Αναγνώριση Προτύπων Ι

Ενότητα 4: Νευρωνικά Δίκτυα στην Ταξιμόμηση Προτύπων

Αν. Καθηγητής Δερματάς Ευάγγελος

Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Πατρών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα

Εισαγωγή (1)

- Ένας 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 προσομοιωθεί με μη γραμμικούς τελεστές που δέχονται σήματα από ένα πεπερασμένο αριθμό εισόδων και διαθέτουν μια μόνο έξοδο.

- Η εγκεφαλική ουσία των έμβιων όντων αποτελείται από εκατομμύρια διασυνδεδεμένους νευρώνες. Οι νευρώνες αυτοί αποτελούν ένα πολύπλοκο και ισχυρά μη γραμμικό νευρωνικό δίκτυο. Η τοπολογία του δικτύου διαφέρει σημαντικά στα έμβια όντα και είναι αυτή που δίνει τα διαφορετικά χαρακτηριστικά αντίληψης των ερεθισμάτων του περιβάλλοντος.

Εισαγωγή (2)

- Έχουν παρατηρηθεί ποσοτικές και ποιοτικές διαφοροποιήσεις στο είδος των συνδέσεων, τον μηχανισμό επεξεργασίας σημάτων και ό πλήθος των εισόδων των νευρικών κυττάρων. Αυτός είναι και ο λόγος που τα νευρικά κύτταρα έχουν ομαδοποιηθεί σε ένα μικρό αριθμό κατηγοριών, ανάλογα με τους μηχανισμούς λειτουργίας τους.

- Η υπάρχουσα τεχνολογία δεν επιτρέπει την προσομοίωση των πολύπλοκων επεξεργασιών που εκτελούνται στα νευρωνικά κύτταρα. Έχει 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 αυτομάτου ελέγχου, κ.α..

Εισαγωγή (3)

• 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 καμία περίπτωση, δεν μπορεί να θεωρηθεί ότι αποτελεί έναν γραμμικό συνδυασμό των εισόδων του.

2. Εκπαίδευση από παραδείγματα. Βασικό στοιχείο των νευρωνικών δικτύων, είναι η ικανότητα τους να εκπαιδεύονται, αντλώντας γνώση και τροποποιώντας τα στοιχεία μνήμης του δικτύου. Προηγούμενες τροποποιήσεις του δικτύου, μεταβάλλουν την συμπεριφορά του, έτσι ώστε όταν ενεργοποιηθεί από ίδια ή παρόμοια σήματα εισόδου, να δίνουν με μεγαλύτερη ακρίβεια την επιθυμητή έξοδο.

Εισαγωγή (4)

3. Προσαρμογή. Τα φυσικά νευρωνικά δίκτυα έχουν την δυνατότητα να αλλάζουν τα δεδομένα των εξόδων τους και να προσαρμόζουν την συμπεριφορά τους, όταν μεγάλης κλίμακας αλλαγές συμβαίνουν στην είσοδο τους. Χαρακτηριστικό παράδειγμα προσαρμογής αποτελεί η μεταβολή της συμπεριφοράς του δικτύου των νευρικών κυττάρων που βρίσκονται πάνω στον αμφιβληστροειδή χιτώνα.

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.

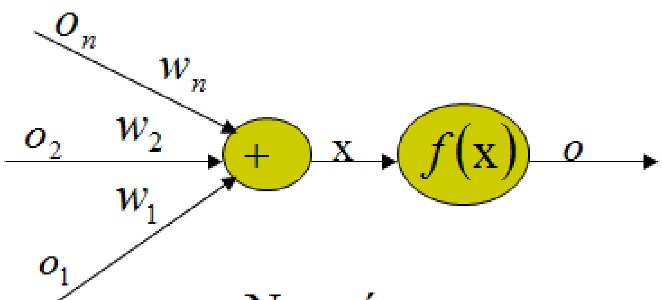
Εισαγωγή (5)

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

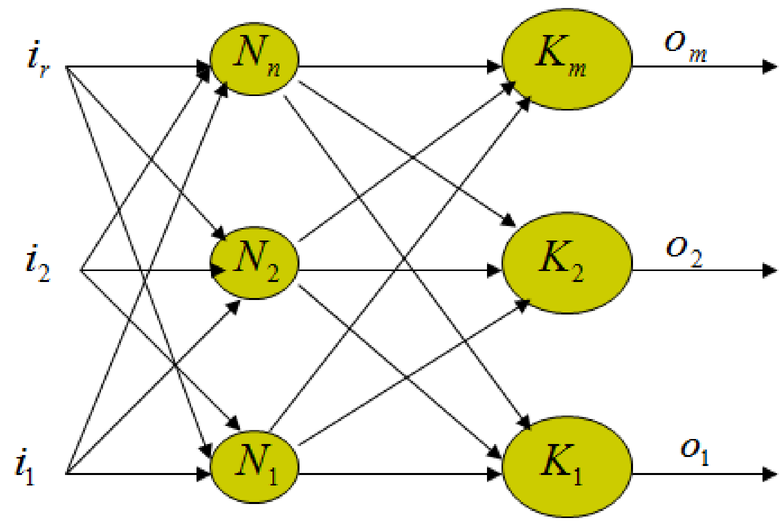
6. Παράλληλη επεξεργασία δεδομένων. Λαμβάνοντας υπόψιν τον τρόπο συνδεσμολογίας των νευρικών κυττάρων, το τεράστιο πλήθος των πληροφοριών που επεξεργάζονται και την σχετικά μεγάλη καθυστέρηση απόκρισης του φυσικού κυττάρου (έχει υπολογιστεί ένας μέσος χρόνος απόκρισης της τάξης των 10^{-3} sec, ενώ η απόκριση ενός ολοκληρωμένου κυκλώματος ανέρχεται σε 10^{-6}sec), 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.

