

ΚΕΦΑΛΑΙΟ 1

ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ ΚΑΙ ΣΥΝΘΕΣΗΣ ΑΠΟΦΑΣΕΩΝ

1.1. Εισαγωγή.

Η ικανότητα των μηχανών να αντιλαμβάνονται το περιβάλλον τους είναι σχετικά περιορισμένη. Με την βοήθεια αισθητήρων (transducers) φυσικά μεγέθη όπως το φως, ο ήχος ή θερμοκρασία κ.λ.π, μπορούν να μετατραπούν σε ηλεκτρικά σήματα. Στην συνέχεια με κατάλληλη επεξεργασία από ηλεκτρονικό σύστημα θα πρέπει να αναλυθούν και να κατανοηθούν. Βοήθεια στη διαδικασία ανάλυσης της πληροφορίας και αναγνώρισης αντικειμένων έχουμε λάβει τα τελευταία χρόνια με την εξέλιξη των υπολογιστών. Η δυνατότητα αποθήκευσης και επεξεργασίας μεγάλου όγκου δεδομένων έχει προσδώσει ανεκτίμητες δυνατότητες στην προσπάθεια εξομοίωσης του ανθρώπινου τρόπου σκέψης από τον ηλεκτρονικό υπολογιστή, προσπάθεια που δεν έχει ουσιαστικά αποδώσει ακόμα.

Η έλλειψη μιας ολοκληρωμένης θεωρίας για τον τρόπο αντίληψης από τον άνθρωπο δεν έχει εμποδίσει, παρόλα αυτά, την επίλυση προβλημάτων μικρής ως μέτριας δυσκολίας. Πολλά από τα προβλήματα αυτά σχετίζονται με την αναγνώριση και ταξινόμηση αντικειμένων (pattern recognition and classification). Η λεπτομερής μελέτη των προβλημάτων ταξινόμησης έχει οδηγήσει από νωρίς σε μαθηματική θεωρία που παρέχει την βάση για τον σχεδιασμό των συστημάτων ταξινόμησης.

Σε αυτό το κεφάλαιο θα παρουσιάσουμε στοιχεία τόσο από την κλασική θεωρία ταξινόμησης όσο και από σύγχρονες διαδικασίες που περιλαμβάνουν κυρίως νευρωνικά δίκτυα. Επίσης θα παρουσιάσουμε τις βασικές αρχές της σύνθεσης πληροφορίας (κυρίως σύνθεσης αποφάσεων) που χρησιμοποιούνται όταν επιθυμούμε να συνδυάσουμε δεδομένα που προέρχονται από διάφορες πηγές. Στόχος είναι να προσδώσουμε μεγαλύτερη αξιοπιστία σε ένα σύστημα σε σχέση με αντίστοιχα που χρησιμοποιούν μια μόνο πηγή πληροφορίας.

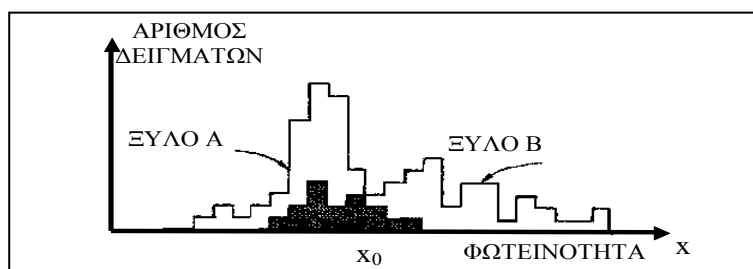
Ένα χαρακτηριστικό παράδειγμα

Στην συνέχεια θα δούμε ένα απλό παράδειγμα ταξινόμησης. Έστω ότι ένα πριονιστήριο που παράγει προϊόντα ξυλείας θέλει να αυτοματοποιήσει την διαδικασία της ταξινόμησης των ξύλων σύμφωνα με τον τύπο τους (διαχωρισμός του ξύλου Α από το ξύλο Β). Ένα σύστημα που θα μπορούσε να εκτελέσει αυτό τον διαχωρισμό φαίνεται στο σχήμα 1.1. Η κάμερα λαμβάνει την εικόνα του ξύλου και την μεταβιβάζει σε ένα εξαγωγέα χαρακτηριστικών (feature extractor), που σκοπό έχει να ελαττώσει την διαθέσιμη πληροφορία μετασχηματίζοντας τα δεδομένα σε μια πιο συνοπτική μορφή (άνυσμα). Τα χαρακτηριστικά εισέρχονται σε κατάλληλο ταξινομητή (classifier) που αποφασίζει σχετικά με τον τύπο του ξύλου.



Σχήμα 1.1. Ένα απλό σύστημα ταξινόμησης.

Ο τρόπος με τον οποίο σχεδιάζονται ο εξαγωγέας χαρακτηριστικών και ο ταξινομητής έχει ως εξής: Έστω ότι το ξύλο Α είναι πιο φωτεινόχρωμο από το ξύλο Β. Τότε είναι προφανές ότι η φωτεινότητα, οι τιμές της οποίας αντιστοιχούν στη μεταβλητή x , αποτελεί ένα χαρακτηριστικό για τον διαχωρισμό του τύπου των ξύλων. Θα ταξινομήσουμε τα ξύλα με βάση το αν η τιμή της φωτεινότητας ξεπερνάει μια τιμή κατωφλίου x_0 . Για να εκλέξουμε την τιμή του x_0 θα πρέπει να μετρήσουμε μερικά δείγματα της φωτεινότητας x και να δούμε τα αποτελέσματα. Το σχήμα 1.2 δείχνει το ιστόγραμμα της φωτεινότητας ή την κατανομή του κάθε τύπου ξύλου ως συνάρτηση της φωτεινότητας (υπό συνθήκη πυκνότητα πιθανότητας).



Σχήμα 1.2. Ιστόγραμμα φωτεινότητας.

Από την μορφή των ιστογραμμάτων συμπεραίνουμε ότι το ξύλο Β είναι πιο φωτεινό από το ξύλο Α. Συμπεραίνουμε επίσης, ότι το χαρακτηριστικό αυτό από μόνο του δεν

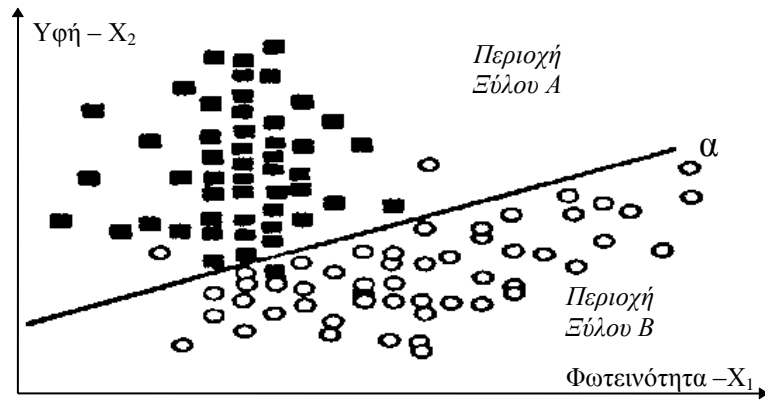
αρκεί για έναν χωρίς λάθη διαχωρισμό των τύπων των ξύλων αφού υπάρχουν περιοχές στις οποίες δεν μπορούμε να αποφασίσουμε καθώς οι τιμές της φωτεινότητας και για τα δύο ξύλα συμπίπτουν.

Στην προσπάθεια για αναζήτηση άλλων χαρακτηριστικών μπορούμε να δώσουμε έμφαση στην προεξέχουσα υφή των ξύλων (τα γνωστά νερά). Το χαρακτηριστικό αυτό είναι πολύ πιο δύσκολο να μετρηθεί από την μέση φωτεινότητα. Είναι όμως λογικό να υποθέσουμε ότι μπορούμε να έχουμε ένα μέτρο σύγκρισης της υφής εξετάζοντας τις μεταβάσεις από φωτεινή σε σκοτεινή περιοχή. Τώρα λοιπόν έχουμε δύο χαρακτηριστικά για την ταξινόμηση του τύπου του ξύλου. Έτσι ο εξαγωγέας χαρακτηριστικών έχει μετατρέψει κάθε εικόνα σε ένα σημείο ή διάνυσμα χαρακτηριστικών \mathbf{x} σε ένα χώρο δυο διαστάσεων όπου:

$$\tilde{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \text{Φωτεινότητα} \\ \text{Υφή} \end{bmatrix}$$

Το πρόβλημα πλέον είναι να διαχωριστεί ο χώρος των χαρακτηριστικών σε δύο περιοχές όπου όλα τα σημεία για τον ένα τύπο ξύλου να είναι σε μία περιοχή ενώ όλα τα σημεία για τον άλλο τύπο ξύλου να είναι σε μια άλλη περιοχή. Ας υποθέσουμε ότι μετρούμε τα διανύσματα χαρακτηριστικών για τα δείγματα εισόδου και λαμβάνουμε τις απεικονίσεις τους στο επίπεδο όπως φαίνεται στο σχήμα 1.3.

Η μορφή του σχήματος μπορεί να δώσει τον επόμενο κανόνα για την ταξινόμηση οποιουδήποτε τύπου ξύλου που μπορεί να εμφανιστεί: Ταξινόμησε το ξύλο ως τύπου A εάν το διάνυσμα χαρακτηριστικών βρίσκεται πάνω από την γραμμή a ή ως τύπου B σε αντίθετη περίπτωση. Το πρόβλημα της ταξινόμησης είναι βασικά ο διαχωρισμός του χώρου των χαρακτηριστικών σε περιοχές όπου η κάθε μια χαρακτηρίζει διαφορετική τάξη. Σε αυτήν την περίπτωση το πρόβλημα της ταξινόμησης γίνεται πρόβλημα στατιστικής θεωρίας αποφάσεων. Είναι φανερό ότι το πρόβλημα της εξαγωγής χαρακτηριστικών εξαρτάται από την εφαρμογή.



