$o = \begin{cases} +1, & y \geq 0 \\ -1, & y < 0 \end{cases} \tag{4.51}$$

Το δίκτυο Hopfield αποτελέι την πιο απλή μορφή επανατροφοδοτούμενου νευρωνικού δικτύου. Η βασική ιδέα του δικτύου συνίσταται στην κατασκευή νευρωνικού δικτύου στο οποίο τα πρωτότυπα των κατηγοριών θα αποθηκεύονται σε κοιλάδες δυναμικού. Οταν το δίκτυο έχει συμμετρικούς συντελεστές βαρύτητας ($w_{ij} = w_{ji}$) τότε μπορεί να αποδειχθεί ότι υπάρχουν σημεία στο χώρο των καταστάσεων τα οποία αντιστοιχούν σε ελάχιστα ενεργειών, τα οποία καλούνται και ελκυστές του δικτύου. Αποδεικνύεται επίσης ότι όταν τοποθετήσουμε μία τυχαία αρχική κατάσταση στο δίκτυο και υπολογίσουμε διαδοχικά τις εξόδους του δικτύου, η τελική τιμή σύγκλισης είναι ένας από τους ελκυστές του δικτύου.

Ενέργεια του δικτύου Hopfield ορίζεται να είναι το ακόλουθο αριθμητικό μέγεθος:

$$E = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} o_i o_j \tag{4.52}$$

Η διαδικασία ενεργοποίησης του δικτύου αποτελείται από τις ακόλουθες ενέργειες:

1. Την τοποθέτηση του διανύσματος του άγνωστου πρότυπου στις εξόδους του δικτύου. Τα διανύσματα εισόδου έχουν διαδικές τιμές, τις -1 και +1.

2. Την μεταφορά της εξόδου μέσω των συνάψεων επανατροφοδότησης στις εισόδους του δικτύου.

3. Τον υπολογισμό των νέων εξόδων του δικτύου.

4. Τον έλεγχο της σύγκλισης. Αν η κατάσταση των εξόδων δεν αλλάξει, τότε η σύγκλιση του δικτύου έχει πραγματοποιηθεί, διαφορετικά επαναλαμβάνονται οι υπολογισμοί από το δεύτερο βήμα.

Η διαδικασία επαναπροσδιορισμού των εξόδων του δικτύου μπορεί να γίνει και με ασύγχρονο τρόπο. Με αυτή την μέθοδο επιλέγουμε με τυχαίο τρόπο έναν νευρώνα και ακολουθούμε την διαδικασία υπολογισμού της νέας εξόδου και της επανατροφοδότησης της εισόδου του δικτύου.

Στην περίπτωση κατά την οποία οι συντελεστες βαρύτητας δεν είναι συμμετρικοί τότε δεν μπορεί να εξασφαλιστεί η σύγκλιση του δικτύου. Υπάρχουν βέβαια περιπτώσεις στις οποίες παρόλο που οι συντελεστες βαρύτητας δεν είναι συμμετρικοί το δίκτυο συγκλίνει σε ελκυστή του.

Η εκπαίδευση του δικτύου επιτυγχάνεται με την στατιστική έκδοση του κανόνα του Hebb. Χρησιμοποιώντας τις αναμενόμενες τιμές των M πρωτότυπων που αποθηκεύονται στο δίκτυο υπολογίζουμε τους συντελεστές βαρύτητας των νευρώνων:

$$w_{ij} = \begin{cases} \frac{1}{N} \sum\limits_{m=1}^{N} x_i{}^{(m)} x_j{}^{(m)}, & i \neq j \\ 0, & i = j \end{cases} \tag{4.53}$$

Οπου $x_i{}^{(m)}$ είναι η i συνιστώσα του διανύσματος του m πρωτότυπου που θέλουμε να αποθηκεύσουμε σε ελκυστή του δικτύου.

