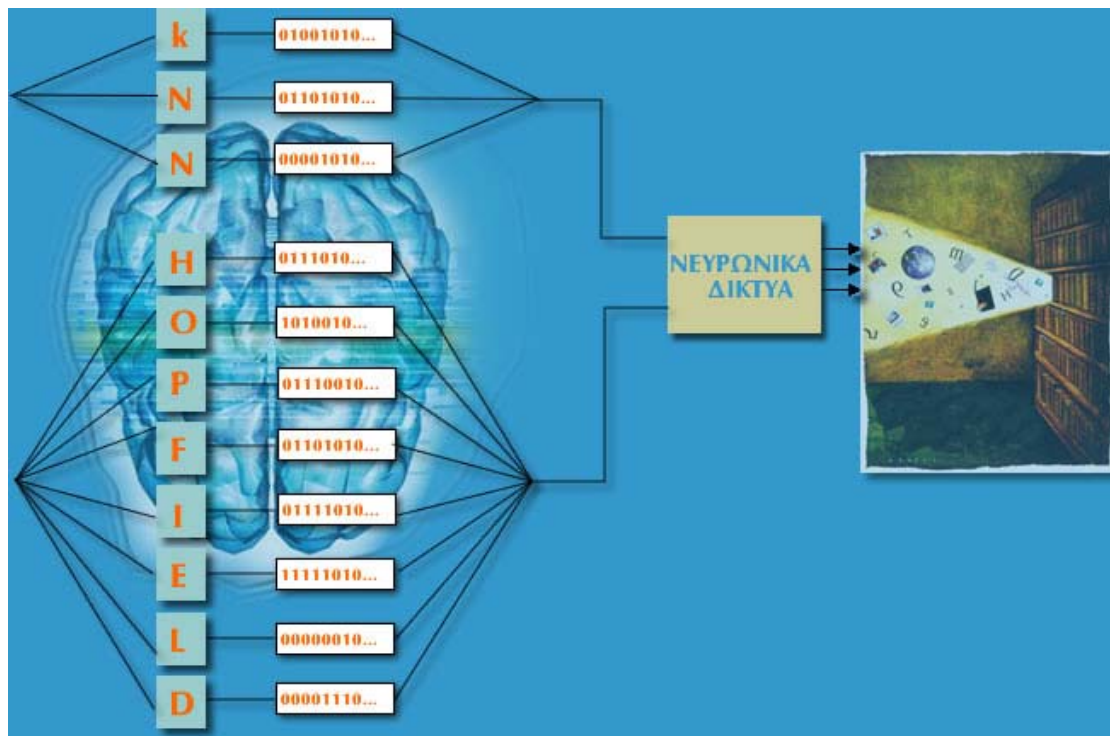


Α.Τ.Ε.Ι. ΗΠΕΙΡΟΥ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΤΗΛΕΠΛΗΡΟΦΟΡΙΚΗΣ & ΔΙΟΙΚΗΣΗΣ

ΤΙΤΛΟΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ
ΒΙΒΛΙΟΘΗΚΗ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ



ΣΥΝΕΡΓΑΤΕΣ : ΚΥΡΙΩΤΗ ΣΤΕΦΑΝΙΑ
ΣΑΡΙΠΑΠΑΖΟΓΛΟΥ ΝΤΟΜΙΝΙΚ

ΥΠΕΥΘΥΝΟΣ ΚΑΘΗΓΗΤΗΣ : Κ. ΤΣΟΥΛΟΣ ΙΩΑΝΝΗΣ

ΑΡΤΑ - ΟΚΤΩΒΡΙΟΣ 2004

Πρόλογος

Η παρούσα εργασία εκπονήθηκε στο Τμήμα Τηλεπληροφορικής και Διοίκησης του Τ.Ε.Ι Ηπείρου από τις σπουδάστριες Κυριώτη Στεφανία και Σαριπαπάζογλου Ντομινίκ. Επιβλέπων καθηγητής ήταν ο κ.Τσούλος Ιωάννης, του οποίου η προσφορά και βοήθεια υπήρξε πολύτιμη. Μας παρείχε καθοδήγηση και άφθονο υλικό σε έντυπη και ηλεκτρονική μορφή, ώστε να προχωρήσουμε στην κατανόηση και συγγραφή της μελέτης των Νευρωνικών Δικτύων. Αρχικά, μας τράβηξε το ενδιαφέρον αυτός ο σύγχρονος και πολύ ενδιαφέρον τομέας της τεχνολογίας των υπολογιστών. Έτσι, αποφασίσαμε να ασχοληθούμε διεξοδικά με το συγκεκριμένο θέμα. Τώρα πια που η εργασία μας έφτασε στο τέλος της, θα θέλαμε να ευχαριστήσουμε τον καθηγητή μας και την επιτροπή για τη στήριξη, αλλά και για τις γνώσεις που μας παρείχαν κατά την διάρκεια της ακαδημαϊκής μας σταδιοδρομίας. Επίσης, ευχαριστούμε τους γονείς μας και όσους μας συμπαράσταθηκαν για να ολοκληρώσουμε με επιτυχία την διπλωματική μας εργασία.

Άρτα Οκτώβριος 2004

Οι συγγραφείς

Κυριώτη Στεφανία
Σαριπαπάζογλου Ντομινίκ

Περίληψη

Η παρούσα διπλωματική εργασία έχει σαν στόχο της, την εισαγωγή σε βασικές έννοιες των Νευρωνικών Δικτύων. Ειδικότερα αναλύονται τα θέματα της μάθησης, της ταξινόμησης, γενίκευσης και της κατηγοριοποίησης. Στο προγραμματιστικό μέρος της εργασίας υλοποιήθηκαν δύο βιβλιοθήκες. Η πρώτη αφορά τον αλγόριθμο των κ-κοντινότερων γειτόνων και η δεύτερη τον αλγόριθμο του νευρωνικού δικτύου ενός επιπέδου. Για να δούμε και στην πράξη την λειτουργία τους, δημιουργήθηκαν παραδείγματα που αναδεικνύουν τις δυνατότητες και τα μειονεκτήματα των αντίστοιχων προγραμμάτων.

Το πρώτο κεφάλαιο ξεκινάει με την παράθεση ορισμών που κατά καιρούς έχουν δοθεί για τα Νευρωνικά Δίκτυα. Στη συνέχεια, παρουσιάζονται οι ιδιότητες και ορισμένες εφαρμογές των δικτύων, καθώς και μία συνοπτική ιστορική αναδρομή.

Η μάθηση είναι μια θεμελιώδης ιδιότητα των νευρωνικών δικτύων και είναι το αντικείμενο του δεύτερου κεφαλαίου. Η έννοια της μάθησης διακρίνεται σε δύο διαφορετικά είδη, την μάθηση με επίβλεψη και την μάθηση χωρίς επίβλεψη. Επίσης, παρουσιάζονται αναλυτικά και οι βασικοί τομείς της ταξινόμησης και γενίκευσης.

Στο τρίτο κεφάλαιο αναπτύσσουμε την έννοια και τον τρόπο λειτουργίας του ταξινομητή των κ-κοντινότερων γειτόνων μέσα από πραγματικές εφαρμογές. Επίσης, υλοποιήθηκαν κάποια παραδείγματα από τα οποία εξάγαμε συμπεράσματα για την λειτουργία του αλγορίθμου.

Στο κεφάλαιο που αναφέρεται στα νευρωνικά δίκτυα ενός επιπέδου, περιγράφεται ο αλγόριθμος εκπαίδευσης του δικτύου και τα προβλήματα στα οποία δίνουν λύσεις. Παράλληλα αναλύονται τα βασικά χαρακτηριστικά τους, όπως τα βάρη και η συνάρτηση ενεργοποίησης. Τα παραδείγματα που παραθέτονται δείχνουν την ικανότητα ή αδυναμία του δικτύου να γενικεύει.

Η πρακτική αξιοποίηση αυτής της εργασίας μπορεί να γίνει και εάν χρησιμοποιήσουμε τα προγράμματα που αναπτύχθηκαν αυτούσια και πολύ περισσότερο με τον περαιτέρω εμπλουτισμό τους.

Το πρόγραμμα των κ-κοντινότερων γειτόνων που δημιουργήσαμε στο πρώτο στάδιο αυτής της εργασίας, ταξινομεί νέα πρότυπα σε κατηγορίες με βάση τους κοντινότερους γείτονες του σημείου. Ο κώδικας αυτού του αλγορίθμου μπορεί να γίνει πολύ χρήσιμος στην αντιμετώπιση προβλημάτων, είτε αυτούσιος είτε με την υλοποίηση πιο σύνθετων συναρτήσεων για την εύρεση των κοντινότερων γειτόνων. Ανάλογα με την πολυπλοκότητα του προβλήματος, επιλέγουμε την κατάλληλη συνάρτηση η οποία αποδίδει καλύτερα.

Από την άλλη, ο αλγόριθμος του νευρωνικού δικτύου ενός επιπέδου που αναπτύχθηκε επιτρέπει την περαιτέρω ανάπτυξη του με σκοπό την υλοποίηση μιας πιο σύνθετης εφαρμογής. Με άλλα λόγια το προγραμματιστικό μέρος της εργασίας μπορεί να χρησιμοποιηθεί στην ανάπτυξη νευρωνικών δικτύων ανώτερου επιπέδου (MLP, Hopfield, κ.α.). Βέβαια ο αλγόριθμος αυτός μπορεί να χρησιμοποιηθεί και στη μορφή που είναι ήδη, για την επίλυση προβλημάτων π.χ. κατηγοριοποίησης. Πραγματικά δεδομένα απαλλαγμένα από θόρυβο θα επιτρέψουν στο δίκτυο να κατηγοριοποιήσει με ικανοποιητικά ποσοστά και να δώσει λύσεις σε πραγματικά προβλήματα.

ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος.....	i
Περίληψη.....	ii
Περιεχόμενα.....	iii

ΜΕΡΟΣ ΠΡΩΤΟ

ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ

1. ΓΕΝΙΚΕΣ ΕΝΟΙΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	1
2. ΙΔΙΟΤΗΤΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	5
3. ΕΦΑΡΜΟΓΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	6
4. ΙΣΤΟΡΙΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ.....	7

ΚΕΦΑΛΑΙΟ 2 ΜΑΘΗΣΗ

2.1 Η ΕΝΝΟΙΑ ΤΗΣ ΜΑΘΗΣΗΣ.....	14
2.2 ΤΑΞΙΝΟΜΗΣΗ.....	18
2.2.1 Η έννοια της ταξινόμησης σε κατηγορίες.....	18
2.2.2 Χρησιμότητα μιας διαδικασίας ταξινόμησης.....	19
2.2.3 Βασικά σημεία μιας διαδικασίας ταξινόμησης.....	20
2.2.4 Η έννοια των κατηγοριών στην ταξινόμηση.....	21
2.2.5 Η έννοια της ακρίβειας.....	22
2.2.6 Παραδείγματα Ταξινομητών.....	23
2.2.7 Μία γενική δομή των προβλημάτων ταξινόμησης.....	27
2.3 ΓΕΝΙΚΕΥΣΗ.....	28
2.3.1 Εισαγωγή στην έννοια της γενίκευσης.....	28
2.3.2 Προϋπόθεσης γενίκευσης.....	29
2.3.3 Μέθοδοι εκτίμησης της ακρίβειας γενίκευσης.....	33
2.4 ΠΟΛΩΣΗ ΜΟΝΤΕΛΩΝ	34
2.4.1 Εισαγωγή στην έννοια της πόλωσης.....	34
2.4.2 Ορισμοί που έχουν δοθεί για την έννοια της πόλωσης.....	34
2.4.3 Το δίλημμα της πόλωσης / μεταβλητότητας.....	35

2.5 ΕΚΜΑΘΗΣΗ ΣΥΝΑΡΤΗΣΕΩΝ.....	37
2.6 ΜΑΘΗΣΗ ΜΕ ΕΠΙΒΛΕΨΗ.....	40
2.7 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ.....	42

ΜΕΡΟΣ ΔΕΥΤΕΡΟ

ΚΕΦΑΛΑΙΟ 3 ΚΟΝΤΙΝΟΤΕΡΟΙ ΓΕΙΤΟΝΕΣ

3.1 ΠΕΡΙΓΡΑΦΗ ΤΕΧΝΙΚΗΣ.....	44
3.2 Η ΑΞΙΟΠΙΣΤΙΑ ΤΑΞΙΝΟΜΗΣΗΣ ΣΤΗ ΜΕΘΟΔΟ K-NN.....	47
3.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΜΕ ΤΗ ΜΕΘΟΔΟ ΤΟΥ k-NN.....	48
3.3.1 Εφαρμογή kNN για την κατηγοριοποίηση ασθενή ως διαβητικού.....	48
3.3.2 Ταξινομητής k-NN για την κατηγοριοποίηση κειμένου.....	51
3.4 ΣΥΜΠΕΡΑΣΜΑΤΑ ΓΙΑ ΤΗ ΜΕΘΟΔΟ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ.....	51
3.5 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ.....	52
3.5.1 Παράδειγμα 1 ^ο	52
3.5.2 Παράδειγμα 2 ^ο	54
3.5.3 Παράδειγμα 3 ^ο	55

ΚΕΦΑΛΑΙΟ 4 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΕΝΟΣ ΕΠΙΠΕΔΟΥ

4.1 ΕΙΣΑΓΩΓΗ.....	57
4.2 ΠΕΡΙΓΡΑΦΗ ΤΕΧΝΙΚΗΣ.....	58
4.2.1 Αλγόριθμος εκπαίδευσης.....	58
4.2.2 Σε ποια προβλήματα μπορεί να χρησιμοποιηθούν τα νευρωνικά δίκτυα ενός επιπέδου.....	59
4.2.3 Συνάρτηση ενεργοποίησης.....	61
4.2.4 Συναπτικά Βάρη νευρώνων δικτύου.....	62
4.2.5 Πόλωση.....	63
4.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ.....	63

4.3.1 Παράδειγμα 1ο	63
4.3.2 Παράδειγμα 2ο	69
4.3.3 Παράδειγμα 3ο.....	72

ΜΕΡΟΣ ΤΡΙΤΟ

ΠΑΡΑΡΤΗΜΑ ΥΛΟΠΟΙΗΣΗ

5.1 ΠΕΡΙΓΡΑΦΗ ΚΑΤΗΓΟΡΙΑΣ ΚΟΝΤΙΝΟΤΕΡΩΝ ΓΕΙΤΟΝΩΝ.....	75
5.2 ΠΕΡΙΓΡΑΦΗ ΚΑΤΗΓΟΡΙΑΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΕΝΟΣ ΕΠΙΠΕΔΟΥ.....	77
5.3 ΕΓΚΑΤΑΣΤΑΣΗ ΠΑΚΕΤΟΥ.....	79
5.3.1 Εγκατάσταση πακέτου στα Windows.....	79
5.3.2 Εγκατάσταση πακέτου στα Linux.....	80
Βιβλιογραφία.....	81

Κεφάλαιο 1

ΕΙΣΑΓΩΓΗ

1.1 ΓΕΝΙΚΕΣ ΕΝΟΙΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Τα **Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ)**, τα οποία συχνά αναφέρονται ως «νευρωνικά δίκτυα», είναι μοντέλα επεξεργασίας πληροφοριών τα οποία αποθηκεύουν εμπειρική γνώση που χρησιμοποιείται για την επίλυση του εκάστοτε προβλήματος. Αποτελούν ένα ισχυρό εργαλείο μοντελοποίησης σύνθετων προβλημάτων (αναγνώριση χαρακτήρων, πρόβλεψη, έλεγχος).

Ένα νευρωνικό δίκτυο είναι ένα παράλληλος κατανεμημένος επεξεργαστής αποτελούμενος από απλές μονάδες επεξεργασίας (νευρώνες) που έχει την φυσική ικανότητα να αποθηκεύει γνώση και να την καθιστά διαθέσιμη για χρήση. Η δομή και η λειτουργία τους παρουσιάζει ομοιότητες με αυτή των βιολογικών νευρωνικών δικτύων στα εξής:

- ✓ Η γνώση αποκτάται από το περιβάλλον μέσω μιας μαθησιακής διαδικασίας (αλγόριθμος εκμάθησης).
- ✓ Η γνώση αποθηκεύεται στα βάρη των συνάψεων(ενώσεων μεταξύ των νευρώνων).

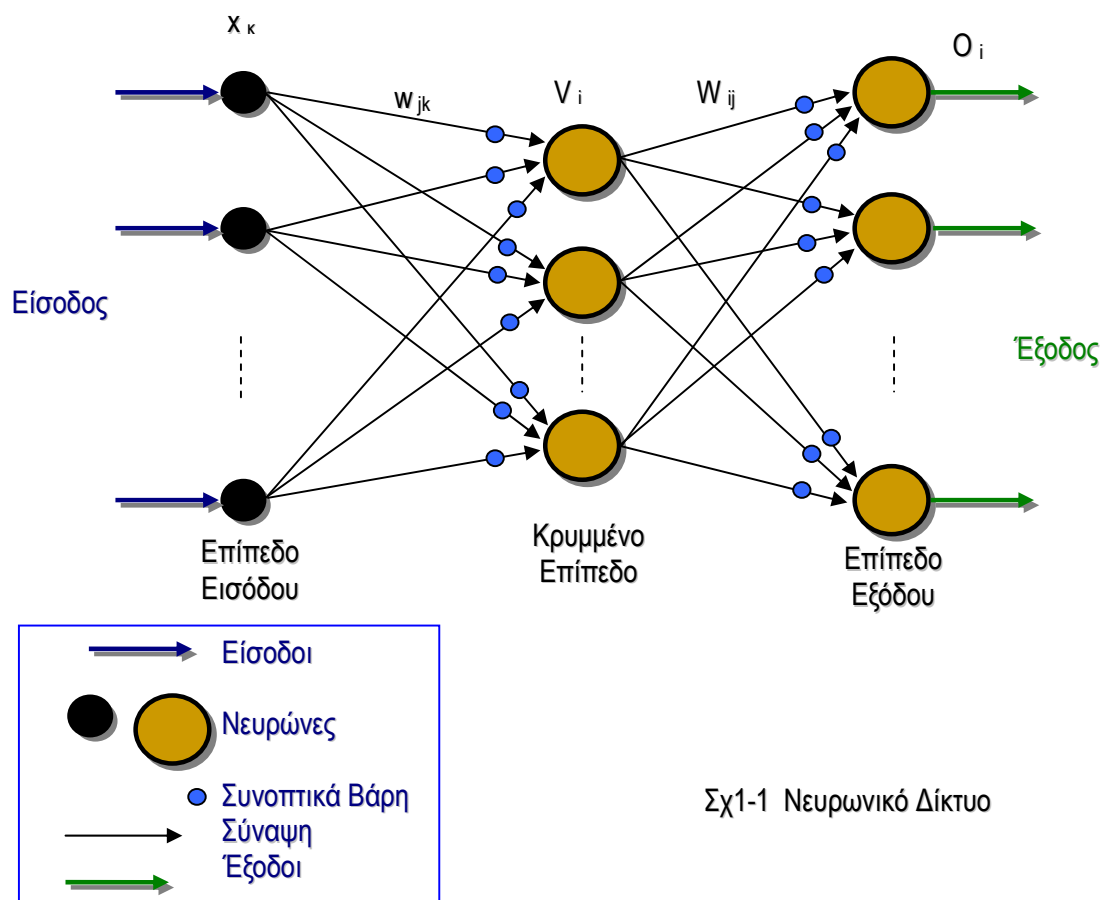
Οι γνώσεις μας όμως, ακόμα και σε κυτταρικό επίπεδο, για τα βιολογικά συστήματα είναι τόσο λίγες, που τα μοντέλα των ΤΝΔ που χρησιμοποιούμε είναι μια υπεραπλούστευση τους. Ένα νευρωνικό δίκτυο μετά το πέρας της εκπαίδευσης του, έχει την ικανότητα να προβλέπει ιδιότητες δεδομένων που δε του ήταν γνωστά κατά τη διάρκεια της μάθησης (γενίκευση).

Αναλυτικότερα, τα νευρωνικά δίκτυα αποτελούνται από απλά στοιχεία (νευρώνες) που διατάσσονται ανάλογα με τις ανάγκες του κάθε προβλήματος. Όπως συμβαίνει και στη φύση, η λειτουργία του νευρωνικού δικτύου προσδιορίζεται σε μεγάλο βαθμό από τον τρόπο με τον οποίο είναι συνδεδεμένα μεταξύ τους τα επιμέρους στοιχεία. Η

αρχιτεκτονική του των νευρωνικών δικτύων χαρακτηρίζεται από ένα γράφημα (Σχ1-1), του οποίου οι κόμβοι κατανέμονται σε:

- α) ένα επίπεδο εισόδου
- β) ένα επίπεδο εξόδου
- γ) ένα ή περισσότερα ενδιάμεσα κρυμμένα επίπεδα.

Καθένας από τους κόμβους εισόδου αναπαριστά και μια μεταβλητή ανεξάρτητη (μεταβλητή εισόδου). Κάθε κόμβος συνδέεται με όλους τους κόμβους στο πρώτο κρυμμένο επίπεδο.



Σχ1-1 Νευρωνικό Δίκτυο

Οι κόμβοι του πρώτου κρυμμένου επιπέδου συνδέονται με τους κόμβους ενός άλλου κρυμμένου επιπέδου ή με τους κόμβους του επιπέδου εξόδου. Οι κόμβοι του νευρωνικού δικτύου ονομάζονται «νευρώνες» ενώ οι δεσμοί μεταξύ των κόμβων ονομάζονται συνάψεις. Κάθε σύναψη χαρακτηρίζεται από ένα βάρος (Συναπτικό βάρος) με το οποίο πολλαπλασιάζεται το σήμα εισόδου.

Η αρχιτεκτονική ενός νευρωνικού δικτύου προσδιορίζεται από τον αριθμό των κόμβων, τον αριθμό των κρυμμένων επιπέδων και τον τρόπο με τον οποίο οι νευρώνες συνδέονται μεταξύ τους. Ο αριθμός των κρυμμένων επιπέδων και των νευρώνων, τα τυχόν όρια μέσα στα οποία θα κυμαίνονται τα συναπτικά βάρη είναι στοιχεία που προσδιορίζονται στο σχεδιασμό του νευρωνικού δικτύου.

Ένα νευρωνικό δίκτυο μπορεί να εκπαιδευτεί ώστε να επιτελεί μια συγκεκριμένη λειτουργία. Αυτό επιτυγχάνεται με κατάλληλη προσαρμογή των συναπτικών βαρών. Η διαδικασία αυτή, ονομάζεται **εκπαίδευση** του νευρωνικού δικτύου. Η εκπαίδευση των νευρωνικών δικτύων γίνεται κατά τέτοιο τρόπο, ώστε σε μια είσοδο προτύπου να αντιστοιχίζεται σε μια ή περισσότερες εξόδους του δικτύου. Η λειτουργία αυτή στηρίζεται στους νευρώνες, τα συστατικά εκείνα του νευρικού συστήματος που συνδέονται και επικοινωνούν μεταξύ τους επιτρέποντας τη μάθηση με παραδείγματα.

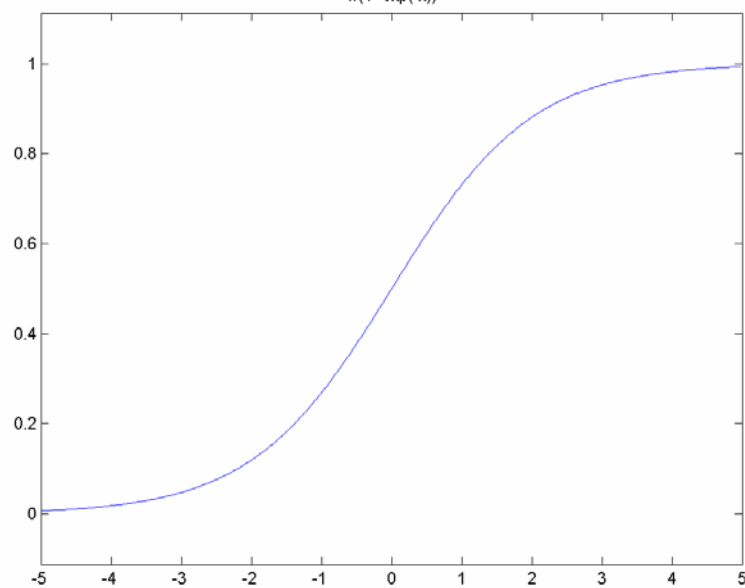
Ορισμοί Νευρωνικών Δικτύων

Μαθηματικός ορισμός :

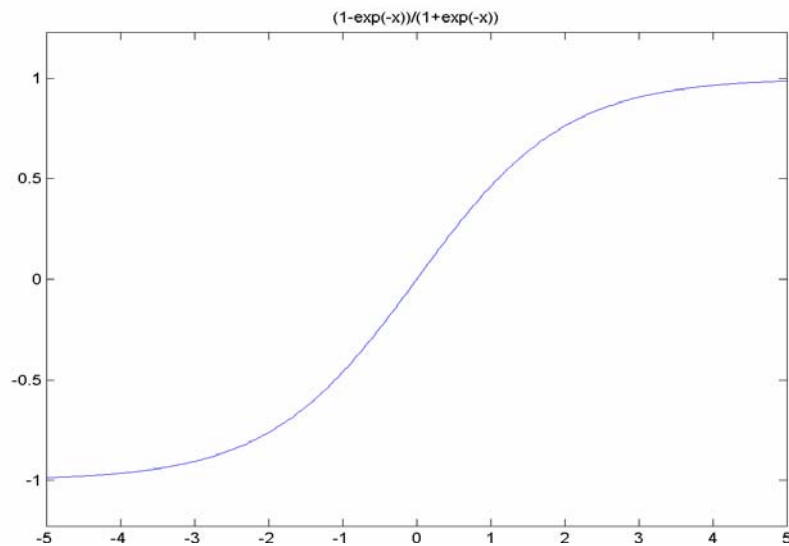
Ένα ΤΝΔ είναι ένας κατευθυντικός γράφος που αποτελείται από κόμβους διασυνδεδεμένους με συναπτικούς κλάδους και κλάδους ενεργοποίησης, με τις ακόλουθες ιδιότητες:

- ❖ Κάθε νευρώνας αποτελείται από ένα σύνολο γραμμικών συναπτικών συνδέσμων, μια σταθερά πόλωσης εξωτερικά εφαρμοζόμενη και έναν κατά τεκμήριο μη γραμμικό κλάδο ενεργοποίησης.
- ❖ Οι συναπτικοί σύνδεσμοι πολλαπλασιάζουν τα βάρη τα σήματα εισόδου τους (Εσωτερικό γινόμενο βάρους* σήμα εισόδου).
- ❖ Το άθροισμα των εσωτερικών γινομένων των σημάτων εισόδου ενός νευρώνα επί τα αντίστοιχα βάρη, αποτελούν το δυναμικό ενεργοποίησης του νευρώνα..

- ❖ Η συνάρτηση ενεργοποίησης (Σχ1-2 Λογιστική συνάρτηση, Σχ1-3 Συνάρτηση υπερβολικής εφαιπτόμενης) περιορίζει το δυναμικό ενεργοποίησης του νευρώνα για να παράγει την έξοδο του νευρώνα.