Σχήμα 1.3. Διάγραμμα διασποράς για τα διανύσματα χαρακτηριστικών.

1.2. Κλασσική θεωρία ταξινόμησης.

1.2.1. Θεωρία αποφάσεων του Bayes.

Η θεωρία αποφάσεων του Bayes αποτελεί τη συμβατική στατιστική προσέγγιση στο πρόβλημα της θεωρίας ταξινόμησης. Για να περιγραφεί με απλό τρόπο η θεωρία του Bayes, θα ξεκινήσουμε την ανάλυση με το παράδειγμα της παραγράφου 1.1. Εκεί είχαμε σκοπό τον σχεδιασμό ενός συστήματος που να διαχωρίζει δυο είδη ξύλων. Υποθέτουμε ότι ένας παρατηρητής που κοιτάει τα διάφορα ξύλα που περνάνε από το πριονιστήριο δεν μπορεί να προβλέψει τι είδους ξύλο θα περάσει μετά, καθώς θεωρεί ότι η ακολουθία των ξύλων είναι τυχαία. Χρησιμοποιώντας ορολογία στατιστικής λέμε ότι καθώς ένα καινούργιο ξύλο μπαίνει στο πριονιστήριο είναι δυνατόν να επιλεγεί μια από τις δύο δυνατές κατηγορίες ξύλου: Ας ονομάσουμε ω την μεταβλητή που απεικονίζει την κατηγορία ξύλου, με $\omega=\omega_1$ να αναπαριστά το είδος του ξύλου Α και $\omega=\omega_2$ να αναπαριστά το είδος του ξύλου Β. Λόγω του ότι η φύση είναι απρόβλεπτη στην εκλογή της λέμε ότι η μεταβλητή ω είναι μια τυχαία μεταβλητή.

Εάν το πριονιστήριο παρήγαγε τόση ξυλεία από το πρώτο είδος όση και από το δεύτερο είδος, θα μπορούσαμε να πούμε ότι το επόμενο κομμάτι που θα εισέλθει στο πριονιστήριο θα είναι είτε του τύπου Α είτε του τύπου Β με πιθανότητα ίδια και για τα δύο ξύλα. Γενικότερα μιλώντας υποθέτουμε ότι υπάρχει μια εκ των προτέρων πιθανότητα (*a-priori*) $P(\omega_1)$ το επόμενο κομμάτι ξύλου να είναι τύπου Α και μια εκ των προτέρων πιθανότητα $P(\omega_2)$ το επόμενο κομμάτι ξύλου να είναι τύπου Β. Οι *a-priori* πιθανότητες αντανακλούν την προηγούμενη γνώση μας ότι θα δούμε ένα από τα δύο είδη πριν την εμφάνιση του ξύλου στο πριονιστήριο. Είναι προφανές ότι $P(\omega_1)$ και $P(\omega_2)$ είναι μη αρνητικές ποσότητες και το άθροισμά τους ισούται με την μονάδα [99].

Ας υποθέσουμε για μια στιγμή ότι είμαστε υποχρεωμένοι να πάρουμε μια απόφαση για τον τύπο του ξύλου που θα εμφανιστεί αμέσως μετά. Αν η μόνη πληροφορία που

έχουμε αφορά τις εκ των προτέρων πιθανότητες $P(\omega_1)$ και $P(\omega_2)$ τότε θα ήτανε αρκετά λογικό να διατυπώσουμε τον ακόλουθο κανόνα αποφάσεως:

Αποφάσισε ω_1
 εάν $P(\omega_1) > P(\omega_2)$;
 αλλιώς
 αποφάσισε ω_2 .

Αυτή η διαδικασία φαίνεται λίγο παράξενη ως προς την άποψη ότι θα κάνουμε πάντα την ίδια εκτίμηση έστω και αν γνωρίζουμε ότι υπάρχουν δύο τύποι ξύλων. Το πόσο καλά αυτή η απόφαση θα δουλέψει εξαρτάται από τις τιμές των a-priori πιθανοτήτων. Εάν η τιμή του $P(\omega_1)$ είναι πολύ μεγαλύτερη από την τιμή του $P(\omega_2)$ η απόφασή μας υπέρ του ω_1 θα είναι σωστή τις περισσότερες φορές. Εάν $P(\omega_1) = P(\omega_2)$ έχουμε πιθανότητα πενήντα τοις εκατό να είμαστε σωστοί.

Γενικότερα μιλώντας η πιθανότητα της λανθασμένης απόφασης είναι η μικρότερη από τις $\{P(\omega_1), P(\omega_2)\}$.

Στις περισσότερες όμως περιπτώσεις δεν θα μας ζητηθεί να αποφασίσουμε με τόσα λίγα στοιχεία. Έτσι μπορούμε να θεωρήσουμε την φωτεινότητα x ως μια συνεχή τυχαία μεταβλητή, η κατανομή της οποίας εξαρτάται από την εκλογή της κατηγορίας του ξύλου. Η πυκνότητα πιθανότητας $p(x/\omega_j)$ αφορά την κατανομή της μεταβλητής x υπό την συνθήκη ότι έχουμε την κατηγορία ω_j (conditional density). Αν οι a-priori πιθανότητες $P(\omega_1)$ και $P(\omega_2)$ και οι υπό συνθήκη πυκνότητες πιθανότητας $p(x/\omega_1)$ και $p(x/\omega_2)$ είναι γνωστές, τότε η μέτρηση της φωτεινότητας x του ξύλου επηρεάζει την γνώμη μας για το πραγματικό είδος του ξύλου σύμφωνα με τον κανόνα του Bayes [1]:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{\sum_{j=1}^2 p(x|\omega_j)P(\omega_j)} \quad (1.1)$$

Ο κανόνας του Bayes δείχνει πως παρατηρώντας την τιμή της μεταβλητής x έχουμε αποδώσει σε κάθε μια από τις δύο κατηγορίες μια πιθανότητα να συμβεί (a-posteriori probability). Τελικά ο κανόνας της απόφασης διαμορφώνεται ως εξής:

Εάν $P(\omega_1|x) > P(\omega_2|x)$ τότε:
 επέλεξε ω_1
 αλλιώς
 επέλεξε ω_2

ή ισοδύναμα:

$$\begin{aligned} & \text{Επέλεξε } \omega_1 \\ & \text{εάν } p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2), \\ & \text{αλλιώς επέλεξε } \omega_2. \end{aligned}$$

1.2.2. Η γενικευμένη περίπτωση στην θεωρία του Bayes.

Οι ιδέες που αναφέραμε παραπάνω γενικεύονται με τις εξής προσθήκες.

- Θα επιτρέψουμε περισσότερα του ενός χαρακτηριστικά (το \mathbf{x} γίνεται άνυσμα).
- Θα επιτρέψουμε περισσότερες από δυο δυνατές κατηγορίες.

Αυτές οι γενικεύσεις και ο επιπλέον συμβολισμός που θα επιφέρουν, δεν πρέπει να μας απομακρύνουν από την γενική ιδέα που αναλύεται με το παράδειγμά μας.

Επιτρέποντας παραπάνω από ένα χαρακτηριστικό απλά αντικαθιστούμε την μεταβλητή x με ένα διάνυσμα χαρακτηριστικών (feature vector) \mathbf{x} (**\mathbf{x} : συμβολισμός με έντονα γράμματα**). Αν ονομάσουμε $\Omega = \{\omega_1, \dots, \omega_S\}$ το πεπερασμένο σύνολο των δυνατών κατηγοριών, τότε η εκ των υστέρων πιθανότητα $P(\omega_j|\mathbf{x})$ μπορεί να υπολογιστεί από την $p(\mathbf{x}|\omega_j)$ με χρήση του κανόνα του Bayes:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{\sum_{j=1}^S p(\mathbf{x}|\omega_j)P(\omega_j)} \quad (1.2)$$

Οι πιθανότητες $P(\omega_j|\mathbf{x})$ έχουν άθροισμα μονάδα.

$$\sum_{j=1}^S P(\omega_j|\mathbf{x}) = 1 \quad (1.3)$$

Θα αποφασίσουμε για την πιθανότερη κατηγορία από την μεγαλύτερη τιμή της $P(\omega_j|\mathbf{x})$ με $j=1, \dots, S$.

