

ΚΕΦΑΛΑΙΟ 7

BAYESIAN ΤΑΞΙΝΟΜΗΤΕΣ

Σκοπός

Σκοπός του κεφαλαίου αυτού είναι να εισαγάγει τον αναγνώστη στις βασικές αρχές σχεδιασμού ενός ταξινομητή και στη συνέχεια να εστιάσει σε μία κατηγορία ταξινομητών, που βασίζονται στο βασικό θεώρημα του Bayes.

Προσδοκώμενα Αποτελέσματα

Όταν θα έχετε τελειώσει τη μελέτη του κεφαλαίου αυτού θα μπορείτε:

- Να γνωρίζετε ποιος είναι ο ρόλος του ταξινομητή σ' ένα σύστημα ταξινόμησης.
- Να ταξινομείτε πρότυπα σύμφωνα με τον Bayesian ταξινομητή.
- Να γνωρίζετε τι είναι ο ταξινομητής ελάχιστης απόστασης.
- Να γνωρίζετε τι είναι τα παράθυρα Parzen.
- Να γνωρίζετε τι είναι ο ταξινομητής Πλησιέστερου Γείτονα.

Έννοιες-κλειδιά

- | | |
|-----------------------|-------------------------------------|
| • Διακλασική Απόσταση | • Ταξινομητής Bayes |
| • Επιφάνεια Απόφασης | • Ταξινομητής Ελάχιστης Απόστασης |
| • Παράθυρο Parzen | • Ταξινομητής Πλησιέστερου Γείτονα. |
| • Συνάρτηση Διάκρισης | |

Εισαγωγικές Παρατηρήσεις

Στο κεφάλαιο αυτό παρουσιάζονται τεχνικές ταξινόμησης που σχετίζονται με το θεώρημα του Bayes. Αυτές οι μέθοδοι είναι εξαιρετικά αποτελεσματικές όταν είναι γνωστές οι πυκνότητες πιθανότητας $p(\mathbf{x} | \omega_i)$, όπου ω_i , $i=1,2,\dots,m$, οι κλάσεις του ταξινομητή και \mathbf{x} το διάνυσμα των τιμών των ουσιωδών χαρακτηριστικών ενός δείγματος. Οι συναρτήσεις αυτές μπορούν να προσεγγιστούν όταν διατίθεται ένας ικανοποιητικός αριθμός από αναγνωρισμένα δείγματα, δηλαδή δείγματα για καθένα από τα οποία είναι γνωστή η κλάση στην οποία ανήκει.

Στην αρχή του κεφαλαίου αυτού, στην Ενότητα 7.1, περιγράφονται τα βασικά στοιχεία ενός ταξινομητή. Στη συνέχεια, Ενότητα 7.2, περιγράφονται οι κατά Bayes ταξινομητές, η λειτουργία των οποίων στηρίζεται στην ακριβή γνώση των $p(\mathbf{x} | \omega_i)$ και με τους οποίους επιτυγχάνεται η ελάχιστη πιθανότητα του σφάλματος ταξινόμησης. Στην Ενότητα

7.3, εισάγονται οι έννοιες της συνάρτησης διάκρισης και της επιφάνειας απόφασης. Εάν δεν είναι γνωστές οι $p(\mathbf{x} | \omega_i)$, η χρήση Bayesian ταξινομητών προϋποθέτει την εκτίμηση των πυκνοτήτων πιθανότητας με βάση τα δείγματα εκπαίδευσης (αναγνωρισμένα δείγματα). Για το σκοπό αυτό, στην Ενότητα 7.4 χρησιμοποιούνται τα παράθυρα Parzen και στην Ενότητα 7.5 χρησιμοποιείται ένας αριθμός από k γείτονες. Τέλος, στην Ενότητα 7.6 παρουσιάζεται ο ταξινομητής του πλησιέστερου γείτονα.

ΕΝΟΤΗΤΑ 7.1 ΒΑΣΙΚΑ ΣΤΑΔΙΑ ΣΧΕΔΙΑΣΜΟΥ ΣΥΣΤΗΜΑΤΟΣ ΤΑΞΙΝΟΜΗΣΗΣ

Ο ταξινομητής, θα μπορούσε να πει κανείς, ότι αποτελεί την καρδιά του συστήματος ταξινόμησης. Όπως έχουμε ήδη αναφέρει στο πρώτο κεφάλαιο, ο σχεδιασμός ενός συστήματος ταξινόμησης θα μπορούσε να διαιρεθεί σε δύο μεγάλα στάδια, το στάδιο της δημιουργίας και της επιλογής των ουσιωδών χαρακτηριστικών και αυτό της σχεδίασης του ταξινομητή.

7.1.1 ΣΤΑΔΙΟ ΔΗΜΙΟΥΡΓΙΑΣ ΚΑΙ ΕΠΙΛΟΓΗΣ ΟΥΣΙΩΔΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.

Στο πρώτο στάδιο παράγονται και στη συνέχεια επιλέγονται τα ουσιώδη χαρακτηριστικά, με τα οποία περιγράφουμε τα πρότυπα. Τέτοια χαρακτηριστικά μπορεί να είναι, για παράδειγμα, οι διάφορες ροπές περιοχής, διάφορα χαρακτηριστικά ιστογράμματος, χαρακτηριστικά περιγράμματος. Από την πληθώρα, στην κυριολεξία, των χαρακτηριστικών, που είναι δυνατόν να παραχθούν, δεν είναι όλα πλούσια σε πληροφορία. Ας πούμε, για παράδειγμα, ότι η διασπορά της έντασης γύρω από τη μέση τιμή σε μια εικόνα, που πρέπει να ταξινομήσουμε σε μία από δύο κλάσεις, παίρνει περίπου την ίδια τιμή για όλες τις εικόνες, ανεξάρτητα από ποια κλάση ανήκουν. Ένα τέτοιο χαρακτηριστικό είναι φτωχό σε πληροφορία ταξινόμησης. Εμείς απ' όλα τα ουσιώδη χαρακτηριστικά, που είναι δυνατόν να παραχθούν, θα θέλαμε να επιλέξουμε εκείνα, των οποίων οι τιμές, για πρότυπα που ανήκουν σε διαφορετικές κλάσεις, διαφέρουν όσο γίνεται περισσότερο. Όπως συνήθως λέγεται, στην κατάλληλη ορολογία, τα ουσιώδη χαρακτηριστικά θα πρέπει να επιλέγονται έτσι ώστε οι τιμές που λαμβάνουν για τα διάφορα πρότυπα να έχουν μεγάλη **δια-κλασική** (*between class*) απόσταση και μικρή **ενδο-κλασική** (*within-class*) απόσταση. Στο Σχ. 7.1 δίνονται δύο απλοποιημένα παραδείγματα για δύο χαρακτηριστικά x_1, x_2 . Με αστερίσκο συμβολίζονται οι τιμές που λαμβάνουν για τη μία κλάση, και με τελεία για τη δεύτερη κλάση. Από τα παραπάνω συνάγεται ότι το x_1 είναι ένα πλούσιο σε πληροφο-

φορία ταξινόμησης χαρακτηριστικό ενώ το x_2 είναι φτωχό, διότι οι τιμές που λαμβάνει για τις δύο κλάσεις επικαλύπτονται.



Σχήμα 7.1: Τιμές χαρακτηριστικών με (a) μεγάλη διακλασική και (b) μικρή διακλασική απόσταση.

Η επιλογή χαρακτηριστικών δεν είναι μία απλή διαδικασία, αλλά συνήθως, το αποτέλεσμα εφαρμογής μιας σειράς μεθοδολογιών επιλογής χαρακτηριστικών. Στο βιβλίο αυτό δεν θα ασχοληθούμε με τέτοιες μεθοδολογίες. Ο ενδιαφερόμενος αναγνώστης μπορεί να αναφερθεί, για παράδειγμα, στο [1].

Στη συνέχεια θα υποθέσουμε ότι έχουμε επιλέξει l ουσιώδη χαρακτηριστικά, x_1, x_2, \dots, x_l . Με τα χαρακτηριστικά αυτά δημιουργούμε τα **ουσιώδη χαρακτηριστικά διανύσματα** (*feature vectors*),

$$\mathbf{x}=(x_1, x_2, \dots, x_l)^T$$

στον l -διάστατο χώρο, δηλαδή $\mathbf{x} \in R^l$. Επιπλέον, υποθέτουμε ότι υπάρχει μια αμφιμονοσήμαντη σχέση μεταξύ προτύπων και διανυσμάτων. Δηλαδή, σε κάθε διάνυσμα αντιστοιχεί ένα και μόνο πρότυπο.

Μια παρατήρηση, που έχει ιδιαίτερη σημασία για μας, είναι ότι οι τιμές που λαμβάνουν διάφορα πρότυπα δεν είναι προβλέψιμες αλλά έχουν ένα βαθμό τυχαιότητας. Για παράδειγμα, μάλλον θα μας εξέπληττε εάν η μέση τιμή των pixels για διαφορετικές ακτινογραφίες υγιών (μη υγιών) ήταν όλες ίδιες. Η τυχαιότητα των τιμών που προκύπτουν οφείλεται και στα όργανα μετρήσεως, αλλά πολύ περισσότερο στις διαφοροποιήσεις που υπάρχουν μεταξύ των ιστών των διαφορετικών ατόμων. Στη συνέχεια, θα θεωρούμε τα ουσιώδη χαρακτηριστικά $x_i, i=1, 2, \dots, l$, ως τυχαίες μεταβλητές και τα αντίστοιχα διανύσματα ως τυχαία διανύσματα \mathbf{x} στον l -διάστατο χώρο ή όπως θα λέμε στο χώρο των (ουσιωδών) χαρακτηριστικών (*feature space*).

7.1.2 ΣΤΑΔΙΟ ΣΧΕΔΙΑΣΗΣ ΤΟΥ ΤΑΞΙΝΟΜΗΤΗ.

Έχοντας επιλέξει τα χαρακτηριστικά, στο δεύτερο στάδιο, σχεδιάζουμε τον ταξινομητή. Ο ταξινομητής δεν είναι τίποτε άλλο από μία συνάρτηση ή ένα σύνολο συναρτήσεων, με ελεύθερη μεταβλητή το τυχαίο χαρακτηριστικό διάνυσμα \mathbf{x} , που διαιρεί τον l -διάστατο χώρο σε περιοχές $R_k, k=1, 2, \dots, m$. Καθεμιά από αυτές τις περιοχές αντιστοιχεί σε μία από τις m -κλάσεις ($m=2$ για ένα πρόβλημα δύο κλάσεων). Ο σχεδιασμός αυτός βασίζεται στην πληροφορία που απορρέει από το σύνολο των διανυσμάτων εκπαίδευσης που έ-

χουμε στη διάθεσή μας. Η σχεδίαση του ταξινομητή θα μας απασχολήσει στο παρόν και στο επόμενο κεφάλαιο.

Με την ολοκλήρωση του σχεδιασμού του ταξινομητή, η ταξινόμηση ενός αγνώστου πρότυπου γίνεται ως εξής. Μετρούμε την τιμή που παίρνει το αντίστοιχο χαρακτηριστικό διάνυσμα για το συγκεκριμένο πρότυπο. Στη συνέχεια ταξινομούμε το πρότυπο στην κλάση που αντιστοιχεί στην περιοχή, R_k , του χώρου όπου βρίσκεται το διάνυσμα \mathbf{x} . Προφανώς μια τέτοια διαδικασία εμπεριέχει και την έννοια του **λανθασμένης ταξινόμησης**. Ένας από τους στόχους στη διαδικασία σχεδιασμού ενός ταξινομητή είναι η ελαχιστοποίηση αυτού του λάθους.