Το πρόβλημα που τίθεται σε αυτή την μορφή του δικτύου, στην οποία ο αριθμός των νευρώνων ισούται με τον αριθμό των διαστάσεων των προτύπων, είναι ο περιορισμένος αριθμός πρωτότυπων που μπορούν να αποθηκευτούν στο δίκτυο. Εχει υπολογιστεί ότι βελτιωμένοι αλγόριθμοι εκπαίδευσης είναι σε θέση να αποθηκεύσουν και να ανακαλέσουν με επιτυχία μέχρι και αριθμό πρωτότυπων ίσο με τον αριθμό των νευρώνων του δικτύου. Σε πρακτικές εφαρμογές συνήθως αποφεύγουμε αυτο το πάνω όριο αυξάνοντας τεχνητά τις διαστάσεις του διανύσματος των προτύπων.

Μειονέκτημα επίσης αποτελεί το γεγονός ότι δεν υπάρχει τρόπος να ορίσουμε έναν μηχανισμό παραμόρφωσης των προτύπων και να ενσωματώσουμε τον αντίστροφο μετασχηματισμό του στην διαδικασία σύγκλισης του δικτύου.

Η υπολογιστική πολυπλοκότητα του δικτύου μέχρι να επιτευχθεί η σύγκλιση είναι επίσης ένα από τα μειονεκτήματα του αλγόριθμου σε σύγκριση με την λύση του πολυεπίπεδου perceptron, των RBF, και των δομικών ή στοχαστικών συστημάτων.

Τα σημαντικά αυτά μειονεκτήματα είναι η αιτία της μειωμένης χρήσης του δικτύου Hopfield σε πρακτικές εφαρμογές.

Πρόγραμμα 11 *Υπολογισμός των συντελεστών βαρύτητας δικτύου Hopfield με την ασύγρονη μέθοδο ενεργοποίησης των νευρώνων.*

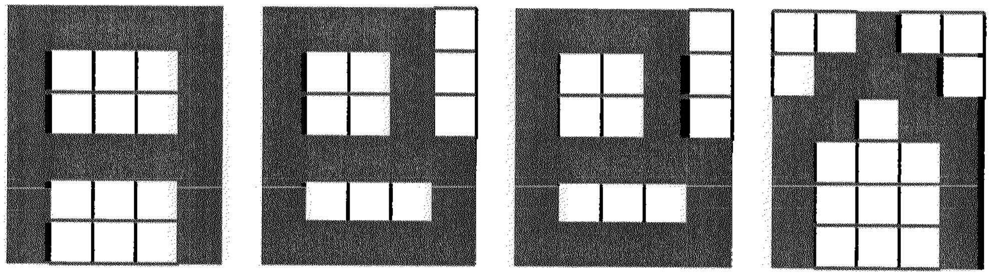
```
function [Rc,Rep] = HopfieldAsync(Pr,x,c)
#
# [Rc,Rep] = HopfieldAsync(Pr,x,c)
#
# Input
#     Pr: Prototypes Vectors
#     x: Pattern Vectors for the first class
#     c: Classes
# Output
#     Rc: Correct classification rate using the C-method
#     Rep: Pattern vectors on each class
#

NumOfClass = columns(Pr) ;
NumOfPatterns = columns(x) ;
Dimens = rows(x) ;
```

```
Rr= 3 * Dimens ;
if ( rows(Pr) != rows(x) )
return ;
endif
Weights = Pr * Pr' ;
for i = 1:Dimens
Weights(i,i) = 0 ;
endfor
Weights = Weights / NumOfPatterns ;

#
#  C-Error Asynchronous mode
#

Rc = zeros(NumOfClass,1) ;
Rep = zeros(NumOfClass,1) ;
for j = 1:NumOfPatterns
k = c(j) ;
Rep(k) = Rep(k) + 1 ;
Out = x(:,j) ;
Inp = Out ;
while( 1 )
Ag = 0 ;
for i = 1:Rr
d = floor( rand() * Dimens ) + 1 ;
Out(d) = Weights(d,:) * Inp ;
if ( Out(d) >= 0 )
Out(d) = 1 ;
else
Out(d) = -1 ;
endif
if ( Inp(d) != Out(d) )
Inp(d) = Out(d) ;
Ag = 1 ;
endif
endfor
if ( Ag == 1 )
break ;
endif
endwhile
Rec = 0 ;
Out'
for i = 1:NumOfClass
if ( Pr(:,i) == Out )
Rec = i ;
break ;
endif
endfor
if ( Rec == k )
Rc(Rec) = Rc(Rec) + 1 ;
endif
endfor
endfunction
```