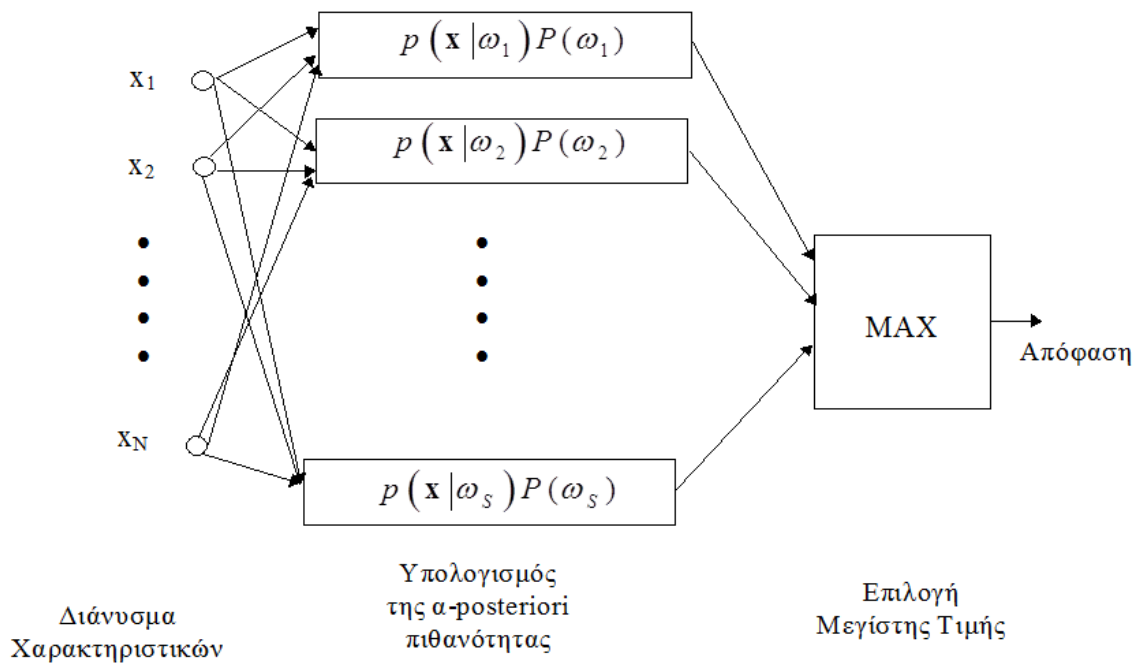
1.2.3. Ταξινομητές και συναρτήσεις διαχωρισμού.

Το σύστημα που δέχεται το άνυσμα των χαρακτηριστικών x , υπολογίζει την $P(\omega_j|\mathbf{x})$ με βάση την $p(x|\omega_j)P(\omega_j)$ και αποφασίζει για την πιθανότερη κατηγορία, καλείται ταξινομητής. Ο ταξινομητής αναθέτει το προς εξέταση διάνυσμα χαρακτηριστικών \mathbf{x} στην κατηγορία ω_i εάν και μόνο εάν:

$$p(x|\omega_i)P(\omega_i) > p(x|\omega_j)P(\omega_j) \text{ για κάθε } j \neq i, \quad j=1, \dots, S.$$

Η αναπαράσταση αυτή του ταξινομητή φαίνεται στο διάγραμμα του σχήματος 1.4. Ο ταξινομητής απεικονίζεται σαν μια μηχανή που υπολογίζει, από το διάνυσμα χαρακτηριστικών \mathbf{x} , S πυκνότητες πιθανότητας $p(x|\omega_i)P(\omega_i)$ και επιλέγει την κατηγορία που αντιστοιχεί στην μεγαλύτερη τιμή. Ο ταξινομητής Bayes απεικονίζεται εύκολα και φυσικά με αυτόν τον τρόπο.

Υπάρχουν πολλοί τρόποι να σχεδιάσουμε τους ταξινομητές. Ένας από αυτούς τους τρόπους περιγράφεται με την βοήθεια των συναρτήσεων διαχωρισμού (discriminant functions) $g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$, $i=1, \dots, s$. Το πρόβλημα της επιλογής της μορφής της συνάρτησης διαχωρισμού δεν είναι μοναδικό. Γενικότερα αν αντικαταστήσουμε την $g_i(\mathbf{x})$ με την $f(g_i(\mathbf{x}))$ όπου f είναι μια μονότονη και αύξουσα συνάρτηση θα έχουμε τα ίδια αποτελέσματα όσο αφορά την ταξινόμηση. Αυτή η ανάλυση συχνά οδηγεί σε σημαντικές απλοποιήσεις.

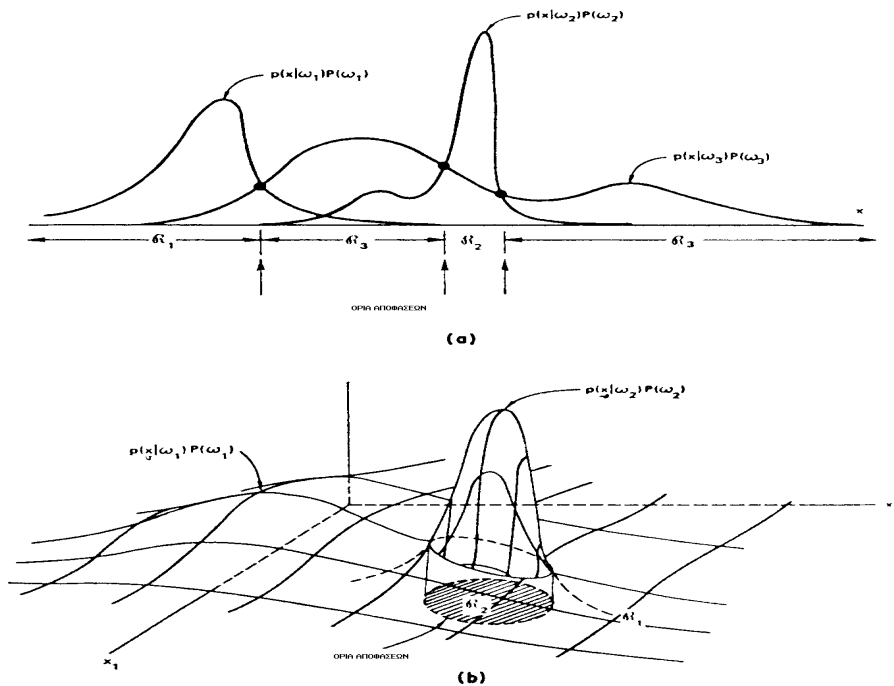


Σχήμα 1.4. Σχηματικό διάγραμμα ταξινομητή.

Το σημαντικό είναι ότι ακόμα και αν οι συναρτήσεις διαχωρισμού μπορούν να γραφτούν με μια ποικιλία μορφών, η ουσία, δηλαδή η επιλογή των κατηγοριών δεν αλλάζει. Ο σκοπός των συναρτήσεων διαχωρισμού είναι να διαιρέσει τον χώρο χαρακτηριστικών σε s -περιοχές αποφάσεων R_1, \dots, R_s . Εάν $g_i(\mathbf{x}) > g_j(\mathbf{x})$ για κάθε $j \neq i$, τότε το διάνυσμα χαρακτηριστικών \mathbf{x} είναι στην περιοχή R_i και ο κανόνας επιλογής

αναθέτει το \mathbf{x} στην κατηγορία ω_i . Οι περιοχές χωρίζονται από τα σύνορα αποφάσεων που είναι υπερεπιφάνειες στον χώρο των χαρακτηριστικών (σχήμα 1.5). Η εξίσωση για την εύρεση των συνόρων προκύπτει από την σχέση:

$$g_i(\mathbf{x}) = g_j(\mathbf{x})$$



Σχήμα 1.5. Παραδείγματα επιφανειών απόφασης.

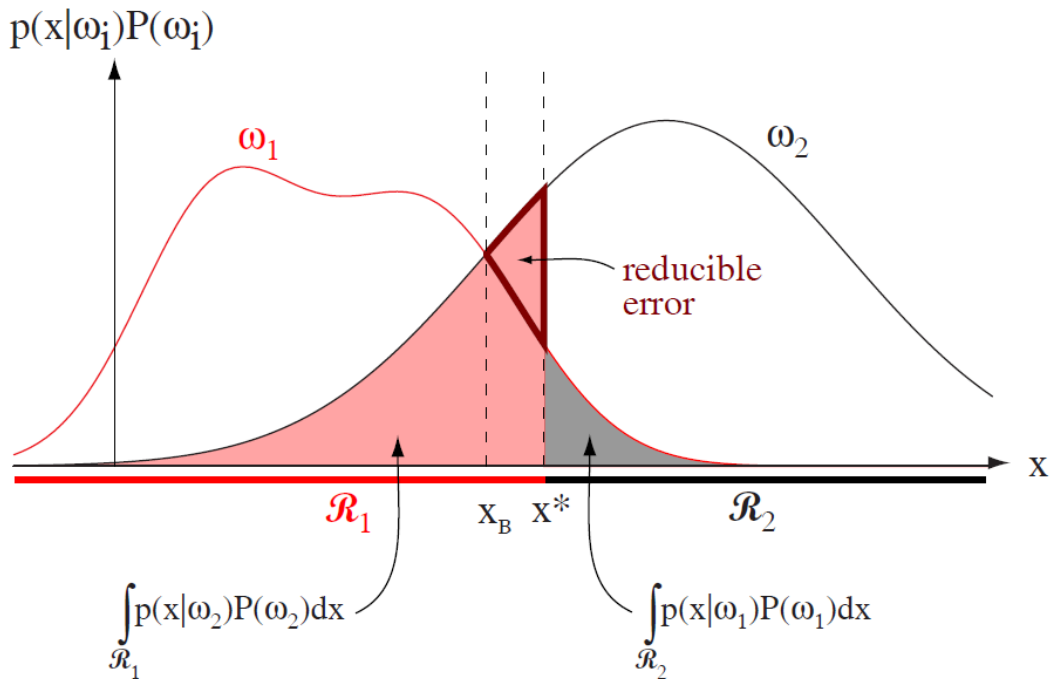
1.2.4. Πιθανότητες σφάλματος και ολοκληρώματα.