Ο Νευρώνας (1)

- Κάθε νευρωνικό δίκτυο αποτελείται από διασυνδεδεμένες υπολογιστικές μονάδες που ονομάζονται **νευρώνες**. Κάθε νευρώνας μετασχηματίζει το διάνυσμα εισόδου του, δίνοντας μια μόνο έξοδο, η οποία συνδέεται με εισόδους άλλων νευρώνων, όπως φαίνεται και στο σχήμα που ακολουθεί.



Νευρώνας



Τυπική μορφή νευρωνικού δικτύου

Ο Νευρώνας (2)

• Οι προσομοιωτές αποτελούν προσέγγιση των μετασχηματισμών που πραγματοποιούνται στο εσωτερικό των φυσικών νευρικών κυττάρων.

➢ Από 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 νευρωνικού κυττάρου με δυο τελεστές, έναν **γραμμικό** και έναν **μη γραμμικό** που είναι συνδεδεμένοι σε σειρά.

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

$$\mathrm{x} = ow = \sum_{i=1}^{N} o_i w_i$$

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

$$o = f(\mathrm{x})$$

Ο Νευρώνας (3)

• 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:

i. **Είναι αύξουσα συνάρτηση**.

$$\forall \mathrm{x}_1, \mathrm{x}_2 \in \Re, \mathrm{x}_1 \rangle \mathrm{x}_2 \Rightarrow f(\mathrm{x}_1) \geq f(\mathrm{x}_2)$$

ii. **Έχει πεπερασμένα απειροστικά όρια**.

$$\lim_{\mathrm{x} \to +\infty} f(\mathrm{x}) = a, a \in \Re - \{-\infty, +\infty\}$$

$$\lim_{\mathrm{x} \to +\infty} f(\mathrm{x}) = b, b \in \Re - \{-\infty, +\infty\}$$

iii. **Έχει πεδίο ορισμού το σύνολο των πραγματικών αριθμών και φραγμένο πεδίο τιμών**.

$$f : \Re \to [a, b]$$

Ο Νευρώνας (4)

• Οι πλέον διαδεδομένοι, σε πρακτικές εφαρμογές, μη γραμμικοί τελεστές είναι οι ακόλουθοι:

i. **Εκθετική σιγμοειδής**. Η συνάρτηση αυτή αποτελεί τον πιο διαδεδομένο μη γραμμικό νευρωνικό τελεστή:

$$f(\mathrm{x}) = \frac{1}{e^{-a\mathrm{x}+b} + c}, a, b \in \Re^{+}$$

ii. **Υπερβολική εφαπτομένη**.

$$o = \frac{e^{a\mathrm{x}} - e^{-a\mathrm{x}}}{e^{a\mathrm{x}} + e^{-a\mathrm{x}}}, a \in \Re^{+}$$

➢ Είναι φανερό, ότι ο υπολογισμός των παραπάνω συναρτήσεων είναι χρονοβόρος και πολλές φορές υπερκαλύπτει τον χρόνο υπολογισμού του γραμμικού τελεστή του νευρώνα. Γι’ αυτό τον λόγο καταφεύγουμε σε απλούστερες υπολογιστικά συναρτήσεις, όπως είναι οι ακόλουθες:

Ο Νευρώνας (5)

iii.Τμηματικά γραμμική συνάρτηση. Χρησιμοποιείται σπάνια διότι δεν αποδίδει αξιόπιστα την μη γραμμική συμπεριφορά των νευρωνικών δικτύων.

$$f(\mathrm{x}) = \begin{cases} \frac{1}{a}\mathrm{x}, -a \leq \mathrm{x} \leq a \\ -1, \mathrm{x} \langle -a \\ 1, \mathrm{x} \rangle a \end{cases} \qquad a \in \Re^{+}$$

iv. Συνάρτηση δυο κλάδων:

$$f(\mathrm{x}) = \begin{cases} \frac{a\mathrm{x}}{\mathrm{x}+b}, \mathrm{x} \geq 0 \\ -\frac{a\mathrm{x}}{\mathrm{x}-b}, \mathrm{x} \langle 0 \end{cases} \qquad a, b \in \Re^{+}$$

v.Βηματική συνάρτηση:

$$f(\mathrm{x}) = \begin{cases} 1, \mathrm{x} \geq 0 \\ 0, \mathrm{x} \langle 0 \end{cases}$$

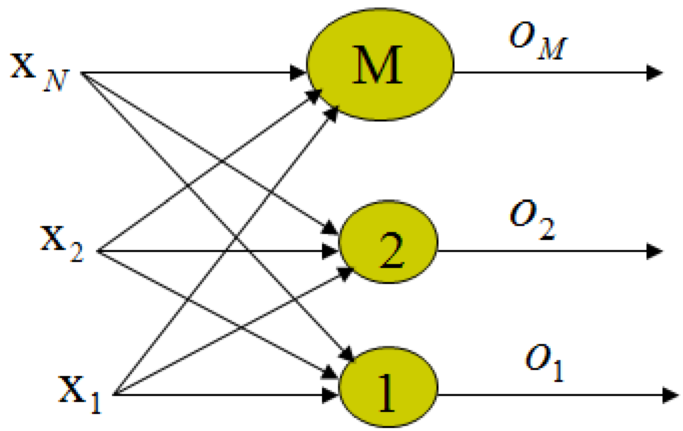
vi. Μη γραμμικός τελεστής του νευρώνα Adaline:

$$f(\mathrm{x}) = \begin{cases} 1, \mathrm{x} \geq 0 \\ -1, \mathrm{x} \langle 0 \end{cases}$$

Παράδειγμα 1ο

Δομή Νευρωνικών Δικτύων (1)

- **<u>Νευρωνικά δίκτυα ενός επιπέδου</u>**. Αποτελεί την πιο 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. Σε μερικές περιπτώσεις, θεωρούμε ότι ο κάθε νευρώνας έχει και μια επιπλέον είσοδο η οποία συνδέεται με μια είσοδο σταθερής στάθμης.