ΕΝΟΤΗΤΑ 7.2 ΤΑΞΙΝΟΜΗΤΕΣ BAYES

Στόχος μας στην ενότητα αυτή είναι να σχεδιάσουμε τον ταξινομητή, αντλώντας έννοιες και γνώσεις από τη θεωρία πιθανοτήτων. Θα ξεκινήσουμε, για χάρη της απλότητας, από το πρόβλημα δύο κλάσεων, που θα συμβολίσουμε ω_1 και ω_2 αντίστοιχα. Υποθέτουμε ότι έχουμε στη διάθεσή μας ένα σύνολο N διανυσμάτων εκπαίδευσης, δηλαδή,

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}, \mathbf{x}_i \in \mathbb{R}^l, i = 1, 2, 3, \dots, N$$

Για καθένα από τα διανύσματα εκπαίδευσης γνωρίζουμε την κλάση από την οποία προέρχεται. Η διαθεσιμότητα αυτής της πληροφορίας θα θεωρείται δεδομένη για το σχεδιασμό οποιουδήποτε ταξινομητή στα πλαίσια του βιβλίου αυτού, όπου διαπραγματευόμαστε το πρόβλημα της ταξινόμησης με εκπαίδευση.

Στη συνέχεια, υποθέτουμε ότι γνωρίζουμε τις πιθανότητες $P(\omega_1)$ και $P(\omega_2)$, τις πιθανότητες δηλαδή κάποιο χαρακτηριστικό διάνυσμα να ανήκει σε μία από τις δύο κλάσεις. Μία τέτοια πληροφορία είναι εύκολα υπολογίσιμη. Για παράδειγμα, εάν N_1 από τα διανύσματα εκπαίδευσης ανήκουν στην κλάση ω_1 και N_2 στην κλάση ω_2 ($N=N_1+N_2$), τότε

$$P(\omega_1) \approx \frac{N_1}{N}, P(\omega_2) \approx \frac{N_2}{N}$$

Μία άλλη πληροφορία που μπορεί επίσης να ποσοτικοποιηθεί, και γι' αυτό θα υποθέσουμε ότι είναι γνωστή, είναι η υπό συνθήκη πυκνότητες πιθανότητας $p(\mathbf{x}|\omega_1)$ και $p(\mathbf{x}|\omega_2)$. Οι πυκνότητες πιθανότητας (pdf) μπορούν, για παράδειγμα, να προσεγγιστούν από τα αντίστοιχα ιστογράμματα. Χρησιμοποιώντας τα N_1 διανύσματα της κλάσης ω_1 μπορούμε να προσεγγίσουμε την $p(\mathbf{x}|\omega_1)$ και με τα N_2 διανύσματα της ω_2 την $p(\mathbf{x}|\omega_2)$. Έχουμε τώρα στη διάθεσή μας όλα τα απαραίτητα υλικά για να σχεδιάσουμε τον ταξινομητή μας. Πίσω από τον σχεδιασμό αυτό υπάρχει ένα πολύ απλό και λογικό σκεπτικό. Όταν μας δοθεί

ένα άγνωστο χαρακτηριστικό διάνυσμα \mathbf{x} (που δεν ανήκει προφανώς στο σύνολο των γνωστών διανυσμάτων εκπαίδευσης, X) ο ταξινομητής θα πρέπει να το ταξινομήσει στην πιο πιθανή κλάση. Δεδομένης, δηλαδή, μιας μέτρησης του \mathbf{x} , που αντιστοιχεί στο άγνωστο πρότυπο, το παραπάνω σκεπτικό μας παραπέμπει στο να υπολογίσουμε τις πιθανότητες $P(\omega_1|\mathbf{x})$ και $P(\omega_2|\mathbf{x})$ και να ταξινομήσουμε το \mathbf{x} στην κλάση που αντιστοιχεί στη μεγαλύτερη τιμή από τις δύο πιθανότητες.⁽¹⁾ Οι $P(\omega_i|\mathbf{x})$, $i=1,2$, είναι οι πιθανότητες το πρότυπο να είναι σε μια από τις δύο κλάσεις, όταν γνωρίζουμε την τιμή του \mathbf{x} και είναι γνωστές ως a-posteriori πιθανότητες, σε αντιδιαστολή με τις $P(\omega_i)$, $i=1, 2$, που είναι γνωστές και ως a-priori πιθανότητες. Τις πιθανότητες όμως $P(\omega_i|\mathbf{x})$, $i=1,2$, δεν τις γνωρίζουμε. Αυτό δεν είναι πρόβλημα, καθώς από τη θεωρία πιθανοτήτων γνωρίζουμε το θεώρημα Bayes που μας λέει ότι:

$$P(\omega_i)p(\mathbf{x}|\omega_i) = p(\mathbf{x})P(\omega_i|\mathbf{x})$$

ή

$$P(\omega_i|\mathbf{x}) = \frac{P(\omega_i)p(\mathbf{x}|\omega_i)}{p(\mathbf{x})} \quad (7.1)$$

Είμαστε τώρα έτοιμοι να διατυπώσουμε τον κανόνα ταξινόμησης:

*Εάν η τιμή του \mathbf{x} είναι τέτοια ώστε: $P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x})$, ταξινομήσε το \mathbf{x} στην κλάση ω_1 .
Αλλιώς ταξινομήσε το \mathbf{x} στην κλάση ω_2 .*

Από τον κανόνα αυτό και με βάση την (7.1) καταλήγουμε στον ταξινομητή για δύο κλάσεις, όπως φαίνεται στον Πίνακα 7.1. Παρατηρήστε ότι η $p(\mathbf{x})$ δεν υπεισέρχεται στους τελικούς υπολογισμούς. Αυτό οφείλεται στο ότι είναι μία θετική ποσότητα ανεξάρτητη των κλάσεων και μπορεί να απαλειφθεί από τις ανισότητες.

Η παραπάνω διαδικασία γενικεύεται και για την περίπτωση m κλάσεων. Όπως φαίνεται στον Πίνακα 7.1, το \mathbf{x} καταχωρείται στην κλάση ω_k για την οποία η συνάρτηση $P(\omega_k)p(\mathbf{x}|\omega_k)$ λαμβάνει τη μεγαλύτερη τιμή.

Στο Σχ. 7.2 δίνονται οι πυκνότητες πιθανότητας $p(\mathbf{x}|\omega_1)$ και $p(\mathbf{x}|\omega_2)$ για ένα παράδειγμα ταξινόμησης, με δύο ισοπίθανες κλάσεις $P(\omega_1)=P(\omega_2)=1/2$. Στο παράδειγμα αυτό έχει επιλεγεί να χρησιμοποιηθούν διανύσματα χαρακτηριστικών με μία μόνο συνιστώσα. Γι αυτό αντί του διανύσματος \mathbf{x} χρησιμοποιείται η βαθμωτή μεταβλητή x και ο χαρακτηριστικός χώρος είναι μονοδιάστατος ($l=1$). Με τα πιο πάνω δεδομένα ο ταξινομητής Bayes ορίζει το σημείο x_0 με το οποίο χωρίζεται ο χώρος στην περιοχή R_1 για κάθε σημείο της

⁽¹⁾ Οι Πυκνότητες πιθανότητας συμβολίζονται με μικρό p και οι πιθανότητες με κεφαλαίο P

Ταξινομητής Bayes (για 2 κλάσεις):

Αν $P(\omega_1)p(\mathbf{x}|\omega_1) > P(\omega_2)p(\mathbf{x}|\omega_2)$
 τότε ταξινόμησε το \mathbf{x} στην κλάση ω_1
 Αλλιώς στην κλάση ω_2 .

Ταξινομητής Bayes (για m κλάσεις):

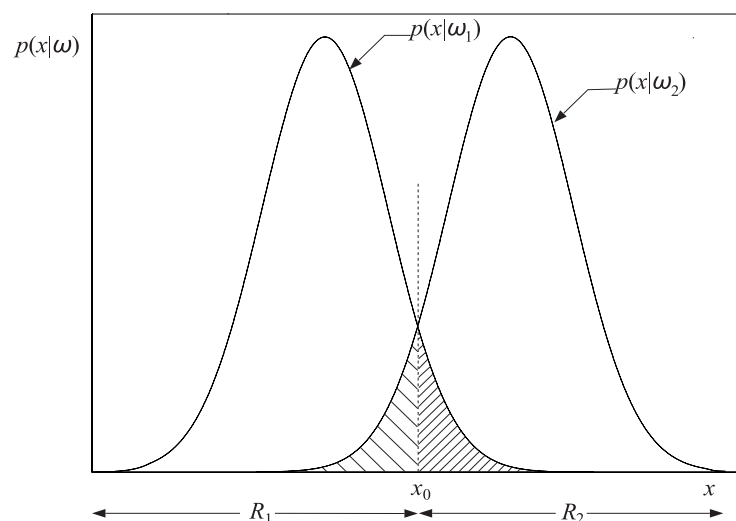
Ταξινόμησε το \mathbf{x} στην κλάση ω_k για την οποία ισχύει:

$$P(\omega_k)p(\mathbf{x}|\omega_k) = \max_{i=1,2,\dots,m} P(\omega_i)p(\mathbf{x}|\omega_i)$$

Πίνακας 7.1 Ταξινομητής Bayes για 2 και για m κλάσεις.

οποίας ισχύει $p(x|\omega_1) > p(x|\omega_2)$ και στην R_2 για την οποία ισχύει αντίστοιχα $p(x|\omega_2) > p(x|\omega_1)$. Είναι ενδιαφέρον να υπολογίσουμε την πιθανότητα P_e με την οποία ένα δείγμα x ταξινομείται σε λάθος κλάση. Ισχύει:

$$\begin{aligned} P_e &= P(\{x \text{ ανήκει στην } \omega_1 \text{ και } x > x_0\} \cup \{x \text{ ανήκει στην } \omega_2 \text{ και } x < x_0\}) = \\ &= P(\{x \text{ ανήκει στην } \omega_1 \text{ και } x > x_0\}) + P(\{x \text{ ανήκει στην } \omega_2 \text{ και } x < x_0\}) = \\ &= P(x \text{ ανήκει στην } \omega_1)P(x > x_0 | x \text{ ανήκει στην } \omega_1) + \\ &\quad P(x \text{ ανήκει στην } \omega_2)P(x < x_0 | x \text{ ανήκει στην } \omega_2) = \\ &= (1/2)P(x > x_0 | x \text{ ανήκει στην } \omega_1) + (1/2)P(x < x_0 | x \text{ ανήκει στην } \omega_2), \text{ ή} \\ P_e &= (1/2)(P_1 + P_2) \end{aligned} \tag{7.2}$$



Σχήμα 7.2: Περιοχές ταξινόμησης που προκύπτουν από τον Bayesian ταξινομητή.