Στην περίπτωση της ταξινόμησης σε δυο κατηγορίες ο ταξινομητής έχει διαιρέσει τον χώρο σε δυο περιοχές τις οποίες ονομάζουμε R_1 και R_2 αντίστοιχα. Υπάρχουν δυο τρόποι για να συμβεί ένα λάθος στην ταξινόμηση ενός αντικειμένου. Είτε μια μέτρηση \mathbf{x} να εμπίπτει στην περιοχή R_2 ενώ η αληθινή κατηγορία να είναι η ω_1 , είτε να εμπίπτει στην περιοχή R_1 ενώ η αληθινή κατηγορία να είναι η ω_2 . Τα δυο αυτά γεγονότα είναι αμοιβαία αποκλειόμενα και διαμερίζουν πλήρως τον χώρο πιθανοτήτων, άρα:

$$\begin{aligned} P(\text{σφάλματος}) &= P(x \in R_2, \omega_1) + P(x \in R_1, \omega_2) = P(x \in R_2 | \omega_1)P(\omega_1) + P(x \in R_1 | \omega_2)P(\omega_2) \\ &= \int_{R_2} p(x \in R_2 | \omega_1)P(\omega_1)dx + \int_{R_1} p(x \in R_1 | \omega_2)P(\omega_2)dx \end{aligned}$$

Το αποτέλεσμα παρουσιάζεται για την μονοδιάστατη περίπτωση στο σχήμα 1.6. Οι δυο όροι στο άθροισμα είναι απλά τα εμβαδά στις ουρές των συναρτήσεων $p(\mathbf{x}|\omega_i)P(\omega_i)$. Λόγω της αυθαίρετης επιλογής των περιοχών R_1 και R_2 η πιθανότητα του λάθους δεν είναι όσο μικρή θα έπρεπε. Ακολουθώντας, μετακινώντας το όριο που σηματοδοτεί την λήψη

απόφασης προς την μια ή την άλλη μεριά μπορούμε να εξαλείψουμε την άνω έντονα γραμμοσκιασμένη περιοχή και να λάβουμε την μικρότερη πιθανότητα του σφάλματος.



Σχήμα 1.6. Οι συνιστώσες της πιθανότητας του σφάλματος.

1.3. Παραμετρικές και μη τεχνικές εκτίμησης των πυκνοτήτων πιθανότητας.

Στις προηγούμενες παραγράφους είδαμε πως θα μπορούσαμε να σχεδιάσουμε ένα ταξινομητή εάν γνωρίζαμε τις εκ των προτέρων πιθανότητες $P(\omega_i)$ και τις πυκνότητες πιθανότητας υπό συνθήκη $p(\mathbf{x}|\omega_i)$. Δυστυχώς σε εφαρμογές ταξινόμησης σπάνια γνωρίζουμε πλήρως την στατιστική δομή του προβλήματος.

Μια προσέγγιση για την αντιμετώπιση του προβλήματος είναι να χρησιμοποιήσουμε τα δείγματα για να εκτιμήσουμε τις άγνωστες πιθανότητες και τις πυκνότητες υπό συνθήκη. Σε μια τυπική εφαρμογή η εκτίμηση των a-priori πιθανοτήτων δεν παρουσιάζει μεγάλες δυσκολίες. Από την άλλη όμως, η εκτίμηση των πυκνοτήτων υπό συνθήκη είναι πιο δύσκολο θέμα. Ο κύριος λόγος συνήθως έχει να κάνει με το πλήθος των δεδομένων. Σχεδόν πάντα ο αριθμός των δεδομένων είναι μικρός, ενώ ανακύπτουν σοβαρά προβλήματα όταν η διαστατικότητα του διανύσματος χαρακτηριστικών είναι σημαντικά μεγάλη. Παρόλα αυτά εάν η γενική γνώση για το πρόβλημα είναι καλή και επιτρέπεται η παραμετροποίηση των υπό συνθήκη πυκνοτήτων πιθανότητας, οι παραπάνω δυσκολίες μπορούν να ελαττωθούν σημαντικά. Για παράδειγμα, έστω ότι η υπό συνθήκη πυκνότητα

$p(\mathbf{x}|\omega_i)$ είναι μια κανονική κατανομή με μέση τιμή μ_j και πίνακα συνδιακύμανσης (covariance matrix) C_j . Ακόμα και αν δεν γνωρίζουμε τις ακριβείς τιμές αυτών των μεγεθών το πρόβλημα απλοποιείται από το να υπολογίσουμε την $p(\mathbf{x}|\omega_i)$ στο να υπολογίσουμε τις παραμέτρους μ_j και C_j .

Το πρόβλημα της εκτίμησης παραμέτρων θεωρείται ένα κλασσικό πρόβλημα στην στατιστική και μπορεί να προσεγγιστεί με πολλούς τρόπους. Οι πιο γνωστές μέθοδοι είναι η εκτίμηση μέγιστης πιθανοφάνειας (maximum likelihood) και η εκτίμηση κατά Bayes [1-4, 99]. Αν και τα αποτελέσματα που δίνουν και οι δυο τρόποι είναι σχεδόν πανομοιότυπα, οι εννοιολογικές αντιλήψεις είναι διαφορετικές.

Είναι σημαντικό να κάνουμε εδώ τον διαχωρισμό μεταξύ επιβλεπόμενης και μη επιβλεπόμενης εκπαίδευσης. Και στις δυο περιπτώσεις, υποθέτουμε για τα δείγματα \mathbf{x} , ότι είναι επιλεγμένα ως εξής: Γνωρίζοντας ότι έχουμε την κατάσταση ω_i της φύσης με πιθανότητα $P(\omega_i)$, επιλέγουμε τα δείγματα μέσα από μια συνάρτηση κατανομής $p(\mathbf{x}|\omega_i)$. Η διαφορά μεταξύ επιβλεπόμενης και μη εκπαίδευσης συνίσταται στο ότι στην πρώτη ξέρουμε την κατάσταση της φύσης ονομαστικά για κάθε δείγμα ενώ στην δεύτερη δεν την ξέρουμε. Όπως αναμένεται, η μη επιβλεπόμενη εκπαίδευση είναι σαφώς πιο δύσκολη.

Στις περισσότερες περιπτώσεις οι παραπάνω προϋποθέσεις (γνώση της κατανομής και εκτίμηση μόνο μερικών παραμέτρων) δεν μπορούν να εφαρμοστούν. Οι συνήθεις παραμετρικές μορφές δεν αποδεικνύονται και πολύ χρήσιμες σε πραγματικές καταστάσεις. Ένα γνωστό πρόβλημα που αντιμετωπίζουμε, είναι το ότι όλες οι κλασσικές παραμετρικές πυκνότητες πιθανότητας έχουν ένα μέγιστο ενώ τα συνήθη προβλήματα έχουν πυκνότητα πιθανότητας με πολλά μέγιστα. Παρακάτω, παρουσιάζονται οι μη παραμετρικές τεχνικές για τον υπολογισμό των συναρτήσεων διαχωρισμού $p(\mathbf{x}|\omega_i)$ που μπορούν να εφαρμοστούν όταν δεν γνωρίζουμε τις μορφές των πιθανοτήτων.

Υπάρχουν αρκετοί τρόποι μη παραμετρικών τεχνικών που μπορούν να εφαρμοστούν σε προβλήματα αναγνώρισης προτύπων. Μια διαδικασία συνίσταται στο να εκτιμήσουμε την πυκνότητα πιθανότητας από τα δείγματα [1-4]. Μια άλλη απευθύνεται σε διαδικασίες που εκτιμούν απ' ευθείας τις posteriori πιθανότητες [1-4]. Στις διαδικασίες αυτές περιλαμβάνονται τα παράθυρα Parzen και η μέθοδος των k-κοντινότερων γειτόνων (k-th nearest neighbours). Αυτές οι διαδικασίες παρακάμπτουν την εκτίμηση των πιθανοτήτων και πηγαίνουν απ' ευθείας στην ουσία της υπόθεσης που είναι να βρεθούν οι συναρτήσεις διαχωρισμού. Τέλος υπάρχουν μη παραμετρικές τεχνικές που μετασχηματίζουν τον χώρο

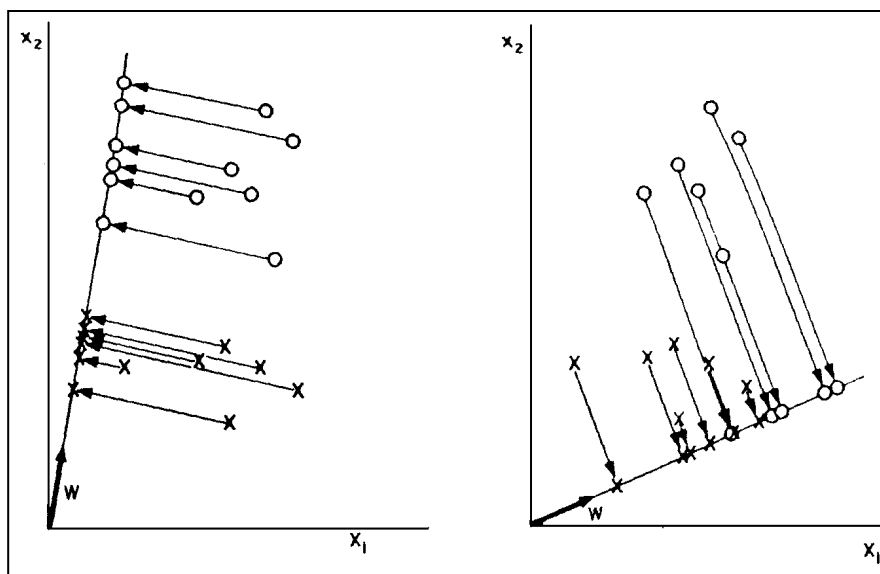
χαρακτηριστικών με την ελπίδα του ότι θα μπορούσαμε να εφαρμόσουμε παραμετρικές τεχνικές στον νέο χώρο. Χαρακτηριστικό παράδειγμα αποτελεί η μέθοδος Fisher [1].