Σχ1-2 Λογιστική συνάρτηση



Σχ1-3 Συνάρτηση υπερβολικής εφαιπτόμενης (tanh)

Σύμφωνα με το **DARPA** [1] :

Ένα Νευρωνικό Δίκτυο είναι ένα σύστημα αποτελούμενο από πολλά «απλά» συστατικά που λειτουργούν με παράλληλες διαδικασίες. Οι λειτουργία τους συνίσταται από τη δομή του νευρωνικού δικτύου, το τρόπο που είναι συνδεδεμένα, από τις δυνάμεις που τα συγκρατούν και από τη διαδικασία επεξεργασίας που συμβαίνει στους κόμβους .

Σύμφωνα με το **Haykin** [2]:

Ένα νευρωνικό δίκτυο είναι ένας παράλληλης επεξεργασίας, επεξεργαστής, ο οποίος έχει την τάση να αποθηκεύει γνώση, απορία παραδειγμάτων και να τη διαθέτει για περαιτέρω χρήση. Μοιάζει με το ανθρώπινο μυαλό με δύο τρόπους:

1. Η γνώση αποκτάται από το δίκτυο μέσω της διαδικασίας της μάθησης.
2. Οι δικτυακές συνδέσεις(δυνάμεις) γνωστές και σαν συναπτικά βάρη χρησιμοποιούνται για να αποθηκεύσουν αυτή τη γνώση.

Τα μεγαλύτερο πλεονέκτημα των ΤΝΔ είναι ότι μπορούν να δώσουν λύση σε προβλήματα που δεν λύνονται με συμβατικές τεχνολογίες λόγω της πολυπλοκότητας τους. Τα νευρωνικά δίκτυα έχουν την ικανότητα να «αποκομίζουν» πληροφορία από δεδομένα περίπλοκα ή μη ακριβή. Ένα εκπαιδευμένο νευρωνικό δίκτυο μπορεί να θεωρηθεί σαν «έμπειρο» στο σύνολο της πληροφορίας που του έχει δοθεί για ανάλυση. Την γνώση λοιπόν αυτή, μπορεί να τη χρησιμοποιήσει προκειμένου να κάνει μία πρόβλεψη σε ένα πρόβλημα στο οποίο παρουσιάζονται νέα και άγνωστα δεδομένα.

1.2 ΙΔΙΟΤΗΤΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

1. **Προσαρμοστικότητα:** Είναι η ικανότητα να μαθαίνουν πως να ενεργούν βασιζόμενα πάντα σε δεδομένα που χρησιμοποίησαν κατά τη διαδικασία της εκπαίδευσης.
 2. **Οργάνωση της πληροφορίας:** Ένα νευρωνικό δίκτυο μπορεί να οργανώνει αλλά και να αποδίδει, χωρίς την ανθρώπινη παρέμβαση, την πληροφορία κατά τη
-

- διάρκεια της εκπαίδευσης του. Με άλλα λόγια είναι σε θέση να αναλύει τα χαρακτηριστικά των προτύπων και να τα κατηγοριοποιεί χωρίς να του δοθεί οποιαδήποτε άλλη πληροφορία
3. **Λειτουργία σε πραγματικό χρόνο:** Τα νευρωνικά δίκτυα μπορούν να εφαρμοστούν σε ειδικό λογισμικό και να δίνουν «απαντήσεις» σε πραγματικό χρόνο. Μετά την εκπαίδευση του ένα δίκτυο έχει αποκτήσει εμπειρία σε ένα σύνολο δεδομένων και μπορεί να δώσει κατηγοριοποιήσει ή να αναγνωρίσει ένα πρότυπο που του δίνεται σε πραγματικό χρόνο.
 4. **Απεικόνιση εισόδου-εξόδου:** Δημιουργία συνάρτησης απεικόνισης π.χ. με επιβλεπόμενη εκπαίδευση.
 - ❖ Τροφοδότηση με παραδείγματα: σήματα εισόδου και επιθυμητές εξοδοι.
 - ❖ Τροποποίηση των συνάψεων για ελαχιστοποίηση της απόστασης πραγματικής-επιθυμητής εξόδου.
 - ❖ Σχέση με μη παραμετρική εκτίμηση.
 5. Ομοιόμορφη ανάλυση και σχεδίαση: καθολικότητα ως επεξεργαστές πληροφορίας
 - ❖ Οι νευρώνες κοινό συστατικό σε όλα τα δίκτυα
 - ❖ Χρήση κοινών τεχνικών εκμάθησης και θεωρητικής ανάλυσης.
 - ❖ Το ίδιο ΤΝΔ μπορεί να εφαρμοστεί σε πολλές εφαρμογές (με κατάλληλη εκμάθηση).
 6. Νευροβιολογικό ανάλογο. Αναζήτηση ιδεών και έμπνευσης από την νευρολογία, και αντιστρόφως

1.3 ΕΦΑΡΜΟΓΕΣ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Έχοντας δώσει τη περιγραφή των νευρωνικών δικτύων και του τρόπου με τον οποίο επεξεργάζονται τη πληροφορία είναι σημαντικό επίσης να περιγράψουμε τις πραγματικές εφαρμογές στις οποίες χρησιμοποιούνται.

Τα νευρωνικά δίκτυα βρίσκουν πολλές εφαρμογές:

- ❖ Στη βιομηχανία (έλεγχος παραγωγικής διαδικασίας)
- ❖ Πρόβλεψη πωλήσεων
- ❖ Επενδύσεις (Πρόβλεψη οικονομικών στοιχείων)
- ❖ Ιατρική (Νευρολογικά μοντέλα νοημοσύνης)
- ❖ Ηλεκτρονικά παιχνίδια
- ❖ Συστήματα ελέγχου
- ❖ Αναγνώριση φωνής στις τηλεπικοινωνίες
- ❖ Διάγνωση ασθενιών
- ❖ Αναγνώριση στις τηλεπικοινωνίες σφαλμάτων λογισμικού
- ❖ Μετάφραση
- ❖ Υποθαλάσσια ανίχνευση ναρκών
- ❖ Ανάλυση υφής, χαρακτήρων και προσώπων
- ❖ Πρόβλεψη φυσικών φαινομένων

1.4 ΙΣΤΟΡΙΑ ΝΕΥΡΩΝΙΚΩΝ ΔΙΚΤΥΩΝ

Μετά από μια περίοδο έντονου ενθουσιασμού ο τομέας των νευρωνικών χρειάστηκε να ξεπεράσει δυσκολίες. Στο χρονικό αυτό διάστημα οι χρηματοδοτήσεις για έρευνα ήταν μηδαμινές και ήταν ελάχιστοι εκείνοι οι επιστήμονες που ασχολήθηκαν με τα νευρωνικά. Αν και λίγοι οι ερευνητές κατάφεραν να αναπτύξουν τη τεχνολογία των νευρωνικών, ώστε να ξεπεράσουν τους περιορισμούς που αρχικά είχαν τεθεί από τους Minsky και Papert.

Οι Minsky και Papert εξέδωσαν ένα βιβλίο(1969) στο οποίο εξέφραζαν ένα γενικό αίσθημα κατά των νευρωνικών δικτύων, αυτό δυστυχώς έγινε αποδεκτό από τους περισσότερους επιστήμονες χωρίς περαιτέρω ανάλυση. Στις μέρες μας βέβαια τα νευρωνικά χαίρουν ιδιαίτερου ενδιαφέροντος για τις δυνατότητες που προσφέρουν και αντίστοιχες είναι και οι χρηματοδοτήσεις για έρευνα .

Αναλυτικότερα η ιστορία των νευρωνικών δικτύων μπορεί να χωριστεί σε περιόδους:

1. **Αρχικές προσπάθειες:** Υπήρξαν μερικές αρχικές εφαρμογές οι McCulloch και Pitts (1943) ανέπτυξαν μοντέλα νευρωνικών δικτύων βασισμένοι στην νευρολογία. Τα μοντέλα αυτά έκαναν «υποθέσεις» σχετικά με το πως λειτουργούν οι βιολογικοί νευρώνες. Τα δίκτυα αυτά μπορούσαν να υλοποιήσουν απλές συναρτήσεις δυαδικές όπως η “or” “and”. Έπειτα δύο ομάδες (Farley και Clark, 1954) και (Rochester, Holland, Haibit και Duda, 1956).
2. **Τα πρώτα δείγματα τεχνολογίας:** Η επιστήμη της νευρολογίας επηρέασε την ανάπτυξη των νευρωνικών δικτύων. Μάλιστα ψυχολόγοι και μηχανικοί λογισμικού συνέβαλαν στην πρόοδο των νευρωνικών. Ο Rosenblatt (1958) δημιούργησε σημαντικό ενδιαφέρον και δραστηριότητα στο πεδίο των νευρωνικών όταν πρωτοσχεδίασε την πιο απλή μορφή νευρωνικού δικτύου (perceptron). Ένα ακόμη σύστημα αναπτύχθηκε το γνωστό ADALINE (ADaptive LInear Element) αναπτύχθηκε στη δεκαετία του 1960 από τους Widrow και Hoff (Πανεπιστήμιο του Stanford). Το ADALINE ήταν μια αναλογική ηλεκτρονική συσκευή αποτελούμενη από απλά συστατικά
3. **Η περίοδος αποδοκιμασίας:** Το 1969 οι Minsk και Paper έγραψαν ένα βιβλίο στο οποίο αναφερόταν στους περιορισμούς του Perceptron ενός επιπέδου αλλά και των συστημάτων με πολλά επίπεδα. Το αποτέλεσμα του βιβλίου αυτού ήταν να ελαχιστοποιηθούν οι χρηματοδοτήσεις της έρευνας στο τομέα των νευρωνικών δικτύων. Και βέβαια ένα γενικό αρνητικό κλίμα δημιουργήθηκε.
4. **Καινοτομία:** Αν και το γενικό ενδιαφέρον και οι χρηματική ενίσχυση στο κλάδο των νευρωνικών ήταν σχεδόν ανύπαρκτη αρκετοί ερευνητές συνέχισαν να εργάζονται στην ανάπτυξη μεθόδων που είχαν τις βάσεις τους στη νευρολογία με χρήση υπολογιστών όπως η αναγνώριση χαρακτήρων Στη διάρκεια αυτής της περιόδου πολλά παραδείγματα που χρησιμοποιούμε ακόμη και σήμερα αναπτύχθηκαν. Οι Steve Grossberg και Gail Carpenter το 1988 ίδρυσαν μια σχολή στην οποία αναπτύχθηκε η ART (Adaptive Resonance Theory) και νευρωνικά δίκτυα βασισμένα σε βιολογικά μοντέλα. Οι Anderson και Kohonen ανέπτυξαν διάφορες τεχνικές. Ο Henry Klopff το 1972, έθεσε τις βάσεις για

- μάθηση στους τεχνητούς νευρώνες βασιζόμενος σε αρχές της βιολογικών νευρώνων. Ενώ ο F. Kunihiro ανέπτυξε ένα «σοφό» νευρωνικό δίκτυο με πολλά επίπεδα νευρώνων για μετάφραση χαρακτήρων γραμμένων με το χέρι (1975).
5. **Ανάπτυξη:** Η πρόοδος από το τέλος της δεκαετίας του 1970 μέχρι τις αρχές του 1980 ήταν πολύ σημαντική για την ενίσχυση του ενδιαφέροντος στο τομέα των νευρωνικών. Πολλοί παράγοντες επηρέασαν αυτή την κίνηση των επιστημόνων. Για παράδειγμα διάφορα βιβλία που εκδόθηκαν και διαλέξεις έδωσαν στους επιστήμονες μια βάση για να ανταλλάξουν απόψεις. Ακαδημαϊκά προγράμματα έρευνας αρχικά αλλά και μαθήματα σε πανεπιστήμια της Αμερικής και της Ευρώπης βοήθησαν ώστε να διαδοθεί η γνώση για τα νευρωνικά.
 6. **Σήμερα:** Ουσιαστική βελτίωση έχει επέλθει στην επιστήμη των νευρωνικών δικτύων και μάλιστα αυτό προσελκύει ερευνητές για περαιτέρω συνέχιση της έρευνας. Μικροτσιπ που βασίζονται σε λειτουργίες των βιολογικών νευρώνων έχουν ήδη κατασκευαστεί. Ουσιαστικά η εποχή για τη πλήρη αναγνώριση των λύσεων που μπορούν να προσφέρουν τα νευρωνικά αλλά και της αξίας τους έχει φτάσει.

Κατηγορίες νευρωνικών δικτύων

Υπάρχουν πολλά είδη ΤΝΔ. Τα δύο κύρια είδη διαχωρίζονται με βάση ορισμένα χαρακτηριστικά:

- Τοπολογία δικτύου
- Χαρακτηριστικά μονάδων
- Λειτουργικότητα δικτύου

Δύο σημαντικά είδη τοπολογίας δικτύων είναι ευθείας τροφοδότησης δίκτυα και τα δίκτυα με ανατροφοδότηση.

- Στα ευθείας τροφοδότησης δίκτυα οι συνδέσεις μεταξύ των μονάδων δεν διαγράφουν κύκλους. Τα δίκτυα αυτής της τοπολογίας παράγουν συνήθως μια απάντηση σε μια είσοδο πολύ γρήγορα. Το περισσότερα δίκτυα εκπαιδεύονται χρησιμοποιώντας μια ευρεία ποικιλία αποδοτικών μεθόδων με χρήση αριθμών .
- Στα δίκτυα με ανατροφοδότηση επιτρέπεται να δημιουργούνται επαναλήψεις στις συνδέσεις. Σε κάποια ανατροφοδότηση ΤΝΔ, κάθε φορά που δίδεται ένα δείγμα για εκπαίδευση το ΤΝΔ πρέπει να επαναλάβει τη διαδικασία για έναν ενδεχομένως ορισμένο χρόνο πριν δώσει μια απάντηση. Το δίκτυο με ανατροφοδότηση συνήθως είναι δυσκολότερο να εκπαιδευθεί από εκείνα της ευθείας τροφοδότησης..

Μερικοί από τους πιο γνωστούς τύπους νευρωνικών δικτύων:

1. Επιβλεπόμενης μάθηση

1.1 Ευθείας τροφοδότησης

- Γραμμικά
 - Δίκτυα που χρησιμοποιούν τον αλγόριθμο Hebb - Hebb (1949), Fausett (1994)
 - Νευρωνικό δίκτυο ενός κρυμμένου επιπέδου - Rosenblatt (1958), Minsky και Papert (1969/1988), Fausett (1994)
 - Νευρωνικό δίκτυο τύπου Adaline - Widrow και Hoff (1960), Fausett (1994)
 - Νευρωνικό δίκτυο τύπου Higher Order - Bishop (1995)
 - Πολυεπίπεδα νευρωνικά δίκτυα (Multilayer perceptron) Bishop (1995), Reed και Marks (1999), Fausett (1994)
-

- Νευρωνικά δίκτυα Οπισθοδρόμησης (Backprop) Rumelhart, Hinton, και Williams (1986)
- Νευρωνικό τύπου Cascade Correlation - Fahlman και Lebiere (1990), Fausett (1994)
- Γρήγορης Οπισθοδρόμησης (Quickprop) - Fahlman (1989)
- Νευρωνικά τύπου RBF - Bishop (1995), Moody και Darken (1989), Orr (1996)
 - Ορθογωνικά - Ελαχίστου τετραγώνου Νευρωνικά Δίκτυα (OLS: Orthogonal Least Squares) - Chen, Cowan και Grant (1991)
- Κατηγοριοποίηση
 - Νευρωνικά Δίκτυα τύπου LVQ: Learning Vector Quantization - Kohonen (1988), Fausett (1994)

1.2 Δίκτυα με ανατροφοδότησης (Feedback) Hertz, Krogh, και Palmer (1991), Medsker και Jain (2000)

- Μηχανή Boltzman (Boltzman Machine) - Ackley et al. (1985), Fausett (1994)
 - Επαλαμβανόμενου χρόνου
 - Νευρωνικά δίκτυα οπισθοδρόμησης μέσω χρόνου (Back propagation through time) - Werbos (1990)
 - Νευρωνικά δίκτυα τύπου Elman - Elman (1990)
 - Νευρωνικά δίκτυα τύπου Jordan - Jordan (1986)
 - Πραγματικού χρόνου δίκτυα επανάληψης (Real-time recurrent network) - Williams and Zipser (1989)
-

- Πραγματικού χρόνου δίκτυα οπισθοδρόμησης (Recurrent back propagation) - Pineda (1989), Fausett (1994)
- Νευρωνικά δίκτυα χρονοκαθυστερήσης (TDNN: Time Delay NN) - Lang, Waibel and Hinton (1990)

1.3 Ανταγωνιστικά

- Δίκτυα τύπου ARTMAP - Carpenter, Grossberg και Reynolds (1991)
- Νευρωνικά δίκτυα τύπου ARTMAP Γκαουσιανά- Williamson (1995)

2. Δίκτυα χωρίς επίβλεψη (Unsupervised) Hertz, Krogh, and Palmer (1991)

2.1 Ανταγωνιστικά

- Κβαντοποίηση διανύσματος
 - Δίκτυα Grossberg - Grossberg (1976)
 - Δίκτυα Kohonen - Kohonen (1984)
- Αυτό-Οργανωμένοι χάρτες.
 - Δίκτυα τύπου Kohonen - Kohonen (1995), Fausett (1994)
 - Τοπικής γραμμικότητας (Local Linear) - Mulier and Cherkassky (1995)
- Θεωρία προσαρμοσμένης ενίσχυσης
 - Νευρωνικά δίκτυα θεωρίας προσαρμοσμένης ενίσχυσης 1 (ART 1) - Carpenter και Grossberg (1987a), Moore (1988), Fausett (1994)

2.2 Μείωση διαστάσεων Dimension Reduction - Diamantaras and Kung (1996)

- Νευρωνικά δίκτυα σύμφωνα με το νόμο του Hebb (Hebbian) - Hebb (1949), Fausett (1994)
- Διαφοροποιημένα νευρωνικά δίκτυα Hebb (Differential Hebbian) - Kosko (1992)

2.3 Συσχετικά (Autoassociation)

- Γραμμικής συσχέτισης δίκτυα (Linear autoassociator) - Anderson (1977), Fausett (1994)
- Νευρωνικά δίκτυα Hopfield - Hopfield (1982), Fausett (1994)

3. Μη γραμμικά

3.1 Νευρωνικά δίκτυα Hopfield - Hertz, Krogh, και Palmer (1991)

Κεφάλαιο 2

ΜΑΘΗΣΗ

2.1 Η ΕΝΝΟΙΑ ΤΗΣ ΜΑΘΗΣΗΣ

Η μάθηση στα νευρωνικά δίκτυα είναι η σημαντικότερη διαδικασία μετά το σχεδιασμό του δικτύου μάλιστα είναι η βασική ιδιότητα των ΤΝΔ και επιτρέπει τη σταδιακή βελτίωση της απόδοσης του.

Μπορεί λοιπόν να ορισθεί πως σαν γενική περίπτωση ότι :

Μάθηση (learning) είναι η διαδικασία της τροποποίησης της τιμής των βαρών του δικτύου μέσω των ερεθισμάτων που δέχεται από το περιβάλλον. Το είδος της μάθησης καθορίζεται από τον τρόπο που λαμβάνει χώρα η μεταβολή των παραμέτρων του δικτύου. Ο αλγόριθμος μάθησης καθορίζει πως θα γίνεται η μεταβολή των συνοπτικών βαρών ώστε δοθέντος συγκεκριμένου διανύσματος εισόδου να παραχθεί συγκεκριμένο διάνυσμα εξόδου (επιθυμητή τιμή εξόδου). Το δίκτυο αποκτά έτσι μεγαλύτερη γνώση του περιβάλλοντος στο οποίο λειτουργεί μετά από κάθε επανάληψη της διαδικασίας μάθησης.

Μερικές φορές, τα πρότυπα εκπαίδευσης μπορούν να περιέχουν λάθη (π.χ. θόρυβος στον πειραματικό προσδιορισμό των τιμών εισαγωγής, ή λανθασμένες ταξινομήσεις). Σε αυτήν την περίπτωση, η τέλεια μάθηση των στοιχείων κατάρτισης μπορεί να καταστήσει την γενίκευση αδύνατη. Η εκπαίδευση σταματά αφού το δίκτυο έχει εκτεθεί αρκετές φορές στο σύνολο εκπαίδευσης ώστε να αποδίδει ικανοποιητικά. Υπάρχει μια σημαντική σχέση μεταξύ της εκμάθησης και της ικανότητας γενίκευσης του δικτύου. Κάθε πλήρες πέρασμα του συνόλου εκπαίδευσης από το δίκτυο ονομάζεται **εποχή**.

Η μάθηση σε υπολογιστή είναι πεδίο της τεχνητής νοημοσύνης που έχει να κάνει με το σχεδίαση συστημάτων αυτοματοποίησης, ικανών να μάθουν μέσω παραδειγμάτων. Η μάθηση μέσω παραδειγμάτων είναι η διαδικασία που επιτρέπει τη λογική κατηγοριοποίηση σε κλάσεις ανάλογα με τις απαιτήσεις που τίθενται, δηλαδή ένας κανόνας σε κάποια δεδομένη γλώσσα αναπαράστασης. Είναι σημαντικό και άξιο

προσοχής να βρεθεί η χρυσή τομή μεταξύ της πολυπλοκότητας των κανόνων μάθησης και των προτύπων μάθησης, ώστε να αποφύγουν η «υπερεκπαίδευση» (overfitting).

Υπάρχουν πολλά είδη μάθηση μερικά από αυτά είναι :

- **Μάθηση με επίβλεψη** (Μάθηση με δάσκαλο)
Με τύπους δικτύων:
 - ✓ Πολυστρωματικά perceptron
 - ✓ Δίκτυα RBF (radial basis functions)
- **Μάθηση χωρίς επίβλεψη** (Μάθηση χωρίς επιτήρηση)
Με τύπους δικτύων:
 - ✓ Ανταγωνιστικά δίκτυα
 - ✓ Αυτοοργανούμενοι χάρτες Kohonen
 - ✓ Δίκτυα ART (adaptive resonance theory)
 - ✓ LVQ (learning vector quantization)
- **Ανταγωνιστική μάθηση** (competitive learning)
Με τύπους δικτύων:
 - ✓ LVQ
 - ✓ SOM
- **Υποστηρικτική μάθηση** (reinforcement learning)
- **Μάθηση βασισμένη στη μνήμη** (memory-based learning)
- **Μάθηση μέσω διόρθωσης σφάλματος** (error-correction learning)
- **Μάθηση σύμφωνα με το αξίωμα του Hebb** (Hebbian learning)

Αναλυτικότερα θα δούμε την μάθηση με επίβλεψη και την μάθηση χωρίς επίβλεψη σε επόμενο εδάφιο.

Ανταγωνιστική μάθηση

Η ανταγωνιστική μάθηση υλοποιείται με τη χρήση ενός επιπέδου νευρώνων που ονομάζεται ανταγωνιστικό επίπεδο. Κάθε φορά που εμφανίζεται κάποιο διάνυσμα εισόδου στο ανταγωνιστικό επίπεδο τότε με βάση κάποιο κριτήριο βγαίνει ένας νικητής νευρώνας για ένα πρότυπο εκπαίδευσης. Ο νικητής νευρώνας είναι αυτός που μεταβάλλει τις τιμές των βαρών του και εναλλακτικά επιτρέπεται μικρή τροποποίηση και των «χαμένων» νευρώνων. Συνήθως το κριτήριο που χρησιμοποιείται για την εύρεση του νικητή νευρώνα είναι η ευκλείδεια απόσταση του διανύσματος εισόδου $\chi=(\chi_1, \dots, \chi_d)$ από το διάνυσμα $w_i=(w_{i1}, \dots, w_{id})$ κάθε νευρώνα i . Κατά τη διάρκεια της διαδικασίας μεταβάλλεται ο ρυθμός μάθησης (φθίνει). Με τη διαδικασία αυτή έχει βρεθεί ότι επιτυγχάνεται αυτοοργάνωση (self-organization) και οι νευρώνες μαθαίνουν από τα πρότυπα εισόδου. Ωστόσο υπάρχει δίλημμα ευστάθειας (διατήρηση της γνώσης) και πλαστικότητας (αντίδραση στις αλλαγές του περιβάλλοντος). Η διαδικασία επαναλαμβάνεται για κάθε νέο πρότυπο που δίδεται σαν είσοδο στο δίκτυο. Παραδείγματα δικτύων που χρησιμοποιούν τη μέθοδο αυτή είναι το LVQ και το SOFM.

Υποστηρικτική μάθηση

Στη περίπτωση αυτή ΤΝΔ μαθαίνει τι να κάνει ώστε να μεγιστοποιήσει ένα σήμα ανταμοιβής. Αυτό γίνεται χωρίς υποβοήθηση από δάσκαλο αλλά δοκιμάζοντας (trial and error). Οι ενέργειες του ΤΝΔ επηρεάζουν την μελλοντική κατάσταση του περιβάλλοντος και συνεπώς τα μελλοντικά σήματα ανταμοιβής. Η αξία μιας ενέργειας δεν υπολογίζεται μόνο βάσει της ανταμοιβής στην τρέχουσα χρονική στιγμή αλλά και σε σχέση με την ανταμοιβή σε μετέπειτα στιγμές.

Μάθηση βασισμένη στη μνήμη

Στη διαδικασία αυτή όλες οι εμπειρίες αποθηκεύονται στη μνήμη σαν ζεύγη εισόδων και (σωστών) εξόδων (x_i, d_i) . Η απόκριση του δικτύου για ένα πρότυπο απαιτεί τη κατάταξη

του προτύπου ανακαλώντας και αναλύοντας τα δεδομένα εκπαίδευσης σε μια γειτονιά του προτύπου.

Στη μάθηση αυτή εφαρμόζεται:

- ✓ Ο κανόνας κοντινότερου γείτονα:
 - ο Εύρεση του κοντινότερου δείγματος εκπαίδευσης στο πρότυπο
 - ο Κατάταξη στην ίδια κλάση με τον κοντινότερο γείτονα
- ✓ Ο κανόνας k-κοντινότερων γειτόνων:
 - ο Εύρεση των k κοντινότερων δειγμάτων εκπαίδευσης στο xtest
 - ο Κατάταξη στην κλάση στην οποία ανήκει η πλειοψηφία των γειτόνων

Μάθηση μέσω διόρθωσης σφάλματος

Στη διαδικασία αυτή θεωρούμε έναν μοναδικό νευρώνα εξόδου k που την χρονική στιγμή n λαμβάνει είσοδο διάνυσμα $x(n)$ από ένα ή περισσότερα στρώματα κρυμμένων νευρώνων και παράγει έξοδο $y_k(n)$. Δίδεται επίσης:

- ο Επιθυμητή έξοδος $dk(n)$ (γνωστή)
 - ο Σήμα λάθους $ek(n) = dk(n) - y_k(n)$
 - ο Ενεργοποιείται ένας μηχανισμός ελέγχου που μεταβάλλει τα βάρη ώστε βηματικά να μικρύνει η διαφορά $dk(n) - y_k(n)$.
 - ο Ελαχιστοποίηση συνάρτησης κόστους
- $$E(n) = 1/2 e^2(n) \text{ στιγμιαία ενέργεια λάθους}$$

Η μεταβολή των βαρών συνεχίζεται μέχρι το TNA να φτάσει σε κατάσταση ισορροπίας

Μάθηση σύμφωνα με το αξίωμα του Hebb

Ο πιο φημισμένος και παλιότερος κανόνας εκπαίδευσης “The organization of behavior”. Η μάθηση σύμφωνα με το αξίωμα του Hebb βασίζεται στον κανόνα μάθησης του Hebb. Έτσι λοιπόν η διαδικασία αυτή στηρίζεται στα συχνότερα εμφανιζόμενα πρότυπα τα οποία έχουν και τη μεγαλύτερη επίδραση στα συναπτικά βάρη των νευρώνων. Εάν δύο νευρώνες που συνδέονται με μία σύναψη ενεργοποιούνται ταυτόχρονα το βάρος της σύναψης αυτής ενισχύεται (αυξάνεται). Εάν δύο νευρώνες που συνδέονται με μία

σύναψη ενεργοποιούνται ασύγχρονα, το βάρος της σύναψης αυτής εξασθενεί (μειώνεται). Τα δίκτυα που εκπαιδεύονται μπορούν να κάνουν αναλύσεις σε κύριες συνιστώσες.