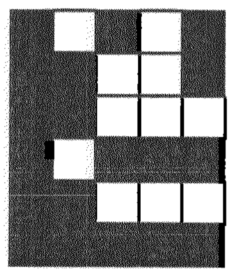
Παράδειγμα 45 *Θέλετε να κατασκευάσετε σύστημα οπτικής ταξινόμησης χαρακτήρων βασιζόμενο στο δίκτυο Hopfield που να αναγνωρίζει τα τέσσερα πρώτα γράμματα της Ελληνικής γλώσσας.*

Υπολογίστε τους συντελεστές βαρύτητας των συνάψεων και ταξινομείστε το σχήμα 4.16 :

Κατασκευάζουμε τα πρωτότυπα διανύσματα που πρόκειται να αποθηκευτούν στο δίκτυο. Στα μαύρα σημεία δίνουμε



Σχήμα 4.15: Οπτική απεικόνιση των γραμμάτων Α, Β, Γ, Δ



Σχήμα 4.16: Οπτική απεικόνιση άγνωστου γράμματος

την τιμή 1 ενώ στα λευκά δίνουμε την τιμή -1. Σαρώνοντας απο επάνω προς τα κάτω και από αριστερά προς τα δεξιά η ισοδύναμη διανυσματική παράσταση των γραμμάτων είναι η ακόλουθη:

$A = (1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,-1,1,1,1,1,1,1,1,-1,-1,-1,1,1,-1,-1,-1,1)$

$B = (1,1,1,1,-1,1,-1,-1,1,-1,1,-1,-1,1,-1,1,1,1,1,1,1,-1,-1,-1,1,1,1,1,1,1)$

$\Gamma = (1,1,1,1,1,1,-1,-1,-1,-1,1,-1,-1,-1,-1,1,-1,-1,-1,-1,1,-1,-1,-1,-1,1,-1,-1,-1,-1)$

$\Delta = (-1,-1,1,-1,-1,-1,1,1,1,-1,1,1,-1,1,1,1,-1,-1,-1,1,1,-1,-1,-1,1,1,-1,-1,-1,1)$

Απο τα διανύσματα εκπαίδευσης υπολογίζουμε τους συντελεστές βαρύτητας των συνάψεων. Στην συνέχεια τοποθετούμε το άγνωστο πρότυπο σαν αρχική κατάσταση εξόδων του δικτύου.

$X = (1,-1,1,-1,1,1,1,-1,-1,1,1,1,-1,-1,-1,1,-1,1,1,1,1,-1,-1,-1,1,1,1,1,1)$

Στο σχήμα που ακολουθεί δίνουμε τις εξόδους για κάθε ένα βήμα του σύγχρονου επαναπροσδιορισμού των εξόδων του δικτύου. Βλέπουμε ότι μετά από τέσσερα βήματα η έξοδος συγκλίνει στον εκλυστή του γράμματος B.

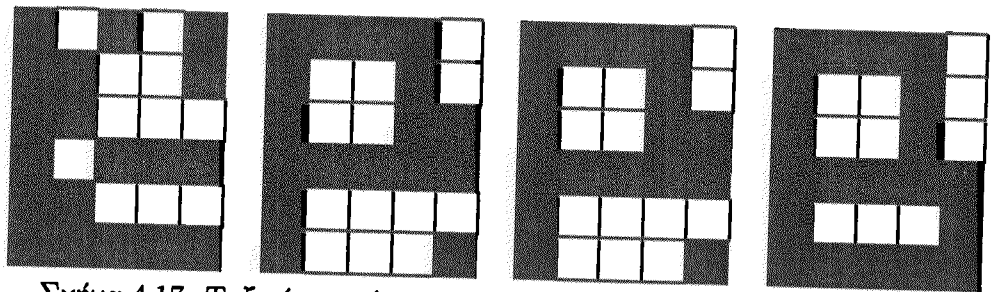
4.9 Λυμένα Προβλήματα

Πρόβλημα 4 *Το δίκτυο Adaline είναι ένα νευρωνικό δίκτυο το οποίο διαθέτει μονάχα ένα επίπεδο νευρώνων οι οποίοι περιγράφονται από τις ακόλουθες εξισώσεις:*