Με την μέθοδο Fisher μπορούμε να ελαττώσουμε τις d -διαστάσεις ενός διανύσματος σε μια διάσταση απλώς προβάλλοντας τα σημεία των d -διαστάσεων πάνω σε μια γραμμή. Βέβαια αν τα δείγματα είναι καλά χωρισμένα, σχηματίζοντας αυτόνομα νέφη στον χώρο των d -διαστάσεων η προβολή τους σε μια αυθαίρετη γραμμή θα προκαλέσει κατά πάσα πιθανότητα μία ανάμειξη σημείων από διαφορετικές τάξεις. Μετακινώντας την γραμμή μπορούμε να βρούμε θέσεις της που να αποδίδουν αρκετά καλά διαχωρίσιμα αποτελέσματα. Κάτι τέτοιο φαίνεται στο σχήμα 1.7 και αυτό αποτελεί τον σκοπό της παρακάτω ανάλυσης.

Ας υποθέσουμε ότι έχουμε ένα d -διαστάτο σύνολο n -σημείων που αποτελείται από τα $\{x_1, x_2, x_3, \dots, x_n\}$. Από αυτά n_1 ανήκουν στο υποσύνολο X_1 που θα καλείται ω_1 και n_2 ανήκουν στο υποσύνολο X_2 που θα καλείται ω_2 . Εάν διαμορφώσουμε ένα γραμμικό συνδυασμό των συνιστωσών του x θα έχουμε το μέγεθος:

$$y = w^t x$$

και τα αντίστοιχα σύνολα των n σημείων y_1, \dots, y_n που χωρίζονται με την σειρά τους στα υποσύνολα Y_1 και Y_2 . Η διεύθυνσή του w είναι όπως καταλαβαίνουμε από το σχήμα 1.7 πολύ σημαντική.



Σχήμα 1.7. Προβολή των δειγμάτων σε μια γραμμή. Στο δεξιό σχήμα τα νέφη χωρίζονται καλά μετά την προβολή τους πάνω στην ευθεία. Αντίθετα στο αριστερό σχήμα αυτό δεν συμβαίνει.

Η γραμμική συνάρτηση διαχωρισμού του Fisher είναι η γραμμική συνάρτηση $\mathbf{w}^t \mathbf{x}$ για την οποία η επόμενη συνάρτηση κριτήριο γίνεται μέγιστη:

$$J(\mathbf{w}) = \frac{|\tilde{\mathbf{m}}_1^2 - \tilde{\mathbf{m}}_2^2|^2}{(\tilde{\mathbf{s}}_1^2 + \tilde{\mathbf{s}}_2^2)} \quad (1.3)$$

όπου,

$$\tilde{\mathbf{m}}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{x \in X_i} \mathbf{w}^t x = \mathbf{w}^t \mathbf{m}_i \quad (1.4)$$

και

$$\tilde{\mathbf{s}}_i^2 = \sum_{y \in Y_i} (y - \tilde{\mathbf{m}}_i)^2 \quad (1.5)$$

Για να λάβουμε το J σαν συνάρτηση του \mathbf{w} ορίζουμε τους πίνακες συνδιακύμανσης S_i και S_w ως εξής:

$$S_i = \sum_{x \in X_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \quad (1.6)$$

και

$$S_w = S_1 + S_2 \quad (1.7)$$

Τότε:

$$\tilde{\mathbf{s}}_1^2 + \tilde{\mathbf{s}}_2^2 = \mathbf{w}^t S_w \mathbf{w} \quad (1.8)$$

Ομοίως:

$$S_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t \quad (1.9)$$

Ο πίνακας S_w καλείτε εσωτερικός πίνακας σκέδασης (συνδιακύμανσης) και είναι ανάλογος του απλού πίνακα συνδιακύμανσης για τα d -διαστάσεων δεδομένα. Είναι συμμετρικός, θετικά ημιορισμένος και είναι συνήθως μη ιδιάζων¹ για $n > d$ [1]. Ο πίνακας S_B ονομάζεται πίνακας σκέδασης (συνδιακύμανσης) μεταξύ των τάξεων. Και αυτός είναι συμμετρικός και θετικά ημιορισμένος, με την μόνη διαφορά του ότι επειδή προέρχεται από το εξωτερικό γινόμενο δυο διανυσμάτων ο βαθμός του είναι το πολύ ένα. Σε όρους των S_B, S_w μπορούμε να γράψουμε την συνάρτηση κριτήριο ως:

$$J(\mathbf{w}) = \frac{\mathbf{w}^t S_B \mathbf{w}}{\mathbf{w}^t S_w \mathbf{w}} \quad (1.10)$$

Το διάνυσμα λύσης δίνεται από την παρακάτω εξίσωση ιδιοτιμών.

$$S_w^{-1} S_B \mathbf{w} = \lambda \mathbf{w} \quad (1.11)$$

¹ Ως μη ιδιάζων ορίζεται ο πίνακας X που $\det(X) \neq 0$.

Τότε έχουμε λάβει την γραμμική διακρίνουσα του Fisher δηλαδή τον μετασχηματισμό πάνω στο \mathbf{w} , που συνήθως είναι υπόχωρος του αρχικού χώρου των χαρακτηριστικών, ο οποίος μεγιστοποιεί τον λόγο εσωτερικής-εξωτερικής απόστασης προς την εσωτερική διακύμανση. Παρόμοια μελέτη γίνεται στην περίπτωση των πολλών τάξεων.

1.4. Γραμμικές συναρτήσεις διαχωρισμού.

Στην παράγραφο 1.3 θεωρήσαμε ότι οι μορφές για τις κατανομές πιθανότητας των νεφών κάθε τάξης ήταν γνωστές και χρησιμοποιήσαμε τα δείγματα για να εκτιμήσουμε τις παραμέτρους των. Σε αυτή την παράγραφο υποθέτουμε ότι γνωρίζουμε τις μορφές των συναρτήσεων διαχωρισμού (discriminant functions) και θα χρησιμοποιήσουμε τα δείγματα για να εκτιμήσουμε τις τιμές των παραμέτρων του ταξινομητή. Θα εξετάσουμε μερικές διαδικασίες για τον καθορισμό των συναρτήσεων διαχωρισμού που ένα μέρος τους είναι στατιστικές ενώ ένα άλλο μέρος τους δεν είναι. Γενικότερα, καμιά από τις μεθόδους δεν απαιτεί γνώση της κατανομής πιθανότητας και μπορούμε να πούμε ότι αποτελούν μέρος των μη παραμετρικών διαδικασιών (§1.4).

Θα ασχοληθούμε σε αυτή την υποπαράγραφο του κεφαλαίου με συναρτήσεις που είτε είναι γραμμικές σε σχέση με το διάνυσμα χαρακτηριστικών \mathbf{x} , είτε είναι γραμμικές σε σχέση με κάποιες άλλες συναρτήσεις του. Οι γραμμικές συναρτήσεις διαχωρισμού έχουν πολλές καλές μαθηματικές ιδιότητες και είναι χρήσιμες για αναλυτική θεώρηση και μελέτη. Όπως είδαμε στην παράγραφο 1.2 μπορούν να είναι βέλτιστες εάν βοηθούν οι πυκνότητες πιθανότητας. Είναι εύκολο να υπολογιστούν, ενώ είναι εύκολο να υλοποιηθεί και ένας ταξινομητής με γνωστή δομή σε ένα ειδικού σκοπού επεξεργαστή. Η προσοχή μας θα στραφεί επίσης σε διαδικασίες σύγκλισης διαφόρων διαδικασιών gradient-descent, καθώς είναι οι πιο κρίσιμες στην ελαχιστοποίηση ποσοτήτων που χρησιμοποιούνται για τον σχεδιασμό συστημάτων ταξινόμησης.

1.5.1. Γραμμικές συναρτήσεις διαχωρισμού και επιφάνειες απόφασης.

Μια γραμμική συνάρτηση διαχωρισμού είναι ένας γραμμικός συνδυασμός των συνιστωσών του διανύσματος \mathbf{x} και μπορεί να γραφεί στην ειδική περίπτωση των δυο κατηγοριών ως:

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0. \quad (1.17)$$

όπου \mathbf{x} είναι ένα διάνυσμα βάρους (weight vector) και w_0 είναι ένα όριο κατωφλίου (threshold weight). Ένας ταξινομητής δύο κατηγοριών ταξινομεί σύμφωνα με τον παρακάτω κανόνα: Αποφάσισε ω_1 εάν $g(\mathbf{x}) > 0$ ειδάλλως αποφάσισε ω_2 . Έτσι το διάνυσμα \mathbf{x} κατατάσσεται στην κατηγορία ω_1 εάν το εσωτερικό γινόμενο $\mathbf{w}^t \mathbf{x}$ υπερβεί μια ορισμένη τιμή κατωφλίου $-w_0$. Εάν η τιμή της $g(\mathbf{x})$ είναι ίση με μηδέν τότε το \mathbf{x} μπορεί να ταξινομηθεί αυθαίρετα.

Η εξίσωση $g(\mathbf{x}) = 0$ μας καθορίζει την επιφάνεια διαχωρισμού που χωρίζει τα σημεία που είναι ταξινομημένα στην περιοχή ω_1 από αυτά που είναι ταξινομημένα στην περιοχή ω_2 . Όταν η $g(\mathbf{x})$ είναι γραμμική αυτή η επιφάνεια απόφασης είναι ένα υπερεπίπεδο διάστασης $N-1$ στον N -διάστατο χώρο χαρακτηριστικών. Εάν \mathbf{x}_1 και \mathbf{x}_2 βρίσκονται στην επιφάνεια απόφασης έχουμε:

$$\mathbf{w}^t \mathbf{x}_1 + w_0 = \mathbf{w}^t \mathbf{x}_2 + w_0 \quad (1.18)$$

ή

$$\mathbf{w}^t (\mathbf{x}_1 - \mathbf{x}_2) = 0 \quad (1.19)$$

από όπου συμπεραίνουμε ότι το διάνυσμα \mathbf{w} είναι κάθετο σε οποιοδήποτε διάνυσμα που κείται στο υπερεπίπεδο. Γενικά το υπερεπίπεδο H διαιρεί τον χώρο χαρακτηριστικών σε δυο υποχώρους, που καλούνται περιοχές απόφασης R_1 για την περίπτωση $\omega = \omega_1$ και R_2 για την περίπτωση $\omega = \omega_2$.