2.2 ΤΑΞΙΝΟΜΗΣΗ

2.2.1 Η έννοια της ταξινόμησης σε κατηγορίες

Η ταξινόμηση ως ενέργεια εντοπίζεται σε ποικίλους τομείς της ανθρώπινης δραστηριότητας. Σε γενικά πλαίσια, ο όρος θα μπορούσε να συμπεριλάβει οποιοδήποτε περιβάλλον στο οποίο λαμβάνεται μία απόφαση ή γίνεται μία πρόβλεψη με βάση κάποιων πληροφοριών. Στη συνέχεια μία διαδικασία ταξινόμησης χρησιμοποιείται σαν μια μέθοδο με την οποία θα μπορούν να λαμβάνονται επανειλημμένα κάποιες αποφάσεις σε καινούργιες περιπτώσεις. Στόχος της ταξινόμησης είναι να μπορούμε να έχουμε μία ομάδα με κοινά χαρακτηριστικά. Δηλαδή να περνάμε από το ειδικό που είναι το ένα αντικείμενο στο γενικό που είναι μία κατηγορία αντικειμένων.

Η δημιουργία μιας διαδικασίας ταξινόμησης από ένα σύνολο δεδομένων για τα οποία οι πραγματικές κατηγορίες είναι γνωστές, έχει ονομαστεί και ως *Αναγνώριση Προτύπων*, *Διάκριση* ή *Επιβλεπόμενη Μάθηση*.

Ένα παράδειγμα όπου χρησιμοποιείται η ταξινόμηση είναι μηχανικές διαδικασίες για τον διαχωρισμό γραμμάτων με βάση τους ταχυδρομικούς κώδικες (οι οποίοι διαβάζονται από μηχανή). Ένα άλλο παράδειγμα είναι ανάθεση ατόμων σε κάποιες κατηγορίες πίστωσης με βάση οικονομικές και άλλες προσωπικές πληροφορίες. Η προκαταρκτική διάγνωση της αρρώστιας ενός ασθενή, ώστε να επιλεγεί άμεσα μία θεραπεία μέχρι να έχουμε τα τελικά αποτελέσματα από τις εξετάσεις. Μάλιστα, μερικά από τα πιο έντονα προβλήματα που υπάρχουν στο χώρο της επιστήμης, της βιομηχανίας και του εμπορίου μπορούν να θεωρηθούν σαν προβλήματα ταξινόμησης χρησιμοποιώντας πολύπλοκα και μεγάλης έκτασης δεδομένα.

Η ταξινόμηση έχει δύο ξεχωριστές έννοιες. Η μία έννοια υποθέτει ότι μας έχει δοθεί ένα σύνολο από παρατηρήσεις με στόχο την εύρεση κατηγοριών στα δεδομένα. Η

δεύτερη έννοια προϋποθέτει ότι γνωρίζουμε σίγουρα την ύπαρξη κατηγοριών κάποιου πλήθους και ο στόχος μας είναι να δημιουργήσουμε έναν κανόνα με τον οποίο μπορούμε να κατηγοριοποιήσουμε μία νέα παρατήρηση σε μία από τις υπάρχουσες κατηγορίες. Ο πρώτος τρόπος ονομάζεται *Εκμάθηση Χωρίς Επίβλεψη ή Ομαδοποίηση* ή και ο δεύτερος *Εκμάθηση Με Επίβλεψη*.

Η ύπαρξη σωστά κατηγοριοποιημένων δεδομένων προϋποθέτει ότι κάποιος (**ο Επιβλέπων**) είναι ικανός να κατηγοριοποιεί χωρίς λάθη. Με τη παρατήρηση αυτή όμως, δημιουργείται αυτόματα η απορία κατά πόσο είναι αναγκαίο να αντικαταστήσουμε την ακριβή κατηγοριοποίηση με κάποια διαδικασία προσέγγισης .

2.2.2 Χρησιμότητα μιας διαδικασίας ταξινόμησης

Υπάρχουν πολλοί λόγοι για τους οποίους θα θέλαμε να εκτελέσουμε μία διαδικασία ταξινόμησης. Μερικοί πιθανοί λόγοι είναι οι παρακάτω :

1. Οι διαδικασίες μηχανικής ταξινόμησης μπορεί να είναι πολύ πιο γρήγορες : για παράδειγμα, οι μηχανές αναγνώρισης ταχυδρομικών κωδικών είναι ικανές να ταξινομούν την πλειονότητα των γραμμάτων, αφήνοντας τις δύσκολες περιπτώσεις στον ανθρώπινο παράγοντα.
2. Μία εταιρεία ταχυδρομικών παραγγελιών πρέπει να πάρει μια απόφαση για την παραχώρηση πίστωσης στηριζόμενη καθαρά στις πληροφορίες που υπάρχουν στην φόρμα αίτησης : στην συγκεκριμένη περίπτωση, ο ανθρώπινος παράγων μπορεί να πάρει αποφάσεις με λάθος κριτήρια και μπορεί να απορρίψει καλούς πελάτες.
3. Στον ιατρικό χώρο, μπορεί να θέλουμε να αποφύγουμε την εγχείριση η οποία θα ήταν ο μόνος σίγουρος τρόπος για μία σωστή διάγνωση, οπότε και ρωτάμε αν μία αξιόπιστη διάγνωση μπορεί να γίνει βασιζόμενοι αποκλειστικά και μόνο σε εξωτερικά συμπτώματα.
4. Ο Επιβλέπων, που αναφέραμε παραπάνω, μπορεί να λαμβάνει μια απόφαση από ιστορικά στοιχεία, όπως συμβαίνει στην μετεωρολογία, στις συναλλαγές μετοχών και στις επενδύσεις ή στις αποφάσεις για παραχώρηση δανείων. Σε αυτή την περίπτωση το θέμα είναι η πρόβλεψη.

2.2.3 Βασικά σημεία μιας διαδικασίας ταξινόμησης

Κάποια από τα θέματα που θα πρέπει να μας απασχολήσουν όταν δημιουργούμε έναν ταξινομητή / κατηγοριοποιητή είναι και τα παρακάτω :

- **Ακρίβεια.** Υπάρχει η αξιοπιστία του κανόνα, που συχνά απεικονίζεται από την αναλογία των σωστών ταξινομήσεων. Παρόλο που κάποια λάθη είναι πιο σοβαρά από άλλα και μπορεί να είναι σημαντικό να ελέγχεται το ποσοστό λάθους για κάποιες κατηγορίες κλειδιά.
- **Ταχύτητα.** Σε κάποιες περιπτώσεις, η ταχύτητα του ταξινομητή είναι ένα πρωταρχικό ζήτημα. Ένας ταξινομητής που είναι 90% ακριβής μπορεί να προτιμηθεί από έναν ο οποίος είναι 95% ακριβής αν είναι 100 φορές πιο γρήγορος στις δοκιμές (και τέτοιες διαφορές στα χρονοδιαγράμματα δεν είναι ασυνήθιστες στα νευρωνικά δίκτυα). Τέτοιες εκτιμήσεις θα ήταν σημαντικές στην αναγνώριση ταχυδρομικών κωδικών ή στην αυτόματη ανίχνευση ελαττωμάτων σε τεμάχια μιας γραμμής παραγωγής για παράδειγμα.
- **Εύκολα Κατανοήσιμο.** Αν ένας ανθρώπινος παράγων θα πρέπει να εφαρμόσει την διαδικασία ταξινόμησης, η διαδικασία πρέπει να είναι εύκολα κατανοητή αλλιώς θα γίνουν λάθη στην εφαρμογή του κώδικα. Είναι επίσης σημαντικό, ο ανθρώπινος χειριστής να πιστεύει ότι το σύστημα είναι αξιόπιστο. Ένα συχνά-αναφερόμενο παράδειγμα είναι η περίπτωση του “Three Mile Island”, όπου οι αυτόματες μηχανές σωστά συνέστησαν την διακοπή λειτουργίας, αλλά οι χειριστές δεν προχώρησαν στην ενέργεια αυτή γιατί δεν την θεώρησαν σωστή. Ένα παρόμοιο γεγονός συνέβη και στην καταστροφή του Chernobyl.
- **Χρόνο για μάθηση.** Ιδίως σε ένα ταχύτατα αναπτυσσόμενο περιβάλλον, μπορεί να είναι αναγκαίο το σύστημα να μάθει γρήγορα έναν κανόνα ταξινόμησης ή να κάνει κάποιες προσαρμογές σε έναν ήδη υπάρχον κανόνα σε πραγματικό χρόνο. Η έννοια “γρήγορα” μπορεί να υπονοεί επίσης ότι χρειαζόμαστε μόνο ένα μικρό αριθμό από παρατηρήσεις για την καθιέρωση του κανόνα.

Ας πάρουμε για παράδειγμα τον αφελή κανόνα του 1-κοντινότερου γείτονα, όπου στο σύνολο εκπαίδευσης ψάχνουμε για τον πιο “κοντινό” (με συγκεκριμένη έννοια κοντινό)

προηγούμενο παράδειγμα, του οποίου η κατηγορία μετά συμπεραίνεται για τη νέα περίπτωση. Αυτό μαθαίνεται πάρα πολύ γρήγορα, αλλά είναι πολύ αργό στην περίπτωση που χρησιμοποιούνται όλα τα δεδομένα (βέβαια αν έχουμε υπολογιστή με παράλληλους επεξεργαστές μπορούμε να αυξήσουμε αρκετά την ταχύτητα). Σε άλλες περιπτώσεις, είναι χρήσιμο να έχουμε μία γρήγορη μέθοδο για τον έλεγχο διαφόρων δεδομένων ή για μία γρήγορη επαλήθευση στα αποτελέσματα μιας διαδικασίας. Για παράδειγμα, ο διευθυντής μιας τράπεζας έχει θέσει ως κανόνα “επιτρέπεται η παραχώρηση πίστωσης μόνο σε υποψήφιους που έχουν τραπεζικό λογαριασμό”. Αν παρατηρηθεί ότι μία νέα υπάλληλος (ή η νέα αυτοματοποιημένη διαδικασία) παραχωρεί πίστωση σε πελάτες που **δεν** έχουν τραπεζικό λογαριασμό, προφανώς θα ήθελε να ελέγχει ότι η νέα υπάλληλος (ή η νέα διαδικασία) κάνει σωστά την δουλειά της.

2.2.4 Η έννοια των κατηγοριών στην ταξινόμηση

Ένα σημαντικό ερώτημα, το οποίο δεν έχει κατανοηθεί σωστά σε πολλές μελέτες ταξινόμησης, είναι η φύση των κατηγοριών και τον τρόπο με τον οποίο ορίζονται. Μπορούμε να διαχωρίσουμε τρεις κοινές περιπτώσεις, με την πρώτη μόνο να οδηγείται σε αυτό που οι στατιστικολόγοι θα ονόμαζαν ταξινόμηση :

1. Οι κατηγορίες αντιστοιχούν σε ετικέτες για διαφορετικούς πληθυσμούς : εδώ δεν μας απασχολεί η ιδιότητα μέλους σε διαφορετικούς πληθυσμούς. Για παράδειγμα, οι σκύλοι και οι γάτες αποτελούν δύο τελείως διαφορετικές κατηγορίες/ πληθυσμούς, και μπορούμε να πούμε με βεβαιότητα αν ένα ζώο είναι σκύλος ή γάτα (ή κανένα από τα δύο). Η ιδιότητα μέλους μιας κατηγορίας ή ενός πληθυσμού καθορίζεται από μία ανεξάρτητη αρχή (τον Επιβλέπων), η κατανομή σε μία κλάση καθορίζεται ανεξάρτητα από οποιαδήποτε μεταβλητή ή ιδιότητα.
2. Οι κατηγορίες είναι απόρροια ενός προβλήματος πρόβλεψης. Εδώ η κατηγορία είναι ουσιαστικά ένα αποτέλεσμα το οποίο πρέπει να προβλεφθεί από την γνώση των ιδιοτήτων. Σε στατιστικούς όρους, η κατηγορία είναι μία τυχαία μεταβλητή. Ένα τυπικό παράδειγμα είναι η πρόβλεψη των επιτοκίων. Συχνά, υπάρχει το εξής ερώτημα: αν τα επιτόκια θα αυξηθούν (κατηγορία=1) ή όχι (κατηγορία=0).

3. Οι κατηγορίες προκαθορίζονται ανάλογα με τις διαφορετικές ιδιότητες των δειγμάτων και με αυτόν τον τρόπο διαιρούμε την επιφάνεια δείγματος. Μπορούμε να πούμε ότι η κατηγορία είναι μια συνάρτηση των ιδιοτήτων. Γι' αυτό το λόγο, ένα προϊόν μπορεί να θεωρηθεί ως ελαττωματικό (να μπει στην κατηγορία των ελαττωματικών) αν μερικές ιδιότητες/χαρακτηριστικά είναι εκτός των προκαθορισμένων ορίων, και αντίστοιχα να θεωρηθεί όχι ελαττωματικό σε αντίθετη περίπτωση. Υπάρχει ένας κανόνας ο οποίος έχει ήδη κατηγοριοποιήσει τα δεδομένα στηριζόμενος στα χαρακτηριστικά : το πρόβλημα είναι η δημιουργία ενός κανόνα ο οποίος μιμείται τον πραγματικό κανόνα όσο πιο πιστά γίνεται. Πολλά σύνολα δεδομένων που χρησιμοποιούνται για εκτίμηση είναι αυτού του τύπου.

Στην πράξη, τα σύνολα δεδομένων μπορεί να είναι συνδυασμός αυτών των τύπων ή να είναι κάτι στο ενδιάμεσο.

2.2.5 Η έννοια της ακρίβειας

Όταν μελετάμε την έννοια της ακρίβειας, πρέπει να έχουμε πάντα υπόψη μας ότι η ακρίβεια που μετράμε στο σύνολο εκπαίδευσης και η ακρίβεια που μετράμε στα νέα δεδομένα (σύνολο δοκιμής) είναι συχνά πολύ διαφορετικά. Είναι συχνό φαινόμενο, ιδίως στις εφαρμογές Εκμάθησης Μηχανής, το σύνολο εκπαίδευσης να είναι τέλεια προσαρμοσμένο, αλλά η απόδοση στο σύνολο δοκιμής να είναι πολύ απογοητευτικό. Η ακρίβεια στα νέα δεδομένα, όταν η πραγματική κατηγοριοποίηση είναι άγνωστη, είναι αυτή που μας ενδιαφέρει και είναι πρακτικής σημασίας. Η κοινώς αποδεκτή μέθοδος για να το εκτιμήσουμε αυτό, χρησιμοποιώντας τα δεδομένα που μας δίνονται, στα οποία υποθέτουμε ότι όλες οι κατηγορίες μέλη είναι γνωστές, είναι η παρακάτω :

- Αρχικά, χρησιμοποιούμε μια σημαντική ποσότητα (το σύνολο εκπαίδευσης) από τα δεδομένα που μας δίνονται για να εκπαιδεύσουμε την διαδικασία. Αυτός ο κανόνας μετά δοκιμάζεται στα υπόλοιπα δεδομένα (σύνολο δοκιμής) και τα αποτελέσματα συγκρίνονται με τις γνωστές κατηγοριοποιήσεις. Η αναλογία των σωστών (κατηγοριοποιήσεων) στο σύνολο δοκιμής είναι μία μη πολωμένη εκτίμηση της

ακρίβειας του παρακάτω κανόνα : ότι το σύνολο εκπαίδευσης δειγματοληπτείται τυχαία από τα δοθέντα δεδομένα.

2.2.6 Παραδείγματα Ταξινομητών

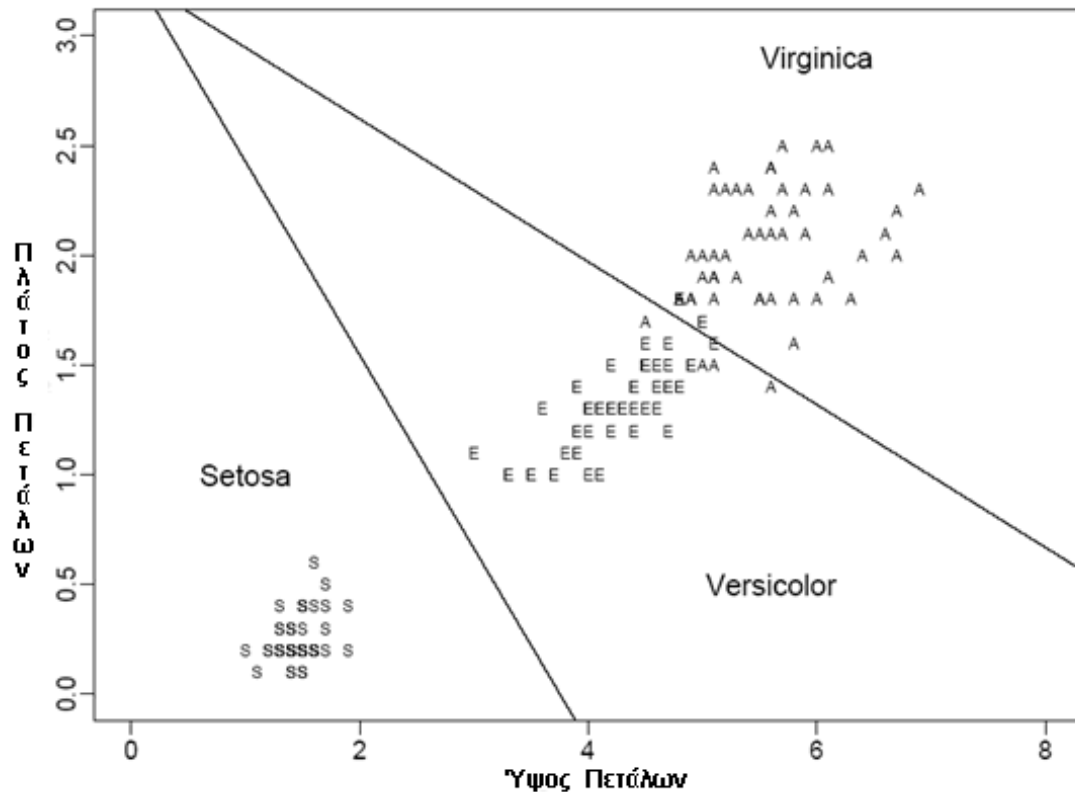
Για να παρουσιάσουμε τους βασικούς τύπους ταξινομητών, θα χρησιμοποιήσουμε το πολύ γνωστό σύνολο δεδομένων Ίριδα (Iris), το οποίο μας έχει δοθεί πλήρες από τους Kendall και Stuart (1983) [6]. Υπάρχουν τριών ειδών ποικιλίες Ίριδας : η Σετόζα, η Τετράχρωμη Ίριδα και η ποικιλία η οποία ευδοκιμεί στην Βιρτζίνια. Το μήκος και το πλάτος και του πετάλου και του σεφάλου μετρήθηκαν σε πενήντα λουλούδια από κάθε ποικιλία. Το αρχικό πρόβλημα είναι να κατηγοριοποιήσουμε ένα νέο λουλούδι Ίριδα σε μία από τις τρεις κατηγορίες με βάση τα τέσσερα αυτά χαρακτηριστικά (το μήκος και το πλάτος του πετάλου και του σέφαλου). Για να κρατήσουμε όμως απλό αυτό το παράδειγμα, θα αναζητήσουμε έναν κανόνα κατηγοριοποίησης με βάση τον οποίο οι ποικιλίες μπορούν να ξεχωρίσουν ανάλογα με τις δύο από τις τέσσερις μετρήσεις, δηλαδή **το Μήκος και το Πλάτος του Πετάλου**. Έχουμε στην διάθεσή μας πενήντα ζευγάρια μετρήσεων από κάθε ποικιλία από τα οποία μπορούμε να μάθουμε τον κανόνα κατηγοριοποίησης στο σύστημα.

1. Γραμμικός Διαχωριστής του Fisher

Αυτή είναι μία από τις παλαιότερες διαδικασίες ταξινόμησης, και εφαρμόζεται πιο συχνά σε λογισμικά υπολογιστών. Η ιδέα είναι να διαιρέσουμε τον χώρο δειγματοληψίας με μία σειρά γραμμών σε δύο διαστάσεις, τα επίπεδα σε τρεις διαστάσεις και γενικά τα υπερεπίπεδα σε πολλές διαστάσεις. Η γραμμή που διαιρεί δύο κατηγορίες σχεδιάζεται για να διχοτομεί την γραμμή που ενώνει τα κέντρα αυτών των κατηγοριών. Η κατεύθυνση αυτής της γραμμής καθορίζεται από την μορφή των ομάδων των σημείων. Για παράδειγμα, για να διαφοροποιήσουμε το Versicolor από το Virginica, εφαρμόζουμε τον παρακάτω κανόνα :

- Αν Πλάτος Πετάλου $< 3.272 - 0.3254 \times \text{Μήκος Πετάλου}$, τότε Versicolor.
- Αν Πλάτος Πετάλου $> 3.272 - 0.3254 \times \text{Μήκος Πετάλου}$, τότε Virginica.

Οι γραμμικοί διαχωριστές Fisher που εφαρμόζονται στα δεδομένα Ίριδας φαίνονται στο Σχ.2-1. Έξι από τις παρατηρήσεις θα ταξινομηθούν λάθος.



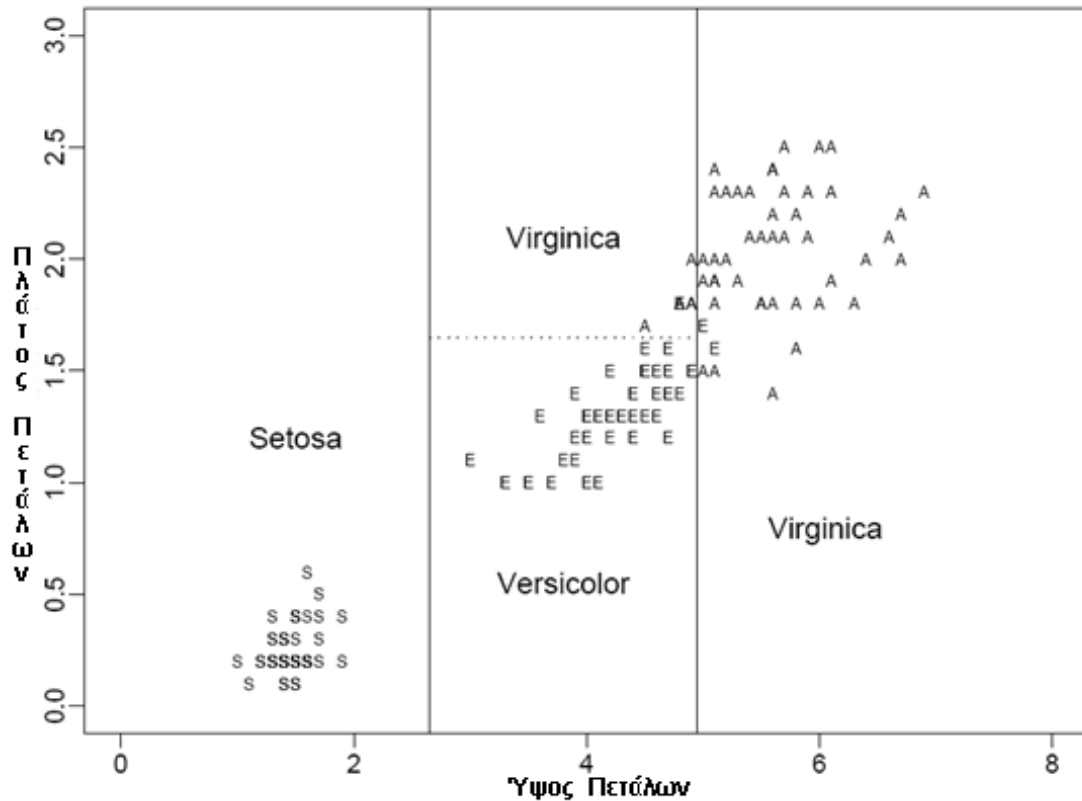
Σχ. 2-1: Ταξινόμηση από γραμμικούς διαχωριστές : Iris δεδομένα.

2. Δένδρα αποφάσεων και Μέθοδοι βασισμένοι σε Κανόνες

Μία κατηγορία από τις διαδικασίες κατηγοριοποίησης βασίζεται σε περιοδικά επαναλαμβανόμενη διαίρεση του χώρου δειγματοληψίας. Ο χώρος χωρίζεται σε κουτιά, και σε κάθε στάδιο της διαδικασίας, κάθε κουτί εξετάζεται για να δούμε αν μπορεί να χωριστεί σε δύο κουτιά. Η διαίρεση συνήθως γίνεται παράλληλα στους συντεταγμένους άξονες. Ένα παράδειγμα για τα δεδομένα Iris ακολουθεί.

- Αν Μήκος Πετάλου < 2.65 , τότε Setosa.
- Αν Μήκος Πετάλου > 4.95 , τότε Virginica.
- Αν $2.65 < \text{Μήκος Πετάλου} < 4.95$, τότε :
 - Αν Μήκος Πετάλου < 1.65 τότε Versicolor
 - Αν Μήκος Πετάλου > 1.65 τότε Virginica.

Το αποτέλεσμα της διαίρεσης φαίνεται στο Σχ.2-2. Παρατηρούμε ότι αυτός ο κανόνας ταξινόμησης έχει τρεις λανθασμένες ταξινομήσεις.