όπου

$$P_1 = P(x > x_0 | x \text{ ανήκει στην } \omega_1) = \int_{x_0}^{+\infty} p(x | \omega_1) dx \quad (7.3)$$

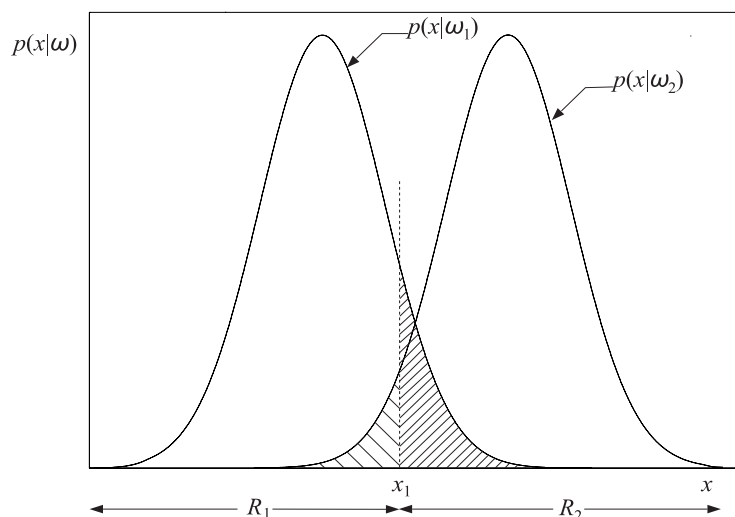
και

$$P_2 = P(x < x_0 | x \text{ ανήκει στην } \omega_2) = \int_{-\infty}^{x_0} p(x | \omega_2) dx \quad (7.4)$$

Από τις (7.2), (7.3) και (7.4) φαίνεται ότι η πιθανότητα σφάλματος P_e είναι ίση με το μισό του ολικού εμβαδού των γραμμοσκιασμένων περιοχών του σχήματος 7.1.

Από το παραπάνω παράδειγμα γίνεται σαφές ότι ο Bayesian ταξινομητής διαιρεί τον χώρο σε περιοχές, έτσι ώστε η περιοχή R_i , όπου αποφασίζεται η κλάση ω_i , είναι εκείνη η περιοχή για την οποία ισχύει $P(\omega_i)p(\mathbf{x} | \omega_i) > P(\omega_j)p(\mathbf{x} | \omega_j)$, $\forall i \neq j$.

Αν και φθάσαμε στον κατά Bayes ταξινομητή μέσα από το απλό σκεπτικό να ταξινομούμε το άγνωστο \mathbf{x} στην πιο πιθανή κλάση, αποδεικνύεται ότι ο ταξινομητής αυτός έχει μια βαθιά θεωρητική υπόσταση. Είναι ο ταξινομητής που δίνει τη μικρότερη πιθανότητα λάθους, P_e , από οποιονδήποτε άλλο ταξινομητή. Με άλλα λόγια, ο κατά Bayes ταξινομητής είναι ο βέλτιστος ταξινομητής. Αυτό μπορεί να το επιβεβαιώσει κανείς και από το σχήμα 7.3. Για οποιοδήποτε άλλη τιμή κατώφλιου $x_1 \neq x_0$, το άθροισμα των εμβαδών των γραμμοσκιασμένων περιοχών γίνεται μεγαλύτερο και, επομένως, μεγαλώνει η P_e . Για μια πιο αναλυτική απόδειξη ο αναγνώστης παραπέμπεται στο [1].



Σχήμα 7.3: Όταν το κατώφλι δεν είναι το βέλτιστο, το ολικό εμβαδόν που αντιστοιχεί στην πιθανότητα λανθασμένης ταξινόμησης, αυξάνει.

Παράδειγμα 1/Κεφαλαίου 7

Σε ένα πρόβλημα τριών κλάσεων χρησιμοποιείται ένα μόνο χαρακτηριστικό για την ταξινόμηση των δειγμάτων. Οι πυκνότητες πιθανότητας $p(x|\omega_i)$ και οι a-priori πιθανότητες

	$P(\omega_i)$	$p(x \omega_i)$
ω_1	1/2	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$
ω_2	1/3	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-1)^2}{2}\right)$
ω_3	1/6	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2}\right)$

$P(\omega_i)$ για τις τρεις κλάσεις καταγράφονται στον παραπλεύρως πίνακα. Η τιμή του χαρακτηριστικού x ενός άγνωστου δείγματος βρέθηκε $x=1.6$. Να ταξινομηθεί το δείγμα αυτό σύμφωνα με τον Bayes ταξινομητή.

Απάντηση

Θα πρέπει να υπολογίσουμε την μέγιστη τιμή $P(\omega_i)p(x|\omega_i), i=1,2,3$, για $x=1.6$.

Έχουμε $P(\omega_1)p(x|\omega_1)=0.0555$, $P(\omega_2)p(x|$

$\omega_2)=0.1111$, $P(\omega_3)p(x|\omega_3)=0.0614$. Σύμφωνα με τη θεωρία που αναπτύξαμε στη ενότητα αυτή, το δείγμα με $x=1.6$ ταξινομείται στην κλάση που αντιστοιχεί στη μέγιστη τιμή, δηλαδή στην κλάση ω_2 .

ΕΝΟΤΗΤΑ 7.3 ΣΥΝΑΡΤΗΣΕΙΣ ΔΙΑΚΡΙΣΗΣ - ΕΠΙΦΑΝΕΙΕΣ ΑΠΟΦΑΣΗΣ - ΤΑΞΙΝΟΜΗΤΗΣ ΕΛΑΧΙΣΤΗΣ ΑΠΟΣΤΑΣΗΣ.

Στην προηγούμενη ενότητα γνωρίσαμε τον κατά Bayes ταξινομητή. Είδαμε ότι η φιλοσοφία διαίρεσης του l -διάστατου χώρου των χαρακτηριστικών στις περιοχές ταξινόμησης $R_i, i=1,2,3,\dots,m$, βασίστηκε σε ένα συγκεκριμένο σύνολο συναρτήσεων, δηλαδή στις συναρτήσεις

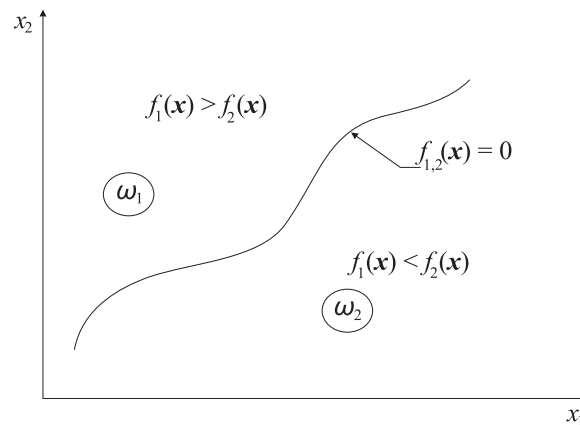
$$h_i(\mathbf{x}) \equiv P(\omega_i)p(\mathbf{x}|\omega_i), i=1,2,\dots,m \quad (7.5)$$

και η διαίρεση του χώρου έγινε σύμφωνα με το πού η καθεμιά από τις παραπάνω m συναρτήσεις παίρνει τιμή μεγαλύτερη από όλες τις άλλες συναρτήσεις. Σε πολλές περιπτώσεις είναι απλούστερο από άποψη μαθηματικού φορμαλισμού και μαθηματικών υπολογισμών να χρησιμοποιούμε συναρτήσεις όπως οι παραπάνω. Με άλλα λόγια, εάν $g(\mathbf{x})$ είναι μια μονότονη συνάρτηση, τότε μπορούμε να ορίσουμε τις

$$f_i(\mathbf{x}) \equiv g(h_i(\mathbf{x})), i=1,2,\dots,m \quad (7.6)$$

και στη θέση των $h_i(\mathbf{x})$ μπορούμε να χρησιμοποιούμε τις $f_i(\mathbf{x})$. Το γεγονός ότι η $g(\mathbf{x})$ είναι μονότονη έχει ως αποτέλεσμα να μην αλλοιώνει τις σχετικές θέσεις των μεγίστων και επομένως η διαίρεση του χώρου είτε με βάση τις h_i ή τις f_i θα είναι ακριβώς η ίδια. Οι συναρτήσεις που προκύπτουν είναι γνωστές και ως **συναρτήσεις διάκρισης**. Ο ένας λόγος που μας οδηγεί στη χρήση συναρτήσεων διάκρισης είναι η απλούστευση των υπολογισμών. Για παράδειγμα, εάν $g(x)=\ln(x)$ (που είναι μονότονη συνάρτηση) και η $p(\mathbf{x}|\omega_i)$ είναι εκθετικής μορφής, όπως στο παράδειγμα 1 του κεφαλαίου 7, η δράση του λογαρίθμου θα έχει ως αποτέλεσμα οι υπολογισμοί να περιορίζονται μόνο στους εκθέτες. Υπάρχει ό-

μως και ένας άλλος ουσιαστικός λόγος. Οι συναρτήσεις διάκρισης γίνονται η αφετηρία “απογαλακτισμού” μας από τον Bayesian ταξινομητή. Ο ταξινομητής αυτός, αν και βέλτιστος, δεν είναι ελεύθερος μειονεκτημάτων. Ο υπολογισμός των $p(\mathbf{x}|\omega_i)$ δεν είναι πάντα μια εύκολη υπόθεση, ιδίως όταν ο αριθμός των δεδομένων εκπαίδευσης, N , είναι σχετικά μικρός. Σε πολλές, επομένως, περιπτώσεις κρίνεται προτιμότερο να διαιρέσει κανείς τον χώρο των χαρακτηριστικών υιοθετώντας μια άλλη κατάλληλη (για κάθε πρόβλημα) επιλογή συναρτήσεων διάκρισης, που να μην σχετίζονται, κατ’ ανάγκη, με τον Bayesian ταξινομητή. Ο ταξινομητής που θα προκύψει δεν θα είναι βέλτιστος, με κριτήριο την πιθανότητα λάθους, αλλά πιθανόν να έχει άλλα προτερήματα που να καθιστούν την χρήση του ελκυστική στην πράξη. Θα συναντήσουμε τέτοιους ταξινομητές σύντομα.



Σχήμα 7.4 Παράδειγμα επιφάνειας απόφασης στον δισδιάστατο χώρο (x_1, x_2) .