Η συνάρτηση διαχωρισμού $g(\mathbf{x})$ μας αποδίδει ένα μέτρο της απόστασης του \mathbf{x} από το υπερεπίπεδο διαχωρισμού. Ίσως, ο ευκολότερος τρόπος να το δούμε είναι να εκφράσουμε το \mathbf{x} ως:

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (1.20)$$

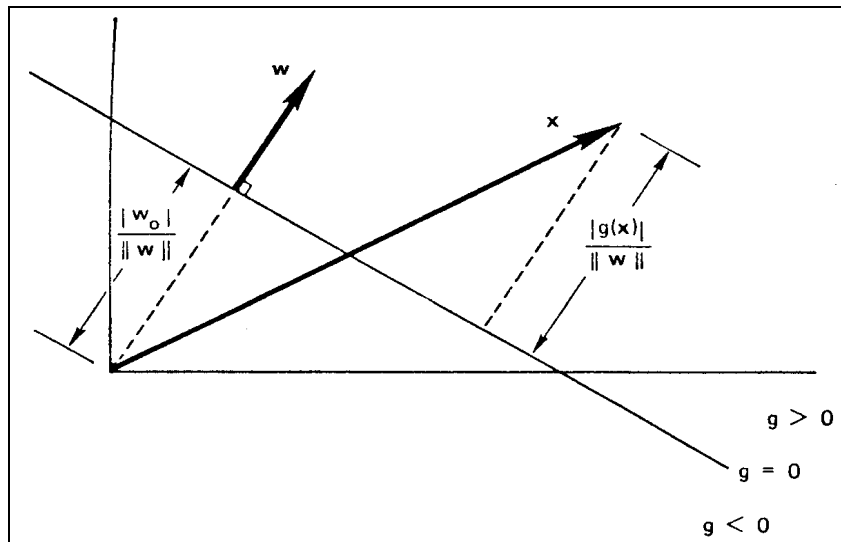
όπου \mathbf{x}_p είναι η προβολή του \mathbf{x} στο H (βλέπε σχήμα 1.8) και r είναι η επιθυμητή αλγεβρική απόσταση, που είναι θετική εάν το \mathbf{x} είναι στον χώρο R_1 και αρνητική εάν το \mathbf{x} είναι στον χώρο R_2 . Τότε καθώς $g(\mathbf{x}_p) = 0$ θα έχουμε:

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0 = r \|\mathbf{w}\| \quad (1.21)$$

ή

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|} \quad (1.22)$$

Ειδικότερα, η απόσταση από την αρχή στο H δίνεται από το μέγεθος $\frac{w_0}{\|\mathbf{w}\|}$. Εάν $w_0 > 0$ τότε η αρχή βρίσκεται στην θετική πλευρά του H ενώ στην αντίθετη περίπτωση βρίσκεται στην αρνητική πλευρά του H . Μια γεωμετρική αναπαράσταση φαίνεται στο σχήμα 1.8.



Σχήμα 1.8. Η γραμμική επιφάνεια απόφασης $g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$.

Ανακεφαλαιώνοντας, μια γραμμική συνάρτηση διαχωρισμού διαιρεί τον χώρο χαρακτηριστικών με την βοήθεια ενός υπερεπιπέδου. Ο προσανατολισμός του επιπέδου καθορίζεται από το διάνυσμα \mathbf{w} ενώ η θέση του υπερεπιπέδου δίνεται από το κατώφλι w_0 . Η συνάρτηση διαχωρισμού $g(\mathbf{x})$ είναι ανάλογη της απόστασης του διανύσματος \mathbf{x} από το υπερεπίπεδο με $g(\mathbf{x}) > 0$ όταν το \mathbf{x} είναι στην θετική πλευρά του και $g(\mathbf{x}) < 0$ όταν το \mathbf{x} είναι στην αρνητική πλευρά του. Όμοιες αναλύσεις γίνονται όταν έχουμε διαμερίσεις σε παραπάνω από δυο κατηγορίες.

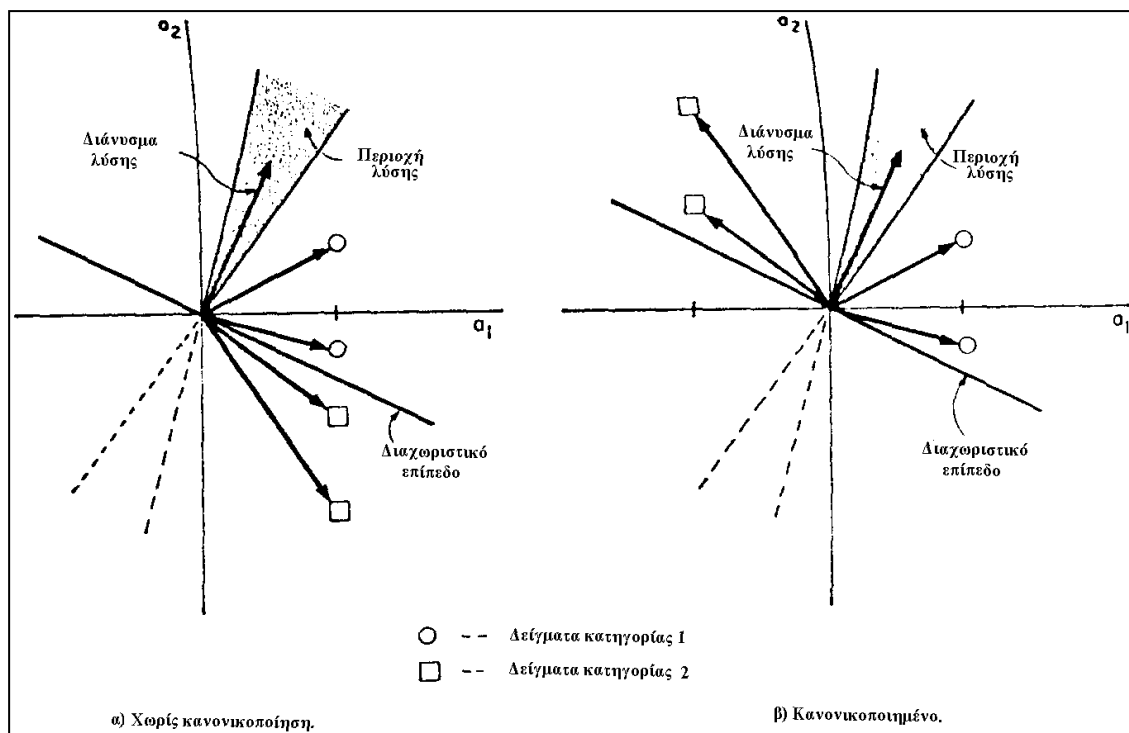
1.5.2. Γραμμικά διαχωρίσιμες προσεγγίσεις στην ταξινόμηση δυο κατηγοριών.

Ας υποθέσουμε ότι διαθέτουμε ένα σύνολο από n -σημεία y_1, \dots, y_n και έστω ότι μερικά ανήκουν στην κατηγορία ω_1 ενώ τα υπόλοιπα στην κατηγορία ω_2 . Θέλουμε να

χρησιμοποιήσουμε αυτά τα δείγματα για να καθορίσουμε το διάνυσμα βάρους σε μια γραμμική συνάρτηση διαχωρισμού $g(\mathbf{x})=\mathbf{a}^t \mathbf{y}$. Ας υποθέσουμε επίσης ότι υπάρχει μια λύση για την οποία η πιθανότητα του σφάλματος είναι πολύ μικρή. Τότε μια λύση θα ήταν να βρούμε ένα διάνυσμα βάρους \mathbf{a} που να κατατάσσει τα δείγματα σωστά. Εάν ένα τέτοιο διάνυσμα υπάρχει τότε θα καλούμε το σύνολο \mathbf{y} *γραμμικά διαχωρίσιμο*.