Δομή Νευρωνικών Δικτύων (2)

• Τα δίκτυα ενός επιπέδου νευρώνων χρησιμοποιούνται, συνήθως, σε απλά προβλήματα διότι έχουν δυο σοβαρά μειονεκτήματα:

I. Έστω ότι το σύστημα που θέλουμε να προσομοιώσουμε έχει N εισόδους και M εξόδους, Τότε, στην καλύτερη των περιπτώσεων (όταν κάθε νευρώνας έχει συνάψεις που συνδέουν όλες τις εισόδους), το σύστημα προσομοίωσης έχει (N+1)*M βαθμούς ελευθερίας, γεγονός που περιορίζει την ικανότητα του δικτύου να προσομοιώνει πολύπλοκες διανυσματικές συναρτήσεις.

Το πρόβλημα αντιμετωπίζεται, επεκτείνοντας με τεχνητό τρόπο το διάνυσμα εισόδου, με αποτέλεσμα την ταυτόχρονη αύξηση των βαθμών ελευθερίας του συστήματος προσομοίωσης. Η επέκταση αυτή μπορεί να επιτευχθεί όταν συνδέσουμε με γραμμικό ή μη γραμμικό τρόπο τις αρχικές εισόδους του συστήματος.

Δομή Νευρωνικών Δικτύων (3)

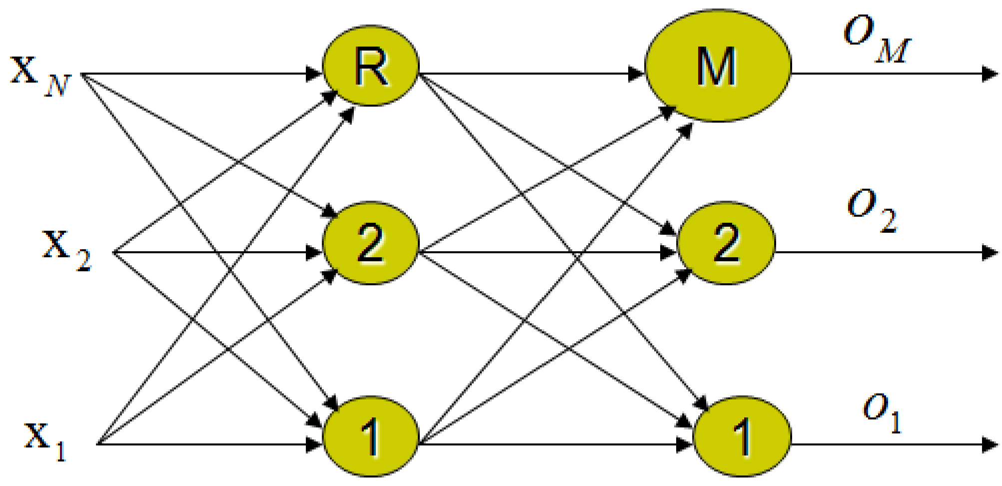
II. Ο δεύτερος περιορισμός αναφέρεται στις περιορισμένες δυνατότητες του δικτύου, όσον αφορά στην προσομοίωση των μη γραμμικών χαρακτηριστικών της συνάρτησης που θέλουμε να προσομοιώσουμε. Έτσι, πιο συγκεκριμένα, μπορούμε να δούμε ότι κάθε έξοδος μπορεί να εκφραστεί σαν συνάρτηση των δεδομένων εισόδου, ως εξής:

$$o_m = f\left(\sum_i \mathrm{x}_i w_{mi}\right)$$

Η έξοδος o_m μπορεί να προσομοιώσει 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.

Δομή Νευρωνικών Δικτύων (4)

• **<u>Πολυεπίπεδα νευρωνικά δίκτυα</u>**. Προσθέτοντας ένα ή περισσότερα κρυφά επίπεδα, μπορούμε να αυξήσουμε απεριόριστα τους βαθμούς ελευθερίας του νευρωνικού δικτύου. Η πιο συνηθισμένη τακτική που ακολουθείται, είναι η σύνδεση των νευρώνων κάθε επιπέδου με τις εξόδους των νευρώνων που βρίσκονται στο προηγούμενο επίπεδο. Όταν ο νευρώνας συνδέεται με όλους τους νευρώνες του προηγούμενου επιπέδου η σύνδεση ονομάζεται πλήρης, διαφορετικά ονομάζεται μερική.



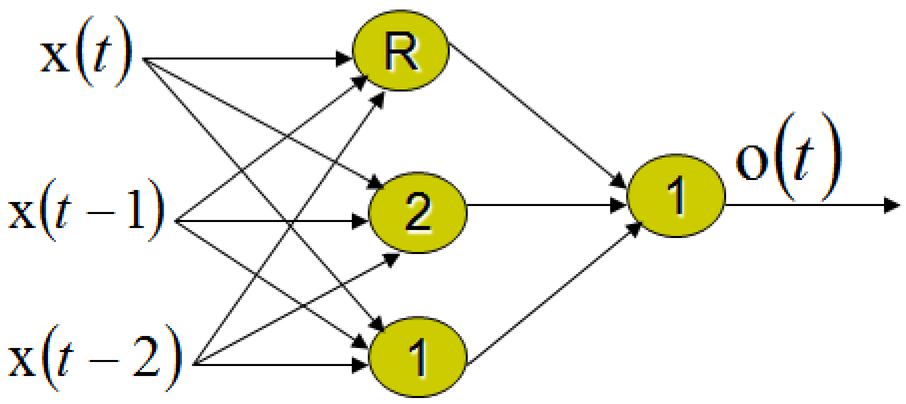
Δομή Νευρωνικών Δικτύων (5)

- 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, 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 δυο επιπέδων (ενός κρυφού και του επιπέδου εξόδου) μπορεί να ταξινομήσει σωστά κατηγορίες μη γραμμικά διαχωρίσιμες, που επιπλέον βρίσκονται σε διαφορετικές νησίδες στον χώρο των μετρήσεων.