Ο γραμμικός τελεστής

$$y = \sum_{j=1}^{N} w_i x_i$$

συνδέεται με σειρά με έναν μη γραμμικό τελεστή:



Σχήμα 4.17: Ταξινόμηση άγνωστου γράμματος με το δίκτυο Hopfield

$$o = \begin{cases} 1, & y \geq 0 \\ -1, & y < 0 \end{cases}$$

Υπολογίστε με την μέθοδο LMS μία επαναληπτική διαδικασία υπολογισμού των συντελεστών βαρύτητας των συνάψεων από παραδείγματα.

Πρόβλημα 5 *Κατασκευάστε σύστημα ταξινόμησης προτύπων δύο κατηγοριών με δίκτυο που περιέχει νευρώνες Adline από τα ακόλουθα παραδείγματα εκπαίδευσης:*

$$\Omega_1 = \{(1,1),(2,2),(1,0),(0,1)\}$$

$$\Omega_2 = (-10,-10),(-9,-9),(-8,-9),(-9,-10)\}$$

Ποιό είναι το ελάχιστο σφάλμα του συστήματος ταξινόμησης που κατασκευάσατε;. Συγκρίνατε τα αποτελέσματά σας με αυτά που προκύπτουν από το γραμμικό φίλτρο Wiener.

Πρόβλημα 6 *Θέλουμε να κατασκευάσουμε σύστημα ταξινόμησης προτύπων τριών κατηγοριών με ένα δίκτυο Hopfield. Διαθέτουμε τρία παραδείγματα από κάθε κατηγορία:*

$$\Omega_1 = \{(1,1,1,1,1),(1,1,1,-1,1),(1,1,-1,1,-1)\}$$

$$\Omega_2 = \{(1,1,1,-1,-1),(1,-1,1,-1,-1),(-1,-1,-1,-1,-1)\}$$

$\Omega_3 = \{(-1,-1,1,-1,-1),(-1,1,1,1,-1),(-1,1,1,1,1)\}$

Υπολογείστε τους συντελεστές βαρύτητας των συνάψεων και στην συνέχεια βρείτε το ελάχιστο σφάλμα του συστήματος ταξινόμησης που κατασκευάσατε.

Γνωρίζουμε ότι στο δίκτυο δεν μπορούμε να αποθηκεύσουμε περισσότερα πρότυπα από την διάσταση του διανύσματος του προτύπου. Αρχικά ας αγνοήσουμε αυτή την διαπίστωση και ας προσπαθήσουμε να αποθηκεύσουμε όλα τα παραδείγματα στο δίκτυο.

Χρησιμοποιώντας όλα τα πρότυπα υπολογίζουμε τους συντελεστές βαρύτητας των συνάψεων:

$$\mathbf{w} = \begin{pmatrix} 0 & 3 & 1 & -1 & 1 \\ 3 & 0 & 3 & 5 & 3 \\ 1 & 3 & 0 & -1 & 1 \\ -1 & 5 & -1 & 0 & 3 \\ 1 & 3 & 1 & 3 & 0 \end{pmatrix}$$

Στην συνέχεια τοποθετούμε τα πρότυπα εκπαίδευσης στο δίκτυο, ενεργοποιούμε το δίκτυο και εξετάζουμε τον ελκυστή της σύγκλισης. Με την διαδικασία αυτή θα υπολογίσουμε το ελάχιστο σφάλμα του συστήματος ταξινόμησης.

Στον πίνακα που ακολουθεί δίνουμε το πρότυπο εισόδου, το πρότυπο σύγκλισης και την ταξινόμηση για κάθε ένα από τα πρότυπα εκπαίδευσης.