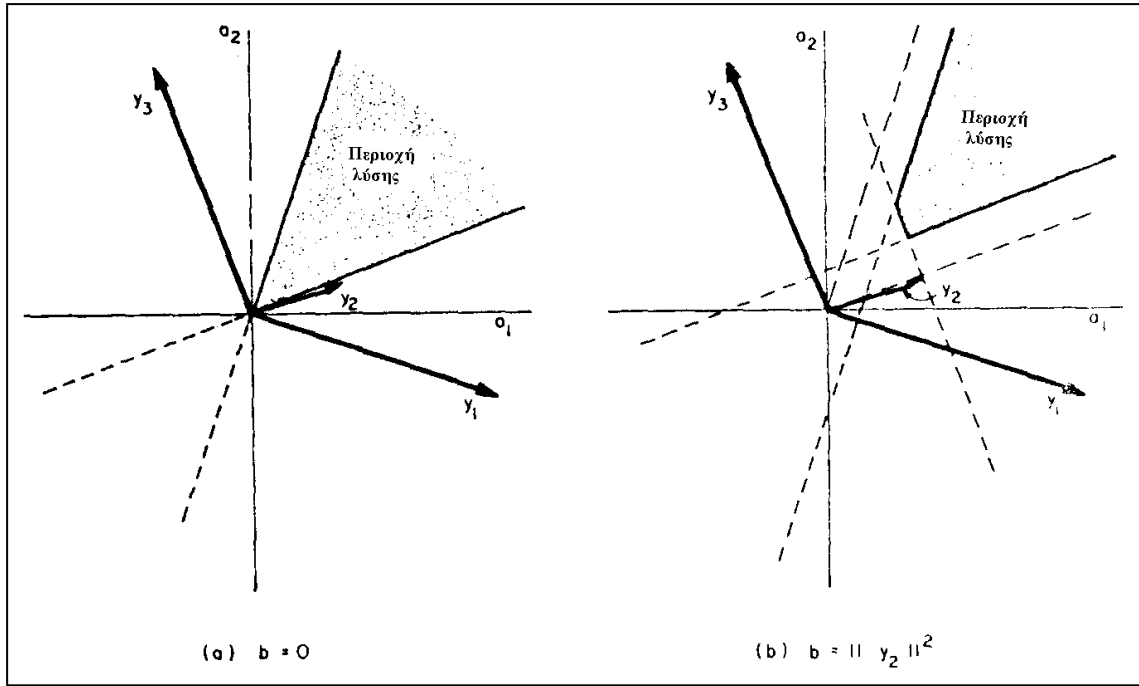
Ένα δείγμα \mathbf{y}_i θα ταξινομείται σωστά μόνο εάν $\mathbf{a}^t \mathbf{y}_i > 0$ και το \mathbf{y}_i ταξινομείται στην κατηγορία ω_1 ή εάν $\mathbf{a}^t \mathbf{y}_i < 0$ και το \mathbf{y}_i ταξινομείται στην κατηγορία ω_2 . Ως ένα σχόλιο πάνω στην δεύτερη περίπτωση μπορούμε να πούμε ότι το \mathbf{y}_i ταξινομείται στην κατηγορία ω_2 όταν $\mathbf{a}^t (-\mathbf{y}_i) > 0$ κάτι που υποδεικνύει μια κανονικοποίηση που μπορούμε να ακολουθήσουμε στην ταξινόμηση σε δυο κατηγορίες. Αυτή συνίσταται στο να αντικαταστήσουμε όλα τα δείγματα \mathbf{y}_i που ανήκουν στην κατηγορία ω_2 με τα $-\mathbf{y}_i$. Με αυτή την κανονικοποίηση μπορούμε να ξεχάσουμε τον φορμαλισμό του να κατηγοριοποιούμε τα δείγματα σε ω_1 και ω_2 καθώς η αναζήτηση για το διάνυσμα \mathbf{a} έτσι ώστε $\mathbf{a}^t \mathbf{y}_i > 0$ για κάθε \mathbf{y}_i θα μας δώσει το \mathbf{a} που εφεξής θα καλείτε *διάνυσμα λύσης*.

Το διάνυσμα λύσης \mathbf{a} μπορεί να αναπαρασταθεί ως ένα διάνυσμα στον χώρο χαρακτηριστικών. Κάθε δείγμα \mathbf{y}_i θέτει μια απαίτηση για την πιθανή θέση του διανύσματος λύσης. Η εξίσωση $\mathbf{a}^t \mathbf{y}_i = 0$ καθορίζει ένα υπερεπίπεδο που περνάει από την αρχή των αξόνων έχοντας το \mathbf{y}_i ως το κάθετο διάνυσμα σε αυτό. Το διάνυσμα \mathbf{a} εάν υπάρχει, πρέπει να βρίσκεται στο θετικό ημιεπίπεδο σύμφωνα με ότι αναφέραμε στην προηγούμενη παράγραφο και την §1.5.1. Έτσι το διάνυσμα λύσης πρέπει να βρίσκεται στην τομή των n -υπερεπιπέδων και κάθε διάνυσμα που βρίσκεται σε αυτή την περιοχή καλείτε ως διάνυσμα λύσης. Ένα γεωμετρικό παράδειγμα για κανονικοποιημένα και μη κανονικοποιημένα δείγματα φαίνεται στο σχήμα 1.9.



Σχήμα 1.9. Γραμμικά διαχωρίσιμα δείγματα και το διάνυσμα λύσης.

Από την παραπάνω ανάλυση φαίνεται ότι το διάνυσμα λύσης δεν είναι μοναδικό. Υπάρχουν αρκετοί τρόποι για να θέσουμε παραπάνω απαιτήσεις στο διάνυσμα λύσης. Μια απαίτηση μπορεί να είναι η εύρεση ενός μοναδιαίου διανύσματος που μεγιστοποιεί την ελάχιστη απόσταση των δειγμάτων από το επίπεδο. Άλλη απαίτηση μπορεί να είναι να βρούμε το \mathbf{a} έτσι ώστε να ικανοποιεί την συνθήκη $\mathbf{a}^t \mathbf{y}_i > \mathbf{b}$ όπου \mathbf{b} είναι μια θετική σταθερά που καλείτε *περιθώριο*. Όπως φαίνεται στο σχήμα 1.10, το διάνυσμα λύσης προκύπτει από την τομή των επιπέδων που ικανοποιούν την σχέση $\mathbf{a}^t \mathbf{y}_i > \mathbf{b} > 0$.



Σχήμα 1.10. Επίδραση του περιθωρίου στην περιοχή λύσης.

1.5.3. Διαδικασίες gradient descent.

Η προσέγγιση που θα ακολουθήσουμε για να βρούμε την λύση σε ένα σύνολο από ανισότητες $\mathbf{a}^t \mathbf{y}_i > 0$ θα είναι να καθορίσουμε μια συνάρτηση κριτήριο $J(\mathbf{a})$ που θα ελαχιστοποιείται εάν το \mathbf{a} είναι ένα διάνυσμα λύσης. Αυτό ελαττώνει το πρόβλημα μας στο να ελαχιστοποιήσουμε μια βαθμωτή συνάρτηση, κάτι που γίνεται με διαδικασίες gradient descent. Η γενική αρχή τους είναι απλή: Ξεκινώντας από ένα τυχαίο διάνυσμα \mathbf{a}_1 υπολογίζουμε το μέγεθος $\nabla J(\mathbf{a}_1)$ και μετά υπολογίζουμε τα $\mathbf{a}_2, \dots, \mathbf{a}_{k+1}$ με τον παρακάτω αλγόριθμο:

$$\mathbf{a}_{k+1} = \mathbf{a}_k - \rho_k \nabla J(\mathbf{a}_k) \quad (1.23)$$

όπου ρ_k είναι μια θετική ποσότητα που θέτει το βήμα του αλγόριθμου. Λόγω της μεγάλης σπουδαιότητας της παραμέτρου ρ_k έχει ερευνηθεί το βέλτιστο της τιμής της. Αποδεικνύεται ότι η βέλτιστη τιμή της παραμέτρου ρ_k είναι η:

$$\rho_k = \frac{\|\nabla J\|^2}{\nabla J^t D \nabla J} \quad (1.24)$$

όπου D είναι ο πίνακας των μερικών παραγώγων $\partial^2 J / \partial a_i \partial a_j$.

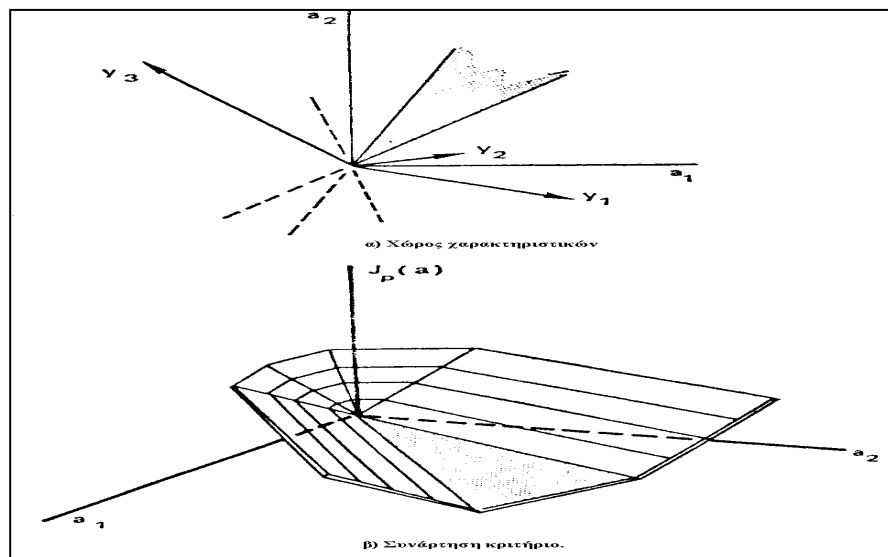
Μία άλλη προσέγγιση μας παρέχει ένα αλγόριθμο που είναι γνωστός ως ο αλγόριθμος του Newton και που χρησιμοποιείται περισσότερο.

$$\mathbf{a}_{k+1} = \mathbf{a}_k - (\mathbf{D}^{-1}\nabla J) \quad (1.25)$$

Όσο αφορά την αναζήτηση της συνάρτησης $J(\mathbf{a})$ η πιο προφανής λύση θα ήταν να έχει την μορφή $J(\mathbf{a}; \mathbf{y}_1, \dots, \mathbf{y}_n)$ του αριθμού των δειγμάτων που δεν ταξινομεί. Μια καλύτερη όμως συνάρτηση είναι η ‘συνάρτηση αντίληψης’ (perceptron function) που ορίζεται ως εξής:

$$J_p(\mathbf{a}) = \sum_{\mathbf{y} \in Y} (-\mathbf{a}^t \mathbf{y}) \quad (1.26)$$

όπου $Y(\mathbf{a})$ είναι το σύνολο των δειγμάτων που δεν ταξινομούνται σωστά. Γεωμετρικά το $J_p(\mathbf{a})$ είναι ανάλογο της απόστασης των μη ταξινομημένων δειγμάτων από την επιφάνεια απόφασης. Το σχήμα 1.11 δείχνει την κατάσταση για ένα απλό παράδειγμα δύο διαστάσεων.