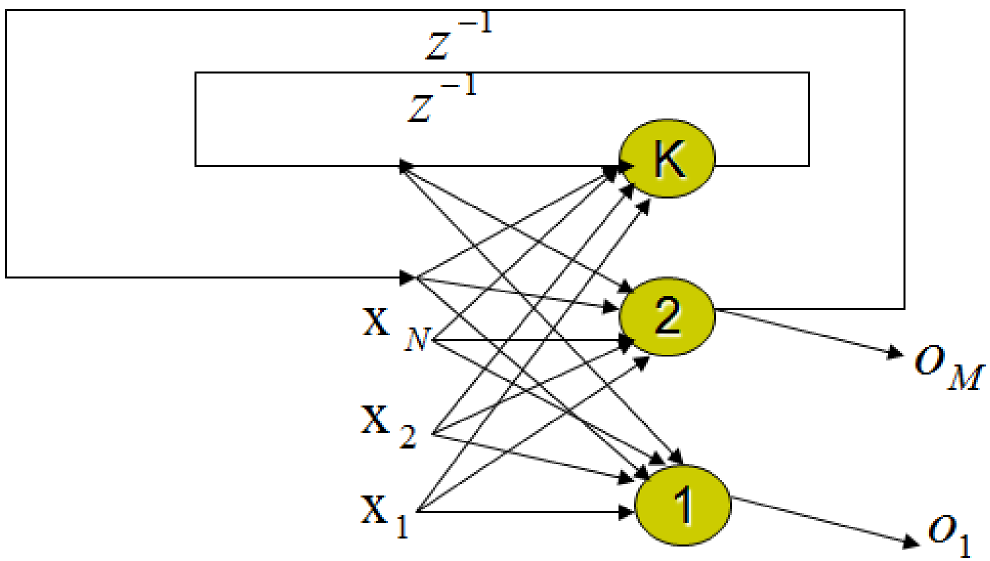
Δομή Νευρωνικών Δικτύων (6)

• 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 στη είσοδο του δικτύου, το οποίο έχει μόνο μια είσοδο, τοποθετηθούν σήματα διακριτού χρόνου που έχουν δειγματοληπτηθεί, κατά αύξουσα χρονική σειρά, τότε το δίκτυο λειτουργεί σαν ένα μη γραμμικό φίλτρο FIR, όπως φαίνεται στο σχήμα που ακολουθεί:



Δομή Νευρωνικών Δικτύων (7)

- **<u>Επανατροφοδοτούμενα νευρωνικά δίκτυα</u>**. Όταν υπάρχει, έστω και μια 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). Η αρχιτεκτονική αυτή προσδίδει στα νευρωνικά δίκτυα κάποια ιδιαίτερα χαρακτηριστικά. Οι χρονικές καθυστερήσεις δίνουν την δυνατότητα στο δίκτυο να προσομοιώσει χρονικά μεταβαλλόμενα πρότυπα.



Εκπαίδευση Νευρωνικών Δικτύων

• Τα βήματα που ακολουθούνται για την κατασκευή των νευρωνικών δικτύων, είναι τα ακόλουθα:

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

II. Υπολογίζονται 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 έξοδος που επιθυμούμε, τότε εφαρμόζεται ένας αλγόριθμος κατευθυνόμενης εκπαίδευσης, διαφορετικά χρησιμοποιείται αλγόριθμος αυτοεκπαίδευσης του δικτύου.

Εκπαίδευση Νευρωνικών Δικτύων-Διόρθωση σφάλματος εξόδου (1)

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

- Έστω σύστημα s(x,w) όπου x είναι η είσοδος του συστήματος και w είναι ένα σύνολο παραμέτρων του συστήματος, των οποίων θέλουμε να υπολογίσουμε την αριθμητική τους τιμή από παραδείγματα. Έστω ότι είναι διαθέσιμα M παραδείγματα. Τοποθετούμε τα παραδείγματα στη είσοδο του συστήματος και υπολογίζουμε την έξοδο για κάθε ένα από αυτά. Έστω ότι x_i,i=1,M είναι οι είσοδοι και t_i,i=1,M οι επιθυμητές έξοδοι. Το μέσο τετραγωνικό σφάλμα του συστήματος (LMS) είναι:

$$\text{Σφάλμα} = \frac{1}{M}\sum_{i=1}^{M}\left(t_i - s(\mathrm{x}_i, w)\right)^T\left(t_i - s(\mathrm{x}_i, w)\right)$$

Εκπαίδευση Νευρωνικών Δικτύων- Διόρθωση σφάλματος εξόδου (2)

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

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

• 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 επιτευχθούν διαδοχικές μειώσεις του συνολικού σφάλματος. Όταν το σφάλμα εκφράζεται σαν το τετράγωνο της διαφοράς της πραγματικής από την επιθυμητή έξοδο του νευρωνικού δικτύου, τότε η μέθοδος καλείται αλγόριθμος LMS (least mean square).

Εκπαίδευση Νευρωνικών Δικτύων-Hebbian Εκπαίδευση (1)

- Ο νευροφυσιολόγος Hebb περιέγραψε την διαδικασία μάθησης των νευρικών κυττάρων στο βιβλίο του « The Organization of Behavior» (1949) ως εξής: 'Όταν ένα νευρωνικό κύτταρο Α διεγείρει συστηματικά ένα νευρωνικό κύτταρο Β, τότε μία διαδικασία μεταβολισμού αλλάζει την συμπεριφορά του κυττάρου Β, έτσι ώστε η διέγερση του κυττάρου Β να προκαλείται πλέον με ευκολότερο τρόπο από το κύτταρο Α'.

- Προφανώς, η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας, που περιέγραψε ο Hebb, λαμβάνει υπόψιν την σχέση ενεργοποίησης που έχουν αλληλοσυνδεόμενοι νευρώνες, οπότε και δεν χρειάζεται να γνωρίζουμε την επιθυμητή έξοδο του δικτύου. Συνεπώς, οι συνάψεις Hebb αποτελούν μια διαδικασία αυτοεκπαίδευσης ενός νευρωνικού συστήματος.