Μια άλλη έννοια με “έντονη παρουσία” στην αναγνώριση προτύπων είναι αυτή της **επιφάνειας απόφασης**. Ας πάρουμε την περίπτωση δύο κλάσεων. Για το απλό αυτό πρόβλημα θα έχουμε δύο συναρτήσεις διάκρισης, ας πούμε την $f_1(\mathbf{x})$ και $f_2(\mathbf{x})$. Στην περίπτωση του Bayesian ταξινομητή αυτές είναι οι $P(\omega_i)p(\mathbf{x}|\omega_i)=f_i(\mathbf{x})$, $i=1,2$. Σύμφωνα με τα όσα έχουμε πει, δεν είναι δύσκολο να δούμε ότι αποφασίζουμε ω_1 εάν $f_1(\mathbf{x})-f_2(\mathbf{x})>0$ και ω_2 εάν $f_1(\mathbf{x})-f_2(\mathbf{x})<0$. Ας ορίσουμε τη συνάρτηση

$$f_{1,2}(\mathbf{x}) \equiv f_1(\mathbf{x}) - f_2(\mathbf{x}) \quad (7.7)$$

Η επιφάνεια που ορίζεται από την εξίσωση

$$f_{1,2}(\mathbf{x})=0 \quad (7.8)$$

χωρίζει τον χώρο των χαρακτηριστικών στις δύο περιοχές R_1 και R_2 με βάση τις οποίες αποφασίζουμε ω_1 και ω_2 αντίστοιχα, και συγκεκριμένα

$$R_1: f_{1,2}(\mathbf{x})>0 \text{ και } R_2: f_{1,2}(\mathbf{x})<0 \quad (7.9)$$

Η επιφάνεια που ορίζει η (7.8) είναι γνωστή ως επιφάνεια απόφασης (Σχ. 7.4). Από την

μια πλευρά της ($f_{1,2}(\mathbf{x}) > 0$) βρίσκεται η περιοχή R_1 και από την άλλη ($f_{1,2}(\mathbf{x}) < 0$) η R_2 . Στην περίπτωση πολλών κλάσεων, επιφάνειες απόφασης έχουν νόημα και ορίζονται από συναρτήσεις διάκρισης μόνο γειτονικών περιοχών.

Στη συνέχεια, θα εφαρμόσουμε τα παραπάνω σε μία ειδική κατηγορία προβλημάτων στα οποία οι πυκνότητες πιθανότητας ακολουθούν Gaussian κατανομές. Από τη θεωρία πιθανοτήτων και στατιστικής γνωρίζουμε ότι η πυκνότητα πιθανότητας μιας βαθμωτής τυχαίας μεταβλητής x που ακολουθεί Gaussian κατανομή δίνεται από τη σχέση

$$p(x) = \frac{1}{\sqrt{2\mu\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (7.10)$$

όπου $\mu = E[x]$, η μέση τιμή δηλαδή της x και $\sigma^2 = E[(x-\mu)^2]$, η διασπορά της.

Η (7.10) γενικεύεται για περισσότερες διαστάσεις. Ας το δούμε αυτό λίγο σταδιακά. Εάν x_1, x_2 είναι δύο στοχαστικά ανεξάρτητες Gaussian μεταβλητές, με την ίδια διασπορά σ^2 και μέσες τιμές μ_1, μ_2 αντίστοιχα, τότε (λόγω ανεξαρτησίας) η από κοινού πυκνότητα πιθανότητας (*joint probability density*) θα είναι

$$p(\mathbf{x}) \equiv p(x_1, x_2) = p(x_1)p(x_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_1-\mu_1)^2}{2\sigma^2}\right) \exp\left(-\frac{(x_2-\mu_2)^2}{2\sigma^2}\right), \text{ ή}$$

$$p(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2}\right) \quad (7.11)$$

Στην περίπτωση που οι μεταβλητές δεν είναι ανεξάρτητες και η διάσταση του προβλήματος είναι l , η (7.11) γενικεύεται στην

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{l/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right) \quad (7.12)$$

όπου $|\Sigma|$ η ορίζουσα του $l \times l$ πίνακα Σ , που ορίζεται ως

$$\Sigma \equiv E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T] \quad (7.13)$$

και είναι γνωστός ως **πίνακας συνδιασποράς**, και $\boldsymbol{\mu} = E[\mathbf{x}]$. Εμείς θα εστιάσουμε σε πίνακες συνδιασποράς της μορφής:

$$\Sigma = \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma^2 & \dots & 0 & \dots & 0 \\ 0 & \dots & \sigma^2 & \dots & 0 \\ 0 & \dots & 0 & \dots & \sigma^2 \end{bmatrix} \quad (7.14)$$

Δεν είναι δύσκολο να δει κανείς, ότι για την περίπτωση αυτή και για διάσταση $l=2$, η (7.12) καταλήγει στην (7.11).

Μετά την μικρή αυτή αναδρομή, ας επιστρέψουμε στο πρόβλημα της ταξινόμησης.

Για λόγους απλότητας θα σταθούμε στο πρόβλημα δύο κλάσεων με διδιάστατο χώρο χαρακτηριστικών. Τα αποτελέσματα γενικεύονται άμεσα σε πολλές διαστάσεις και πολλές κλάσεις. Υιοθετούμε τις ακόλουθες υποθέσεις:

- Κλάσεις ισοπίθανες, $P(\omega_1)=P(\omega_2)$.
- Τα χαρακτηριστικά είναι στοχαστικά ανεξάρτητες τυχαίες μεταβλητές με την ίδια διασπορά, σ^2 .
- Σε κάθε κλάση τα χαρακτηριστικά ακολουθούν Gaussian κατανομή.
- Η διασπορά είναι ίδια για όλες τις κλάσεις.

Γράφοντας λίγο διαφορετικά την (7.11), οι παραπάνω υποθέσεις κωδικοποιούνται ως

$$p(\mathbf{x} | \omega_i) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2}\right) \quad i = 1,2 \quad (7.15)$$

όπου $\|\cdot\|$ η Ευκλείδεια απόσταση μεταξύ του \mathbf{x} και $\boldsymbol{\mu}_i$ και $\boldsymbol{\mu}_i$ η μέση τιμή της κλάσης ω_i , δηλαδή $\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 = (x_1 - \mu_{i1})^2 + (x_2 - \mu_{i2})^2$.

Υιοθετώντας στην θέση της $g(\mathbf{x})$ την λογαριθμική συνάρτηση, οι αντίστοιχες συναρτήσεις διάκρισης για την περίπτωση του Bayesian ταξινομητή γίνονται

$$f_i(\mathbf{x}) = \ln(p(\mathbf{x} | \omega_i)) = -\ln(2\pi\sigma^2) - \frac{\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2}$$

και επειδή το σ είναι σταθερά, για τον υπολογισμό της μέγιστης τιμής μεταξύ $f_1(\mathbf{x})$ και $f_2(\mathbf{x})$ αρκεί να υπολογίσουμε την ελάχιστη (λόγω του αρνητικού προσήμου) Ευκλείδεια απόσταση του \mathbf{x} από τις δύο μέσες τιμές $\boldsymbol{\mu}_1$ και $\boldsymbol{\mu}_2$ δηλαδή την ελάχιστη των τιμών $\|\mathbf{x} - \boldsymbol{\mu}_1\|$ και $\|\mathbf{x} - \boldsymbol{\mu}_2\|$. Τελικά το δείγμα ταξινομείται στην κλάση της οποίας η μέση τιμή βρίσκεται πλησιέστερα στο \mathbf{x} . Ο ταξινομητής αυτός είναι γνωστός και ως **ταξινομητής ελάχιστης Ευκλείδειας απόστασης**.

Μερικές φορές ο ταξινομητής αυτός χρησιμοποιείται έστω και εάν οι υποθέσεις που διατυπώθηκαν παραπάνω δεν ισχύουν. Βέβαια, στις περιπτώσεις αυτές ο ταξινομητής δεν είναι Bayesian και, επομένως, δεν είναι βέλτιστος ως προς την πιθανότητα λάθους.

Ας μείνουμε λίγο ακόμα στον απλό αυτό ταξινομητή. Από τον ορισμό του ταξινομητή ελάχιστης απόστασης και από τις (7.8) και (7.9), έπεται ότι η επιφάνεια απόφασης περιγράφεται από την

$$f_{1,2}(\mathbf{x}) = \|\mathbf{x} - \boldsymbol{\mu}_1\|^2 - \|\mathbf{x} - \boldsymbol{\mu}_2\|^2 \quad (7.17)$$

$$\text{ή} \quad f_{1,2}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_2)^T (\mathbf{x} - \boldsymbol{\mu}_2) - (\mathbf{x} - \boldsymbol{\mu}_1)^T (\mathbf{x} - \boldsymbol{\mu}_1) = 0 \quad (7.16)$$

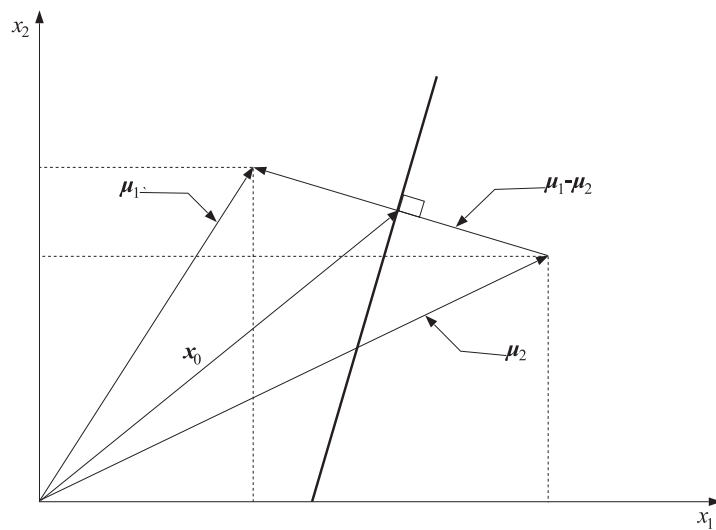
και μετά από λίγες πράξεις καταλήγουμε στη σχέση

$$f_{1,2}(\mathbf{x}) = \mathbf{w}^T (\mathbf{x} - \mathbf{x}_0) = 0 \quad (7.18)$$

όπου

$$\mathbf{w} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \quad \text{και} \quad \mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

Η 7.18 είναι η εξίσωση μιας ευθείας. Από το σχήμα 7.5 φαίνεται ότι η ευθεία αυτή είναι η μεσοκάθετος του ευθύγραμμου τμήματος που συνδέει τα σημεία που αντιστοιχούν στις μέσες τιμές. Πράγματι, όλα τα σημεία πάνω στη γραμμή αυτή απέχουν ίση απόσταση



Σχήμα 7.5: Η επιφάνεια απόφασης που αντιστοιχεί στον ταξινομητή ελάχιστης Ευκλείδειας απόστασης είναι υπερεπίπεδο.

από τις μέσες τιμές. Τα σημεία στην αριστερή πλευρά είναι πλησιέστερα στην $\boldsymbol{\mu}_1$ (ω_1) και αυτά στη δεξιά πλευρά πλησιέστερα στην $\boldsymbol{\mu}_2$ (ω_2). Με άλλα λόγια, η επιφάνεια απόφασης είναι ευθεία (επίπεδο / υπερεπίπεδο, για μεγαλύτερη του $l=2$ διάσταση). Ο ταξινομητής, δηλαδή, είναι γραμμικός. Αυτό συμβαίνει γιατί υποθέσαμε ότι για όλες τις κλάσεις οι διασπορές είναι ίδιες. Δεν είναι δύσκολο να διαπιστώσετε ότι, εάν οι διασπορές δεν είναι ίδιες σε όλες τις κλάσεις, τότε οι επιφάνειες απόφασης που προκύπτουν δεν είναι υπερεπίπεδα αλλά τετραγωνικής μορφής, π.χ., (υπερ)παραβολοειδή, (υπερ)ελλειψοειδή. Για περισσότερες λεπτομέρειες ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο [1].

Παράδειγμα 2/Κεφαλαίου 7.