Πίνακας 4.11: Ελάχιστο σφάλμα συστήματος ταξινόμησης βασισμένο στο δίκτυο Hopfield

Πρότυπο	(1 1 1 1 1)	(1 1 1 -1 1)	(1 1 -1 1 -1)	(1 1 1 -1 -1)	(1 -1 1 -1 -1)
Σύγκλιση	(1 1 1 1 1)	(1 1 1 1 1)	(1 1 1 1 1)	(-1 1 -1 -1 1)	(-1 -1 -1 -1 -1)
Ταξινόμηση	1	1	1	-	6
Κατηγορία ω_1	ω_1	ω_1	-	ω_2	
Πρότυπο	(-1-1-1-1-1)	(-1-11-1-1)	(-1111-1)	(-11111)	
Σύγκλιση	(-1 -1 -1 -1 -1)	(-1 -1 -1 -1 -1)	(1 1 1 1 1)	(1 1 1 1 1)	
Ταξινόμηση	6	6	1	1	
Κατηγορία ω_2	ω_2	ω_1	ω_1		

Βλέπουμε ότι για δύο πρότυπα εκπαίδευσης το δίκτυο δεν συγκλίνει σε ελκυστή ο οποίος συμπεριλαμβάνεται στα πρότυπα εκπαίδευσης. Αυτό το φαινόμενα παρουσιάζεται όταν προσπαθούμε να αποθηκεύσουμε στο δίκτυο περισσότερα πρότυπα από τις δυνατότητες του. Ενα άλλο φαινόμενο που παρατηρούμε είναι η σύγκλιση του δικτύου σε ελκυστή διαφορετικό από αυτόν που τίθεται στην είσοδο του δικτύου.

Παρόλα αυτά παρατηρούμε ότι το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων που κατασκευάσαμε είναι $4/9 \times 100 = 44.44\%$.

Για να βελτιώσουμε την απόδοση του συστήματος μπορούμε είτε να μειώσουμε τον αριθμό των προτύπων που αποθηκεύουμε στο σύστημα, είτε να επεκτείνουμε το διάνυσμα κάθε πρότυπου ή να χρησιμοποιήσουμε και τις δύο τεχνικές ταυτόχρονα.

Σε αυτή την άσκηση θα επεκτείνουμε το διάνυσμα των προτύπων προσθέτοντας τρείς νέες διαστάσεις.

Η πρώτη επέκταση τίθεται στο -1 αν στο σύνολο των αρχικών παραμέτρων του πρότυπου υπερτερούν οι τιμές -1. Η δεύτερη επέκταση τίθεται στο -1 αν στο σύνολο των τριών πρώτων παραμέτρων του πρότυπου υπερτερούν οι τιμές -1. Η τρίτη επέκταση τίθεται -1 αν στο σύνολο των τριών τελευταίων παραμέτρων του πρότυπου υπερτερούν οι τιμές -1.

Προσθέτοντας τις νέες συνιστώσες τα διανύσματα εκπαίδευσης γίνονται:

$$\Omega_1 = \{(1,1,1,1,1,1,1,1),(1,1,1,-1,1,1,1,1),(1,1,-1,1,-1,1,1,-1)\}$$

$$\Omega_2 = \{(1,1,1,-1,-1,1,1,-1),(1,-1,1,-1,-1,-1,1,-1),(-1,-1,-1,-1,-1,-1,-1,-1)\}$$

$$\Omega_3 = \{(-1,-1,1,-1,-1,-1,-1,-1),(-1,1,1,1,-1,1,1,1),(-1,1,1,1,1,1,1,1)\}$$

Χρησιμοποιώντας όλα τα πρότυπα υπολογίζουμε τους συντελεστές βαρύτητας των συνάψεων:

$$\mathbf{w} = \begin{pmatrix} 0 & 3 & 1 & -1 & 1 & 3 & 5 & -1 \\ 3 & 0 & 3 & 5 & 3 & 9 & 7 & 5 \\ 1 & 3 & 0 & -1 & 1 & 3 & 5 & 3 \\ -1 & 5 & -1 & 0 & 3 & 5 & 3 & 5 \\ 1 & 3 & 1 & 3 & 0 & 3 & 1 & 7 \\ 3 & 9 & 3 & 5 & 3 & 0 & 7 & 5 \\ 5 & 7 & 5 & 3 & 1 & 7 & 0 & 3 \\ -1 & 5 & 3 & 5 & 7 & 5 & 3 & 0 \end{pmatrix}$$

Με τον ίδιο τρόπο υπολογίζουμε το ελάχιστο σφάλμα του συστήματος ταξινόμησης. Στον πίνακα που ακολουθεί δίνουμε το πρότυπο εισόδου, το πρότυπο σύγκλισης και την διαδικασία ταξινόμησης για κάθε ένα από τα πρότυπα εκπαίδευσης.