Εκπαίδευση Νευρωνικών Δικτύων-Hebbian Εκπαίδευση (2)

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

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

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

Εκπαίδευση Νευρωνικών Δικτύων-Hebbian Εκπαίδευση (3)

Αν δώσουμε μια 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:

$$\Delta w_{ij} = H_{\varepsilon}\left(o_i, o_j\right)$$

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 τομών το $\Re^{+}$, γεγονός που περιορίζει τις δυνατότατες της σύναψης στον πρώτο κανόνα εκπαίδευσης. Η απλούστερη μορφή της συνάρτησης εκπαίδευσης που ικανοποιεί τους περιορισμούς που τέθηκαν είναι:

$$\Delta w_{ij} = a o_i o_j$$

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

Εκπαίδευση Νευρωνικών Δικτύων-Hebbian Εκπαίδευση (4)

- Το σημαντικότερο μειονέκτημα των συνάψεων Hebb είναι ο κορεσμός που μπορεί να επέλθει στον συντελεστή βαρύτητας των συνάψεων, όταν επαναλαμβάνεται η διαδικασία επαναληπτικού επαναπροσδιορισμού, λόγω διαδοχικών ενεργοποιήσεων των δυο 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. Το φυσικό νόημα αυτού του επιπρόσθετου όρου, είναι η αφαίρεση της πληροφορίας που έχει συσσωρευτεί στην σύναψη του νευρώνα:

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

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

$$\Delta w_{ij} = a o_i o_j - \beta o_i w_{ij} = \beta o_i (\gamma o_j - w_{ij}), \gamma = \frac{a}{\beta}$$

Εκπαίδευση Νευρωνικών Δικτύων-Hebbian Εκπαίδευση (5)

- Η στατιστική προσέγγιση των μηχανισμών 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:

$$\Delta w_{ij} = aConv(o_i, o_j) = aE\left[(o_i - E[o_i])(o_j - E[o_j])\right] = aE[o_i o_j] - aE[o_i]E[o_j]$$

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

i. Όταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης anti-Hebbian είναι ενεργοποιημένοι ταυτόχρονα, τότε η συνδετική ικανότητα της σύναψης μειώνεται.

ii. Όταν οι νευρώνες που βρίσκονται στα άκρα μιας σύναψης anti-Hebbian είναι ενεργοποιημένοι ταυτόχρονα, τότε η συνδετική ικανότητα της σύναψης μειώνεται.

Παράδειγμα 2ο

Εκπαίδευση Νευρωνικών Δικτύων-Ανταγωνιστική Εκπαίδευση (1)

- Η δεύτερη μέθοδος αυτοεκπαίδευσης, που θα εξεταστεί, είναι η ανταγωνιστική εκπαίδευση που εφαρμόζεται σε ανεξάρτητες ομάδες νευρώνων του δικτύου.

- 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 δυο 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:

Εκπαίδευση Νευρωνικών Δικτύων-Ανταγωνιστική Εκπαίδευση (2)

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

$$\sum_i w_{ij} = 1 \quad \text{ή} \quad \sum_i {w_{ij}}^2 = 1$$

II. Επιλέγουμε τυχαία ένα πρότυπο εκπαίδευσης και με αυτό ενεργοποιούμε το δίκτυο.
III. Βρίσκουμε τον νευρώνα j της ομάδας Β, ο οποίος βρίσκεται στην υψηλότερη στάθμη, έχει δηλαδή την υψηλότερη τιμή εξόδου. Τον νευρώνα αυτό, τον ονομάζουμε νικητή (winning neuron).
IV. Επαναπροσδιορίζουμε όλες τις συνάψεις του νικητή νευρώνα, σύμφωνα με την σχέση:

$$\Delta w_{ij} = a\left(o_i - w_{ji}\right)$$

όπου α είναι ένας 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.
V. Οι συντελεστές βαρύτητας των συνάψεων του νικητή νευρώνα κανονικοποιούνται.

Παράδειγμα 3ο

Το γραμμικό φίλτρο Wiener (1)

- Έστω 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:

$$o = \sum_{i=1}^{N} \mathrm{x}_i w_i$$

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

- Έστω d η επιθυμητή έξοδος για το πρότυπο της εισόδου: $\mathrm{x} = \left(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_{\mathrm{N}}\right)^T$
Το σφάλμα της εξόδου ισούται με την διαφορά της επιθυμητής από την πραγματική έξοδο: $e = d - o$

- Όταν έχουμε στην διάθεση μας έναν αριθμό παραδειγμάτων, τότε ορίζουμε σαν συνολικό σφάλμα το μισό του στατιστικά αναμενόμενου τετραγώνου του σφάλματος:

- Όταν υπολογιστούν οι συντελεστές βαρύτητας του γραμμικού δικτύου, που ελαχιστοποιεί το στατιστικά αναμενόμενο σφάλμα, τότε θα λέμε ότι έχουμε κατασκευάσει ένα φίλτρο Wiener.

$$\Sigma\phi\acute{\alpha}\lambda\mu\alpha = \frac{1}{2} E\left[e^2\right]$$

Το γραμμικό φίλτρο Wiener (2)

• Οι συντελεστές βαρύτητας των συνάψεων μπορούν να υπολογιστούν είτε με επαναληπτικό είτε με αναλυτικό τρόπο. Στην συνέχεια ακολουθεί η αναλυτική μέθοδος που δίνει την βέλτιστη λύση και έχει την μικρότερη υπολογιστική πολυπλοκότητα.