Σε ένα πρόβλημα χρησιμοποιείται ένα μόνο χαρακτηριστικό για την ταξινόμηση των δειγμάτων σε δύο κλάσεις ω_1 και ω_2 με a-priori πιθανότητες $P(\omega_1)$ και $P(\omega_2)$. Οι υπό συνθήκη πιθανότητες ακολουθούν Gaussian κατανομές με μέσες τιμές 0 και 1 αντίστοιχα για

κάθε κλάση. Η διασπορά είναι σ^2 και για τις δύο κλάσεις. Να υπολογιστεί η τιμή καταωφλίου x_0 που διαχωρίζει τις περιοχές R_1 και R_2 , όπου αποφασίζουμε ω_1 και ω_2 αντίστοιχα.

Απάντηση

Σύμφωνα με τον κατά Bayes ταξινομητή οι περιοχές καθορίζονται από:

$$R_1: P(\omega_1)p(x|\omega_1) > P(\omega_2)p(x|\omega_2) \quad \text{και} \quad R_2: P(\omega_1)p(x|\omega_1) < P(\omega_2)p(x|\omega_2)$$

Στην τιμή καταωφλίου οι δύο ποσότητες είναι ίσες και οι αντίστοιχες καμπύλες τέμνονται (Σχ. 7.2). Άρα η τιμή καταωφλίου x_0 προκύπτει ως λύση της εξίσωσης

$$\frac{P(\omega_1)}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x_0^2}{2\sigma^2}\right) = \frac{P(\omega_2)}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x_0-1)^2}{2\sigma^2}\right).$$

Λαμβάνοντας το λογάριθμο και απαλείφοντας τις ίσες ποσότητες έχουμε:

$$-\frac{x_0^2}{2\sigma^2} + \ln(P(\omega_1)) = -\frac{(x_0-1)^2}{2\sigma^2} + \ln(P(\omega_2)), \quad \text{ή} \quad x_0 = \frac{1}{2} - \sigma^2 \ln\left(\frac{P(\omega_2)}{P(\omega_1)}\right)$$

Στην περίπτωση που το πρόβλημα ήταν τέτοιο ώστε οι καμπύλες να είχαν περισσότερα από ένα σημείο τομής, η εξίσωση θα είχε πολλαπλή λύση.

Παράδειγμα 3/Κεφαλαίου 7.

Σε ένα πρόβλημα με δύο ισοπίθανες κλάσεις ω_1 και ω_2 και δύο ουσιώδη χαρακτηριστικά x_1, x_2 , στο δισδιάστατο χώρο, τα χαρακτηριστικά διανύσματα σε κάθε κλάση ακολουθούν τις κατανομές

$$p(\mathbf{x} | \omega_1) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \quad \text{και} \quad p(\mathbf{x} | \omega_2) = \frac{1}{2\pi} \exp\left(-\frac{(x_1-1)^2 + (x_2-1)^2}{2}\right)$$

Δίνεται ότι ένα πρότυπο χαρακτηρίζεται από το διάνυσμα χαρακτηριστικών $\mathbf{x}=(0.3, 0.6)^T$. Σε ποια κλάση ανήκει, σύμφωνα με τον Bayesian ταξινομητή;

Απάντηση

- Παρατηρούμε ότι οι κλάσεις είναι ισοπίθανες ($P(\omega_1)=P(\omega_2)$) και ότι τα διανύσματα ακολουθούν Gaussian κατανομή και στις δύο κλάσεις. Επίσης, όπως προκύπτει άμεσα από τη σύγκριση των πυκνοτήτων πιθανότητας του παραδείγματος με την (7.11), το μητρώο συνδιασποράς Σ είναι το ίδιο και στις δύο κλάσεις και μάλιστα ισχύει:

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Άρα ο Bayesian ταξινομητής ισοδυναμεί με τον ταξινομητή ελάχιστης Ευκλείδειας απόστασης, όπως ήδη εξηγήσαμε στην ενότητα 7.3. Σύμφωνα με τις κατανομές του παραδείγματος έχουμε:

$$\boldsymbol{\mu}_1 = (0, 0)^T \text{ και } \boldsymbol{\mu}_2 = (1, 1)^T$$

Η Ευκλείδεια απόσταση του χαρακτηριστικού διανύσματος \mathbf{x} του του άγνωστου πρω-τύπου από τις μέσες τιμές είναι $d_{\epsilon_1} = \sqrt{0.3^2 + 0.6^2} = 0.671$ και $d_{\epsilon_2} = \sqrt{0.7^2 + 0.4^2} = 0.806$. Άρα το δείγμα ταξινομείται στην κλάση ω_1 , η μέση τιμή της οποίας ευρίσκεται πλησιέστερα στο \mathbf{x} .

Άσκηση Αυτοαξιολόγησης 1/Κεφαλαίου 7

Θεωρείστε ένα πρόβλημα μίας διάστασης ($l=1$) και δύο ισοπίθανων κλάσεων. Σε κάθε μία από τις κλάσεις το χαρακτηριστικό ακολουθεί την Rayleigh κατανομή

$$p(x | \omega_i) = \begin{cases} \frac{x}{\sigma_i^2} \exp\left(-\frac{x^2}{2\sigma_i^2}\right), & x > 0 \\ 0, & x < 0 \end{cases}$$

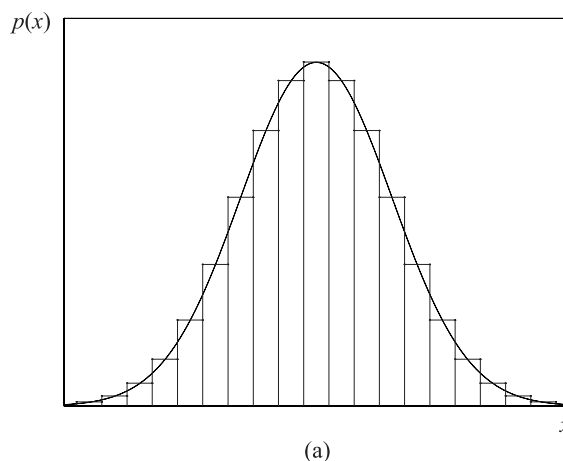
Να υπολογιστεί το σημείο x_0 του κατωφλίου που διαχωρίζει τις δύο περιοχές R_1 και R_2 .

ΕΝΟΤΗΤΑ 7.4 ΠΑΡΑΘΥΡΑ PARZEN

Στις προηγούμενες ενότητες θεωρήσαμε ότι οι πυκνότητες πιθανότητας $p(\mathbf{x} | \omega_i)$ για κάθε κλάση είναι γνωστές. Στην ενότητα αυτή θα μας απασχολήσει το πώς οι συναρτήσεις αυτές εκτιμώνται από τα N διανύσματα του συνόλου εκπαίδευσης X . Ταυτόχρονα, η μεθοδολογία αυτή θα μας ανοίξει το δρόμο για να ορίσουμε ένα νέο ταξινομητή.

Η εκτίμηση των άγνωστων συναρτήσεων $p(\mathbf{x} | \omega_i)$ δεν ακολουθεί κάποιο μονόδρομο και υπάρχει πληθώρα μεθόδων και φιλοσοφιών προσέγγισης του προβλήματος. Στο [1] υπάρχει μια πιο εκτεταμένη παρουσίαση του προβλήματος. Στην παρούσα ενότητα θα προσεγγίσουμε το πρόβλημα με σημείο εκκίνησης το γνωστό μας ιστόγραμμα, στο μονο-διάστατο χώρο.

Έστω ότι θέλουμε να προσεγγίσουμε την πυκνότητα πιθανότητας $p(x)$, της τυχαίας μεταβλητής x με δεδομένο το σύνολο N μετρήσεων του x , το $X = \{x_1, \dots, x_N\}$. Όπως φαίνεται στο Σχ. 7.6 χωρίζουμε τον άξονα των πραγματικών αριθμών σε διαδοχικά διαστήματα εύρους h . Στη συνέχεια μετράμε πόσα από τα N στοιχεία του X εμπίπτουν στο κάθε διάστημα και κατασκευάζουμε το ιστόγραμμα. Ο κάθετος άξονας αντιστοιχεί στον αριθμό των δειγμάτων ανά διάστημα.



Σχήμα 7.6 Προσέγγιση πυκνότητας πιθανότητας με ιστόγραμμα.

Εάν k είναι ο αριθμός δειγμάτων που βρίσκονται σε κάποιο διάστημα, τότε η πιθανότητα να προκύψει τιμή του x στο συγκεκριμένο αυτό διάστημα θα είναι, κατά προσέγγιση, ίση με

$$P \approx k/N \quad (7.19)$$

Στη συνέχεια, θεωρούμε ότι η πυκνότητα πιθανότητας στο διάστημα είναι σταθερή για όλα τα σημεία που ανήκουν στο ίδιο διάστημα. Άρα, η πυκνότητα πιθανότητας για το παραπάνω διάστημα προσεγγίζεται από

$$p(x) \approx \hat{p}(x) = \hat{p}(\hat{x}) = \frac{1}{h} \frac{k}{N}, \quad \hat{x} - \frac{h}{2} \leq x \leq \hat{x} + \frac{h}{2} \quad (7.20)$$

όπου \hat{x} το μέσον του διαστήματος. Η παραπάνω διαδικασία επαναλαμβάνεται για όλα τα διαστήματα. Αποδεικνύεται ότι, εάν: α) η $p(x)$ είναι συνεχής συνάρτηση, β) το εύρος h τείνει στο μηδέν, γ) τα N και k τείνουν στο ∞ με τρόπο ώστε ο λόγος k/N να τείνει στο μηδέν, τότε η προσέγγιση $\hat{p}(x)$ τείνει στην $p(x)$, $\forall x \in \mathbb{R}$.

Οι παραπάνω υποθέσεις είναι πολύ λογικές. Μας λένε ότι το εύρος του διαστήματος πρέπει να τείνει στο μηδέν (ώστε να έχει νόημα η υπόθεση σταθερής τιμής σε όλο το διάστημα για συνεχή $p(x)$), ο αριθμός των δειγμάτων N να είναι πολύ μεγάλος και ταυτόχρονα ο αριθμός των δειγμάτων k σε κάθε διάστημα να είναι επίσης μεγάλος, ώστε η εκτίμηση των πιθανοτήτων να είναι καλή. Δεδομένου βέβαια ότι ο αριθμός των διαστημάτων είναι πολύ μεγάλος ($h \rightarrow 0$), ο αριθμός των δειγμάτων που αντιστοιχεί σε κάθε διάστημα είναι ένα μικρό ποσοστό του N ($k/N \rightarrow 0$).

Ένας άλλος τρόπος για να γράψουμε την 7.20, που θα μας διευκολύνει στη συνέχεια για γενικεύσεις, είναι ο εξής. Ορίζουμε τη συνάρτηση $\varphi(x)$

$$\varphi(x) = \begin{cases} 1, & |x| \leq \frac{1}{2} \\ 0, & \text{αλλού} \end{cases} \quad (7.21)$$

Τότε η προσέγγιση στην 7.20 γράφεται

$$p(x) \approx \frac{1}{h} \left(\frac{1}{N} \sum_{i=1}^N \varphi \left(\frac{x_i - x}{h} \right) \right) \quad (7.22)$$

Πράγματι, από τον ορισμό 7.21 έχουμε ότι

$$\varphi \left(\frac{x - x_i}{h} \right) = \begin{cases} 1, & \text{για } x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2} \\ 0, & \text{αλλού} \end{cases}$$

Με άλλα λόγια, αντικαθιστούμε στην (7.22) την τιμή του x , όπου θέλουμε να εκτιμήσουμε την $p(x)$. Τότε το άθροισμα ισούται με τον αριθμό των δειγμάτων του X που βρίσκονται στο διάστημα εύρους h , το οποίο έχει ως μέσον το συγκεκριμένο x .