Σχήμα 2-2 : Ταξινόμηση με δένδρα αποφάσεων : Δεδομένα Ίριδας

Οι Weiss και Karouleas (1989) [6] δίνουν έναν εναλλακτικό κανόνα ταξινόμησης για τα δεδομένα Ίριδας που σχετίζονται άμεσα με το Σχ. 2-2. Ο δικός τους κανόνας είναι το Σχ.2-2 αν συνεχίσουμε την διακεκομμένη γραμμή στα αριστερά και μπορεί συνεπώς να καθορισθεί ως εξής :

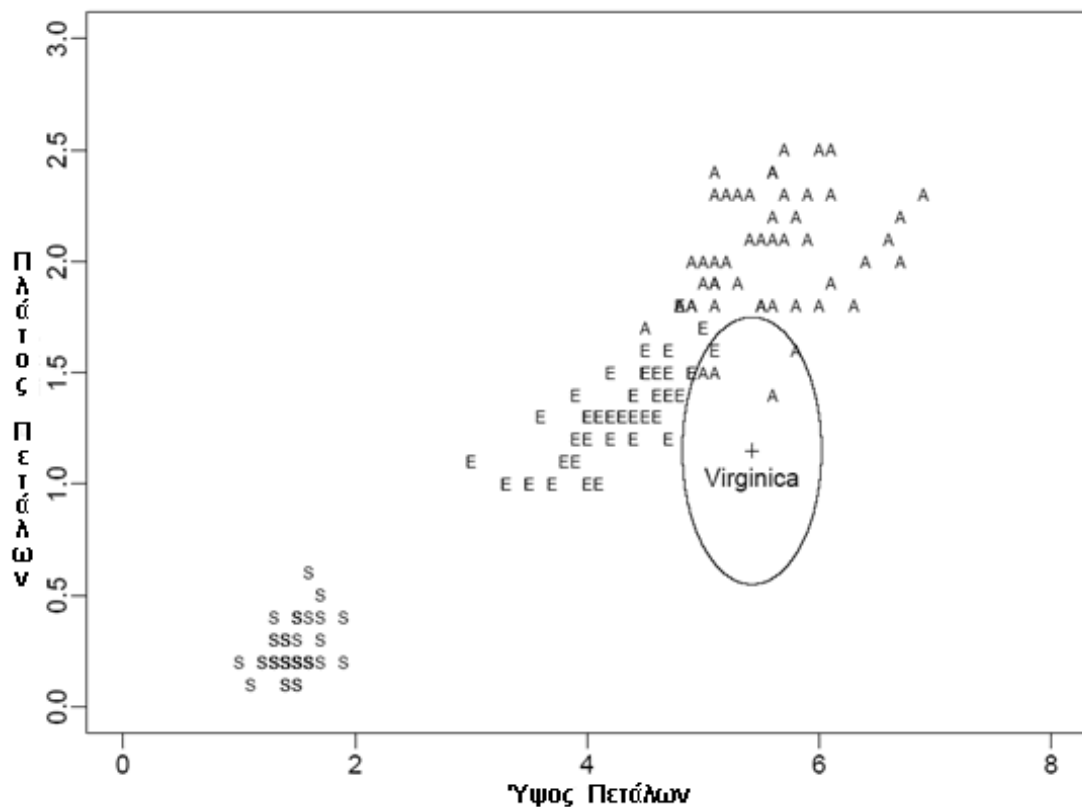
- Αν Μήκος Πετάλου < 2.65 , τότε Setosa.
- Αν Μήκος Πετάλου > 4.95 , ή Πλάτος Πετάλου > 1.65 , τότε Virginica.
- Σε κάθε άλλη περίπτωση Versicolor.

Παρατηρούμε ότι αυτός ο κανόνας, παρόλο που είναι ισοδύναμος με τον κανόνα του Σχ. 2-2, είναι διατυπωμένος πιο περιληπτικά, και συνεπώς μπορεί να προτιμηθεί γι' αυτό το λόγο. Επίσης βλέπουμε ότι αυτός ο κανόνας είναι διφορούμενος αν Μήκος Πετάλου < 2.65 και Πλάτος Πετάλου > 1.65 . Οι κανόνες που παραθέτουμε εδώ μπορεί να ξεκαθαριστούν, αν τους εφαρμόσουμε με την σειρά που μας δίνονται, και τότε είναι απλά μία επαναδιατύπωση του προηγούμενου δένδρου αποφάσεων. Ο κανόνας που αναφέρουμε σ' αυτό το σημείο αποτελεί ένα στιγμιότυπο μιας μεθόδου βασισμένης σε κανόνες. Τέτοιες μέθοδοι έχουν πολύ στενούς δεσμούς με τα δένδρα αποφάσεων.

3. κ-Κοντινότεροι Γείτονες

Επεξηγούμε αυτή την τεχνική στα δεδομένα Ίριδας. Υποθέτουμε ότι θέλουμε να κατηγοριοποιήσουμε μία νέα ποικιλία Ίριδας. Η κεντρική ιδέα είναι ότι το πιο πιθανό είναι ότι θα βρίσκεται κοντά σε παρατηρήσεις του δικού της πληθυσμού. Συνεπώς παρατηρούμε τις πέντε (για παράδειγμα) κοντινότερες παρατηρήσεις από όλες τις προηγούμενες Ίριδες που καταγράψαμε, και ταξινομούμε την παρατήρηση σύμφωνα με την πιο συχνή κατηγορία μεταξύ των γειτόνων της. Στο Σχ.2-3, η νέα παρατήρηση σημειώνεται με a^+ , και οι πέντε κοντινότερες παρατηρήσεις υπάρχουν μέσα στον **κύκλο** συγκεντρωμένα στο $+$. Το φαινομενικά ελλειπτικό σχήμα οφείλεται στις διαφοροποιημένες οριζόντιες και κάθετες κλίμακες. Η σωστή κλιμάκωση των παρατηρήσεων αποτελεί μία σημαντική δυσκολία αυτής της μεθόδου.

Αυτό φαίνεται στο Σχ.2-3, όπου μια παρατήρηση εστιασμένη στο $+$ θα ταξινομηθεί ως *Virginica* μιας και έχει 4 *Virginica* μεταξύ των πέντε κοντινότερων γειτόνων της.



Σχ.2-3 : Ταξινόμηση με βάση των 5-Κοντινότερων-Γειτόνων : Δεδομένα Ίριδας

2.2.7 Μία γενική δομή των προβλημάτων ταξινόμησης

Υπάρχουν τρία βασικά μέρη σε ένα πρόβλημα ταξινόμησης.

1. Η σχετική συχνότητα με την οποία οι κατηγορίες συναντώνται στον πληθυσμό που μας ενδιαφέρει. Επίσης, αυτό μπορεί να εκφραστεί και ως προηγούμενη πιθανότητα διανομής.
2. Ένα κριτήριο, το οποίο δηλώνεται ρητά εξ αρχής ή υπονοείται, για τον διαχωρισμό των κατηγοριών. Μπορούμε να θεωρήσουμε μία βασική σχέση εισόδου/ εξόδου που χρησιμοποιεί ιδιότητες για να διαχωρίσει μία τυχαία εγγραφή/ αντικείμενο από κάθε κατηγορία.
3. Το κόστος όταν κάνουμε μία λάθος ταξινόμηση.

Οι περισσότερες τεχνικές αναμφίβολα συνδυάζουν αυτά τα μέρη και, για παράδειγμα, παράγουν έναν κανόνα ταξινόμησης που προέρχεται από μία συγκεκριμένη προηγούμενη διανομή και δεν μπορεί πολύ εύκολα να υιοθετηθεί σε μία αλλαγή στην συχνότητα κατηγορίας. Ωστόσο, στην θεωρία καθένα από αυτά τα μέρη μπορεί να μελετηθεί ατομικά και μετά τα αποτελέσματα να συνδυαστούν και να παραχθεί ένας κανόνας ταξινόμησης.

2.3 ΓΕΝΙΚΕΥΣΗ

2.3.1 Εισαγωγή στην έννοια της γενίκευσης

Η ικανότητα μάθησης και γενίκευσης είναι θεμελιώδης σε κάθε μηχανή μάθησης. Συγκεκριμένα, η γενίκευση είναι ένα από τα πιο σημαντικά θέματα σε οποιοδήποτε κατηγοριοποιητή που είναι εκπαιδευμένος μη παραμετρικά (π.χ. νευρωνικά δίκτυα). Τα Τεχνητά Νευρωνικά Δίκτυα χρησιμοποιούν επαγωγική μάθηση για να βρουν γενικές έννοιες από τα συγκεκριμένα παραδείγματα που τους εμφανίζονται. Η διαδικασία μάθησης έχει ως στόχο την εξαγωγή κανόνων στους οποίους στηρίζεται το πρόβλημα. Η αύξηση της ακρίβειας μάθησης (και αντίστοιχα η μείωση του λάθους γενίκευσης) ενός κατηγοριοποιητή είναι ακόμη ένα ενεργό πεδίο έρευνας.

Η τελική κατάσταση μιας οποιαδήποτε μηχανής μάθησης, συμπεριλαμβανομένων και των νευρωνικών δικτύων, αντιπροσωπεύει την έννοια που προέρχεται από το σύνολο εκπαίδευσης. Η γενίκευση (δηλαδή να μπορούμε να προβλέπουμε και εκτός του συνόλου εκπαίδευσης) μπορεί να επιτευχθεί μέσα από συγκεκριμένες μόνο περιπτώσεις εκμάθησης όπως και μέσα από εξαντλητική εκμάθηση. Με άλλα λόγια, το μέγεθος του συνόλου εκπαίδευσης και η επιλογή της συνάρτησης διανομής, με βάση την οποία τα δεδομένα εκπαίδευσης έχουν επιλεγεί, επηρεάζει απευθείας την απόδοση οποιασδήποτε μηχανής εκμάθησης.

Τα τελευταία χρόνια, διάφορες μέθοδοι έχουν αναπτυχθεί για την βελτίωση της ακρίβειας γενίκευσης. Αυτές οι μέθοδοι ανήκουν σε δύο βασικές κατηγορίες : **α)** σε μεθόδους μείωσης της μεταβλητότητας και **β)** σε μεθόδους μείωσης της πόλωσης. Οι μέθοδοι της δεύτερης κατηγορίας κατηγοριοποιούνται ως εξής :

- Μοντέλο ενός επιπέδου εκμάθησης : σε αυτό το μοντέλο ο στόχος είναι η επιλογή του πιο κατάλληλου αλγορίθμου (μοντέλου) για μία δοθείσα αποστολή.
- Μοντέλο πολλαπλών επιπέδων εκμάθησης : αυτός ο τομέας έρευνας επικεντρώνεται στο να συνδυάζει προβλέψεις από πολλά και διαφορετικά μοντέλα (αλγόριθμοι) με σκοπό να επιλέγει την πιο κατάλληλη πόλωση για την ελαχιστοποίηση του λάθους γενίκευσης.

Η έννοια της γενίκευσης συμπυκνώνεται στον εξής ορισμό : η ικανότητα παραγωγής επιθυμητών εξόδων, για εισόδους που δε συμπεριλαμβάνονταν στα δεδομένα εκπαίδευσης.

2.3.2 Προϋποθέσεις γενίκευσης

Κατά την διάρκεια της εκμάθησης, οι εξοδοί ενός νευρωνικού δικτύου με επιβλεπόμενη μάθηση τείνουν να πλησιάζουν τις επιθυμητές τιμές με βάση τις τιμές εισόδου στο σύνολο εκπαίδευσης. Αυτή η ιδιότητα μπορεί να είναι χρήσιμη από μόνη της, αλλά πιο συχνά ο σκοπός που χρησιμοποιούμε ένα νευρωνικό δίκτυο είναι η γενίκευση – π.χ. να έχουμε τις εξόδους του δικτύου να πλησιάζουν τις τιμές- στόχους όταν μας δίνονται τιμές εισόδου οι οποίες δεν υπάρχουν μέσα στο σύνολο εκπαίδευσης. Η γενίκευση δεν είναι πάντα δυνατή παρά το γεγονός ότι πολλοί συγγραφείς αυτό υποστηρίζουν. Για παράδειγμα, οι Caudill και Butler (1990) υποστηρίζουν ότι “Ένα νευρωνικό δίκτυο είναι ικανό να γενικεύει”, αλλά δεν παρέχουν καμία δικαιολόγηση γι αυτή την υποστήριξή τους και αγνοούν εντελώς τα περίπλοκα θέματα που εμπεριέχονται για να πετύχουμε μία καλή γενίκευση [8].

Η γενίκευση απαιτεί προηγούμενη γνώση, όπως τόνισε ο Hume (1739/1978), ο Russell (1948), και ο Goodman (1954/1983) και με πάθος απέδειξε ο Wolpert (1995a, 1996a, 1996b) .[8] Για οποιαδήποτε πρακτική εφαρμογή, πρέπει να γνωρίζουμε ποιές είναι οι σχετικές εισοδοί (δεν μπορούμε έτσι απλά να συμπεριλάβουμε οποιαδήποτε φανταστική είσοδο). Πρέπει να γνωρίζουμε μία περιορισμένη κατηγορία συναρτήσεων

εισόδων-εξόδων που εμπεριέχουν μία επαρκή κλίση προς την συνάρτηση που θέλουμε να μάθει το δίκτυο (δεν μπορούμε να χρησιμοποιήσουμε μία μέθοδο εκμάθησης που είναι ικανή να προσαρμόζει οποιαδήποτε συνάρτηση μπορούμε να φανταστούμε) . Πρέπει επίσης να λαμβάνουμε υπόψιν μας ότι οι περιπτώσεις που θέλουμε να γενικεύσουμε, πρέπει να έχουν σε κάποια ομοιότητα με τις περιπτώσεις εκπαίδευσης.

Παρόλαυτα, υπάρχουν τρεις προϋποθέσεις που είναι απαραίτητες τυπικά - όχι όμως και αρκετές - για να πετύχουμε μία καλή γενίκευση :

1. Η πρώτη απαραίτητη προϋπόθεση είναι οι εισοδοί στο δίκτυο να περιέχουν επαρκή πληροφορία που σχετίζεται με το στόχο, έτσι ώστε να υπάρχει μία μαθηματική συνάρτηση που συσχετίζει σωστές εξόδους με εισόδους με το επιθυμητό βαθμό ακριβείας. Δεν μπορούμε να αναμένουμε από ένα δίκτυο να μάθει μία συνάρτηση που δεν υπάρχει - τα νευρωνικά δίκτυα δεν έχουν μαντικές ικανότητες! Για παράδειγμα, αν θέλουμε να προβλέψουμε την τιμή μιας μετοχής, ένα ιστορικό με τιμές της συγκεκριμένης μετοχής είναι σπάνια αρκετό σαν είσοδο. Χρειαζόμαστε τόσο λεπτομερή πληροφόρηση για την οικονομική κατάσταση της εταιρείας όσο και πληροφορίες για τις γενικές οικονομικές συνθήκες που επικρατούν, και για να αποφύγουμε άσχημες εκπλήξεις, πρέπει επίσης να συμπεριλάβουμε εισόδους που μπορούν με ακρίβεια να προβλέψουν πολέμους στην Μέση Ανατολή και σεισμούς στην Ιαπωνία. Το να βρούμε σωστές εισόδους για ένα δίκτυο και να συλλέξουμε αρκετά δεδομένα για εκπαίδευση είναι συνήθως πιο χρονοβόρο από την ίδια την διαδικασία εκπαίδευσης του δικτύου.
2. Η δεύτερη απαραίτητη προϋπόθεση είναι ότι η συνάρτηση που προσπαθούμε να μάθουμε (η οποία συσχετίζει εισόδους με σωστές εξόδους) πρέπει να είναι ομαλή. Με άλλα λόγια, μία μικρή αλλαγή στις εισόδους, πρέπει, τις περισσότερες φορές, να προκαλέσει μικρές αλλαγές στις εξόδους. Για συνεχείς εισόδους και στόχους, η ομαλότητα της συνάρτησης υποδηλώνει συνέχεια και περιορισμούς στο μεγαλύτερο μέρος της επιφάνειας εισόδου. Μερικά νευρωνικά δίκτυα μπορούν να μάθουν τις ασυνέχειες όσο η συνάρτηση περιέχει έναν περιορισμένο αριθμό συνεχόμενων μερών. Συναρτήσεις οι οποίες δεν είναι ομαλές, όπως αυτές που παράγονται με ψευδοτυχαίους αριθμούς και αλγορίθμους κρυπτογράφησης, δεν μπορούν να γενικευθούν από νευρωνικά δίκτυα. Συχνά μία μη-γραμμική μεταβολή της

επιφάνειας εισόδου μπορεί να αυξήσει την ομαλότητα της συνάρτησης και να βελτιώσει την γενίκευση.

Για την κατηγοριοποίηση, αν δεν χρειάζεται να εκτιμήσουμε μεταγενέστερες πιθανότητες, τότε η ομαλότητα δεν είναι θεωρητικά απαραίτητη. Συγκεκριμένα, τα δίκτυα ευθείας τροφοδότησης ενός επιπέδου που έχουν εκπαιδευτεί για την μείωση του ποσοστού λάθους είναι καθολικά συνεπείς κατηγοριοποιητές αν ο αριθμός των κρυμμένων μερών αναπτύσσεται σε ένα ποσοστό σύμφωνο με τον αριθμό των περιπτώσεις εκπαίδευσης (Devroye, Györfi και Lugosi, 1996) [8]. Παρόλαυτα, το πιθανότερο είναι να πάρουμε καλύτερη γενίκευση με ρεαλιστικά μεγέθη δειγμάτων αν τα όρια της γενίκευσης είναι πιο ομαλά.

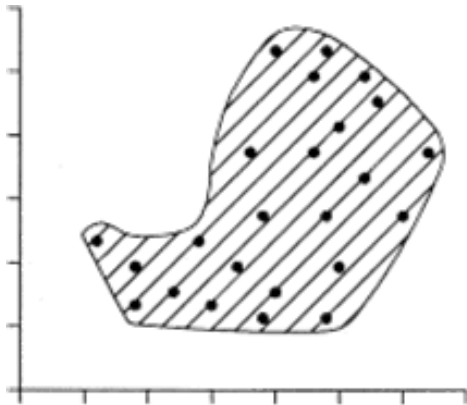
Για λογικές συναρτήσεις, η έννοια της ομαλότητας υπονοείται και δεν είναι απόλυτα ξεκάθαρη. Φαίνεται διαισθητικά καθαρά ότι ένα λογικό δίκτυο με έναν μικρό αριθμό κρυμμένων νευρώνων και μικρά βάρη θα υπολογίσει μία πιο ομαλή συνάρτηση εισόδου-εξόδου από ότι ένα δίκτυο με πολλούς κρυμμένους νευρώνες και μεγάλα βάρη.

3. Η τρίτη απαραίτητη προϋπόθεση για καλή γενίκευση είναι οι περιπτώσεις εκπαίδευσης να αποτελούν ένα επαρκώς μεγάλο και αντιπροσωπευτικό υποσύνολο (“δείγμα” σε ορολογία στατιστικής) του συνόλου όλων των περιπτώσεων στα οποία θέλουμε να γενικεύσουμε (ο “πληθυσμός” σε ορολογία στατιστικής). Η σημασία αυτής της υπόθεσης σχετίζεται με το γεγονός ότι υπάρχουν, εν γένει, δύο διαφορετικά είδη γενίκευσης: η γενίκευση με παρεμβολή δεδομένων και η γενίκευση με εξαγωγή συμπερασμάτων από ιστορικά δεδομένα. Η παρεμβολή δεδομένων εφαρμόζεται σε περιπτώσεις που βρίσκονται λίγο έως πολύ κοντά σε περιπτώσεις εκπαίδευσης, ενώ οι περιπτώσεις που είναι εκτός των ορίων των δεδομένων εκπαίδευσης απαιτούν το δεύτερο είδος γενίκευσης. Οι περιπτώσεις που είναι μέσα σε μεγάλα “κενά” στα δεδομένα εκπαίδευσης μπορεί επίσης να απαιτούν συμπερασματική γενίκευση. Η μέθοδος παρεμβολής μπορεί συχνά να γίνει αξιόπιστη, αλλά το συμπερασματική γενίκευση είναι γνωστό για την αναξιопιστία του. Συνεπώς είναι σημαντικό να έχουμε επαρκή δεδομένα εκπαίδευσης για να αποφύγουμε την ανάγκη για συμπερασματική γενίκευση.

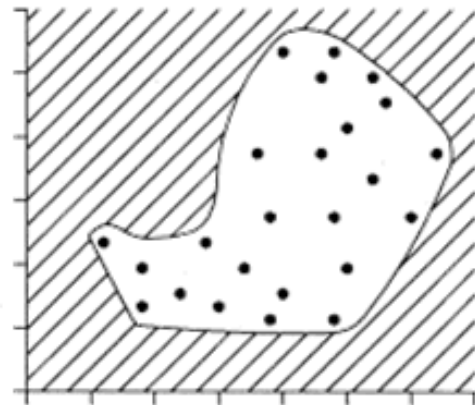
Γι' αυτό το λόγο, για μία ομαλή συνάρτηση εισόδου-εξόδου, αν έχουμε μία δοκιμαστική περίπτωση που είναι κοντά σε κάποια δεδομένα εκπαίδευσης, η σωστή έξοδος για την περίπτωση που δοκιμάζουμε θα βρίσκεται κοντά στις σωστές εξόδους για αυτά τα δεδομένα εκπαίδευσης. Αν έχουμε ένα επαρκές δείγμα για το σύνολο εκπαίδευσης, κάθε περίπτωση στον πληθυσμό θα είναι κοντά σε ένα αρκετό πλήθος δεδομένων εκπαίδευσης. Γι' αυτό το λόγο, κάτω από αυτές τις συνθήκες και με σωστή εκπαίδευση, ένα νευρωνικό δίκτυο θα είναι ικανό να γενικεύει αξιόπιστα στον πληθυσμό.

Αν έχουμε περισσότερες πληροφορίες για μία συνάρτηση, για παράδειγμα ότι οι εξοδοί θα πρέπει να σχετίζονται με τις εισόδους γραμμικά, μπορούμε να εκμεταλλευτούμε αυτή την πληροφορία βάζοντας περιορισμούς στο δίκτυο ή προσαρμόζοντας ένα πιο συγκεκριμένο μοντέλο, όπως ένα γραμμικό μοντέλο, για την βελτίωση της γενίκευσης. Η συμπερασματική γενίκευση είναι πολύ πιο αξιόπιστη σε γραμμικά μοντέλα απ' ό,τι σε ευμετάβλητα μη-γραμμικά μοντέλα, αλλά πάντα η γενίκευση με παρεμβολή είναι πιο ασφαλής. Μπορούμε επίσης να χρησιμοποιήσουμε τέτοιου τύπου πληροφορίες για να επιλέξουμε τις περιπτώσεις εκπαίδευσης πιο αποτελεσματικά. Για παράδειγμα, με ένα γραμμικό μοντέλο, θα πρέπει να επιλέξουμε περιπτώσεις εκπαίδευσης κοντά στα όρια του χώρου δεδομένων εισόδου αντί να τα μοιράσουμε με ισοδύναμο τρόπο μέσα στον χώρο των δεδομένων εισόδου.

Βασικά, όπως φαίνεται και από το Σχ.2-4, ο χώρος των δεδομένων, τον οποίο θα επεξεργαστεί ένα εκπαιδευμένο νευρωνικό δίκτυο, χωρίζεται σε δύο περιοχές : **(1)** την περιοχή όπου η πυκνότητα δεδομένων του συνόλου εκπαίδευσης είναι μεγαλύτερη του μηδενός, και **(2)** οποιοδήποτε άλλο μέρος του χώρου δεδομένων όπου η πυκνότητα των δεδομένων εκπαίδευσης είναι ίση με μηδέν (ή κοντά στο μηδέν). Για άγνωστα σημεία δεδομένων που εμπίπτουν στην πρώτη περιοχή, χρησιμοποιούμε γενίκευση με παρεμβολή. Για όλα τα υπόλοιπα σημεία, ο υπολογισμός απαιτεί συμπερασματική γενίκευση.



Γενίκευση με Παρεμβολή



Συμπερασματική Γενίκευση

Σχ.2-4 : Δύο τύποι γενίκευσης : (1) Γενίκευση με παρεμβολή και (2) Συμπερασματική Γενίκευση

2.3.3 Μέθοδοι εκτίμησης της ακρίβειας γενίκευσης

Στην επαγωγική εκμάθηση, μία μηχανή εκμάθησης αναμένεται να εξάγει κανόνες από ένα πεπερασμένο σύνολο παραδειγμάτων. Είναι συχνό το γεγονός, η μηχανή να απομνημονεύει καλά τα δεδομένα εκπαίδευσης, αλλά να αδυνατεί να παράγει σωστή έξοδο για κάποια από τα νέα δεδομένα που του παρέχονται σαν εκπαίδευση. Γι'αυτό το λόγο, είναι επιθυμητό να δημιουργήσουμε έναν κατηγοριοποιητή που αποδίδει καλά στο πρόβλημα επιτυγχάνοντας την μέγιστη ακρίβεια γενίκευσης. Υπάρχουν πολλές μέθοδοι για την εκτίμηση της ακρίβειας γενίκευσης όπως διαχωριστής-δειγμάτων, επικύρωση με διασταύρωση, η μέθοδος μπάγκινγκ, η μέθοδος “στοίβας και ενίσχυσης”.

Σύμφωνα με έναν απλό κανόνα της θεωρίας στατιστικής, όταν υπάρχουν πολλαπλές εκτιμήσεις, ο καλύτερος υπολογισμός γίνεται με έναν συνδυασμό των μεμονωμένων αυτών εκτιμήσεων. Μία κοινή προσέγγιση για να εξασφαλίσουμε έναν ταξινομητή με τον μικρότερο πιθανό ρυθμό λάθους ταξινόμησης είναι να δημιουργήσουμε και να συνδυάσουμε πολλαπλούς ταξινομητές. Η γενίκευση στοίβας που προτάθηκε από τον Wolpert, η γενίκευση μπάγκινγκ του Breiman και η γενίκευση ενίσχυσης του Schapire και Freud, όλες ακολουθούν τον ίδιο κανόνα (δημιουργία και συνδυασμό πολλαπλών

εκτιμητών). Αυτές οι μέθοδοι είναι πλάνα υψηλού επιπέδου που συνδυάζουν αυτόνομους γενικευτές για να επιτύχουν μεγαλύτερη ακρίβεια πρόβλεψης [13] .

2.4 ΠΟΛΩΣΗ ΜΟΝΤΕΛΩΝ

2.4.1 Εισαγωγή στην έννοια της πόλωσης

Προκειμένου να αναλύσουμε την μορφή του σφάλματος ενός εκτιμητή είναι χρήσιμο να αποσυνθέσουμε το σφάλμα εκτίμησης στα συστατικά της *πόλωσης και μεταβλητότητας* και να χειριστούμε καθένα ξεχωριστά.

- *Πόλωση* είναι η ικανότητα του μοντέλου να ταιριάζει τα δεδομένα. Δηλώνει πόσο ευέλικτο είναι το μοντέλο και όσες περισσότερες παραμέτρους έχει το μοντέλο τόσο καλύτερα προσεγγίζει τη λύση. Ένα απλό και μη ευέλικτο μοντέλο έχει μεγάλη πόλωση.
- *Μεταβλητότητα* είναι η εξάρτηση του μοντέλου από το σύνολο εκπαίδευσης. Αν έχουμε μικρές διαταραχές στο σύνολο δεδομένων, τότε μπορεί να οδηγηθούμε σε σημαντικά διαφορετικά μοντέλα. Ένα ευέλικτο μοντέλο έχει μεγάλη μεταβλητότητα.

Η καλύτερη γενικευτική ικανότητα πετυχαίνεται με μικρή πόλωση και μικρή διακύμανση.