Στα νέα αριθμητικά αποτελέσματα βλέπουμε ότι για όλα τα παραδείγματα έχουμε σύγκλιση σε κάποιο από τα πρότυπα εκπαίδευσης.

Παρόλα αυτά παρατηρούμε ότι το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων που κατασκευάσαμε δεν είναι 0% αλλά έχουμε πάλι τέσσερα πρότυπα τα οποία ταξινομούνται λανθασμένα. Το ελάχιστο σφάλμα παραμένει $4/9 \times 100 = 44.44\%$.

Πίναχας 4.12: Ελάχιστο σφάλμα συστήματος ταξινόμησης με επέκταση του διανύσματος των προτύπων στο δίκτυο Hopfield

Πρότυπο	1	2	3	4	5	6	7	8	9
Σύγκλιση	1	1	1	1	6	6	6	1	1
Κατηγορία	ω_1	ω_1	ω_1	ω_1	ω_2	ω_2	ω_2	ω_1	ω_1

Η αιτία παραμένει ο μεγάλος αριθμός των προτύπων που θέλουμε να αποθηκεύσουμε. Ενώ έχουμε δίκτυο που περιέχει οκτώ νευρώνες θέλουμε να αποθηκεύσουμε εννέα πρότυπα σε ισάριθμους ελκυστές του δικτύου.

4.10 Αλυτα Προβλήματα

Ασκηση 15 *Κατασκευάστε σύστημα ταξινόμησης διδιάστατων προτύπων δύο κατηγοριών όταν είναι γνωστά τα αχόλουθα παραδείγματα εκπαίδευσης:*

$$\Omega_1 = \{(1.0, 0), (1.5, -1), (0, 0.5), (1, 1)\}$$

$\Omega_2 = \{(0, 0), (0, -1), (-1, 1), (1, 0)\}$

1. Υπολογίστε τους συντελεστές βαρύτητας των συνάψεων ενος δικτύου ακτινικών συναρτήσεων το οποίο διαθέτει δύο κρυφούς νευρώνες σημεία και υπολογίστε το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων.

2. Αυξήστε τους κρυφούς νευρώνες σε τέσσερεις και υπολογίστε ξανά το νέο ελάχιστο σφάλμα του συστήματος ταξινόμησης. Τι παρατηρείτε και πως μπορείτε να δικαιολογήσετε τα αριθμητικά αποτελέσματα;

Ασκηση 16 *Επαναλάβατε την λύση της προηγούμενης άσκησης χρησιμοποιώντας τις εξισώσεις Wiener-Hopf. Δικαιολογήστε τις διαφορές των αποτελεσμάτων στο ελάχιστο σφάλμα των συστήματων ταξινόμησης προτύπων που κατασκευάσατε.*

Πως μπορείτε να βελτιώσετε τα αποτελέσματα του γραμμικού συστήματος ταξινόμησης προτύπων;

Ασκηση 17 *Προσπαθείστε να ελαττώσετε το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων που κατασκευάστηκε με την βοήθεια του δικτύου **Hopfield** στην τρίτη λυμένη άσκηση.*

Ασκηση 18 *Κατασκευάστε σύστημα ταξινόμησης διδιάστατων προτύπων δύο κατηγοριών με την βοήθεια του δικτύου **Hopfield** όταν τα παραδείγματα εκπαίδευσης έχουν την μορφή είσοδος-έξοδος μιας πύλης XOR.*

Ασκηση 19 *Υπάρχει διαφορά στο ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων που σχεδιάσατε για την προηγούμενη άσκηση αν οι νευρώνες του δικτύου ενεργοποιούνται ασύγχρονα;*