Η 7.22 γενικεύεται άμεσα για την περίπτωση των πολλών διαστάσεων, δηλαδή

$$p(\mathbf{x}) \approx \frac{1}{h^l} \left(\frac{1}{N} \sum_{i=1}^N \varphi \left(\frac{\mathbf{x}_i - \mathbf{x}}{h} \right) \right) \quad (7.23)$$

$$\text{όπου } \varphi(\mathbf{x}) = \begin{cases} 1 & \text{για } |x_j| \leq \frac{1}{2}, j = 1, 2, \dots, l \\ 0 & \text{αλλού} \end{cases} \quad (7.24)$$

και x_j η j -διάσταση του \mathbf{x} . Η (7.23) και η (7.24) είναι απόρροια του τρόπου που κατασκευάζουμε ιστογράμματα στον πολυδιάστατο χώρο. Τη θέση διαστημάτων εύρους h την παίρνουν υπερκύβου ακμής h και όγκου h^l . Κατ' αναλογία με όσα είπαμε παραπάνω, οι (7.23) και (7.24) μας λένε ότι, για να προσεγγίσουμε την $p(\mathbf{x})$, καθιστούμε το \mathbf{x} κέντρο ενός υπερκύβου ακμής h , μετρούμε πόσα από τα σημεία του συνόλου εκπαίδευσης βρίσκονται εντός του κύβου αυτού και η $p(\mathbf{x})$ δίνεται από το πηλίκο του αριθμού αυτού με το γινόμενο Nh^l .

Μπορεί να αποδειχθεί ότι η προσέγγιση (7.22) γενικεύεται αν στη θέση της $\varphi(\cdot)$ στην (7.21) χρησιμοποιηθούν και άλλες συναρτήσεις, υπό την προϋπόθεση ότι πληρούν ορισμένες συνθήκες. Οι συναρτήσεις αυτές είναι γνωστές και ως παράθυρα Parzen. Η Gaussian συνάρτηση με μέση τιμή 0 και διασπορά 1 είναι μια τέτοια συνάρτηση. Για περισσότερες λεπτομέρειες ο αναγνώστης παραπέμπεται στο [1].

Για να δούμε τώρα πώς όλα τα παραπάνω ενσωματώνονται στο πρόβλημα της ταξινόμησης. Στον πίνακα 7.2 δίνεται ο αλγόριθμος ταξινόμησης ενός προτύπου, με διάνυσμα χαρακτηριστικών \mathbf{x} , με τη μέθοδο των παραθύρων Parzen. Το πρόβλημα περιλαμβάνει δύο κλάσεις, τις ω_1 και ω_2 με N_1 και N_2 αντίστοιχα διανύσματα εκπαίδευσης.

Η επιλογή του h αποτελεί ένα κρίσιμο βήμα του αλγορίθμου. Όπως γνωρίζουμε ήδη η

- Επιλέγουμε κατάλληλα το h
- Υπολογίζουμε τις αποστάσεις του \mathbf{x} απ' όλα τα διανύσματα εκπαίδευσης, και κρατάμε εκείνα τα οποία βρίσκονται εντός ενός υπερκύβου ακμής h με κέντρο το \mathbf{x} . Με άλλα λόγια εκείνα τα \mathbf{x}_i για τα οποία ισχύει $|x_{ij} - x_j| \leq h/2$, $j = 1, 2, \dots, l$. Από αυτά ας υποθέσουμε ότι βρέθηκαν k_1 να προέρχονται από την κλάση ω_1 και k_2 από την κλάση ω_2 .
- Εάν $P(\omega_1)(k_1/N_1) > (k_2/N_2)P(\omega_2)$, ταξινομούμε το \mathbf{x} στην κλάση ω_1 , διαφορετικά στη κλάση ω_2 .

Πίνακας 7.2 Αλγόριθμος Ταξινόμησης με τη μέθοδο των παραθύρων Parzen.

Πρόβλημα με δύο κλάσεις.

τιμή του πρέπει να τείνει στο μηδέν, αλλά τότε ο αριθμός των δειγμάτων εκπαίδευσης που περιλαμβάνονται στον αντίστοιχο υπερκύβο θα είναι μηδενικός (δεδομένου ότι στην πράξη το N είναι πάντα πεπερασμένο). Το h λοιπόν επιλέγεται όσο γίνεται πιο μικρό αλλά αρκετά μεγάλο ώστε τα k_1, k_2 να έχουν ικανοποιητική τιμή.

Εάν υιοθετούνται γενικευμένες συναρτήσεις ως παράθυρα, τότε χρησιμοποιείται η 7.23 με N_1 και N_2 στη θέση του N , για την $p(\mathbf{x}|\omega_1)$ και $p(\mathbf{x}|\omega_2)$, αντίστοιχα.

Άσκηση Αυτοαξιολόγησης 2/Κεφαλαίου 7

Δίνεται το διάνυσμα $\mathbf{x}=(0.5, 0.6)^T$. Ποια από τα παρακάτω διανύσματα ανήκουν στον 'υπερκύβο' (τετράγωνο. Στην περίπτωση αυτή $l=2$) με ακμή $h=0.4$ και κέντρο το \mathbf{x} .

$$\mathbf{x}_1 = (0.6, 0.5)^T, \mathbf{x}_2 = (0.7, 0.9)^T$$

$$\mathbf{x}_3 = (0.1, 0.6)^T, \mathbf{x}_4 = (0.8, 0.1)^T$$

ΕΝΟΤΗΤΑ 7.5 ΥΠΟΛΟΓΙΣΜΟΣ ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ ΑΠΟ ΤΟΥΣ k ΠΛΗΣΙΕΣΤΕΡΟΥΣ ΓΕΙΤΟΝΕΣ

Στη μέθοδο παραθύρων Parzen, η προσέγγιση της πυκνότητας πιθανότητας σε κάποιο σημείο \mathbf{x} έγινε με βάση τον αριθμό των δειγμάτων εκπαίδευσης, από κάθε κλάση, που εμπεριέχονται σε ένα σταθερό όγκο h^l γύρω από το \mathbf{x} . Στην παρούσα ενότητα θα αντιστρέψουμε τους ρόλους. Αντί να χρησιμοποιήσουμε μία περιοχή σταθερού όγκου, αυξάνουμε σταδιακά τον όγκο της περιοχής γύρω από το \mathbf{x} , μέχρις ότου εντοπιστούν εντός αυ-

τής k διανύσματα εκπαίδευσης. Το k είναι ένας ακέραιος που προεπιλέγεται. Εάν V είναι ο όγκος της περιοχής που περιλαμβάνει τα k διανύσματα, τότε προσεγγίζουμε την άγνωστη τιμή της πυκνότητας πιθανότητας $p(\mathbf{x})$ ως

$$p(\mathbf{x}) \approx \frac{k}{NV(\mathbf{x})}$$

όπου N ο αριθμός των δειγμάτων εκπαίδευσης για την κατανομή $p(\mathbf{x})$ και $V(\mathbf{x})$ ο αντίστοι-

- Μας δίδονται N_1 διανύσματα εκπαίδευσης από την ω_1 και N_2 από την ω_2 .
- Επιλέγουμε το k
- Με κέντρο το \mathbf{x} αυξάνουμε σταδιακά τον όγκο της περιοχής ώστε να συμπεριληφθούν k δείγματα από την κλάση ω_1 και k δείγματα από την κλάση ω_2 . Οι αντίστοιχοι όγκοι θα είναι V_1 και V_2 .
- Εάν $P(\omega_1)(k/N_1V_1) > P(\omega_2)(k/N_2V_2)$ ταξινομούμε το \mathbf{x} στην ω_1 , διαφορετικά στην ω_2 .

Πίνακας 7.3 *Αλγόριθμος Ταξινόμησης, βασισμένος στην εκτίμηση των πυκνοτήτων πιθανότητας με τη μέθοδο των k πλησιέστερων γειτόνων.*

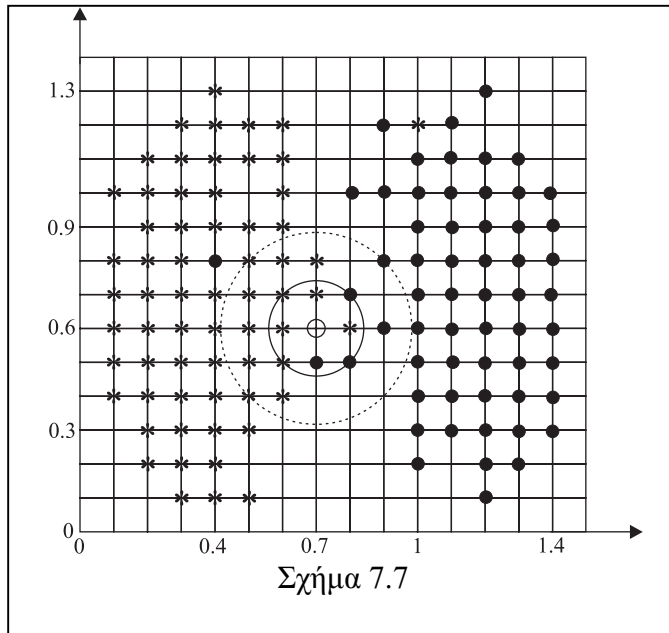
χος όγκος, που δίνεται ως συνάρτηση του \mathbf{x} για να δηλώσει ότι ο όγκος αυτός εξαρτάται από την εκάστοτε θέση. Αποδεικνύεται ότι όταν το $k \rightarrow \infty$ και $(k/N) \rightarrow 0$, η προσέγγιση αυτή συγκλίνει στην τιμή $p(\mathbf{x})$ ([1]).

Εξειδικεύοντας τα παραπάνω στο πρόβλημα της ταξινόμησης, και για την περίπτωση δύο κλάσεων (η γενίκευση για περισσότερες κλάσεις είναι προφανής) προκύπτει ο αλγόριθμος του Πίνακα 7.3. Για την εφαρμογή του αλγορίθμου διαθέτουμε N_1 και N_2 δείγματα εκπαίδευσης από τις κλάσεις ω_1 , ω_2 αντίστοιχα.

Παρατήρηση:

1. Μπορούμε να επιλέξουμε διαφορετικές τιμές του k για τις δύο κλάσεις.
2. Η περιοχή που επιλέγεται δεν είναι ανάγκη να είναι υπερκύβος. Μπορεί να έχει και άλλες μορφές, όπως για παράδειγμα υπερσφαίρα. Αυτό διευκολύνει στην υλοποίηση του αλγορίθμου. Υπολογίζουμε τις Ευκλείδειες αποστάσεις του \mathbf{x} απ' όλα τα διανύσματα κάθε κλάσης, κατατάσσουμε τα τελευταία κατά αύξουσα σειρά απόστασης, και υπολογίζουμε τους αντίστοιχους όγκους της υπερσφαίρας που περιλαμβάνουν k διανύσματα για καθεμιά από τις κλάσεις.