2.4.2 Ορισμοί που έχουν δοθεί για την έννοια της πόλωσης

- **Πόλωση** : Επιπλέον είσοδος που χρησιμοποιείται προκειμένου να αυξήσει ή να μειώσει το σταθμισμένο άθροισμα ανάλογα με το αν αυτό είναι θετικό ή αρνητικό
-

- **Στατιστική πόλωση** : η διαφορά μεταξύ της μέσης τιμής ενός εκτιμητή και της σωστής τιμής.
 - ο **Ανεπαρκής προσαρμογή** : Η όχι τόσο καλή προσαρμογή παράγει υπερβολική πόλωση στις εξόδους, ενώ
 - ο **Υπερπροσαρμογή** : η υπερβολική προσαρμογή παράγει υπερβολική μεταβλητότητα.

Η πόλωση είναι μη-μηδενική για όλα τα υποσύνολα που δεν είναι “καλά”, αλλά είναι μηδέν για όλα τα “καλά” υποσύνολα, μιας και υποθέτουμε ότι η συνάρτηση που επρόκειτο να εκπαιδύσουμε είναι γραμμική. Γι’ αυτό το λόγο, το λάθος γενίκευσης των “καλών” υποσυνόλων θα διαφέρει μόνο στην μεταβλητότητα.

2.4.3 Το δίλημμα της πόλωσης / μεταβλητότητας

Ένα από τα πιο σοβαρά προβλήματα που ανακύπτουν στην μάθηση με συνδέσμους από τα νευρωνικά δίκτυα είναι η υπερπροσαρμογή από τα παρεχόμενα παραδείγματα εκπαίδευσης. Αυτό σημαίνει ότι η συνάρτηση που μαθαίνεται ταιριάζει πολύ με τα δεδομένα εκπαίδευσης παρ’όλ’ αυτά δεν γενικεύει καλά και συνεπώς δεν μπορεί να μοντελοποιήσει επαρκώς καλά δεδομένα που δεν έχουν ξαναεμφανιστεί.

Μιας και ο απόλυτος στόχος πίσω από την εκπαίδευση μηχανής είναι η επίτευξη γενίκευσης, τέτοια προβλήματα υπερπροσαρμογής πρέπει να ανιχνεύονται προσεκτικά.

Κριτήρια για την αποφυγή υπερπροσαρμογής με τα δεδομένα εκπαίδευσης και για την αύξηση της γενίκευσης, δίνονται από τις θεωρίες στατιστικής και γενίκευσης.

Ορισμοί

Αυτά τα κριτήρια χρησιμοποιούνται για να ισορροπήσουν την **στατιστική πόλωση** και την **στατιστική μεταβλητότητα** όταν εκτελείται μάθηση νευρωνικού δικτύου με σκοπό την επίτευξη μικρότερου μέσου όρου λάθους γενίκευσης (σε δεδομένα που δεν έχουν ξαναπαρουσιαστεί).

Στατιστική πόλωση είναι ο περιορισμός της πολυπλοκότητας που επιβάλλει η αρχιτεκτονική του νευρωνικού δικτύου στο βαθμό ακριβούς προσαρμογής στην συνάρτηση στόχο. Η στατιστική πόλωση παίζει ρόλο μόνο στον βαθμό προσαρμογής των δοθέντων δεδομένων εκπαίδευσης, αλλά όχι στο επίπεδο γενίκευσης.

Στατιστική μεταβλητότητα είναι η απόκλιση της αποδοτικότητας μάθησης του νευρωνικού δικτύου από το ένα δείγμα δεδομένων στο άλλο, το οποίο θα μπορούσε να περιγραφεί από το ίδιο μοντέλο συνάρτησης στόχου. Αυτή η μεταβλητότητα αναφέρεται στην γενίκευση για το αν το νευρωνικό δίκτυο προσαρμόζει / ταιριάζει τα παραδείγματα χωρίς να λάβει υπόψιν τις ιδιαιτερότητες των παρεχόμενων δεδομένων.

Όταν το δίκτυο μαθαίνει από ένα σταθερό και πεπερασμένο σύνολο παραδειγμάτων, το σταθερό λάθος μπορεί να είναι μικρό αλλά αυτό δεν είναι αρκετό για μία υψηλού βαθμού γενίκευση μιας και τα παραδείγματα δεν είναι αξιόπιστα εξαιτίας του θορύβου και των αβεβαιοτήτων. Τέτοιες δυσκολίες μπορούν να ξεπεραστούν αν αναλύσουμε το λάθος E_D του νευρωνικού δικτύου $f(\mathbf{x})$ σε δύο μέρη:

$$E_D[(f(\mathbf{x}) - E[y|\mathbf{x}])^2] = (E_D[f(\mathbf{x})] - E[y|\mathbf{x}])^2 + E_D[(f(\mathbf{x}) - E_D[f(\mathbf{x})])^2] =$$

$$\text{BIAS}^2(f(\mathbf{x})) + \text{VAR}(f(\mathbf{x}))$$

Όπου :

- η συνάρτηση $f(\mathbf{x})$ είναι μία προσέγγιση της πραγματικής συνάρτησης $f^*(\mathbf{x}) = E[y|\mathbf{x}]$ λαμβάνοντας υπόψιν τα δεδομένα εκπαίδευσης D .
- $\text{BIAS}(f(\mathbf{x})) = E_D[f(\mathbf{x})] - E[y|\mathbf{x}]$ είναι η στατιστική πόλωση
- και $\text{VAR}(f(\mathbf{x})) = E_D[(f(\mathbf{x}) - E_D[f(\mathbf{x})])^2]$ και η στατιστική μεταβλητότητα

Η απόδοση του νευρωνικού δικτύου μπορεί να βελτιωθεί αν μειώσουμε και την στατιστική πόλωση και την στατιστική μεταβλητότητα. Παρόλαυτα υπάρχει μία φυσική αλληλεπίδραση μεταξύ πόλωσης και μεταβλητότητας, το οποίο είναι γνωστό ως το δίλημμα πόλωσης / μεταβλητότητας :

Ένα νευρωνικό δίκτυο που προσαρμόζει προσεκτικά τα παρεχόμενα παραδείγματα εκπαίδευσης, έχει χαμηλή πόλωση αλλά υψηλή μεταβλητότητα. Αν μειώσουμε την μεταβλητότητα του δικτύου αυτό θα οδηγήσει σε μία μείωση στο επίπεδο προσαρμογής των δεδομένων.

Μία πιθανή στρατηγική για την παράλληλη μείωση στατιστικής πόλωσης και μεταβλητότητας είναι να θεωρήσουμε περισσότερα σημεία δεδομένων ή να αυξήσουμε τα παραδείγματα εκπαίδευσης.

Μία πιθανή στρατηγική για την μείωση της στατιστικής πόλωσης είναι η ανάπτυξη του νευρωνικού δικτύου.

Μία πιθανή στρατηγική για την μείωση της στατιστικής μεταβλητότητας είναι ο περιορισμός του νευρωνικού δικτύου.

2.5 ΕΚΜΑΘΗΣΗ ΣΥΝΑΡΤΗΣΕΩΝ

Η γνώση του πώς να χτιστεί ένα νευρωνικό δίκτυο είναι κάτι σημαντικό, αλλά η μετατροπή του σε μια χρήσιμη εφαρμογή είναι κάτι σημαντικότερο. Το αγκάλιασμα της νευρωνικής τεχνολογίας δικτύων, χωρίς όμως προσεκτικό προγραμματισμό, μπορεί να οδηγήσει σε λιγότερο χρήσιμες εφαρμογές.

Τα νευρωνικά δίκτυα χρησιμοποιούν μια πολυστρωματική προσέγγιση όσον αφορά τις σύνθετες μαθηματικές λειτουργίες για να επεξεργαστούν τα στοιχεία τους. Αποτελούνται από πολλούς κόμβους επεξεργασίας που λειτουργούν παράλληλα. Οι κόμβοι συνδέονται ο ένας με τον άλλον στα στρώματα, και τα στρώματα διασυνδέονται μεταξύ τους. Αυτοί οι κόμβοι εκτελούν απλές μαθηματικές λειτουργίες, οι συνδέσεις μεταξύ αυτών των κόμβων, που σταθμίζουν τον μετασχηματισμό στοιχείων από κάθε κόμβο και στέλνουν τις πληροφορίες στο επόμενο στρώμα κόμβων ή παραγωγής, είναι η «νοητική» λειτουργία των νευρωνικών δικτύων.

Για να «μάθει» μια συνάρτηση ένα νευρωνικό δίκτυο, ο υπεύθυνος για την εκπαίδευση, του παρέχει ποικίλα από πραγματικά παραδείγματα, αποκαλούμενα σύνολα εκπαίδευσης. Το σύνολο προτύπων περιέχει δεδομένα εισόδου και δεδομένα εξόδου. Το

νευρωνικό δίκτυο μαθαίνει τα πρότυπα βασιζόμενο σε αυτά τα σύνολα δεδομένων εισόδου και εξόδου. Κάθε πρότυπο δημιουργεί μια μοναδική διαμόρφωση της δομής δικτύων με ένα μοναδικό σύνολο δυνάμεων ή βαρών σύνδεσης.

Η διαδικασία είναι ανάλογη με την ανθρώπινη. Οι άνθρωποι «μαθαίνουν» στη καθημερινή τους ζωή. Για παράδειγμα « Σε έναν αγώνα καλαθοσφαίρισης, ένας παίκτης προσπαθεί να σκοράρει, αν η μπάλα μπει στο στεφάνι, τότε ο στόχος έχει εκτελεσθεί μέσα στις ανοχές του παιχνιδιού δεν δημιουργείται αντίδραση. Εάν όμως χτυπήσει κάπου στο πλαίσιο, ή στη στεφάνι ή παρά μόνο το δίχτυ, την επόμενη φορά ο παίκτης θα αλλάξει ίσως τη δύναμη ή τη γωνία με την οποία εκτέλεσε τη βολή προς το καλάθι». Ομοίως, όταν το δίκτυο φθάσει σε μια τελική απόφαση, τη συγκρίνει με την απάντηση που του παρέχεται στο σύνολο εκπαίδευσης. Εάν αυτή έχει μια απόκλιση από την επιθυμητή βρίσκεται όμως μέσα στα όρια της ανοχής του δικτύου τότε το νευρωνικό δίκτυο αποθηκεύει τα βάρη σύνδεσης. Εάν η έξοδος του δικτύου είναι έξω από την ανοχή του δικτύου, τότε γίνεται μια ανακύκλωση και δίνει εκ νέου μια βελτιωμένη έξοδο. Το δίκτυο μπορεί να ανακυκλώσει πολλές φορές μέχρι να επιτευχθεί η έξοδος στα όρια ανοχής.

Αναλυτικότερα, τα νευρωνικά δίκτυα μαθαίνουν μια συνάρτηση $y = f(x)$, όπου το y είναι ένα διάνυσμα εξόδου, x είναι ένα διάνυσμα εισόδων, και το $f()$ είναι η συνάρτηση που ζητάμε από το δίκτυο να μάθει. Αυτή η διαδικασία ακολουθείται στην μάθηση με επίβλεψη. Αντίστοιχα υπάρχει και η αντίστροφη απαίτηση να μπορεί το δίκτυο από ένα διάνυσμα εξόδου δηλαδή από ένα y να βρει τα x ώστε να ισχύει $y = f(x)$, σε αυτή τη περίπτωση μπορεί να υπάρξουν πολλοί διαφορετικοί συνδυασμοί x που ικανοποιούν την εξίσωση $y = f(x)$ ο τρόπος αυτό βρίσκει εφαρμογή στην μη επιβλεπόμενη μάθηση.

Για να περιγράψουμε το τρόπο με τον οποίο ένα νευρωνικό δίκτυο μαθαίνει μια συνάρτηση θα αναλύσουμε ορισμένα παραδείγματα. Στη ρομποτική⁶, το διάνυσμα x περιγράφει τις θέσεις των ενώσεων στον βραχίονα ενός ρομπότ, ενώ το y θα περιέγραφε την θέση του χεριού του ρομπότ. Υπάρχουν απλοί τύποι για να υπολογίσουν την θέση το χέρι δεδομένων των θέσεων των ενώσεων. Αλλά δεν υπάρχει κανένας απλός τύπος για το αντίστροφο πρόβλημα για να υπολογίσει τις θέσεις των ενώσεων που παράγουν μια δεδομένη θέση για το χέρι. Επιπλέον, εάν ο βραχίονας έχει διάφορες ενώσεις, θα υπάρξουν συνήθως πολλές διαφορετικές θέσεις των ενώσεων που παράγουν την ίδια

θέση του χεριού. Επιλογή οποιουδήποτε x έτσι ώστε $y = f(x)$ είναι εντάξει εάν ο μόνος στόχος είναι να τοποθετηθεί το χέρι στο y . Εντούτοις εάν ο στόχος είναι να παραχθεί μια σειρά σημείων για να κινηθεί το χέρι μέσω ενός τόξου αυτό μπορεί να είναι ανεπαρκές. Σε αυτήν την περίπτωση οι σειρές x πρέπει να είναι στον ίδιο "κλάδο" του διαστήματος λειτουργίας. Προσοχή πρέπει να ληφθεί για να αποφύγει τις λύσεις που παράγουν τις ανεπαρκείς ή αδύνατες μετακινήσεις του βραχίονα.

Ένα άλλο παράδειγμα, μια βιομηχανική διαδικασία στην οποία το x αντιπροσωπεύει τις τοποθετήσεις των μεταβλητών ελέγχου που επιβάλλονται από έναν χειριστή, και το y αντιπροσωπεύει τις μετρήσεις του προϊόντος της βιομηχανικής διαδικασίας. Η λειτουργία $y = f(x)$ μπορεί να μαθευτεί από ένα ΤΝΔ χρησιμοποιώντας τις συμβατικές μεθόδους κατάρτισης. Αλλά ο στόχος της ανάλυσης μπορεί να είναι να βρεθούν οι εισόδοι x που παράγουν ένα προϊόν με τις συγκεκριμένα χαρακτηριστικά y , οπότε σ'αυτή την περίπτωση ένα αντίστροφο της $f(x)$ απαιτείται. Στις βιομηχανικές εφαρμογές, οι οικονομικοί περιορισμοί πρέπει να λαμβάνονται υπόψη, δηλαδή πρέπει να βρούμε τις εισόδους x που ικανοποιούν τη συνάρτηση περιορισμού $f(x)$ και παράγουν το επιθυμητό αποτέλεσμα y . Βέβαια θεωρώντας και περιορισμούς όπως το κόστος του x ή την κατανάλωση ενέργειας, των πρώτων υλών, κ.λ.π., οπότε σ'αυτή την περίπτωση θα ζητάμε από το δίκτυο να βρει τα x που ελαχιστοποιούν την $f(x)$ και ικανοποιούν διάφορους περιορισμούς.

Ο προφανής τρόπος να διδάξουμε σε ένα δίκτυο την αντίστροφη λειτουργία είναι να παραχθεί ένα σύνολο στοιχείων εκπαίδευσης από μια δεδομένη μπροστινή λειτουργία, αλλά υποδεικνύει το y ως εισαγωγή και το x ως έξοδο κατά τη εκπαίδευση του δικτύου. Χρησιμοποιώντας μια συνάρτηση ελέγχου λάθους με τη χρήση του ελαχίστου τετραγώνου, το δίκτυο θα αποτύγχανε εάν η συσχέτιση του $f()$ είναι πολλά προς ένα. Το πρόβλημα είναι ότι για μια είσοδο y , το δίκτυο δεν θα μάθει οποιοδήποτε ενιαίο x έτσι ώστε $y = f(x)$, αλλά θα μάθει αντ' αυτού τον αριθμητικό μέσο όρο όλου του x στο σύνολο κατάρτισης που ικανοποιεί την εξίσωση (Bishop, 1995, pp. 207-208). Μια λύση σε αυτήν την δυσκολία είναι να κατασκευαστεί ένα δίκτυο που μαθαίνει με χρήση προσέγγισης το διάνυσμα x δεδομένου του y (Bishop, 1995, pp. 212-221). Από την άλλη η μέθοδος αυτή παρουσιάζει μειονεκτήματα σε περιπτώσεις που ένα διάνυσμα x είναι

πολυδιάστατο όπως $Y = X_1^2 + X_2^2$, δεδομένου ότι ο αριθμός των στοιχείων που θα πρέπει να γνωρίζουμε μπορεί να αυξηθεί εκθετικά με την διάσταση του X .

2.6 ΜΑΘΗΣΗ ΜΕ ΕΠΙΒΛΕΨΗ

Η πλειοψηφία των τεχνητών νευρωνικών δικτύων έχει εκπαιδευθεί με την μάθηση με επίβλεψη. Ο τρόπος εκπαίδευσης αυτός, στήρίζεται στο ότι η έξοδος που δίνει το δίκτυο συγκρίνεται με την επιθυμητή τιμή που περιμένουμε να έχει το δίκτυο. Τα συναπτικά βάρη, που είναι συνήθως τυχαία αρχικά, ρυθμίζονται έπειτα με το δίκτυο έτσι ώστε στην επόμενη επανάληψη, ή κύκλο, που θα κάνει το δίκτυο να παραγάγει μια πιο κοντινή αντιστοιχία μεταξύ της επιθυμητής και πραγματικής εξόδου του δικτύου. Η μέθοδος εκμάθησης προσπαθεί με άλλα λόγια να ελαχιστοποιήσει τα λάθη όλων των στοιχείων επεξεργασίας (νευρώνων). Αυτή η μείωση λάθους δημιουργείται κατά τη διάρκεια της μάθησης με συνεχείς τροποποίηση των βαρών έως ότου επιτυγχάνεται το δίκτυο έχει τη δυνατότητα να προβλέψει ή να κατηγοριοποιήσει μέσα σε αποδεκτά ποσοστά αποτυχίας. [3].

Αυτή η διαδικασία μπορεί να συγκριθεί με ένα νέο άνθρωπο μικρής ηλικίας, του οποίου εγκέφαλος έχει αναπτυχθεί πλήρως αλλά δεν είναι ικανός να λειτουργήσει να διαχωρίσει έννοιες, επειδή δεν έχει την εμπειρία από οποιοδήποτε ερέθισμα. Η μεθοδολογία που ακολουθείται είναι ανάλογη με τον τρόπο που διδάσκατε ένα παιδί να διαβάσετε ή να μετρά, δηλαδή με τη χρήση παραδειγμάτων. Με άλλα λόγια θα μπορούσαμε να παρουσιάσουμε σε ένα παιδί 5 ετών μια εικόνα ενός «Α» και θα μπορούσαμε έπειτα να του πούμε ότι είναι ένα "Α'Άλφα".

Διδάσκουμε ένα νευρωνικό δίκτυο με τον ίδιο ακριβώς τρόπο, δηλαδή θα δώσουμε ένα σύνολο αριθμών (στη βιβλιοθήκη μας μεταξύ 0 και 1), και το δίκτυο θα σας δώσει ένα αποτέλεσμα στο στην έξοδο. Δεδομένου ότι τα βάρη των συνδέσεων στο δίκτυο είναι αρχικά τυχαία αρχικοποιημένα, αυτό το αποτέλεσμα σίγουρα στη αρχή δεν θα είναι ικανοποιητικό, ωστόσο με το πέρασ της εκπαίδευσης θα επιτευχθεί ένα καλύτερο αποτέλεσμα. Η «καλύτερη» μάθηση επιτυγχάνεται με τη χρήση πολλών προτύπων εκπαίδευσης που διοχετεύονται στο δίκτυο έτσι ώστε να καταφέρει να αντιμετωπίσει πολλές διαφορετικές περιπτώσεις προτύπων. Το σύνολο των προτύπων που

χρησιμοποιούνται στην επιβλεπόμενη εκπαίδευση ονομάζεται (training set = σύνολο εκπαιδευσεως). Βέβαια αν είναι εφικτό πριν στην εκπαίδευση του δικτύου, να προστεθεί θόρυβος, δηλαδή παραδείγματα που δεν είναι ξεκάθαρα. Αυτό βοηθά το νευρωνικό δίκτυο να εκτιμά το θόρυβο και τη φυσική μεταβλητότητα των δεδομένων, έτσι δημιουργούμε ένα αξιόπιστο νευρωνικό δίκτυο. Είναι εξίσου σημαντικό να επιλεχθούν τα παραδείγματα που έχουν σημαντικά χαρακτηριστικά γνωρίσματα με τα πρότυπα που θα αποδοθούν για γενίκευση.

Η εκπαίδευση θεωρείται πλήρης όταν φθάνει το νευρωνικό δίκτυο σε επίπεδο απόδοσης που έχει καθορίσει ο χρήστης. Το επίπεδο αυτό δηλώνει ότι το δίκτυο έχει επιτύχει την επιθυμητή στατιστική ακρίβεια, δεδομένου ότι παράγει τα απαραίτητα αποτελέσματα για μια δεδομένη ακολουθία εισαγωγών. Όταν η διαδικασία της μάθησης δεν είναι απαραίτητο να συνεχιστεί, τα βάρη διατηρούνται σταθερά. Αν και υπάρχουν μερικοί τύποι δικτύων που είναι σχεδιασμένα έτσι ώστε να έχουν μια συνεχή εκπαίδευση, σε μια πολύ πιο αργή διαδικασία, ενώ χρησιμοποιούνται.

Η εκπαίδευση απαιτεί ένα αρκετά μεγάλο σύνολο δεδομένων (πρότυπα εκπαίδευσης), ώστε να περιέχει όλες τις αναγκαίες πληροφορίες για το δίκτυο προκειμένου να μάθει χαρακτηριστικά γνωρίσματα και σχέσεις που είναι σημαντικές. Εάν το δίκτυο εκπαιδεύεται μόνο με ένα παράδειγμα τη φορά, όλα τα βάρη καθορίζονται τόσο σχολαστικά για ένα γεγονός και έτσι θα μπορούσαν να αλλάξουν δραστικά στην εκμάθηση του επόμενου γεγονότος. Τα προηγούμενα γεγονότα θα μπορούσαν να ξεχαστούν στην εκμάθηση του νέου. Κατά συνέπεια, το σύστημα πρέπει να μάθει ταυτόχρονα πολλές περιπτώσεις, να βρει έτσι τις καλύτερες τιμές για τα βάρη. Παραδείγματος χάριν, στη διδασκαλία ενός συστήματος για να αναγνωρίσουν τα δέκα ψηφία, εάν υπήρξαν είκοσι παραδείγματα κάθε ψηφίου, όλα τα παραδείγματα ενός ψηφίου δεν πρέπει να παρουσιαστούν συγχρόνως.

«Ένα διάσημο παράδειγμα είναι ένα σύστημα που αναπτύχθηκε για της ανάγκες του αμερικανικού στρατού. Ένα σύστημα που αναγνώριζε εχθρικά οχήματα. Στη διάρκεια της επιβλεπόμενης εκμάθησης στο νευρωνικό δόθηκαν παραδείγματα των σοβιετικών τανκ από πολλές διαφορετικές αποστάσεις και οπτικές γωνίες μια φωτεινή ηλιόλουστη ημέρα, και παραδείγματα των αμερικανικών τανκ μια νεφελώδη ημέρα. Το αποτέλεσμα ήταν το συγκεκριμένο δίκτυο να μπορούσε να αναγνωρίζει τις καιρικές συνθήκες και όχι

τα εχθρικά οχήματα». Το νευρωνικό δίκτυο σε αυτή τη περίπτωση δεν σχεδιάστηκε σωστά ώστε να μπορέσει να «μάθει» τα διάφορα χαρακτηριστικά των στρατιωτικών οχημάτων. Βέβαια και τα πρότυπα που χρησιμοποιήθηκαν δεν ήταν «ξεκάθαρα». Με άλλα λόγια το νευρωνικό δεν μπόρεσε να προσδιορίσει τα χαρακτηριστικά στα οποία έπρεπε να εστιάσει. Δεν ήταν σαφής η σημαντικότητα του κάθε στοιχείου που αποτελούσε το πρότυπο εκπαίδευσης (απόσταση, στρατιωτικό όχημα, καιρικές συνθήκες, οπτική γωνία).

Τέλος, όταν ένα νευρωνικό αποδίδει καλά στα στοιχεία κατάρτισης, κατόπιν είναι σημαντικό να φανεί ότι τι μπορεί να κάνει με τα στοιχεία αυτό δεν έχει δει πριν. Εδώ περνάμε στη διαδικασία ελέγχου του δικτύου. Τα πρότυπα που του δίδονται ονομάζονται “test data”. Στη περίοδο ελέγχου αποφασίζεται αν έχει τελειώσει η εποχή κατάρτισης. Αυτή η δοκιμή είναι κρίσιμη για να ασφαλίσει ότι το δίκτυο δεν έχει απομνημονεύσει απλά ένα δεδομένο σύνολο στοιχείων αλλά έχει μάθει τα γενικά σχέδια που περιλαμβάνονται μέσα σε αυτά.

2.7 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ

Η μη επιβλεπόμενη μάθηση είναι μια μεγάλη υπόσχεση του μέλλοντος. Είναι η εποχή εκείνη στην οποία οι υπολογιστές θα μπορούν να λειτουργούν με «ανθρώπινη» υπόσταση στο θέμα της νοημοσύνης. Στις μέρες μας , αυτή η μέθοδος εκπαίδευσης νευρωνικών δικτύων περιορίζεται στα δίκτυα γνωστά ως αυτορογανωμένοι χάρτες χαρακτηριστικών. Βέβαια αυτά τα είδη δικτύων δεν είναι διαδεδομένη ακόμη, βρίσκονται όμως σε διαρκή μελέτη και ερευνά.

Αυτό το είδος μάθηση ονομάζεται και προσωπική μάθηση. Τα δίκτυα σε αντίθεση με τη τεχνική που περιγράψαμε σε προηγούμενο εδάφιο δεν χρησιμοποιεί καμία εξωτερική επιρροή για να προσαρμόσει για παράδειγμα τα βάρη. Αντ' αυτού, ελέγχουν εσωτερικά την απόδοσή τους. Αυτά τα δίκτυα ψάχνουν τυχόν ομοιότητες, τάσεις ή χαρακτηριστικά των στοιχείων που επεξεργάζεται και αυτοπροσαρμόζεται σύμφωνα και με τη συνάρτηση που το διέπει. Το δίκτυο δέχεται ένα πλήθος πληροφοριών που αφορούν το τρόπο λειτουργίας και αντιμετώπισης από την ίδια του τη δομή και από τους κανόνες μάθησης.

Ένας αλγόριθμος μάθησης χωρίς επίβλεψη βασίζεται στις σχέσεις των νευρώνων στις ομάδες (συστάδες) που σχηματίζουν, στο τρόπο που συνδέονται και στο πως αλληλεπιδρούν μεταξύ τους. Αναλυτικότερα αν κάποια εξωτερική εισαγωγή (πρότυπο μαθήσεως) ενεργοποιούσε οποιοδήποτε κόμβο της ομάδας του, η δραστηριότητα της ομάδας συνολικά θα μπορούσε να αλλάξει και να αυξηθεί. Το αντίστοιχο συμβαίνει και στην αντίθετη περίπτωση αν η εισαγωγή σε ένα κόμβο της συστάδας μειώσει την ενεργητικότητα της τότε, αυτή θα μπορούσε υιοθετήσει μια ανασταλτική επίδραση σε ολόκληρη συστάδα.