• Αντικαθιστώντας τις αναλυτικές εκφράσεις, στην συνάρτηση του στατιστικά αναμενόμενου σφάλματος, προκύπτει:

$$\text{Σφάλμα} = \frac{1}{2}E\left[d^2\right] - E\left[\sum_{i=1}^{N} \mathrm{x}_i w_i d\right] + \frac{1}{2}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \mathrm{x}_i \mathrm{x}_\mathrm{j}\right]$$

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

$$\text{Σφάλμα} = \frac{1}{2}E\left[d^2\right] - \sum_{i=1}^{N} w_i E\left[\mathrm{x}_i d\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j E\left[\mathrm{x}_i \mathrm{x}_j\right]$$

Το γραμμικό φίλτρο Wiener (3)

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

$$\frac{\partial}{\partial w_i} \Sigma\phi\acute{\alpha}\lambda\mu\alpha = 0$$

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

$$\sum_{j=1}^{N} w_j E\left[\mathrm{x}_i \mathrm{x}_j\right] = E\left[\mathrm{x}_i d\right]$$

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

Παράδειγμα 4ο

Πολυεπίπεδο Δίκτυο Perceptron (1)

• Το πλέον διαδεδομένο νευρωνικό δίκτυο συνεχών τιμών εισόδου και εξόδου, είναι το πολυεπίπεδο δίκτυο Perceptron. Ένα τυπικό δίκτυο αποτελείται από νευρώνες του τύπου Perceptron, οι οποίοι διαθέτουν έναν γραμμικό και έναν μη-γραμμικό τελεστή συνδεδεμένους σε σειρά:

$$y = \sum_{j=1}^{N} w_i \mathrm{x}_i$$

$$o = f(y)$$

• Το δίκτυο αποτελείται από περισσότερες των δυο ομάδων νευρώνων Perceptron που ονομάζονται και επίπεδα δικτύου. Ένα 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.

Πολυεπίπεδο Δίκτυο Perceptron (2)

• 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 γίνεται, εκ των προτέρων, να οριστεί η ακριβής τοπολογία του δικτύου. Η γενική αυτή τοπολογία καλύπτει και την πλειονότητα των εφαρμογών.

• Η ευρεία διάδοση του πολυεπίπεδου δικτύου Perceptron, οφείλεται σε δυο κυρίως λόγους:

i. Η ισχυρή πολυπλοκότητα και η μη γραμμικότητα του δικτύου, παρέχουν την δυνατότητα να προσομοιώσουμε με ικανοποιητική ακρίβεια, συνεχείς μη γραμμικές διανυσματικές συναρτήσεις.
ii. Ο αλγόριθμος εκπαίδευσης LMS δίνει την δυνατότητα υπολογισμού των συντελεστών βαρύτητας των συνάψεων από παραδείγματα. Η δυνατότητα αυτή επιτρέπει να χρησιμοποιηθεί το δίκτυο σε πρακτικές εφαρμογές.

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (1)

- Ο αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος, όπως και η μέθοδος της κατευθυνόμενης εκπαίδευσης LMS, έχει σαν σκοπό την εύρεση των συντελεστών βαρύτητας των συνάψεων, με την διαφορά ότι στο πολυεπίπεδο δίκτυο Perceptron, η αναλυτική λύση δεν είναι εφικτή.

- Αυτός είναι και ο λόγος που ακολουθείται η μέθοδος 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 πιο διαδεδομένη τεχνική προσδιορισμού της μνήμης πολύπλοκων νευρωνικών δικτύων, ονομάζεται **αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος**.

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (2)

- 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 της εξόδου, υπολογίζοντας την τιμή της εξόδου των νευρώνων στα κρυφά επίπεδα, η εκπαίδευση εκτελεί την αντίστροφά 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.

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (3)

• 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 (α,b),

$$a \in \Re^N,\ b \in \Re^M.$$

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

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

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

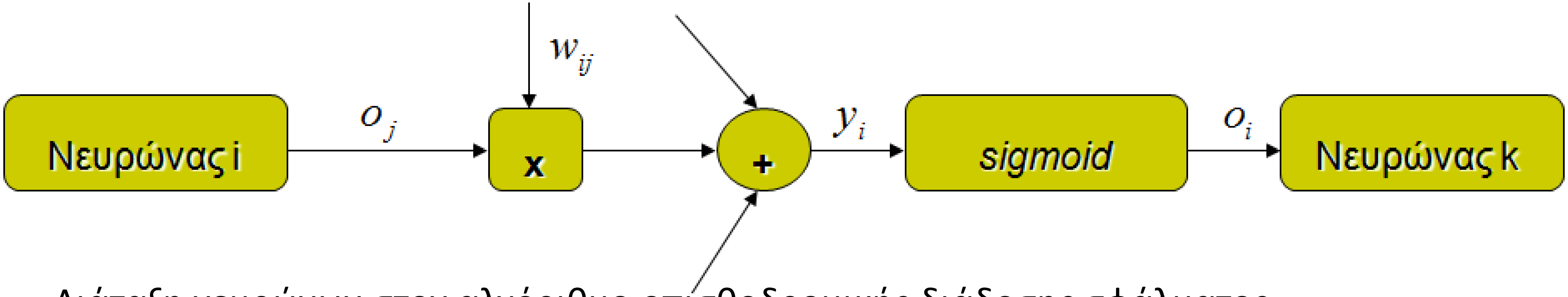
$$E[\Sigma\varphi\acute{\alpha}\lambda\mu\alpha] = \frac{1}{2}E\left[\sum_{i=1}^{M}(b_i - o_i)^2\right] = \frac{1}{2}\sum_{i=1}^{M}E\left[(b_i - o_i)^2\right] = \frac{1}{2Q}\sum_{i=1}^{M}\sum_{j=1}^{Q}\left(b_{ij} - o_{ij}\right)^2$$

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (4)

- Ο αλγόριθμος της οπισθοδρομικής διάδοσης του σφάλματος επαναπροσδιορίζει την αριθμητική τιμή του συντελεστή βαρύτητας των συνάψεων, σύμφωνα με την ακόλουθη σχέση:

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

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



Διάταξη νευρώνων στον αλγόριθμο οπισθοδρομικής διάδοσης σφάλματος.

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (5)

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

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

- 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:

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

και

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

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (6)

- Για να πετύχουμε μια αναλυτική έκφραση της συνάρτησης επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων, χωρίζουμε τους νευρώνες σε δυο κατηγορίες:

I. Νευρώνες εξόδου.
Η αρνητική τιμή της παραγώγου του σφάλματος ως προς την εσωτερική κατάσταση του νευρώνα (δ_i), γίνεται:

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

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:

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

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (7)

ΙΙ. Κρυφοί νευρώνες.