Παράδειγμα 4/ Κεφάλαιο 7



Στο Σχ. 7.7 δίνεται η γεωμετρία κατανομής των χαρακτηριστικών διανυσμάτων από δυο κλάσεις. Τα διανύσματα της ω_1 συμβολίζονται με ‘*’ και της ω_2 με ‘•’. Το πλέγμα έχει πλευρά 0.1. Να ταξινομηθεί το διάνυσμα $(0.7, 0.6)^T$, σύμφωνα με τη μέθοδο εκτίμησης των πυκνοτήτων πιθανότητας με βάση τους 5 πλησιέστερους γείτονες. Δίνεται $P(\omega_1) = P(\omega_2) = 1/2$.

Απάντηση

Υπολογίζουμε τις αποστάσεις όλων των διανυσμάτων της ω_1 από το $(0.7, 0.6)^T$. Τα πέντε πλησιέστερα είναι τα: $(0.8, 0.6)^T$, $(0.7, 0.7)^T$, $(0.6, 0.7)^T$, $(0.6, 0.6)^T$, $(0.6, 0.5)^T$ και βρίσκονται εντός του κύκλου που έχει κέντρο το σημείο $(0.7, 0.6)$ και ακτίνα $\sqrt{0.1^2 + 0.1^2} = \sqrt{2} \times 0.1 \equiv \rho$. Επαναλαμβάνουμε τη διαδικασία για τα πέντε πιο κοντινά σημεία της κλάσης ω_2 . Αυτά είναι τα: $(0.7, 0.5)^T$, $(0.8, 0.5)^T$, $(0.9, 0.6)^T$, $(0.8, 0.7)^T$, $(0.9, 0.8)^T$ και βρίσκονται εντός του κύκλου με κέντρο το σημείο $(0.7, 0.6)$ και με ακτίνα $\rho' = \sqrt{0.2^2 + 0.2^2} = \sqrt{2} \times 0.2 = 2\rho$. Τα διανύσματα εκπαίδευσης που διαθέτουμε είναι $N_1=63$ για την κλάση ω_1 και $N_2=59$ για την κλάση ω_2 . Έχουμε:

$$\frac{1}{2} \frac{5}{N_1 \pi \rho^2} = \frac{1}{2} \cdot \frac{5}{63 \cdot \pi \cdot 2 \cdot 0.01} = 0.63. \quad \text{και} \quad \frac{1}{2} \frac{5}{N_2 \pi (2\rho)^2} = \frac{1}{2} \cdot \frac{5}{59 \cdot \pi \cdot 4 \cdot 2 \cdot 0.01} = 0.17.$$

Δηλαδή το $(0.7, 0.6)^T$ ταξινομείται στην κλάση ω_1 .

ΕΝΟΤΗΤΑ 7.6 ΤΑΞΙΝΟΜΗΤΕΣ ΠΛΗΣΙΕΣΤΕΡΗΣ ΓΕΙΤΝΙΑΣΗΣ

Μια παραλλαγή της προηγούμενης μεθόδου μας οδηγεί σ' ένα νέο ταξινομητή. Στις προηγούμενες μεθόδους, το σημείο αφετηρίας ήταν πάντα ο Bayesian ταξινομητής και οι διάφορες τεχνικές ήταν διαφορετικές απόπειρες εκτίμησης των εμπλεκόμενων pdf's. Στην ενότητα αυτή θα “απελευθερωθούμε”, για πρώτη φορά, από το βέλτιστο ταξινομητή. Το σκεπτικό που δεσπόζει στο νέο ταξινομητή είναι απλό. Για την ταξινόμηση του άγνωστου προτύπου ακολουθούμε τα βήματα:

- Επιλέγουμε ένα k , περιττό για την περίπτωση των δύο κλάσεων.
- Με κέντρο το \mathbf{x} αυξάνουμε τον όγκο της περιοχής έως ότου περιλάβει συνολικά k δια-

νύσματα εκπαίδευσης. Μετρούμε πόσα από αυτά ανήκουν στην κλάση ω_1 , ας πούμε k_1 και πόσα στην κλάση ω_2 , έστω k_2 .

- Ταξινομούμε το \mathbf{x} στην κλάση ω_1 εάν $k_1 > k_2$ και διαφορετικά στην κλάση ω_2 . Για περισσότερες κλάσεις, ταξινομούμε το \mathbf{x} στην κλάση που εκπροσωπείται με το μεγαλύτερο αριθμό διανυσμάτων από τα k συνολικά.

Η απλούστερη εκδοχή της μεθόδου αυτής είναι όταν $k=1$. Με άλλα λόγια, όταν μας δίνεται το \mathbf{x} , υπολογίζουμε την απόστασή του απ' όλα τα διανύσματα εκπαίδευσης του X , και το ταξινομούμε στην κλάση, στην οποία ανήκει το πλησιέστερο στο \mathbf{x} από τα διανύσματα εκπαίδευσης. Αποδεικνύεται ότι, ο απλός αυτός ταξινομητής πλησιέστερου γείτονα μπορεί να οδηγήσει σε ικανοποιητικά αποτελέσματα όταν ο αριθμός N των δειγμάτων εκπαίδευσης είναι αρκετά μεγάλος. Πράγματι, αν θεωρήσουμε ότι το $N \rightarrow \infty$ και συμβολίσουμε με P_B την ελάχιστη πιθανότητα λάθους που αντιστοιχεί, βέβαια, στον Bayesian ταξινομητή, τότε η πιθανότητα λάθους του ταξινομητή πλησιέστερου γείτονα ($k=1$), P_{NN} , φράσσεται [1, 2] από:

$$P_B \leq P_{NN} \leq 2P_B$$

Δηλαδή, το λάθος που θα κάνουμε θα είναι το πολύ διπλάσιο του Bayesian ταξινομητή. Για μικρές τιμές του P_B αυτό δεν είναι κακό. Για παράδειγμα, εάν $P_B=0.005$ τότε $P_{NN} \leq 0.01$. Λάθος της τάξης 1% δεν είναι κακό. Θα πρέπει, βέβαια, να τονιστεί ότι λάθη αυτής της τάξης αντιστοιχούν σε “εύκολα” προβλήματα, δηλαδή σε εύκολα διαχωρίσιμες κλάσεις.

Τέλος, για την περίπτωση όπου $k > 1$ υπάρχουν, επίσης, ενδιαφέροντα αποτελέσματα. Για $N \rightarrow \infty$ μπορεί να αποδειχθεί ότι η πιθανότητα λάθους του ταξινομητή k -πλησιέστερων γειτόνων, P_{kNN} , φράσσεται από [3].

$$P_B \leq P_{kNN} \leq P_B + \sqrt{2P_{NN}/k}$$

Προφανώς, για $k \rightarrow \infty$ $P_{kNN} \rightarrow P_B$. Όλα αυτά, βέβαια, απαιτούν μεγάλα N και k , που, βέβαια, στην πράξη κάθε άλλο παρά τον κανόνα αποτελούν. Ταυτόχρονα δημιουργούν πολύ υψηλές υπολογιστικές απαιτήσεις, δεδομένου ότι κάθε φορά θα πρέπει να υπολογίζονται οι πλησιέστεροι γείτονες μέσα από ένα μεγάλο αριθμό δειγμάτων. Για μια εκτενέστερη συζήτηση ο ενδιαφερόμενος αναγνώστης παραπέμπεται στα [1, 2, 4].

Παράδειγμα 5/Κεφαλαίου 7

Τα διανύσματα εκπαίδευσης για δύο κλάσεις ω_1, ω_2 είναι:

Για την ω_1 :

$\mathbf{x}_1=(3.1,2.5)^T$, $\mathbf{x}_2=(5.6,7.1)^T$, $\mathbf{x}_3=(8.2,2.4)^T$, $\mathbf{x}_4=(3.4, 3)^T$, $\mathbf{x}_5=(4.3,8)^T$, $\mathbf{x}_6=(10,11.1)^T$
 $\mathbf{x}_7=(5.5,3.5)^T$, $\mathbf{x}_8=(5.7,2.5)^T$, $\mathbf{x}_9=(5,3)^T$.

Για την ω_2 :

$\mathbf{x}_{10}=(3.1,3.9)^T$, $\mathbf{x}_{11}=(4,1)^T$, $\mathbf{x}_{12}=(7,2)^T$, $\mathbf{x}_{13}=(3.5, 3.5)^T$, $\mathbf{x}_{14}=(4.5,3.8)^T$, $\mathbf{x}_{15}=(5.2,2.6)^T$
 $\mathbf{x}_{16}=(7.5,2.8)^T$, $\mathbf{x}_{17}=(5.5,2.5)^T$, $\mathbf{x}_{18}=(6,4)^T$.

Να ταξινομηθεί το $\mathbf{x}=(4,3)^T$ σύμφωνα με τον ταξινομητή $k=3$ πλησιέστερων γειτόνων.

Απάντηση:

Υπολογίζουμε τις αποστάσεις του \mathbf{x} απ' όλα τα διανύσματα των κλάσεων ω_1 ω_2 . Τα τρία πλησιέστερα στο \mathbf{x} διανύσματα είναι τα \mathbf{x}_4 (ω_1), \mathbf{x}_{13} (ω_1) και \mathbf{x}_{14} (ω_2), με Ευκλείδειες αποστάσεις $d_4=0.60$, $d_{13}=0.71$ και $d_{14}=0.94$, αντίστοιχα. Οι αποστάσεις όλων των άλλων σημείων είναι μεγαλύτερες. Σύμφωνα με τον ταξινομητή $k=3$ πλησιέστερων γειτόνων ταξινομούμε το \mathbf{x} στην κλάση ω_2 , εφόσον οι δύο από τους πλησιέστερους γείτονες ανήκουν στην ω_2 .

Άσκηση Αυτοαξιολόγησης 3/Κεφαλαίου 7

Για το πρόβλημα του παραδείγματος 7.5, να ταξινομηθεί το $\mathbf{x}=(6.2,4.3)^T$ σύμφωνα με τον ταξινομητή πλησιέστερης γειτνίασης $k=1$.

Άσκηση Αυτοαξιολόγησης 4/Κεφαλαίου 7

	Σωστό	Λάθος
1. Η επιλογή των ουσιωδών χαρακτηριστικών πρέπει να γίνεται με τέτοιο τρόπο ώστε αυτά να έχουν την ίδια τιμή για όλα τα στοιχεία μιας κλάσης.		
2. Για να λυθεί το πρόβλημα ταξινόμησης δύο κλάσεων χωρίζουμε πάντα τον δισδιάστατο χώρο, το επίπεδο, σε περιοχές καθεμιά από τις οποίες θα αντιστοιχεί σε μία κλάση.		
3. Ένας κατά Bayes ταξινομητής στη λειτουργία του παρουσιάζει σφάλματα, όπως συμβαίνει και με οποιοδήποτε άλλο ταξινομητή. Η πιθανότητα όμως να συμβεί ένα σφάλμα είναι η ελάχιστη δυνατή.		
4. Ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης είναι μια παραλλαγή του Bayesian ταξινομητή, γι' αυτό επιτυγχάνει πάντα πιθανότητα σφάλματος P_{ED} ελάχιστη, ίση με του ταξινομητή Bayes, P_B .		
1. 5. Η ταξινόμηση με την τεχνική των παραθύρων Parzen δεν απαιτεί κανένα άμεσο υπολογισμό για στατιστικά στοιχεία των κλάσεων που ταξινομούμε.		