Ο ανταγωνισμός μεταξύ των στοιχείων επεξεργασίας θα μπορούσε επίσης να αποτελέσει μια βάση για μάθηση. Η εκπαίδευση των ανταγωνιστικών ομάδων νευρώνων ενισχύει τις ενέργειες τους σε συγκεκριμένα ερεθίσματα. Ουσιαστικά η διαδικασία μάθηση ενεργοποιεί και επιτρέπει την τροποποίηση βαρών που αντιστοιχούν στην ισχυρότερη κάθε φορά ομάδα νευρώνων. Σήμερα, μάθηση χωρίς επίβλεψη αποτελεί ακόμα το αντικείμενο έρευνας.

Κεφάλαιο 3

ΚΟΝΤΙΝΟΤΕΡΟΙ ΓΕΙΤΟΝΕΣ

ΠΕΡΙΓΡΑΦΗ ΤΕΧΝΙΚΗΣ

Οι κ-Κοντινότεροι Γείτονες (k-NN) είναι μία μη-παραμετρική αλλά αποτελεσματική μέθοδος για ταξινόμηση.

Στην τεχνική του κοντινότερου γείτονα, έχουμε κάποια δεδομένα (με κάποιες προκαθορισμένες διαστάσεις) τα οποία ανήκουν σε κάποιες κατηγορίες αντίστοιχα. Σκοπός μας είναι να μπορεί το δίκτυο να ταξινομεί μία νέα εγγραφή δεδομένων $\mathbf{t}$ στην σωστή κατηγορία. Αυτό γίνεται βρίσκοντας τους k κοντινότερους γείτονες αυτής της εγγραφής με υπολογισμό των αποστάσεων του $\mathbf{t}$ σε σχέση με όλες τις υπόλοιπες εγγραφές. Με αυτόν τον τρόπο δημιουργούμε μία γειτονιά του $\mathbf{t}$. Σ' αυτή τη γειτονιά, αν η πλειοψηφία των εγγραφών ανήκει σε μία συγκεκριμένη κατηγορία, τότε ταξινομούμε και τη νέα εγγραφή στην κατηγορία αυτή.

Τα πιο σημαντικά μειονεκτήματα αυτής της μεθόδου είναι η χαμηλή αποδοτικότητα του – μιας και είναι μία μέθοδος που μαθαίνει αργά, δεν είναι κατάλληλη για πολλές εφαρμογές όπως εξόρυξη δεδομένων στο διαδίκτυο που αφορά μεγάλες ποσότητες δεδομένων. Επίσης, μειονέκτημα είναι και η εξάρτησή του από την επιλογή μιας “καλής τιμής” για το k . Ένας τρόπος για την βελτίωση της αποτελεσματικότητας αυτής της μεθόδου είναι να βρούμε μερικούς αντιπροσώπους για να αντιπροσωπεύσουμε ολόκληρο το σύνολο δεδομένων εκπαίδευσης.

Συνεπώς, για να εφαρμόσουμε το k-NN πρέπει να επιλέξουμε μία κατάλληλη τιμή για το k , και η επιτυχία της ταξινόμησης εξαρτάται σε μεγάλο βαθμό από αυτή την τιμή. Η μέθοδος kNN διαφοροποιείται ανάλογα με το k . Υπάρχουν πολλοί τρόποι να επιλέξουμε την τιμή του k , αλλά μία απλή μέθοδος είναι να εκτελέσουμε τον αλγόριθμο πολλές φορές δίνοντας διαφορετικές τιμές στο k κάθε φορά και να διαλέξουμε την τιμή που μας δίνει την καλύτερη απόδοση. Βέβαια, η επιλογή αυτή εξαρτάται πάντα και από τη φύση του προβλήματος στο οποίο εφαρμόζεται η μέθοδος του κοντινότερου γείτονα.

Για να επιτευχθεί ένα δίκτυο k -NN λιγότερο εξαρτημένο από την επιλογή του k , ο Wang πρότεινε να εξετάζονται πολλαπλά σύνολα από κοντινότερους γείτονες παρά μόνο ένα σύνολο.[11] Η προτεινόμενη φόρμα βασίζεται σε σχετική πιθανότητα, και η κεντρική ιδέα είναι να αθροίζονται τα πολλαπλά σύνολα κοντινότερων γειτόνων για διάφορες κατηγορίες για να δίνεται μία πιο αξιόπιστη τιμή, η οποία απεικονίζει την πραγματική κατηγορία του t . Ωστόσο, στην βασική του μορφή η μέθοδος είναι σχετικά αργή, ενώ ταυτόχρονα είναι λιγότερο εξαρτημένη από το k και είναι ικανή να επιτύχει απόδοση στην ταξινόμηση κοντά στο καλύτερο k .

Το k NN έχει υψηλό κόστος όταν ταξινομεί νέα στιγμιότυπα, γιατί σχεδόν όλοι οι υπολογισμοί λαμβάνουν χώρα την στιγμή της ταξινόμησης και όχι όταν τα παραδείγματα εκπαίδευσης πρωτοσυναντώνται. Παρόλο που το k NN εφαρμόστηκε στον τομέα της ταξινόμησης κειμένου από τις πρώτες μέρες της μελέτης αυτού του τομέα και αποδείχθηκε μία από τις πιο αποτελεσματικές μεθόδους, έχει ένα σοβαρό μειονέκτημα. Είναι μία μέθοδος οκνηρής εκμάθησης χωρίς μοντελοποίηση εκ των προτέρων και συνεπώς η αποδοτικότητά του το εμποδίζει να εφαρμοστεί σε τομείς όπου απαιτείται δυναμική ταξινόμηση για μεγάλη ποσότητα δεδομένων.

Επειδή ο k NN ταξινομητής απαιτεί την αποθήκευση ολόκληρου του συνόλου εκπαίδευσης, γεγονός που μπορεί να είναι πολύ σοβαρό μειονέκτημα όταν το σύνολο είναι μεγάλο, πολλοί μελετητές αποπειράθηκαν να ξεφορτωθούν τον πλεονασμό του συνόλου εκπαίδευσης για να λύσουν αυτό το πρόβλημα. Ο Hart πρότεινε μία υπολογιστικά απλή μέθοδο τοπικής αναζήτησης που λέγεται Condensed Nearest Neighbour (CNN) ελαχιστοποιώντας τον αριθμό των αποθηκευμένων προτύπων και αποθηκεύοντας μόνο ένα υποσύνολο του συνόλου εκπαίδευσης για ταξινόμηση. Η κεντρική ιδέα είναι ότι τα πρότυπα στο σύνολο εκπαίδευσης μπορεί να μοιάζουν πολύ μεταξύ τους, μερικά μάλιστα να μην προσθέτουν επιπλέον πληροφορία και συνεπώς να μπορούν να παραβλεφθούν. Ο Gate πρότεινε τον κανόνα Reduced Nearest Neighbour που έχει ως στόχο την περαιτέρω μείωση του αποθηκευμένου υποσυνόλου μετά την εφαρμογή του CNN. Απλά αφαιρεί αυτά τα στοιχεία από το υποσύνολο που δεν θα προκαλέσουν κάποιο λάθος. Ο Alpaydin μελέτησε κάποιες δομές ψηφοφορίας για πολλαπλά συστήματα εκμάθησης με σκοπό την βελτίωση της ακρίβειας ταξινόμησης. Επίσης, ο Kubat πρότεινε μία προσέγγιση η οποία επιλέγει τρία πολύ μικρά σύνολα

παραδειγμάτων τέτοια ώστε, όταν χρησιμοποιείται σαν 1-NN υποταξινομητής, καθένα να τείνει να κάνει λάθος σε διαφορετικό σημείο του χώρου. Η απλή ψηφοφορία στη συνέχεια διορθώνει πολλές αποτυχίες των μεμονωμένων υποταξινομητών. [11]

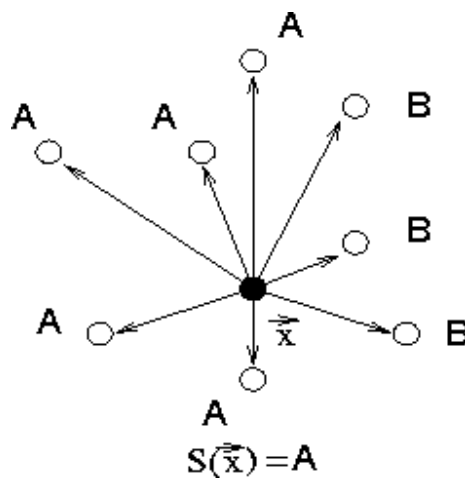
Είναι προφανές ότι η χρήση αυτού του κανόνα περιλαμβάνει και την ανάγκη επιλογής ενός κατάλληλου μετρικού, για παράδειγμα πώς θα υπολογιστεί η απόσταση στα κοντινότερα σημεία; Σε μερικά σύνολα δεδομένων δεν υπάρχει πρόβλημα. Σε δεδομένα όμως με πολλές μεταβλητές, όπου οι μετρήσεις έχουν γίνει σε διαφορετικές κλίμακες, απαιτείται κάποια τυποποίηση. Αυτή θεωρείται συνήθως ως η τυπική απόκλιση ή η διακύμανση της μεταβλητής. Αν υπάρχουν μεταβλητές-δείκτες (όπως συμβαίνει με ονομαστικά δεδομένα), τότε τα δεδομένα αυτά συνήθως μετασχηματίζονται ώστε όλες οι παρατηρήσεις να βρίσκονται μέσα στο υπερσύνολο της μονάδας. Επίσης, πρέπει να τονιστεί ότι η μονάδα μέτρησης μπορεί να είναι και εξαρτημένη από κατηγορία, έτσι ώστε να αποκτά μία απόσταση σχετική με την κατηγορία. Αυτό θα αυξήσει τον χρόνο επεξεργασίας και ταξινόμησης, αλλά θα οδηγήσει σε μία σημαντική αύξηση στην απόδοση.

Για να επιταχύνουμε την διαδικασία εύρεσης των κοντινότερων γειτόνων, πολλές προσεγγίσεις έχουν κατά καιρούς προταθεί. Οι Fukunaka και Narendra (1975) χρησιμοποίησαν έναν αλγόριθμο διαχωρισμού και περιορισμού για να αυξήσουν την ταχύτητα υπολογισμού του κοντινότερου γείτονα. Η γενική ιδέα βασίζεται στην διαίρεση του χώρου ταξινόμησης σε περιοχές και εξερεύνηση μιας περιοχής μόνο όταν υπάρχουν πιθανότητες να βρεθεί εκεί ένας κοντινότερος γείτονας. Οι περιοχές αναλύονται ιεραρχικά σε υποσύνολα, υπό-υποσύνολα και ούτω καθεξής. Άλλοι τρόποι επιτάχυνσης της διαδικασίας είναι να χρησιμοποιήσουμε έναν συμπτυκνωμένο κανόνα κοντινότερου γείτονα (Hart 1968), έναν απλοποιημένο κανόνα κοντινότερου γείτονα (Gates, 1972) ή τον αλλαγμένο κανόνα κοντινότερου γείτονα (Hand και Batchelor, 1978). Όλες αυτές οι μέθοδοι μειώνουν το σύνολο εκπαίδευσης διατηρώντας αυτές τις παρατηρήσεις οι οποίες χρησιμοποιούνται για να κατηγοριοποιήσουν σωστά τα απορριφθέντα σημεία. Με αυτόν τον τρόπο αυξάνεται η ταχύτητα της διαδικασίας ταξινόμησης. [6]

Η επιλογή του k μπορεί να γίνει με μεθόδους ελέγχου διασταύρωσης (cross-validation) με την οποία τα δεδομένα εκπαίδευσης χωρίζονται, και το δεύτερο μέρος κατηγοριοποιείται με τον κανόνα του k -NN. Ωστόσο, όταν πρόκειται για μεγάλα σύνολα

δεδομένων, η μέθοδος αυτή είναι πολύ χρονοβόρα για $k > 1$, μιας και όλα τα δεδομένα εκπαίδευσης πρέπει να αποθηκεύονται και να εξετάζονται για κάθε ταξινόμηση. Οι Etnas και Choy (1986), εξετάσανε αυτό το πρόβλημα σε μία μελέτη προσομοίωσης και πρότειναν κανόνες για τον υπολογισμό του k σε πρόβλημα δύο κατηγοριών. [6]

3.2 Η ΑΞΙΟΠΙΣΤΙΑ ΤΑΞΙΝΟΜΗΣΗΣ ΣΤΗ ΜΕΘΟΔΟ K-NN



Σχ. 3-1 : Εικόνα της k-NN προσέγγισης για τον καθορισμό της αξιοπιστίας ταξινόμησης $q(\vec{x})$. Το δείγμα που ερευνούμε φαίνεται με την μαύρη τελεία.

Δοθέντος ενός συγκεκριμένου χώρου με καθορισμένη μονάδα μέτρησης απόστασης, π.χ. η Ευκλείδεια απόσταση, ο εκτιμητής πυκνότητας των k -Κοντινότερων Γειτόνων (k -NN) μπορεί να χρησιμοποιηθεί για την εκτίμηση της αξιοπιστίας του δείγματος που μελετάται:

$$\hat{q}_{knn}(\vec{x}) = \frac{N_{S(\vec{x})=\omega}}{k}$$

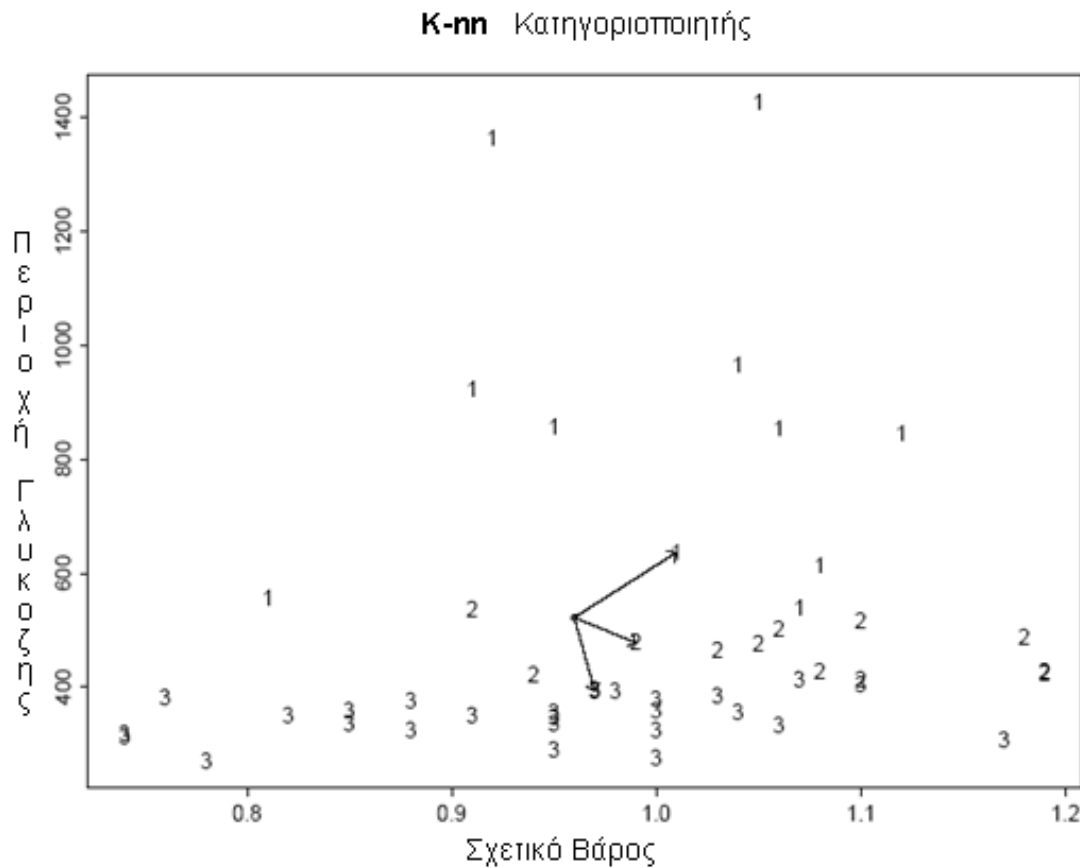
Όπου $N_{S(\vec{x})=\omega}$ είναι ο αριθμός των δειγμάτων μεταξύ των k κοντινότερων γειτόνων που μοιράζονται την ίδια ετικέτα/ κατηγορία με αυτή του δείγματος που μελετάμε. Οι

ετικέτες/κατηγορίες των γειτονικών δειγμάτων είναι γνωστές (π.χ. από το σύνολο εκπαίδευσης), ενώ η ετικέτα $\vec{x}$ προσδιορίζεται από τον ταξινομητή S . Επιπλέον, όσο περισσότερα δείγματα “συμφωνούν” με την ετικέτα του $\vec{x}$, τόσο πιο σίγουροι είμαστε ότι η ανάθεση της ετικέτας είναι σωστή. Αυτή η συμπεριφορά φαίνεται στο Σχ. 3-1. Ας υποθέσουμε ότι θέλουμε να μάθουμε την αξιοπιστία της ανάθεσης ετικέτας/κατηγορίας της μαύρης τελείας. Επιπλέον θεωρούμε ότι τα περιβάλλοντα δείγματα τους έχουν ανατεθεί μία ετικέτα κατηγορίας **A** ή **B**. Χρησιμοποιώντας την προσέγγιση των 8-NN, η αξιοπιστία που θα πρέπει να έχει για σωστή ταξινόμηση είναι 5/8, μιας και 5 από τους 8 γείτονες έχουν ετικέτα κατηγορίας **A**. Το πρόβλημα επιλογής του k παραμένει, αλλά ένας γενικός κανόνας είναι να χρησιμοποιήσουμε $k = \sqrt{N_{\mathcal{L}}}$, όπου $N_{\mathcal{L}}$ είναι ο αριθμός των δειγμάτων εκμάθησης.

Ένα μειονέκτημα αυτής της μεθόδου είναι το γεγονός ότι είναι υπολογιστικά πολύ απαιτητικό για μεγάλα σύνολα δεδομένων.

3.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΜΕ ΤΗ ΜΕΘΟΔΟ ΤΟΥ k -NN

3.3.1 Εφαρμογή k NN για την κατηγοριοποίηση ασθενή ως διαβητικού

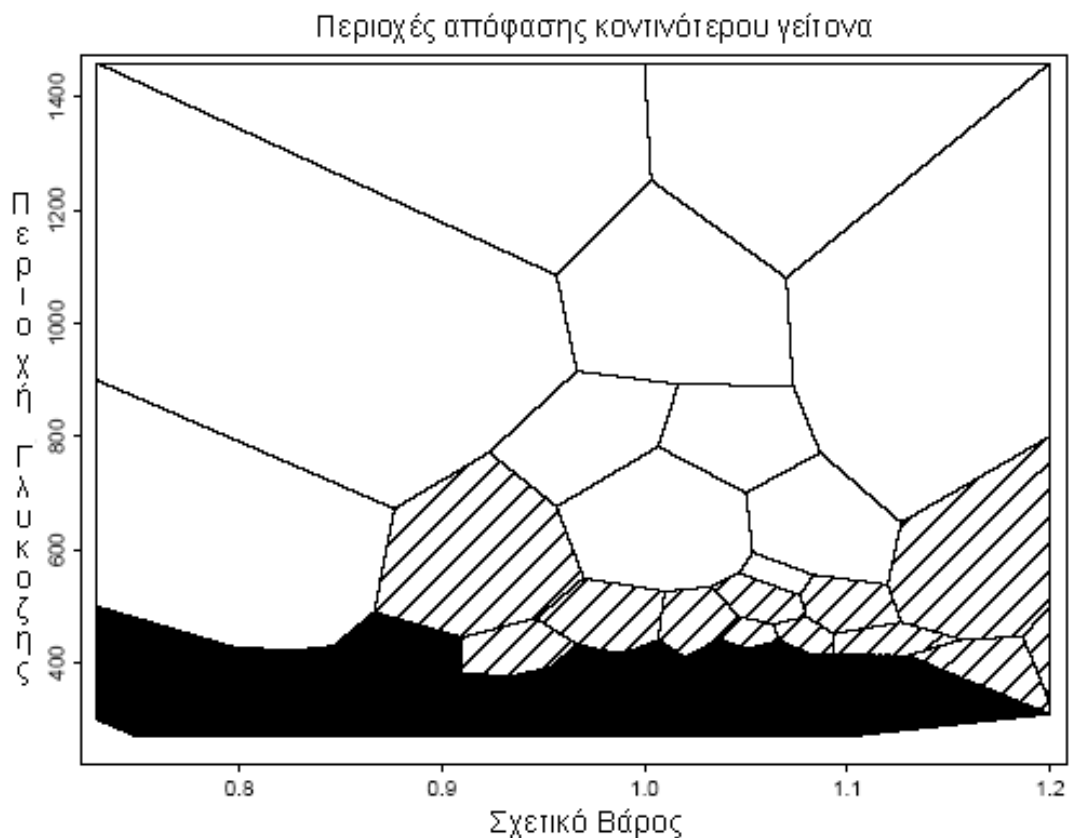


Σχ.3-2 : Ταξινομητής κοντινότερου γείτονα για ένα τεστ παράδειγμα

Το παράδειγμα που ακολουθεί δείχνει πώς λειτουργεί ο ταξινομητής κοντινότερου γείτονα ($k=1$). Τα δεδομένα είναι ένα τυχαίο υποσύνολο του συνόλου δεδομένων 36 στην μελέτη των Andrews και Herzberg (1985) [6] η οποία εξετάζει την σχέση μεταξύ του διαβήτη σε λανθάνουσα μορφή και του αποδεδειγμένου διαβήτη σε 145 ασθενείς. Για ευκολία παρουσίασης, χρησιμοποιήσαμε μόνο 50 ασθενείς και δύο από τις έξι μεταβλητές, το Σχετικό Βάρος και την Περιοχή της Γλυκόζης, και τα δεδομένα φαίνονται στο Σχ.3-2. Οι ταξινομήσεις των 50 ασθενών γίνονται σε μία από τις παρακάτω κατηγορίες : (1) αποδεδειγμένου διαβητικού, (2) λανθάνοντος διαβητικού, (3)

όχι διαβητικού. Ένας νέος ασθενής, του οποίου η κατάσταση υποτίθεται ότι είναι άγνωστη, κατηγοριοποιείται όπως ο κοντινότερός του γείτονας στο σχήμα. Η απόσταση, όπως μετρήθηκε σε κάθε σημείο, πρέπει να βαθμονομηθεί με τέτοιο τρόπο ώστε να ληφθεί υπόψιν και για διαφορετικές διακυμάνσεις σε διαφορετικές κατευθύνσεις. Σε αυτή την περίπτωση ο ασθενής ταξινομείται στην κατηγορία 2, και ταξινομείται σωστά.

Οι περιοχές απόφασης για τον κοντινότερο γείτονα συνθέτονται με γραμμικά περιοριστικά όρια και οι οποίες μπορεί να είναι αποσυνδεδεμένες περιοχές. Αυτές οι περιοχές είναι η ένωση των Dirichlet κελιών; κάθε κελί αποτελείται από σημεία τα οποία είναι κοντινότερα (με βάση κάποια συγκεκριμένη μονάδα) σε μία δοσμένη παρατήρηση από οποιαδήποτε άλλη. Για τα παραπάνω δεδομένα σκίασαμε κάθε κελί σύμφωνα με την κατηγορία του κέντρου του, και ως αποτέλεσμα έχουμε τις περιοχές αποφάσεις του Σχ.3-3.



Σχ.3-3 : Περιοχές απόφασης για τον ταξινομητή κοντινότερου γείτονα.

3.3.2 Ταξινομητής k-NN για την κατηγοριοποίηση κειμένου

Η κατηγοριοποίηση κειμένου είναι η διαδικασία ανάθεσης των κατάλληλων κατηγοριών σε ένα έγγραφο κειμένου. Αυτή η ταξινόμηση έχει πολλές εφαρμογές όπως η δρομολόγηση εγγράφων, η διαχείριση εγγράφων ή η διάδοση εγγράφων. Συνήθως, κάθε έγγραφο που εισέρχεται, αναλύεται και κατηγοριοποιείται χειροκίνητα από εμπειρογνώμονες με βάση το περιεχόμενο του εγγράφου. Μία μεγάλη ποσότητα ανθρώπινου δυναμικού ασχολείται για την ολοκλήρωση αυτής της εργασίας. Για την διευκόλυνση της κατηγοριοποίησης ενός κειμένου, απαιτούνται κάποιες μέθοδοι αυτόματης κατηγοριοποίησης. Ο στόχος της κατηγοριοποίησης κειμένου είναι η εκμάθηση τέτοιων μεθόδων κατηγοριοποίησης που μπορούν να χρησιμοποιηθούν για να ταξινομούμε έγγραφα κειμένου αυτόματα.

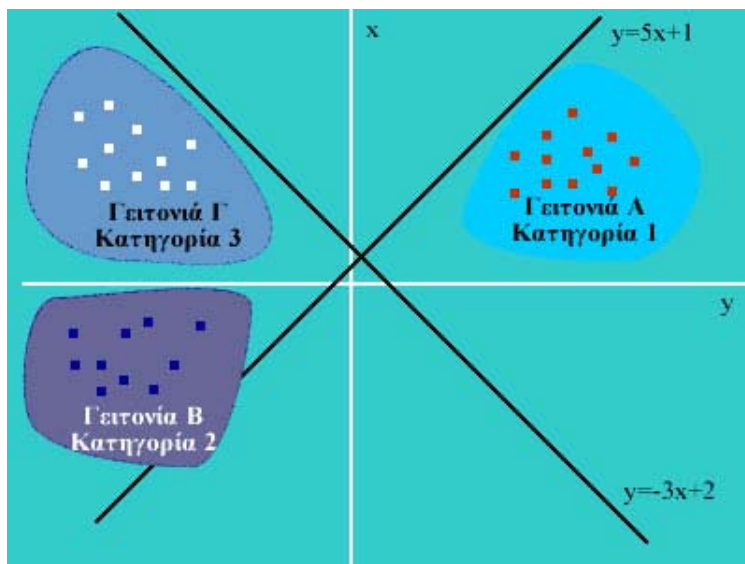
Υπάρχουν πολλές μέθοδοι κατηγοριοποίησης για την διεργασία της ταξινόμησης κειμένου. Μερικές από αυτές είναι και οι παρακάτω : το Νευρωνικό Δίκτυο, Ταξινομητές Απόφασης Δένδρου, Κανόνες Απόφασης, Ταξινομητές κ-Κοντινότερων Γειτόνων.

3.4 ΣΥΜΠΕΡΑΣΜΑΤΑ ΓΙΑ ΤΗ ΜΕΘΟΔΟ ΤΟΥ ΚΟΝΤΙΝΟΤΕΡΟΥ ΓΕΙΤΟΝΑ

Η μέθοδος k-NN είναι γενικά πολύ αργή για πολύ μεγάλα σύνολα δεδομένων. Ωστόσο, είναι γνωστό (Hart, 1968), ότι μπορεί να επιτευχθεί σημαντική εξοικονόμηση χρόνου, με κόστος μια μικρή απώλεια στην ακρίβεια. Αυτό μπορεί να γίνει με χρήση μιας περιληπτικής έκδοσης των δεδομένων εκπαίδευσης. Το πεδίο που απαιτεί επιπλέον μελέτη είναι η κατάλληλη επιλογή, σε γρήγορες μεθόδους που βασίζονται σε δεδομένα, των μονάδων μέτρησης απόστασης, της επιλογής μεταβλητής και του κατάλληλου αριθμού γειτόνων.