Σχήμα 2.5.4. Η συνάρτηση ‘αντίληψης’.

Για περισσότερες πληροφορίες όσο αφορά τις διαδικασίες σύγκλισης ο αναγνώστης μπορεί να αναφερθεί στις.

1.5.4. Μη διαχωριζόμενη συμπεριφορά.

Οι παραπάνω διαδικασίες μας δίνουν απλά αποτελέσματα όταν τα δείγματα είναι γραμμικά διαχωρίσιμα. Όλες αυτές οι διαδικασίες χαρακτηρίζονται ως ‘διαδικασίες διόρθωσης σφάλματος’ γιατί μεταβάλλουν το διάνυσμα λύσης μόνο όταν έχει συμβεί ένα

σφάλμα στην ταξινόμηση. Τις περισσότερες φορές όμως τα δείγματα δεν είναι γραμμικά διαχωρίσιμα, ειδικά όταν πρόκειται για μεγάλα σύνολα. Έτσι είναι σημαντικό να ξέρουμε την συμπεριφορά τέτοιων αλγορίθμων.

Ενώ στην περίπτωση που έχουμε γραμμικώς διαχωρίσιμα δείγματα ο αλγόριθμος κάποτε θα σταματήσει, στην άλλη περίπτωση δεν θα τερματιστεί αφού πάντα θα έχουμε σφάλματα στην ταξινόμηση. Η ακριβής μελέτη τέτοιων φαινομένων έχει μελετηθεί μόνο για μερικές περιπτώσεις. Συνήθως ακολουθούνται εμπειρικοί και ευριστικοί κανόνες για τον τερματισμό του αλγορίθμου. Ο βασικότερος τρόπος να επεξεργαστούμε δεδομένα με τους παραπάνω αλγόριθμους είναι να επιλέξουμε την τιμή του ρ_k έτσι ώστε να το κάνουμε συνάρτηση της συνολικής απόδοσης του συστήματος. Άλλος τρόπος είναι να επιλέξουμε το ρ_k ως $\rho_k = \rho_1/k$. Τέλος, μπορούν να ακολουθηθούν διαδικασίες ελαχιστοποίησης του σφάλματος βασισμένες σε μια μορφή πινάκων (ψευδοανάστροφοι). Είναι φυσικό να αναρωτηθούμε ποιος είναι ο καλύτερος, η αλήθεια είναι όμως ότι κανένας δεν κυριαρχεί ολοκληρωτικά πάνω στους άλλους. Εξαρτώνται από παραμέτρους όπως ο αριθμός των δειγμάτων, η διαστατικότητα τους, και η ευκολία στον προγραμματισμό.

1.5. Επιλογή των χαρακτηριστικών.

Στις προηγούμενες παραγράφους παρουσιάστηκαν μερικές από τεχνικές για την ταξινόμηση και την αναγνώριση προτύπων. Την σχεδίαση όμως του συστήματος αναγνώρισης προηγείται η εξαγωγή και ανάλυση των χαρακτηριστικών. Κάθε αντικείμενο για να αναγνωριστεί και να ταξινομηθεί, πρέπει να διαθέτει ένα ικανό αριθμό από χαρακτηριστικά που να το διαχωρίζουν από τα άλλα αντικείμενα. Το πρώτο βήμα σε κάθε διαδικασία αναγνώρισης που πραγματοποιείται από μια μηχανή (ή ένα άνθρωπο) είναι να αναζητήσει τα καλύτερα χαρακτηριστικά που μπορούν να αποδοθούν στο αντικείμενο και να τα μετρήσει (εξάγει). Συνήθως το πρόβλημα της επιλογής και εξαγωγής χαρακτηριστικών είναι σύνθετο και εξαρτάται σε μεγάλο βαθμό από παραμέτρους όπως η ευκολία εξαγωγής, και το κόστος των διαδικασιών που ακολουθούν. Το σημαντικό λοιπόν στην προεπεξεργασία χαρακτηριστικών είναι το να εξάγουμε τα σημαντικότερα χαρακτηριστικά για να πραγματοποιήσουμε ένα ακριβές και αποδοτικό σύστημα ταξινόμησης.

Η προεπεξεργασία των χαρακτηριστικών περιλαμβάνει κυρίως δύο διαδικασίες: α) τους μετασχηματισμούς ομαδοποίησης (clustering transforms) και β) τους μηχανισμούς επιλογής χαρακτηριστικών (feature selection). Είναι λογικό να αναμένεται ότι οι

μετρήσεις ενός αντικειμένου οι οποίες αναπαρίστανται από τους διαφορετικούς άξονες x_k στον χώρο των χαρακτηριστικών συνήθως έχουν διαφορετική επιρροή στην ταξινόμηση. Η διαδικασία επιβολής ‘βαρών’ στους διάφορους άξονες μπορεί να πραγματοποιηθεί μέσα από ένα γραμμικό μετασχηματισμό (με χρήση ενός πίνακα W) που θα ομαδοποιήσει τα μετασχηματισμένα διανύσματα σε ένα νέο χώρο, και ενός μετασχηματισμού (χρήση ενός άλλου πίνακα A) που φανερώνει ποιες συνιστώσες έχουν μικρή διακύμανση και συνεπώς μεγάλη σημασία.

Ένα κεντρικό πρόβλημα στην αναγνώριση προτύπων είναι η ανάπτυξη συναρτήσεων αποφάσεων από ένα πεπερασμένο σύνολο δειγμάτων έτσι ώστε, κάθε συνάρτηση να διαμερίζει τον χώρο χαρακτηριστικών σε υποχώρους, που να περιέχουν δείγματα από μια κατηγορία. Αυτή η ανάλυση οδηγεί στην έννοια του μετασχηματισμού ομαδοποίησης που εφαρμόζεται στον χώρο μετρήσεων για να ομαδοποιήσει τα σημεία που αντιστοιχούν σε μια συγκεκριμένη κατηγορία. Ένας τέτοιος μετασχηματισμός θα μεγιστοποιήσει την απόσταση μεταξύ των νεφών (intersset distance) και θα ελαχιστοποιήσει την εσωτερική διασπορά των νεφών (intraset distance). Όσο αφορά την επιλογή χαρακτηριστικών, μπορούμε να επιτύχουμε ελάττωση διαστάσεων σε ένα μεγάλο μήκους διάνυσμα χαρακτηριστικών επιλέγοντας τα πιο σημαντικά. Η βέλτιστη επιλογή χαρακτηριστικών γίνεται με βάση την μεγιστοποίηση ή την ελαχιστοποίηση μια συνάρτησης κριτηρίου. Μια τέτοια εφαρμογή χαρακτηρίζεται ως ‘απόλυτη επιλογή χαρακτηριστικών’. Μια διαφορετική επιλογή χαρακτηριστικών έχει σχέση με την απόδοσή τους. Αυτή σχετίζεται άμεσα με την απόδοση του συστήματος ταξινόμησης και εκφράζεται με όρους όπως είναι η πιθανότητα της σωστής αναγνώρισης. Οι πιο διαδεδομένες μέθοδοι επιλογής των χαρακτηριστικών είναι οι εξής:

Επιλογή χαρακτηριστικών με ελαχιστοποίηση της συνάρτησης εντροπίας.

Η έννοια της εντροπίας μπορεί να χρησιμοποιηθεί στην επιλογή των βέλτιστων χαρακτηριστικών. Τα χαρακτηριστικά που θα ελαττώσουν την αβεβαιότητα μιας κατάστασης θεωρούνται ως έχοντα περισσότερη πληροφορία σε σχέση με αυτά που αυξάνουν την αβεβαιότητα μιας κατάστασης.

Επιλογή χαρακτηριστικών με ανάπτυξη σε ορθογώνιες σειρές.

Η διαδικασία επιλογής με την ελαχιστοποίηση της εντροπίας βασίζεται στην υπόθεση ότι οι τάξεις ακολουθούν μια κανονική κατανομή. Όταν αυτή η υπόθεση δεν ισχύει η μέθοδος της ανάπτυξης σε ορθογώνιες σειρές προσφέρει μια εναλλακτική προσέγγιση

στην επιλογή και εξαγωγή χαρακτηριστικών. Χαρακτηριστική περίπτωση αποτελεί ο μετασχηματισμός Karhunen-Loeve (K-L).

Επιλογή χαρακτηριστικών με συναρτησιακή προσέγγιση.

Εάν τα χαρακτηριστικά μιας κατηγορίας μπορούν να χαρακτηριστούν από μια συνάρτηση $f(\mathbf{x})$ που καθορίζεται με βάση τα δεδομένα παρατήρησης, τότε μπορούμε να επιλέξουμε χαρακτηριστικά μέσα από μια διαδικασία συναρτησιακής προσέγγισης. Υπάρχουν αρκετές μέθοδοι για την συναρτησιακή προσέγγιση. Χαρακτηριστικότερες είναι οι ανάπτυξη συναρτήσεων (functional expansion), η μέθοδος του πυρήνα (Kernel approximation) και η στοχαστική προσέγγιση (stochastic approximation).

Η έννοια της απόκλισης.

Η απόκλιση είναι ένα μέτρο της ‘απόστασης’ η ‘ανομοιότητας’ μεταξύ δύο κατηγοριών. Μπορεί να χρησιμοποιηθεί για να κατατάξει τα χαρακτηριστικά και να υπολογίσει την αποτελεσματικότητα του διαχωρισμού του χώρου χαρακτηριστικών. Ένα σημαντικό πλεονέκτημα είναι ότι χρησιμοποιείται όταν έχουμε αυθαίρετες κατανομές. Η επιλογή των καλύτερων χαρακτηριστικών γίνεται με την μεγιστοποίηση της συνάρτησης απόκλισης.