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

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

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

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

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

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

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (8)

- Επειδή ο υπολογισμός των συντελεστών **δ** είναι αναδρομικός, από τους νευρώνες υψηλότερων επιπέδων προς χαμηλότερα επίπεδα, η διαδικασία επαναπροσδιορισμού των συντελεστών βαρύτητας των συνάψεων στο πολυεπίπεδο δίκτυο Perceptron, ακολουθεί τα παρακάτω βήματα:

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

II. 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.

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (9)

III. <u>Υπολογισμός των συντελεστών **δ**</u>. Αντιστρέφοντας την φορά των υπολογισμών, υπολογίζουμε τους συντελεστές **δ** από υψηλότερα σε χαμηλότερα επίπεδα.

IV. <u>Επαναπροσδιορισμός των συντελεστών βαρύτητας των συνάψεων</u>. Έχοντας 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.

V. <u>Έλεγχος σύγκλισης</u>. Υπάρχουν διάφορα κριτήρια ελέγχου της σύγκλισης του αλγορίθμου. Οι συνήθεις έλεγχοι που γίνονται είναι οι εξής:

Πολυεπίπεδο Δίκτυο Perceptron-Αλγόριθμος Οπισθοδρομικής Διάδοσης του Σφάλματος (10)

1. Υπολογίζεται η συνολική απόλυτη μεταβολή των συντελεστών βαρύτητας σε σχέση με έναν πού μικρό θετικό αριθμό, το κριτήριο σύγκλισης.
2. Υπολογίζεται ο μέσος ρυθμός μεταβολής του σφάλματος στα παραδείγματα. Αν η μεταβολή αυτή είναι πολύ μικρή, τότε ο αλγόριθμος τερματίζει.
3. 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.

VI. Επανάληψη του αλγορίθμου. Τοποθετούμε νέο παράδειγμα στην είσοδο του δικτύου και επαναλαμβάνουμε την διαδικασία από το δεύτερο βήμα.

Παράδειγμα 5ο
Παράδειγμα 6ο

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (1)

Παρατηρήσεις

● Ο αλγόριθμος οπισθοδρομικής διάδοσης του σφάλματος εξασφαλίζει την σύγκλιση σε ένα τοπικά ελάχιστο σημείο, χωρίς βεβαία αυτό να σημαίνει ότι εξασφαλίζεται η σύγκλιση και για το συνολικά ελάχιστο.

● 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 έχουμε σύγκλιση του αλγορίθμου στην επιθυμητή λύση. Η συνθήκη αυτή δεν μπορεί να εξασφαλιστεί στην περίπτωση του πολυεπίπεδου δικτύου, τύπου Perceptron.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (2)

- Η λανθασμένη επιλογή του ρυθμού εκπαίδευσης καταλήγει στα δυο παρακάτω, αντίθετα μεταξύ τους, αποτελέσματα:

i. Αν ο συντελεστής εκπαίδευσης έχει μεγάλη αριθμητική τιμή, τότε υπάρχει περίπτωση να μην επιτευχθεί σύγκλιση του αλγορίθμου σε κάποιο τοπικά ελάχιστο σημείο του σφάλματος, αλλά να παρατηρηθούν, κατά την διάρκεια της εκπαίδευσης, ταλαντώσεις γύρω από ένα ή περισσότερα τοπικά ελάχιστα. Οι ταλαντώσεις αυτές είναι μεγαλύτερες, όσο μεγαλύτερη είναι και η αριθμητική τομή του ρυθμού εκπαίδευσης.

ii. 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 αλγορίθμου μέχρι την τελική σύγκλιση.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (3)

- Ακολουθούν μερικές χρήσιμες οδηγίες για την βελτίωση της απόδοσης του αλγορίθμου οπισθοδρομικής διάδοσης του σφάλματος.

i. Τα παραδείγματα πρέπει να επιλέγονται με τυχαίο τρόπο. Μια καλή στρατηγική είναι να αναδιαταχθούν με τυχαίο τρόπο τα παραδείγματα και στην συνέχεια να εφαρμοστεί μια κυκλική διαδοχή χρήσης των παραδειγμάτων στον αλγόριθμο.

ii. Η συνάρτηση σφάλματος που ελαχιστοποιείται, αναφέρεται στο στιγμιαίο σφάλμα που ορίζεται σαν την διαφορά της επιθυμητής 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.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (4)

ΙΙ. (συνέχεια) Αυτή η παραλλαγή του αλγορίθμου αποφεύγει, σε σημαντικό βαθμό, 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, ένα μεγάλο αριθμό νευρώνων και πολλά παραδείγματα, η παραλλαγή του αλγορίθμου, που προτάθηκε πιο πάνω, δύσκολα εφαρμόζεται στη πράξη.
Συνοπτικά, η **στατιστική εκδοχή του αλγορίθμου οπισθοδρομικής διάδοσης του σφάλματος** έχεις εξής:

1. Τοποθετούμε αρχικές τιμές για τον ρυθμό εκπαίδευσης και τους συντελεστές βαρύτητας των νευρώνων.

2. Τοποθετούμε ένα παράδειγμα στην είσοδο του δικτύου και υπολογίζουμε την πραγματική του έξοδο για τη τρέχουσα εκτίμηση των συντελεστών βαρύτητας των νευρώνων.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (5)

3. Κατά την αντίστροφη φορά υπολογίζουμε τους συντελεστές **δ**.

4. Επιλέγουμε το επόμενο παράδειγμα και επαναλαμβάνουμε τον αλγόριθμο από το δεύτερο βήμα.

5. Όταν 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.

6. Ελέγχουμε την συνθήκη σύγκλισης. Αν δεν επιτευχθεί σύγκλιση ο αλγόριθμος επαναλαμβάνεται από την αρχή.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (6)

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

IV. Μια καλή επιλογή των αρχικών τιμών των συντελεστών βαρύτητας των νευρώνων, θα ήταν μικροί πραγματικοί αριθμοί στην περιοχή του μηδενός. Οι αριθμοί αυτοί πρέπει να προσεγγίζουν την τιμή του μηδενός.