Όταν οι αναλογίες των μεταβλητών δεν είναι σημαντικές, όπως σε σύνολα δεδομένων για αναγνώριση αντικειμένων, η μέθοδος k-NN είναι η καλύτερη επιλογή.

ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ

3.5.1 Παράδειγμα 1^ο

Σχήμα 3-4

Χρησιμοποιούμε τη συνάρτηση $y=5x+1$ για να πάρουμε τιμές από το **χώρο Α** (κατηγορία 1).

Για το **χώρο Β** (κατηγορία 2) χρησιμοποιούμε πάλι την ίδια συνάρτηση αλλά για διαφορετικές τιμές του x .

Στη συνέχεια με την συνάρτηση $y = -3x+2$, παίρνω σημεία στο **χώρο Γ** (κατηγορία 3).

Τέλος, εισάγουμε στο σύστημα κάποια άγνωστα σημεία και τα ταξινομούμε με βάση τις παραπάνω γειτονιές. Τα αποτελέσματα είναι τα παρακάτω :

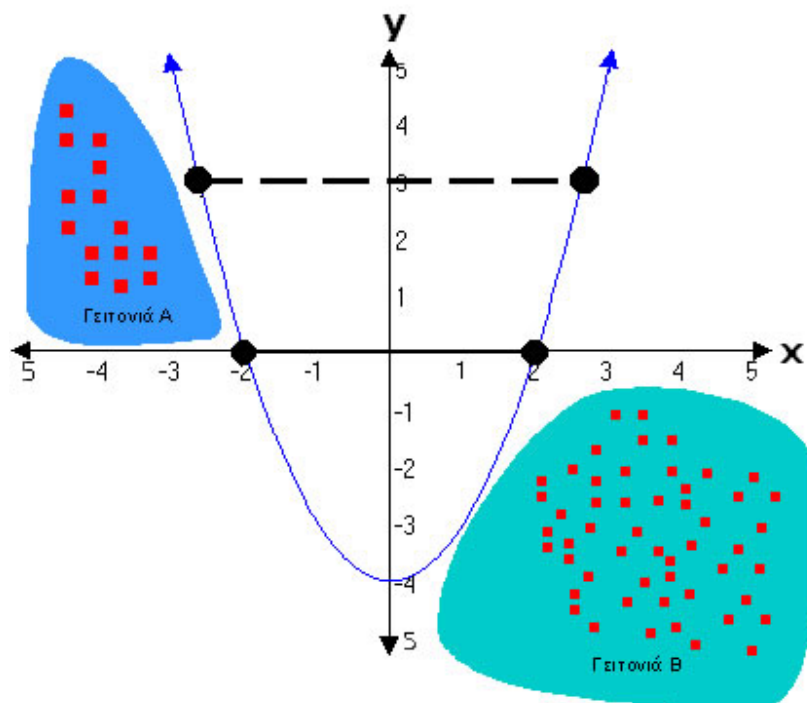
Αποτελέσματα ταξινόμησης

Πρότυπο Βάσης		Κατηγορία
X	Y	

17	80	1
9	41	1
-16	-19	2
-4	-13	2
12	49	1
6	24	1
20	80	1
-10	-4	2
-2	-11	2
-5	-18	2
-3	10	3
-17	30	3
-4	18	3
-19	40	3
-21	80	3

Τα αποτελέσματα της ταξινόμησης είναι σωστά. Αυτό μπορεί να φανεί και διαισθητικά από το σχήμα 3-4.

3.5.2 Παράδειγμα 2^ο



Σχ. 3-5

Η συνάρτηση που χρησιμοποιώ είναι η $y=x^2-4$ (παραβολή)

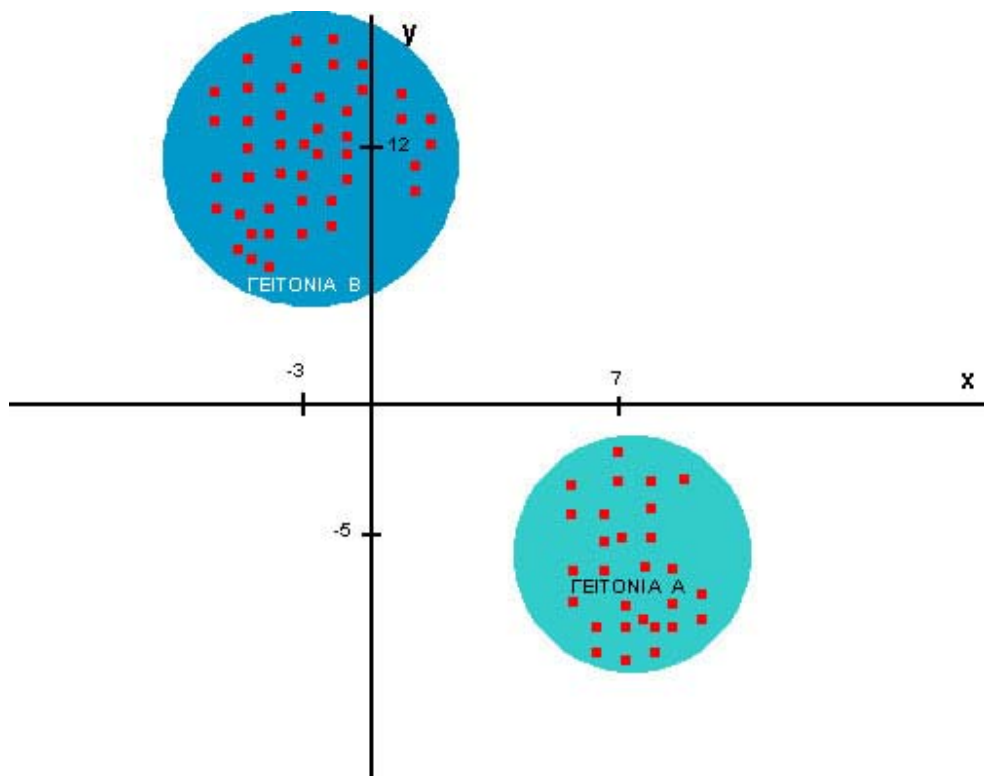
Αποτελέσματα ταξινόμησης

Πρότυπο Βάσης		Κατηγορία
X	Y	
-7	25	1
-8	12	1
-9	33	1
-11	21	1
-15	5	1
6	-7	2
13	-20	2

30	-41	2
27	-10	2
8	-24	2

Τα αποτελέσματα της ταξινόμησης είναι σωστά, όπως μπορούμε να διαπιστώσουμε και από το σχήμα 3-5.

3.5.3 Παράδειγμα 3^ο



Σχήμα 3-6

Σε αυτό το παράδειγμα χρησιμοποιώ την συνάρτηση $(x-7)^2 + [(y-(-5))]^2 = 4^2$ που είναι η εξίσωση ενός κύκλου και αναπαριστά την γειτονιά Α (κατηγορία 1).

Για την γειτονιά B (κατηγορία 2), κάνω χρήση της εξίσωσης κύκλου $[x-(-3)]^2 + (y-12)^2 = 6^2$

Αποτελέσματα ταξινόμησης

Πρότυπο Βάσης		Κατηγορία
X	Y	
5	-3	1
7	-8	1
9	-3	1
4	-6	1
5	-6	1
-5	10	2
-3	14	2
-4	9	2
-1	11	2
-2	13	2

Τα αποτελέσματα της ταξινόμησης είναι σωστά, όπως μπορούμε να δούμε και στο Σχ3-6.

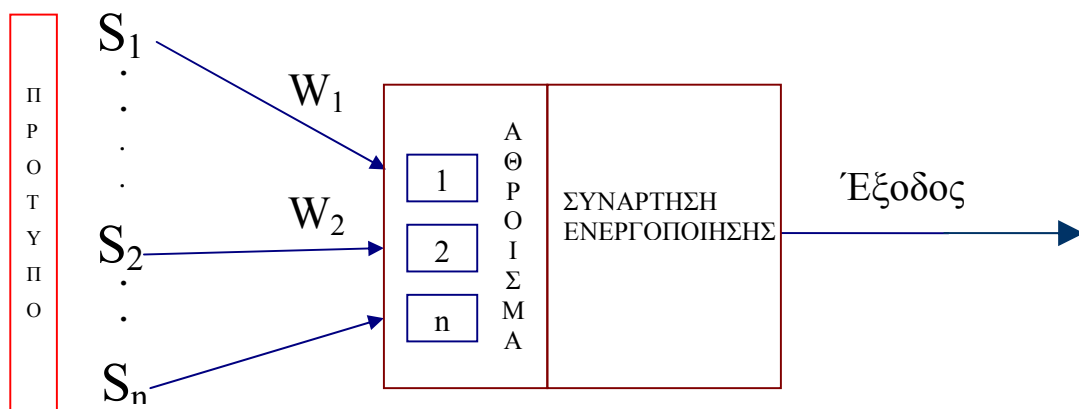
Κεφάλαιο 4

ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ ΕΝΟΣ ΕΠΙΠΕΔΟΥ

4.1 ΕΙΣΑΓΩΓΗ

Τα νευρωνικά δίκτυα ενός επιπέδου αποτελούν μαζί με το τεχνητό νευρώνα (Σχ4) την πιο απλή μορφή νευρωνικών δικτύων που χρησιμοποιείται για ταξινόμηση προτύπων τα οποία είναι γραμμικά διαχωρίσιμα.. Ο τεχνητός νευρώνας σχεδιάστηκε με σκοπό να μιμείται και να προσομοιώνει τα χαρακτηριστικά. Αντίστοιχα τα νευρωνικά δίκτυα ενός επιπέδου λειτουργούν με τον ίδιο τρόπο με τη διαφορά ότι μπορούν και χειρίζονται διανύσματα. [4]

Η τοπολογία δικτύου ενός κρυφού επιπέδου είναι αυτή που φαίνεται στο παρακάτω σχήμα (Σχ.4-1).



Σχ.4-1 Τεχνητός Νευρώνας

Κάθε πρότυπο (διάνυσμα αποτελούμενο από συνιστώσες (S_i)) εφαρμόζεται στην είσοδό και ακολουθεί την διαδικασία που επιβάλλει ο **αλγόριθμος εκπαίδευσης** του δικτύου. Από το σχήμα είναι εμφανές πως η έξοδος του δικτύου επηρεάζεται από τα

συναπτικά βάρη (W_i) των νευρώνων και τη **συνάρτηση ενεργοποίησης**. Παρά την ποικιλία που υπάρχει στα παραδείγματα των νευρωνικών δικτύων. Σχεδόν όλα στηρίζονται σε αυτό τον τρόπο δομής, ακολουθώντας διαφορετική τοπολογία.

4.2 ΠΕΡΙΓΡΑΦΗ ΤΕΧΝΙΚΗΣ

4.2.1 Αλγόριθμος εκπαίδευσης

Ο αλγόριθμος εκπαίδευσης που ακολουθεί το νευρωνικό για την εκπαίδευση του μπορεί να χωριστεί σε βήματα:

Βήμα 1^ο: Θεωρείτε το σύνολο των προτύπων εκπαίδευσης που αποτελείται από n συνιστώσες το καθένα.

Κάθε πρότυπο υπακούει στις απαιτήσεις μιας συνάρτησης. Αυτό επιτρέπει μια εκ των προτέρων γνώση της κατηγορίας στην οποία ανήκει το πρότυπο. Χαρακτηριστικό της μάθησης με επίβλεψη.

Βήμα 2^ο: Αρχικοποιούνται τα βάρη των νευρώνων στο διάστημα $[0,1]$ ή στο $[-1,1]$. Η αρχικοποίηση επιτρέπει στο δίκτυο να ξεκινήσει την εκπαίδευση με τα βάρη τυχαία αρχικοποιημένα. Τα βάρη είναι εκείνα που αλλάζουν όταν κάποιο πρότυπο εκπαίδευσης το επιβάλει. Μάλιστα, τα βάρη είναι εκείνα που ελέγχονται στο τέλος κάθε εποχής προκειμένου να καθοριστεί το τέλος της διαδικασίας της εκπαίδευσης.

Βήμα 3^ο: Έπειτα τα πρότυπα εκπαίδευσης (Z_i) οδηγούνται στο δίκτυο:

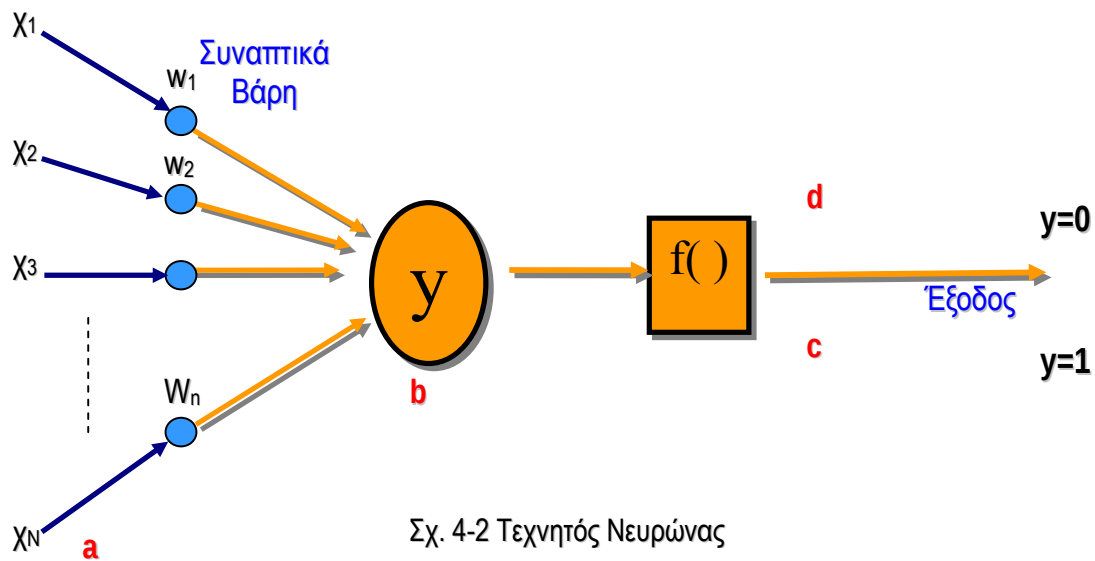
- a. Επιλέγεται ένα πρότυπο τυχαία
- b. Και υπολογίζεται το άθροισμα

$$f(y) = W * X = y = \chi_1 * w_1 + \chi_2 * w_2 + \dots + \chi_n * w_n$$
- c. Αν το $f > 0$ τότε $f = 1$ αλλιώς $y = 0$
- d. Γίνεται προσαρμογή των συναπτικών βαρών των νευρώνων με βάση το τύπο $W_i = W_i + n(f - k)Z_i$.

Εάν υπάρχει μεταβολή στα βάρη (W_i) από το πέρασμα του προτύπου τότε επαναλαμβάνεται η διαδικασία από το 3^ο βήμα και η εκπαίδευση τελειώνει όταν

επηρεάσουν όλα τα πρότυπα το δίκτυο. Το n αντιπροσωπεύει το ρυθμό μάθησης του δικτύου, συνήθως παίρνει τιμές μεταξύ (0,1).

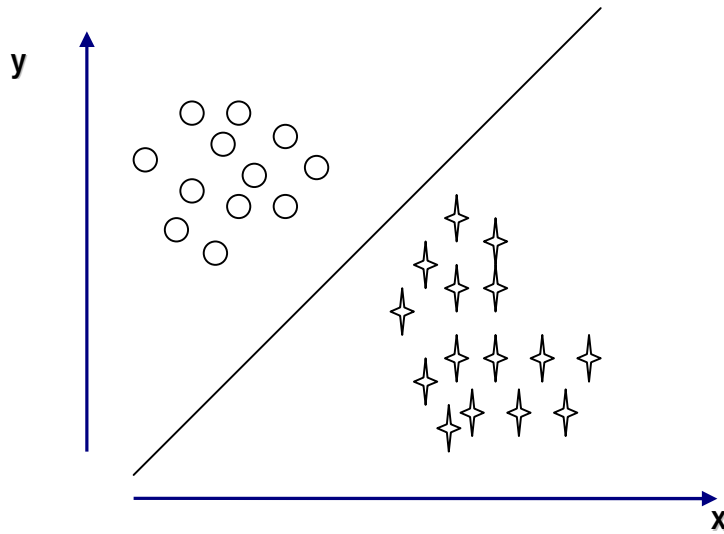
Τα πρότυπα (διανύσματα) δημιουργούν το άθροισμα y το οποίο επηρεάζει την $f(y)$ (συνάρτηση ενεργοποίησης). Η οποία επιστρέφει 0 ή 1 ανάλογα με τη τιμή του y . Έτσι επηρεάζονται και τα βάρη σύμφωνα με το τύπο που δόθηκε στο d. του 3^ο βήματος.



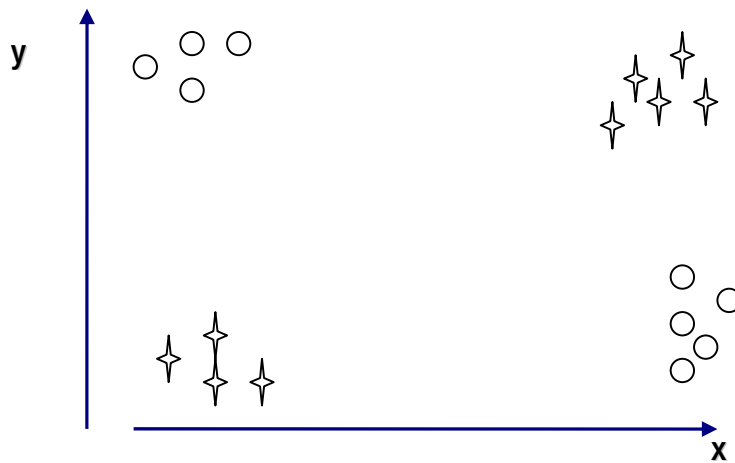
Ο αλγόριθμος εκπαίδευσης δεν χρησιμοποιείται για όλες τις περιπτώσεις δικτύων και προβλημάτων. Είναι συγκεκριμένα τα παραδείγματα στα οποία μπορεί να γενικεύσει ένα δίκτυο που εκπαιδεύτηκε με τον αλγόριθμο εκπαίδευσης που περιγράψαμε.

4.2.2 Σε ποια προβλήματα μπορεί να χρησιμοποιηθούν τα νευρωνικά δίκτυα ενός επιπέδου

Από πολύ νωρίς κάπου στα 1969 ο Marvin Minsky και ο Seymour Papert απέδειξαν πως τα νευρωνικά δίκτυα ενός επιπέδου δεν μπορούσαν να μάθουν για παράδειγμα την σύνθετη συνάρτηση XOR. [5] Απέδειξαν μάλιστα πως τα δίκτυα αυτής της τοπολογίας είναι σε θέση να κατηγοριοποιήσουν πρότυπα που είναι γραμμικά διαχωρίσιμα.



Σχ.4-3 Γραμμικά διαχωρίσιμο πρόβλημα



ΣΧ.4-4 Γραμμικά μη διαχωρίσιμο πρόβλημα XOR

Παρατηρήσεις

1. Τα νευρωνικά δίκτυα ενός επιπέδου επιλέγουμε να εκπαιδεύονται σταδιακά. Αν εκπαιδεύτούν κατά δέσμες, τότε οι προσαρμογές των συναπτικών βαρών γίνονται αφού περάσουν στο δίκτυο όλα τα παραδείγματα του συνόλου εκπαίδευσης και πραγματοποιούνται με συνυπολογισμό (άθροισμα) των επιμέρους

αλλαγών που επιφέρουν τα παραδείγματα όταν εμφανίζονται στην είσοδο του δικτύου. Δεν είναι όμως σίγουρο ότι ένα τέτοιος αλγόριθμος εκπαίδευσης θα συγκλίνει.

2. Η έξοδος μπορεί να πάρει μόνο τις τιμές 0 ή 1 λόγω της συνάρτησης κατωφλίου που χρησιμοποιεί ως συνάρτηση μεταβίβασης στους νευρώνες και ιδιαίτερα στο επίπεδο εξόδου. Το γεγονός αυτό περιορίζει τα δίκτυα σε προβλήματα ταξινόμησης, στα οποία ο χώρος των διανυσμάτων εισόδου είναι γραμμικά διαχωρίσιμος.
3. Η ταχύτητα σύγκλισης του αλγόριθμου εκπαίδευσης επηρεάζεται από την ύπαρξη ακραίων διανυσμάτων εισόδου στο σύνολο εκπαίδευσης. Ως ακραία εννοούμε εδώ διανύσματα των οποίων τα επιμέρους στοιχεία (συνιστώσες) διαφέρουν πολύ από τις τιμές των υπολοίπων διανυσμάτων του συνόλου εκπαίδευσης. Αυτό συμβαίνει γιατί, όταν ένα ακραίο διάνυσμα εμφανίζεται στην είσοδο του δικτύου επιφέρει σημαντικές αλλαγές στα συναπτικά βάρη, σε τέτοιο μάλιστα βαθμό που τα άλλα διανύσματα χρειάζεται να εμφανισθούν αρκετές φορές στην είσοδο του δικτύου (σε πολλούς κύκλους εκπαίδευσης) προκειμένου να επαναπροσαρμόσουν τα συνοπτικά βάρη στα μέτρα τους. Ένας τρόπος αντιμετώπισης του φαινομένου αυτού είναι η χρήση του κανονικοποιημένου κανόνα μάθησης, σύμφωνα με τον οποίο η μεταβολή στα συνοπτικά βάρη ορίζεται από τη σχέση

$$\Delta W = e(P^T / \|P\|)$$

Βέβαια σε πολλές άλλες περιπτώσεις η απλή εξαίρεση ακραίων προτύπων βελτιώνει την απόδοση του δικτύου.

4.2.3 Συνάρτηση ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης χρησιμοποιούνται από τα νευρωνικά δίκτυα και λέγονται και αλλιώς συναρτήσεις μεταφοράς. Οι συναρτήσεις τροφοδοτούνται με δεδομένα που προκύπτουν από το εσωτερικό γινόμενο του βάρους και του προτύπου. Ανάλογα με το είδος της συνάρτησης που έχουμε επιλέξει παρατηρούμε και την αντίστοιχη έξοδο.

Βηματική Συνάρτηση 0/1:

$$f(x)=0 \quad \text{αν } x \leq 0$$

$$f(x)=1 \quad \text{αν } x > 0$$

Βηματική Συνάρτηση -1/1:

$$f(x)= -1 \quad \text{αν } x \leq 0$$

$$f(x)=1 \quad \text{αν } x > 0$$

Συμμοειδής Συνάρτηση:

$$f(x)=1/(1+e^{-x})$$

Υπερβολική Εφαπτόμενη:

$$f(x)=\tanh(x)=(1-e^{-x})/(1+e^{-x})$$

Γραμμική Συνάρτηση:

$$f(x)=x$$

4.2.4 Συναπτικά Βάρη νευρώνων δικτύου

Όπως έχουμε αναφέρει ήδη και στη περιγραφή του αλγορίθμου εκπαίδευσης του δικτύου τα βάρη αρχικοποιούνται τυχαία στο διάστημα $[0,1]$ ή στο $[-1,1]$. Επηρεάζονται κατά τη διάρκεια της εκπαίδευσης από τα πρότυπα. Κάθε πρότυπο μεταβάλλει τα βάρη αν το δίκτυο δεν έχει ολοκλήρωση την εκπαίδευση του. Τα πρότυπα όπως έχουμε αναλύσει αποτελούνται από συνιστώσες και σύμφωνα με τον αλγόριθμο εκπαίδευσης, πολλαπλασιάζονται με τα αντίστοιχα βάρη και το γινόμενο υπόκειται μια σειρά από ελέγχους. Ανάλογα με το αποτέλεσμα των ελέγχων αυτών επηρεάζονται και τα βάρη (σύμφωνα με το τύπο $\mathbf{W}_i = \mathbf{W}_i + \mathbf{n}(\mathbf{f} - \mathbf{k})\mathbf{Z}_i$).

Από τη διαδικασία που περιγράψαμε φαίνεται πως τα βάρη καθορίζουν την ολοκλήρωση της εκπαίδευσης και βέβαια την απόδοση του δικτύου. Αφού τα βάρη μετά το τέλος της εκπαίδευσης χρησιμοποιούνται στη λειτουργία του δικτύου.

4.2.5 Πόλωση

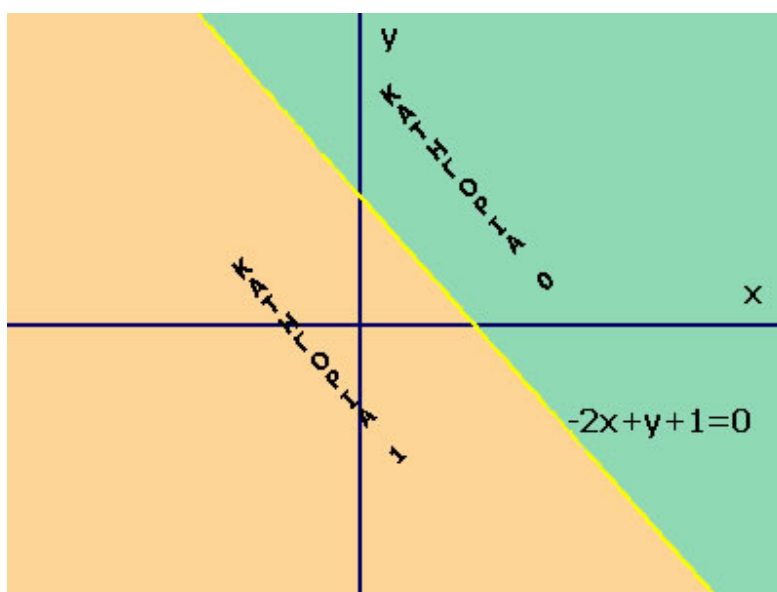
Ένα από τα σοβαρότερα προβλήματα που προκύπτει στη διαδικασία της εκμάθησης στα νευρωνικά δίκτυα ονομάζεται «υπερεκπαίδευση» (overfitting). Το πρόβλημα αυτό δημιουργείται σε περιπτώσεις όπου το δίκτυο δεν μπορεί να γενίκευση σε πρότυπα που δεν του έχουν δοθεί στη διάρκεια της εκπαίδευσης. Με άλλα λόγια μπορεί να αποδώσει μόνο σε ήδη γνωστό σύνολο προτύπων σε αυτό. Δεδομένου ότι η βασικότερη λειτουργία των νευρωνικών δικτύων είναι η γενίκευση, τέτοια προβλήματα είναι καλό να ανιχνεύονται και να εξετάζονται προσεκτικά.

Ένα ακόμη πολύ σημαντικό χαρακτηριστικό των νευρωνικών δικτύων όπως έχουμε αναφέρει και σε άλλο εδάφιο αναλυτικότερα είναι η πόλωση του δικτύου. Με άλλα λόγια αναφερόμαστε στην ικανότητα του νευρωνικού να προσεγγίζει τα πρότυπα και αναφερόμαστε στις παραμέτρους των προτύπων ώστε να εκπαιδεύεται σωστά και βέβαια να αποδίδει στο στάδιο της γενίκευσης.