6. Η τιμή της πιθανότητας σφάλματος P_{NN} στους ταξινομητές πλησιέστερου γείτονα είναι πάντα ίση με την πιθανότητα σφάλματος P_B του Bayes ταξινομητή.		
---	--	--

Σύνοψη Κεφαλαίου

Στο κεφάλαιο αυτό έγινε μια συνοπτική περιγραφή του τρόπου επιλογής των ουσιωδών χαρακτηριστικών με τα οποία περιγράφονται τα πρότυπα ενός συστήματος ταξινόμησης. Στη συνέχεια παρουσιάστηκε η κατηγορία των τεχνικών ταξινόμησης που στηρίζονται σε στατιστικές μεθόδους. Στα πλαίσια των τεχνικών αυτών γνωρίσαμε τον ταξινομητή Bayes, καθώς και τους ταξινομητές ελάχιστης απόστασης. Επίσης γνωρίσαμε τεχνικές προσέγγισης των συναρτήσεων πυκνότητας πιθανότητας οι οποίες οδηγούν σε μεθόδους ταξινόμησης, όπως τη χρήση των k πλησιέστερων γειτόνων ή του ενός πλησιέστερου γείτονα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] S. Theodoridis, K. Koutroumbas: “Pattern Recognition”, Academic Press, 1998
- [2] R. Duda, P.E. Hart: “Pattern Classification and Scene Analysis”, John Wiley, 1973
- [3] L. Derroye, L. Gyorfí, G. Lugosi: “A Probabilistic Theory of Pattern Recognition”, Springer-Verlag, 1996
- [4] F. Fukunaga: “Introduction to Statistical Pattern Recognition”, 2nd ed., Academic Press, 1990

ΑΠΑΝΤΗΣΕΙΣ ΑΣΚΗΣΕΩΝ ΑΥΤΟΑΞΙΟΛΟΓΗΣΗΣ

7.1

Σύμφωνα με τον Bayesian ταξινομητή οι δύο περιοχές R_1 , R_2 καθορίζονται ανάλογα με το ποια από τις ποσότητες $P(\omega_1)p(x|\omega_1)$ και $P(\omega_2)p(x|\omega_2)$ είναι μεγαλύτερη. Στην τιμή κατωφλίου, x_0 , οι δυο ποσότητες είναι ίσες και οι αντίστοιχες καμπύλες τέμνονται. Η τιμή κατωφλίου προκύπτει επομένως από τη λύση της εξίσωσης:

$$\frac{x_0}{\sigma_1^2} \exp\left(-\frac{x_0^2}{2\sigma_1^2}\right) = \frac{x_0}{\sigma_2^2} \exp\left(-\frac{x_0^2}{2\sigma_2^2}\right)$$

και λαμβάνοντας τους λογαρίθμους αμφοτέρων των πλευρών προκύπτει:

$$\ln(x_0) + \ln\left(\frac{1}{\sigma_1^2}\right) - \frac{x_0^2}{2\sigma_1^2} = \ln(x_0) + \ln\left(\frac{1}{\sigma_2^2}\right) - \frac{x_0^2}{2\sigma_2^2}, \text{ ή } x_0^2 \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2}\right) = \ln\left(\frac{1}{\sigma_2^2}\right) - \ln\left(\frac{1}{\sigma_1^2}\right)$$

και τελικά

$$x_0 = \sqrt{\frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 - \sigma_2^2} \ln \frac{\sigma_1^2}{\sigma_2^2}}$$

Συγχαρητήρια σε όποιον έχει λύσει με επιτυχία την άσκηση. Αν λύσατε την άσκηση, αλλά υπήρξαν λογιστικά λάθη να είστε προσεκτικοί στις επόμενες ασκήσεις και όλα θα πάνε καλά. Όποιος δεν κατάφερε να προχωρήσει καθόλου στην άσκηση ας μην απελπίζεται. Να διαβάσει προσεκτικά την Ενότητα 7.3 και να λύσει τα παραδείγματα 7.1 και 7.2. Τότε σίγουρα θα είναι σε θέση να αντιμετωπίζει τέτοιες ασκήσεις.

7.2

	$ x_{i1}-x_1 $	$ x_{i2}-x_2 $	$ x_{i1}-x_1 <h/2$ και $ x_{i2}-x_2 <h/2$
\mathbf{x}_1	0.1	0.1	Αληθής
\mathbf{x}_2	0.2	0.3	Ψευδής
\mathbf{x}_3	0.4	0.0	Ψευδής
\mathbf{x}_4	0.3	0.5	Ψευδής

Για να βρίσκεται ένα διάνυσμα \mathbf{x}_i εντός του τετραγώνου με κέντρο το \mathbf{x} και ακμή $h=0.4$, πρέπει και οι δύο συνιστώσες του, x_{i1} και x_{i2} να πληρούν τη

σχέση $|x_{ij}-x_j|<h/2$, $j=1,2$ όπου x_1, x_2 οι συνιστώσες του \mathbf{x} . Εκτελώντας τους υπολογισμούς για τα τέσσερα διανύσματα της άσκησης, όπως φαίνεται και στον πίνακα, διαπιστώνουμε ότι μόνο ένα από αυτά, το \mathbf{x}_1 , βρίσκεται εντός του τετραγώνου.

Όποιος έλυσε σωστά την άσκηση σίγουρα έχει καταλάβει την ταξινόμηση με τη μέθοδο των παραθύρων Parzen. Αν έγινε λάθος που δεν οφείλεται σε πράξεις μελετήστε ξανά προσεκτικά την ενότητα 7.4, πριν συνεχίσετε τη μελέτη του υπόλοιπου κεφαλαίου.

7.3

Σύμφωνα με τις αποστάσεις, όπως υπολογίστηκαν και στο παράδειγμα 7.5, το πλησιέστερο διάνυσμα στο \mathbf{x} είναι το \mathbf{x}_{18} που ανήκει στην κλάση ω_2 . Σύμφωνα, επομένως, με τον ταξινομητή πλησιέστερου γείτονα, ταξινομούμε το \mathbf{x} στην κλάση ω_2 .

Αν δεν βρήκατε σωστά τη λύση μελετήστε ξανά την ενότητα 7.6 και σίγουρα δεν θα επαναλάβετε το λάθος.

7.4

	Σωστό	Λάθος
1. Η επιλογή των ουσιωδών χαρακτηριστικών πρέπει να γίνεται με τέτοιο τρόπο ώστε αυτά να έχουν την ίδια τιμή για όλα τα στοιχεία μιας κλάσης.		✓
Οι τιμές των ουσιωδών χαρακτηριστικών μεταβάλλονται από πρότυπο σε πρότυπο λόγω των σφαλμάτων που δημιουργούνται κατά τη μέτρησή τους και διαφοροποιήσεις των ίδιων των προτύπων (Υποενότητα 7.1.1).		

<p>2. Για να λυθεί το πρόβλημα ταξινόμησης δύο κλάσεων χωρίζουμε πάντα τον διδιάστατο χώρο, το επίπεδο, σε περιοχές καθεμιά από τις οποίες θα αντιστοιχεί σε μία κλάση.</p> <p>Πρέπει να χωριστεί ο χώρος των ουσιωδών χαρακτηριστικών σε περιοχές που να αντιστοιχούν στις κλάσεις. Η διάσταση αυτού είναι l και όχι απαραίτητα 2 ώστε να είναι επίπεδο. (Υποενότητα 7.1.2)</p>		✓
<p>3. Ένας κατά Bayes ταξινομητής στη λειτουργία του παρουσιάζει σφάλματα, όπως συμβαίνει και με οποιοδήποτε άλλο ταξινομητή. Η πιθανότητα όμως να συμβεί ένα σφάλμα είναι η ελάχιστη δυνατή.</p> <p>Πράγματι! Σε οποιοδήποτε ταξινομητή μερικά πρότυπα ταξινομούνται εσφαλμένα. Στον Bayesian ταξινομητή, όμως, η πιθανότητα σφάλματος είναι η ελάχιστη δυνατή. Το πρόβλημα όμως με τον ταξινομητή αυτό είναι ότι για την κατασκευή του πρέπει να γνωρίζουμε επακριβώς τις συναρτήσεις πυκνότητας πιθανότητας, $p(\mathbf{x} \omega_i)$, $i=1,2,\dots,m$, πράγμα που δεν συμβαίνει πάντα στην πράξη. (Ενότητα 7.2)</p>	✓	
<p>4. Ο ταξινομητής ελάχιστης Ευκλείδειας απόστασης είναι μια παραλλαγή του Bayesian ταξινομητή, γι' αυτό επιτυγχάνει πάντα πιθανότητα σφάλματος P_{ED} ελάχιστη, ίση με του ταξινομητή Bayes, P_B.</p> <p>Θα προσέξατε ότι στην τάξη των ταξινομητών αυτών δεν χρησιμοποιούνται οι $p(\mathbf{x} \omega_i)$, $i=1,2,\dots,m$, αλλά μόνο η μέση τιμή των ουσιωδών χαρακτηριστικών. Για το λόγο αυτό στη γενική περίπτωση ισχύει $P_{ED} > P_B$. Μόνο σε περιπτώσεις με ειδικές συμμετρίες, όπως στην περίπτωση που όλες οι κατανομές είναι Gaussian με τον ίδιο διαγώνιο πίνακα συνδιασποράς $\sigma^2 I$, ισχύει $P_{ED} = P_B$. (Ενότητα 7.3)</p>		✓
<p>5. Η ταξινόμηση με την τεχνική των παραθύρων Parzen δεν απαιτεί κανένα άμεσο υπολογισμό για στατιστικά στοιχεία των κλάσεων που ταξινομούμε.</p> <p>Στην πραγματικότητα στην τεχνική αυτή χρησιμοποιούνται δείγματα εκπαίδευσης και από τις τιμές των χαρακτηριστικών τους προσεγγίζονται οι $p(\mathbf{x} \omega_i)$, $i=1,2,\dots,m$. Αν χρησιμοποιηθεί πολύ μεγάλος αριθμός τέτοιων δειγμάτων το σφάλμα του ταξινομητή πλησιάζει σημαντικά την ελάχιστη τιμή του Bayes (Ενότητα 7.4).</p>	✓	

6. Η τιμή της πιθανότητας σφάλματος P_{NN} στους ταξινομητές πλησιέστερου γείτονα είναι πάντα ίση με την πιθανότητα σφάλματος P_B του Bayes ταξινομητή.		✓
Στην περίπτωση που το πλήθος των δειγμάτων εκπαίδευσης είναι πολύ μεγάλο, ισχύει $P_B \leq P_{NN} \leq 2P_B$ επομένως για μικρές τιμές του P_B η P_{NN} παραμένει μικρή. (Ενότητα 7.6)		

Συγχαρητήρια αξίζουν σε όσους απάντησαν σωστά και στις έξι ερωτήσεις. Ικανοποιημένοι είμαστε και από όσους απάντησαν στις πέντε. Όσοι έκαναν περισσότερα από δύο λάθη δεν πρέπει να απογοητευτούν. Να διαβάσουν ξανά προσεκτικά τις αντίστοιχες ενότητες και θα είναι σε θέση να συνεχίσουν τη μελέτη του επόμενου κεφαλαίου.