V. Η αριθμητική τιμή του ρυθμού εκπαίδευσης (**n**) επιλέγεται, τις περισσότερες φορές, εμπειρικά. Συνήθως, οι τιμές που εξασφαλίζουν καλή ταχύτητα σύγκλισης και αποφεύγουν τις ταλαντώσεις, γύρω από τοπικά βέλτιστα, είναι θετικοί πραγματικοί αριθμοί.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (7)

VI. 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.

VII. Η σύγκλιση του αλγορίθμου επιβραδύνεται σημαντικά με την αύξηση του αριθμού των επιπέδων του δικτύου. Για τον λόγο αυτό, 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 επιθυμητή τιμή, τότε πρέπει να δοκιμάσουμε να προσθέσουμε ένα επίπεδο νευρώνων.

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

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (8)

IX. Θα ήταν καλό να επιλέξουμε έναν μη γραμμικό τελεστή που να ικανοποιεί τη συνθήκη:

$$f(-x) = -f(x)$$

Έχει παρατηρηθεί ότι στις περισσότερες περιπτώσεις ο αλγόριθμος ‘μαθαίνει’ γρηγορότερα, όταν ο μη γραμμικός τελεστής ικανοποιεί την παραπάνω σχέση, Τυπικό παράδειγμα μη γραμμικού νευρωνικού τελεστή που ικανοποιεί αυτή την συνθήκη, είναι η συνάρτηση της υπερβολικής εφαπτομένης.

X. Αυξημένες δυνατότητες αποφυγής τοπικά ελάχιστων σημείων, παρουσιάζονται αν σε κάθε νευρώνα προστεθεί μια 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 θεωρείται ότι αυτή είναι η κύρια αιτία ελάττωσης της πιθανότητας σύγκλισης σε τοπικά ελάχιστο σημείο της συνάρτησης του σφάλματος.

Πολυεπίπεδο Δίκτυο Perceptron-Βελτιώσεις του Αλγορίθμου Οπισθοδρομικής Διάδοσης του Σφάλματος (9)

XI. Μια έξυπνη τεχνική αύξησης της ταχύτητας εκπαίδευσης του δικτύου, είναι η 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 αλγορίθμου με διαφορετικά παραδείγματα. Αντίθετα, όταν έχουμε φαινόμενα αλλαγής πρόσημου (πιθανή ύπαρξη φαινομένων ταλάντωσης γύρω από κάποιο τοπικά ελάχιστο) ο συντελεστής πρέπει να ελαττώνεται.

XII. Σε πολλές εφαρμογές, έχει παρατηρηθεί βελτίωση των χαρακτηριστικών σύγκλισης του αλγορίθμου, 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 αλγορίθμου στο προηγούμενο παράδειγμα:

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

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

Παράδειγμα 7ο

Δίκτυα Ακτινικών Συναρτήσεων (1)

- Υπάρχει μια 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 μια συνάρτηση που περιέχει έκφραση της γενικευμένης Ευκλείδειας απόστασης του διανύσματος εισόδου με το αντίστοιχο διάνυσμα των συντελεστών βαρύτητας των συνάψεων του νευρώνα. Οι νευρώνες εξόδου δεν διαθέτουν μη γραμμικό τελεστή και γι’αυτό τον λόγο εκτελούν μόνο ένα γραμμικό μετασχηματισμό των εξόδων των νευρώνων του κρυφού επιπέδου.

Δίκτυα Ακτινικών Συναρτήσεων (2)

• Παρακάτω ακολουθεί η διαδικασία εκπαίδευσης από παραδείγματα. Η διαδικασία αυτή 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:

$$o = \sum_{i=1}^{N} w_i G(\mathrm{x}, q_i)$$

όπου **x** είναι η είσοδος του δικτύου, $\mathbf{q_i}$ είναι ένα από τα **N** σταθερά σημεία στον χώρο των προτύπων και $\mathbf{w}_i$, 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 επιθυμητής από την πραγματική έξοδο του δικτύου.

Δίκτυα Ακτινικών Συναρτήσεων (3)

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

$$\Sigma\varphi\acute{\alpha}\lambda\mu\alpha = (b - Gw)^T (b - Gw) = |b - Gw|^2 = \sum_{j=1}^{N} \left(b_j - \sum_{i=1}^{N} w_i G(\mathrm{x}, q_i) \right)^2$$

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

$$G = \begin{pmatrix} G(\mathrm{x}_1, q_1) & G(\mathrm{x}_1, q_2) & \cdots & G(\mathrm{x}_1, q_N) \\ G(\mathrm{x}_2, q_1) & G(\mathrm{x}_2, q_2) & \cdots & G(\mathrm{x}_2, q_N) \\ \cdots & \cdots & \cdots & \cdots \\ G(\mathrm{x}_M, q_1) & G(\mathrm{x}_M, q_2) & \cdots & G(\mathrm{x}_M, q_N) \end{pmatrix}$$

Δίκτυα Ακτινικών Συναρτήσεων (4)

- Η LMS λύση δίνεται από την εξίσωση:

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

Ο πίνακας G ονομάζεται ψευδοαντίστροφος πίνακας της G^{+} διότι η διαδικασία υπολογισμού της έχει μοιάζει με τον τρόπο που υπολογίζεται ο αντίστροφος ενός τετραγωνικού πίνακα.

- Από τα παραπάνω παρατηρούμε 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 δικτύου Perceptron. Επίσης έχει αποδειχθεί ότι τα δίκτυα ακτινικών συναρτήσεων μπορούν να προσομοιώσουν, με την επιθυμητή ακρίβεια, οποιαδήποτε συνεχή διανυσματική συνάρτηση.

Τέλος Ενότητας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Σημείωμα Αναφοράς

Copyright Πανεπιστήμιο Πατρών,Δερματάς Ευάγγελος 2015. «Αναγνώριση Προτύπων Ι». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: https://eclass.upatras.gr/courses/EE652/.