4.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΗΣ

4.3.1 Παράδειγμα 1^ο

Είδος προβλήματος: Κατηγοριοποίηση σημείων που ανήκουν σε επίπεδο το οποίο χωρίζει μια ευθεία σε δύο χωρία (Δύο διαφορετικές κατηγορίες). Όπως φαίνεται στο (Σχ. 4-5). Η κατηγορίες ορίζονται από μια ευθεία την $f(x) = -2x + y - 1 = 0$ η οποία χωρίζει το καρτεσιανό επίπεδο σε δύο χωρία (Κατηγορία 0 και 1).



Σχ. 4-5

Διάσταση προβλήματος	200 πρότυπα εκπαίδευση. 100 γενίκευσης
Ρυθμός Μάθησης	1 ^η 0,01 - 2 ^η 0,5 - 3 ^η 0,9
Συνάρτηση Ενεργοποίησης	Βηματική συνάρτηση

Ο πίνακας δείχνει αναλυτικά τα πρότυπα (x,y) που δόθηκαν με τυχαία σειρά στο σύστημα και την αντίστοιχη έξοδο του συστήματος.

Με έντονο είναι οι «λάθος εκτιμήσεις» του συστήματος για τον αντίστοιχο ρυθμό μάθησης ενώ στο τέλος υπάρχει και το ποσοστό του συνολικού σφάλματος του συστήματος για τον κάθε ρυθμό.

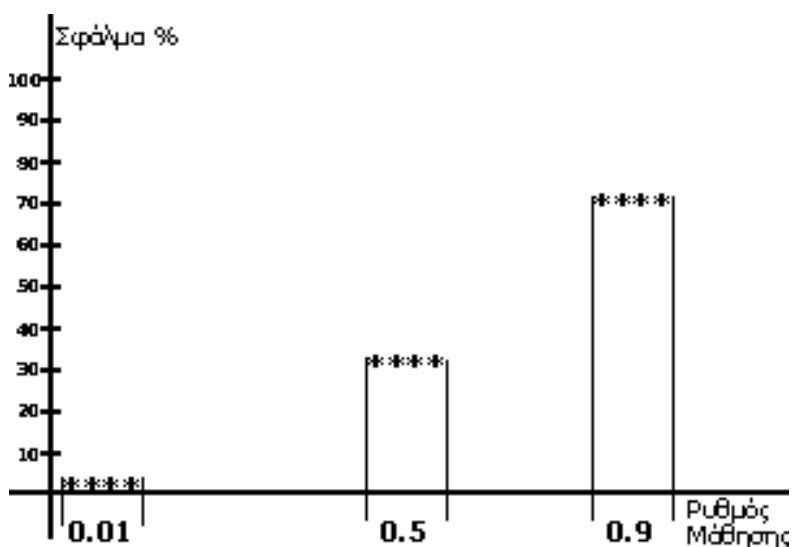
0,01	0,5	0,9
1	1	1
0	1	1
0	1	1
0	0	0
0	0	0
0	0	0

0	0	0
0	0	0
0	0	0
0	0	0
1	1	1
1	1	1
0	1	1
0	1	1
0	1	1
0	1	1
0	0	1
0	0	1
0	0	1
0	1	1
0	0	0
0	0	1
0	0	1
1	1	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
0	0	1
1	1	1
0	1	1

1	1	1
1	1	1
1	0	0
1	0	0
1	0	0
1	0	0
1	1	1
1	0	0
1	1	1
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	0	0
1	1	0
1	1	0
1	1	0
1	1	0
1	1	0
1	1	1
1	1	0
1	0	0
1	1	0
1	0	0
1	1	0
1	1	0

1	0	0
1	0	0
1	0	0
6/100=0,06 6%	33/100=0,33 33%	71/100=0,71 71%

Παρατηρούμε πως ο μικρότερος ρυθμός μάθησης από τους τρεις (0,01) παρουσιάζει και το μικρότερο σφάλμα. Γενικά ο «μικρός» ρυθμός μάθησης επιτρέπει στο δίκτυο να «μαθαίνει σταδιακά» μεταβάλλοντας κάθε φορά ελάχιστα τα βάρη. Αυτό επιτρέπει στο δίκτυο να προσαρμόζεται καλύτερα, ώστε στο τέλος της εκπαίδευσης του να έχει φτάσει σε αρκετή μεγάλη ακρίβεια πρόβλεψης. Βέβαια είναι σημαντικό να λαμβάνονται υπόψη τα στοιχεία του συστήματος. Τα πρότυπα εκπαίδευσης και τα βάρη. Μάλιστα η έντονη διαφοροποιήσεις των βαρών οφείλεται στα πρότυπα. Καταλήγοντας να πούμε πως η επιλογή του ρυθμού μάθησης είναι πολύ σημαντικός παράγοντας και καθορίζει τη δυνατότητα του νευρωνικού δικτύου να γενικεύει σωστά. Και ανάλογα με τα όρια ανοχής σφάλματος που θέτουμε επιλέγουμε ρυθμό μάθησης.

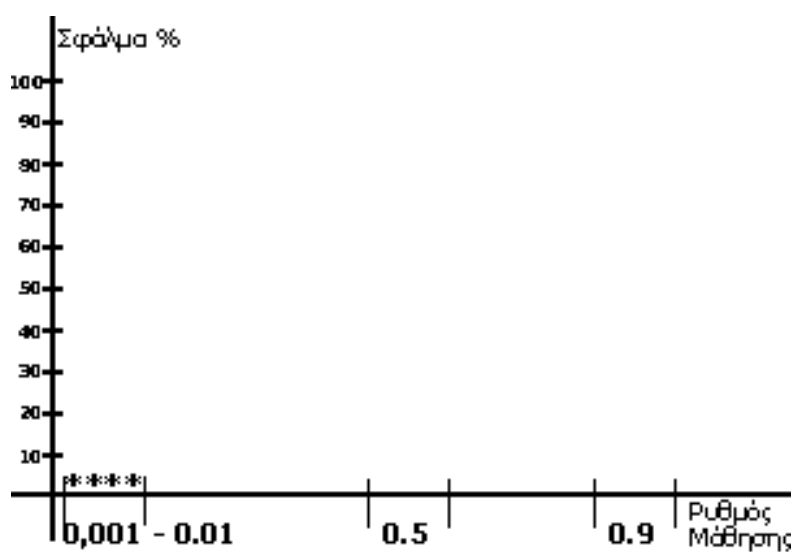


Σχ. 4-6 Αποτελέσματα 1^ο παραδείγματος

4.3.2 Παράδειγμα 2^ο

1	1	1	1
1	1	1	1
1	1	1	1
0	1	1	0
2%	0	0	2%

Παρατηρούμε πως το μικρότερο σφάλμα δημιουργείται με τους δυο μεγαλύτερους ρυθμούς μάθησης (0,5 – 0,9) και μάλιστα στο συγκεκριμένο σύνολο προτύπων έχουμε μηδενικό σφάλμα.. Το αποτέλεσμα αυτό είναι απόρροια του συνόλου εκπαίδευσης και των βαρών. Από το παράδειγμα αυτό καταλαβαίνουμε πως ο ρυθμός μάθησης δεν υπακούει σε «φόρμουλες», αλλά πρέπει σε κάθε περίπτωση να επιλέγεται εκείνος που δίνει το μικρότερο σφάλμα και επιτρέπει καλύτερη απόδοση δικτύου.



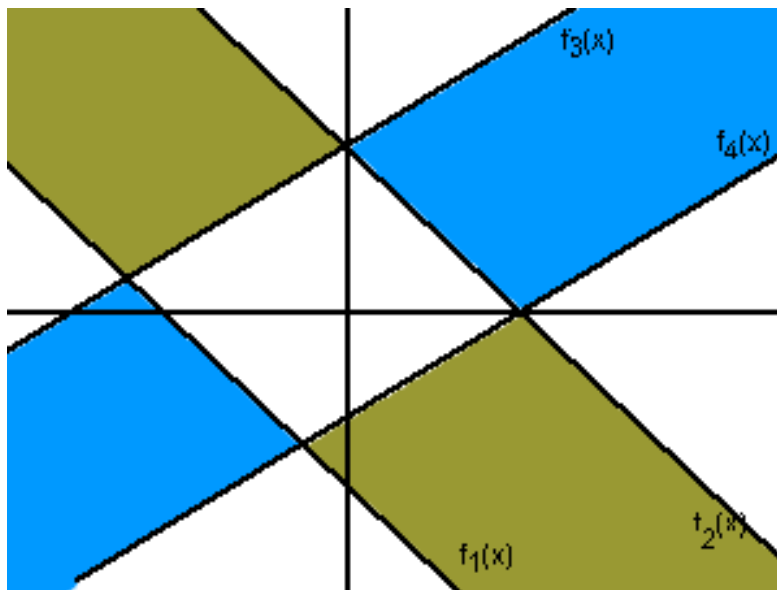
Σχ. 4-7 Αποτελέσματα 2^ο παραδείγματος

4.3.3 Παράδειγμα 3^ο

Το πρόβλημα είναι ο ορισμός προτύπων που ανήκουν σε κατηγορίες μη γραμμικά διαχωρίσιμες. Όπως φαίνεται και από το σχήμα τα πρότυπα που έχουν επιλεγεί βρίσκονται στα 4 διαφορετικά χωρία τα οποία αν δύο δημιουργούν μια κατηγορία. Οι συναρτήσεις που τα ορίζουν είναι οι :

- ✓ $f_1(x)=5x-2$,
- ✓ $f_2(x)=5x+10$
- ✓ $f_3(x)=-2x+1$
- ✓ $f_4(x)=-2x-10$

Όπως φαίνονται και στο σχήμα.



Σχ 4-8

Διάσταση προβλήματος	60 πρότυπα εκπαίδευση. 40 γενίκευσης
Πρότυπα, Βάρη	Παράρτημα
Ρυθμός Μάθησης	1 ^η 0,01 - 2 ^η 0,5 - 3 ^η 0,9
Συνάρτηση Ενεργοποίησης	Βηματική συνάρτηση

1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	0
1	0	0
1	0	0
1	0	0
0	0	0
0	0	0
0	0	0
0	0	0
35%	42.5%	45%

Παρατηρούμε τα ποσοστά σφάλματος είναι πολύ μεγάλα. Αυτό συμβαίνει γιατί το παράδειγμα αυτό χειρίζεται πρότυπα τα οποία δεν είναι γραμμικά διαχωρίσιμα. Μάλιστα ομοιάζει με ένα πρόβλημα που έχουμε αναφέρει (XOR) και στο οποίο τα νευρωνικά δίκτυα ενός επιπέδου δεν μπορούν να αποδώσουν. Το αποτέλεσμα λοιπόν του σφάλματος είναι αναμενόμενο.

ΠΑΡΑΡΤΗΜΑ

ΥΛΟΠΟΙΗΣΗ

5.1 ΠΕΡΙΓΡΑΦΗ ΚΑΤΗΓΟΡΙΑΣ ΚΟΝΤΙΝΟΤΕΡΩΝ ΓΕΙΤΟΝΩΝ

Η υλοποίηση της μεθόδου κατηγοριοποίησης των κ- Κοντινότερων Γειτόνων έγινε σε κώδικα C++.

- Ο κώδικας αποτελείται από ένα αρχείο επικεφαλίδας (header file) το οποίο ονομάζεται **knn.h**. Σ' αυτό έχουμε συμπεριλάβει μια κατηγορία που ονομάζεται KNN.

Στα ιδιωτικά μέλη της έχουμε δηλώσει τα παρακάτω πεδία :

1. Πεδίο **“protupa”** που είναι τύπου ακεραίου και σ' αυτό αποθηκεύεται κάθε φορά το τρέχον πλήθος των προτύπων.
2. Πεδίο **“diastasi”** επίσης τύπου ακεραίου όπου αποθηκεύεται ο αριθμός των συνιστωσών του κάθε προτύπου. Όταν αναφερόμαστε σε συνιστώσες εννοούμε π.χ. το πλήθος των διαστάσεων ενός σημείου που είναι το πρότυπο. Σε πιο πρακτικό επίπεδο, οι συνιστώσες είναι τα διαφορετικά χαρακτηριστικά ενός αντικειμένου.
3. Πεδίο **“geitones”** τύπου ακεραίου η οποία δηλώνει το πλήθος των κοντινότερων γειτόνων σε σχέση με το πρότυπο που έχουμε ορίσει σαν βάση.
4. Πεδίο **“categories”** τύπου ακεραίου με την οποία υποδηλώνουμε το πλήθος των κατηγοριών μεταξύ των οποίων μπορούν να καταταχθούν τα πρότυπα.

Εκτός από τα τέσσερα βασικά αυτά πεδία, έχουμε δηλώσει και έξι πίνακες vectors.

1. Τον πίνακα **“pin”** όπου αποθηκεύονται τα πρότυπα με τις συνιστώσες τους.
 2. Τον πίνακα **“dist”** όπου αποθηκεύονται οι αποστάσεις του σημείου βάση σε σχέση με όλα τα υπόλοιπα πρότυπα που έχουμε.
 3. Τον πίνακα **“categ”** όπου διατηρούμε τις κατηγορίες στις οποίες ανήκουν τα αρχικά σημεία του προγράμματος.
-

Οι υπόλοιποι τρεις πίνακες είναι βοηθητικοί και χρησιμοποιούνται στην πορεία του κώδικα.

Στα δημόσια μέλη της έχουμε δηλώσει τις παρακάτω συναρτήσεις :

1. Συνάρτηση δημιουργίας
 2. Συνάρτηση δημιουργίας απο αρχείο
 3. Συνάρτηση δημιουργίας αντιγράφου
 4. Συνάρτηση αποδιάθρωσης
 5. Συνάρτησεις για την προσπέλαση των ιδιωτικών μεταβλητών *protupa*, *diastasi*, *geitones*, *categories*.
 6. Μία συνάρτηση *TaksinomisiApoArxeio* όπου εκτελείται ο βασικός κώδικας. Εδώ βρίσκουμε τους *k* - κοντινότερους γείτονες του προτύπου που ψάχνουμε και εντέλει και την κατηγορία στην οποία ανήκει.
 7. Μία συνάρτηση *EisagwgiProtStoTelos* όπου εισάγουμε ένα επιπλέον πρότυπο.
 8. Μία συνάρτηση *DiagrafiProtupou* που χρησιμοποιείται για να διαγράψουμε κάποιο πρότυπο.
 9. Μία συνάρτηση *EpistrefeiCategProt* που επιστρέφει την κατηγορία κάποιου απο τα ήδη υπάρχοντα πρότυπα.
 10. Μία συνάρτηση *ApothikeusiKNN* που αποθηκεύει τις συνιστώσες των προτύπων και τις αντίστοιχες κατηγορίες τους σε ένα αρχείο κειμένου.
- a. Επίσης έχουμε ένα αρχείο κειμένου (***data.txt***) όπου έχουμε αποθηκεύσει τα πρότυπα με τις αντίστοιχες συνιστώσες τους.
- b. Στο αρχείο ***knn.cpp*** βρίσκεται όλος ο κώδικας που εκτελείται κάθε φορά που θέλουμε να κατηγοριοποιήσουμε ένα σημείο. Η διαδικασία ταξινόμησης περιλαμβάνει τα παρακάτω στάδια :
1. Αρχικά γεμίζουμε τον πίνακα προτύπων με τα πρότυπα που παίρνουμε από το αρχείο. Αντίστοιχα τον πίνακα κατηγοριών αποθηκεύοντας την κατηγορία στην οποία ανήκει το κάθε πρότυπο.
-

2. Στη συνέχεια βρίσκουμε την απόσταση του προτύπου βάση σε σχέση με όλα τα υπόλοιπα πρότυπα-σημεία.
3. Έχοντας φτάσει ως εδώ, για να βρούμε τους 3 στο συγκεκριμένο παράδειγμα κοντινότερους γείτονες, ταξινομούμε τον πίνακα των αποστάσεων από το μικρότερο στο μεγαλύτερο. Όποια αλλαγή επέλθει στον πίνακα των αποστάσεων, κάνουμε την αντίστοιχη (από πλευράς θέσης) στον πίνακα των κατηγοριών.



Παρατηρούμε ότι ταξινομώντας τους πίνακες παράλληλα, μπορούμε αμέσως με βάση τον πίνακα των κατηγοριών και το πλήθος των κοντινότερων γειτόνων να δώσουμε μία απάντηση στο ερώτημα ποια είναι η κατηγορία του προτύπου που αναζητούμε.

4. Ποτέ δεν λαμβάνουμε υπόψιν το πρώτο στοιχείο στους δύο τελικούς πίνακες, γιατί αφορά την απόσταση του σημείου από τον εαυτό του. Συνεπώς για 3 κοντινότερους γείτονες, παρατηρούμε το 2^ο, 3^ο και 4^ο στοιχείο αντίστοιχα στον πίνακα των κατηγοριών. Με βάση τον κανόνα πλειοψηφίας, δηλαδή ποια κατηγορία εμφανίζεται τις περισσότερες φορές, κατηγοριοποιούμε το πρότυπο βάση στην κατάλληλη κατηγορία.

5.2 ΠΕΡΙΓΡΑΦΗ ΚΑΤΗΓΟΡΙΑΣ ΝΕΥΡΩΝΙΚΟΥ ΔΙΚΤΥΟΥ ΕΝΟΣ ΕΠΙΠΕΔΟΥ

Η υλοποίηση της κατηγορίας έγινε με χρήση γλώσσας C++.

Η κατηγορία περιέχει ένα αρχείο επικεφαλίδας (header file) το **hopfield.h**. Σ' αυτό έχουμε συμπεριλάβει μια κατηγορία που ονομάζεται hopfield.

Στο πεδίο δηλώσεων των ιδιωτικών μελών (private) έχουμε τα παρακάτω μέλη :

- ✓ **diastasi** τύπου ακεραίου (int): Περιγράφει την διάσταση του προτύπου δηλαδή το συνολικό αριθμό συνιστώσεων κάθε προτύπου
- ✓ **protupa** τύπου ακεραίου (int): Περιγράφει το σύνολο των προτύπων που χρησιμοποιεί η εφαρμογή για την εκπαίδευση της .
- ✓ **rithmos** τύπου ακεραίου (int): Περιγράφει το ρυθμό μάθησης του δικτύου είναι ένας αριθμός στο διάστημα (0,1)
- ✓ **categories** τύπου ακεραίου (int): Περιγράφει το σύνολο των διαφορετικών κατηγοριών στις οποίες ανήκουν τα πρότυπα
 - ο Πίνακας (vector) **pin** τύπου ακεραίου (int): Κρατάει στις θέσεις μνήμης του τις συνιστώσες των προτύπων. Η διάσταση του είναι (πρότυπα * συνιστώσες)
 - ο Πίνακας (vector) **categ** τύπου ακεραίου (int): Κρατάει τις κατηγορίες των προτύπων . Η διάσταση του είναι (πρότυπα)
 - ο Πίνακας (vector) **varos** τύπου δεκαδικού (float). Διατηρεί τα βάρη για το κάθε πρότυπο. Εχει ίδια διάσταση με το pin (πρότυπα * συνιστώσες)
 - ο Πίνακας (vector) **protupo_vasi** τύπου ακεραίου (int): Κρατάει τις συνιστώσες του προτύπου γενίκευσης, του οποίου τη κατηγορία ψάχνουμε.

Στο πεδίο δηλώσεων των δημοσίων μελών (public) έχουμε τα παρακάτω μέλη :

- ✓ **hopfield ()**: Συνάρτηση δημιουργίας (constructor) που αρχικοποιεί σε NULL τα πεδία της κατηγορίας.
 - ✓ **~hopfield()**: Συνάρτηση αποδιάρθρωσης (destructor) που διαγράφει το αντικείμενο της κατηγορίας.
 - ✓ **hopfield (char * arxeio, char * arxeion)**: Συνάρτηση δημιουργίας (constructor) που διαβάζει δύο αρχεία που αρχικοποιεί της μεταβλητές του δικτύου και τα δεδομένα τα λαμβάνει από τα δύο αρχεία εισόδου.
-

- ✓ **hopfield(const hopfield &test):** Συνάρτηση δημιουργίας (constructor) που αντιγράφει από ένα «άλλο» hopfield τις τιμές των πεδίων του.
- ✓ **int append(char * arxeio, char * arxeiop):** Μέθοδος προσθήκης ενός προτύπου σε ένα άλλο αρχείο
- ✓ **int epis_kat_prot(int prot):** Μέθοδος που δέχεται σαν είσοδο έναν αριθμό που αντιστοιχεί σε κάποιο πρότυπο και επιστρέφει την κατηγορία στην οποία καταλήγει το νευρωνικό δίκτυο ότι ανήκει. Σύμφωνα πάντα με την εκπαίδευση που είχε.
- ✓ **void training():** Μέθοδος που εφαρμόζει τον αλγόριθμο εκπαίδευσης που έχουμε περιγράψει.
- ✓ **void apothikeusi_hopfield(string arxeio):** Μέθοδος που αποθηκεύει τα μέλη ενός αντικειμένου hopfield σε ένα αρχείο για μελλοντική χρήση ή και για έλεγχο.
- ✓ **void digrafi_protupou(int m):** Μέθοδος που διαγράφει ένα πρότυπο από το αρχείο των προτύπων

5.3 ΕΓΚΑΤΑΣΤΑΣΗ ΠΑΚΕΤΟΥ

5.3.1 Εγκατάσταση πακέτου στα Windows

Μετά την δημιουργία του dll πακέτου από τα αρχεία cpp και h, ουσιαστικά έχουμε στην διάθεσή μας μία συλλογή συναρτήσεων μέσα σε ένα μόντουλο το οποίο μπορεί να χρησιμοποιηθεί από άλλα προγράμματα. Για την εγκατάσταση της βιβλιοθήκης αυτό που πρέπει απλά να κάνουμε είναι να αντιγράψουμε τα αρχεία .dll και .lib που δημιουργήσαμε στον φάκελο όπου βρίσκονται όλα τα αρχεία πηγαίου κώδικα του προγράμματος που θέλει να καλέσει την βιβλιοθήκη μας. Στη συνέχεια η διαδικασία που ακολουθούμε για να χρησιμοποιήσουμε την βιβλιοθήκη είναι απλή. Μέσα από τον κώδικα του προγράμματος καλούμε / προσθέτουμε το όνομα της βιβλιοθήκης και με αυτόν τον τρόπο μπορούμε να χρησιμοποιήσουμε τις συναρτήσεις σε να τις είχαμε δηλώσει.

5.3.2 Εγκατάσταση πακέτου στα Linux

Με αντίστοιχο τρόπο γίνεται και η εγκατάσταση και χρήση της βιβλιοθήκης από εφαρμογές που εκτελούνται στο περιβάλλον Linux.

Βιβλιογραφία

Βιβλιογραφία

1. DARPA Neural Network Study (1988, AFCEA International Press, p. 60)
2. Haykin, S. (1994), Neural Networks: A Comprehensive Foundation, NY: Macmillan, p. 2
3. Training and memory of a neural network, Daniel Franklin (2003)
4. Hinton, G.E. (1992), "How Neural Networks Learn from Experience", Scientific American, 267 (September), 144-151.
5. Definition of Neural network - wordIQ Dictionary & Encyclopedia.htm
6. Machine Learning, Neural and Statistical Classification, Φεβρουάριος 17, 1994-Συντάκτες : D.Michie, D.J. Spiegelhalter, C.C. Taylor,
7. Generalization in Neural Networks, Συντάκτες : Ali A. Ghorbani, Kiarash Owrangh.
8. Wolpert, D.H. (ed.) (1995b), The Mathematics of Generalization: The Proceedings of the SFI/CNLS Workshop on Formal Approaches to Supervised Learning, Santa Fe Institute Studies in the Sciences of Complexity, Volume XX, Reading, MA: Addison-Wesley.
9. Ελαχιστοποίηση της μεταβλητότητας με περιορισμούς στο θόρυβο και στην πόλωση, Σεβεντεκίδης Κώστας, Τζίκας Δημήτρης, Ιωάννινα 28-01-2003.
10. Neural Networks and the Bias/Variance Dilemma, *Neural Computation*, S. Geman, E. Bienenstock and R. Doursat (1992).
11. KNN Model-Based Approach in Classification.
Συντάκτες : Gongde Guo, Hui Wang, David Bell, Yaxin Bi και Kieran Greer.
12. Generalisation in Feed Forward Neural Classifiers, Aarnoud Hoekstra.
13. Stacked Generalization in Neural Networks, Ali A. Ghorbani και Kiarash Owrangh.

URL's

1. <http://www.uab.edu/cogdev/index.htm>
2. <http://www-cse.stanford.edu/classes/sophomore-college/projects-00/neural-networks/History/index.html>
3. <http://www-cse.stanford.edu/classes/sophomore-college/projects-00/neural-networks/Future/index.html>
4. <http://www.dacs.dtic.mil/techs/neural/neural8.html>
5. <ftp://ftp.sas.com/pub/neural/FAQ.html>
6. <http://www.neural-networks-at-your-fingertips.com/hopfield.html>
7. [http://koti.mbnet.fi/~phodju/nenet/NeuralNetworks/Hopfield/Networks Information Homepage.htm](http://koti.mbnet.fi/~phodju/nenet/NeuralNetworks/Hopfield/Networks%20Information%20Homepage.htm)
8. <http://www.montefiore.ulg.ac.be/~lwh/Papers/EIS1994/node5.html>
9. http://www.fact-index.com/u/un/unsupervised_learning.html

10. <http://www.cs.bham.ac.uk/~jxb/inn.html>
11. http://en.wikipedia.org/w/wiki.phtml?title=Analytical_learning&action=edit
12. http://en.wikipedia.org/wiki/Instantaneously_trained_neural_networks
13. <http://www.mathworks.com/access/helpdesk/help/toolbox/nnet/netob13a.html>
14. http://ieee.uow.edu.au/~daniel/software/libneural/BPN_tutorial/BPN_English/BPN_English/node5.html

Ηλεκτρονικά βιβλία

1. Stacked Generalization in Neural Networks:Generalization on Statistically Neutral Problems.pdf , Ali A. Ghorbani, Kiarash Owrangh1 School of Computer Engineering Iran University of Science and Tech
 2. Το δίκτυο Hopfield.pdf, Κώστας Διαμαντάρας Τμήμα Πληροφορικής, ΤΕΙ Θεσσαλονίκης
 3. Comparison of genetic algorithm, Hopfield and MLP Neural Network techniques for a constrained optimization problem.pdf ,H. Altun and T. Yalcinoz Dept. of Electrical & Electronic Eng., Nigde University, Nigde 51100, Turkey
 4. An Improved *k*-Nearest Neighbor Algorithm for Text Categorization.pdf ,Li Baoli1, Yu Shiwen1, and Lu Qin Institute of Computational Linguistics Department of Computer Science and Technology
 5. Java Neural Networks with Java.pdf
 6. An introduction to Neural Networks.pdf Benne Krose and Patrick van der Smagt
 7. Introduction to Neural Networks.pdf 2nd Year UG, Msc in Computer Sience
 8. Machine Learning, Neural and Statistical Classification.pdf , D. Michie, D.J. Spiegelhalter, C.C. Taylor February 17, 1994
-