Κεφάλαιο 3

Στοχαστικά συστήματα

3.1 Εισαγωγή

Στο κεφάλαιο αυτό περιγράφονται, αναλύονται στα μέρη που τα αποτελούν και μελετώνται συστήματα ταξινόμησης προτύπων που περιέχουν στις βαθμίδες του στοχαστικές επεξεργασίες.

Στο προηγούμενο κεφάλαιο μελετήσαμε τα δομικά συστήματα ταξινόμησης προτύπων τα οποία βασίζονται είτε σε μέτρηση της απόστασης του άγνωστου πρότυπου με το πρωτότυπο ή τα πρωτότυπα των κατηγοριών είτε χρησιμοποιήσαμε συναρτήσεις απόφασης ή συναρτήσεις διάκρισης. Τα συστήματα αυτά είναι σε θέση να δώσουν υψηλή αξιοπιστία ταξινόμησης αλλά δεν μπορούν να αντιμετωπίσουν με επιτυχία την παρουσία θορύβου στις μετρήσεις των προτύπων, είτε αυτή προέρχεται από ασαφείς μετρήσεις, είτε με προσθήκη θορύβου από τηλεπικοινωνιακά συστήματα μετάδοσης πληροφοριών.

Τα στοχαστικά συστήματα ταξινόμησης προτύπων μπορούν να αντιμετωπίσουν με βέλτιστο τρόπο το πρόβλημα της ταξινόμησης των προτύπων. Μπορούν δηλαδή να ελαχιστοποιήσουν την πιθανότητα του σφάλματος στην ταξινόμησης, να αντιμετωπίσουν προβλήματα παρεμβολής θορύβου στις μετρήσεις των προτύπων και επιπρόσθετα μπορούν να εκμεταλευτούν δύο επιπρόσθετες πληροφορίες:

1. Την συχνότητα εμφάνισης των προτύπων κάποιας κατηγορίας. Η πληροφορία αυτή είναι σημαντική διότι μπορούμε να κατασκευάσουμε το στοχαστικό σύστημα ταξινόμησης προτύπων έτσι ώστε να αναγνωρίζει με μικρότερη πιθανότητα σφάλματος τα πρότυπα των συχνότερα εμφανιζόμενων κατηγοριών, έτσι ώστε η συνολική πιθανότητα σφάλματος να ελαττώνεται.

2. Τα στοχαστικά συστήματα ταξινόμησης μπορούν να ενσωματώσουν την έννοια του κόστους ή της επικινδυνότητας (risk) γιά την απόφαση που λαμβάνεται. Σε πρακτικές εφαρμογές το είδος του σφάλματος που μπορεί να συμβεί με την λήψη της απόφασης απο το σύστημα ταξινόμησης έχει διαφορετική σημασία και συνεπώς διαφορετική επικινδυνότητα. Χαρακτηριστικό παράδειγμα διαφορετικής επικινδυνότητας που μπορεί να έχουν τα λάθη της ταξινόμησης είναι οι εφαρμογές επιβεβαίωσης ταυτότητας σε συστήματα ασφαλείας. Φανταστείτε πόσο σημαντικό είναι το σφάλμα που θα συμβεί όταν το σύστημα πιστοποιεί ότι οι μετρήσεις του πρότυπου προέρχεται από άνθρωπο στον οποίο επιτρέπεται η πρόσβαση σε σύστημα ασφαλείας, ενώ στην πραγματικότητα ο άνθρωπος δεν επιτρέπεται να έχει πρόσβαση. Η αντίθετη περίπτωση σφάλματος, το σύστημα να απορρίπτει την αίτηση πρόσβασης ενώ ο άνθρωπος να έχει εξουσιοδότηση πρόσβασης, είναι πολύ λιγότερο σημαντική διότι ο εξεταζόμενος μπορεί απλά να προσπαθήσει ξανά να αποκτήσει πρόσβαση στο σύστημα ασφάλειας.

Τα παραπάνω χαρακτηριστικά προσδίδουν στα συστήματα ταξινόμησης με στοχαστικές μεθόδους υψηλή αξιοπιστία, ιδιαίτερα στις εφαρμογές όπου παρατηρείται σημαντική διαφορά τόσο στην συχνότητα εμφάνισης των προτύπων των κατηγοριών όσο και στην επικινδυνότητα των σφαλμάτων του συστήματος ταξινόμησης.

Παράδειγμα 25 *Εστω σύστημα μετάδοσης δεδομένων διαδικής μορφής από τηλεπικοινωνιακό δορυφόρο σε επίγειο σταθμό. Το σήμα που λαμβάνεται είναι ένας πραγματικός αριθμός. Εχει μετρηθεί ότι η ατμόσφαιρα προσθέτει θόρυβο στο σήμα που μεταδίδεται. Ο θόρυβος κατά προσέγγιση έχει κανονική πυκνότητα πιθανότητας με διαφορετική διασπορά στη στάθμη 0 από ότι στην στάθμη 1. Το σήμα που λαμβάνεται στον επίγειο σταθμό (x) έχει την ακόλουθη πυκνότητα πιθανότητας ανάλογα με το αν ο δορυφόρος έστειλε 0 ή 1:*

$$p(x|\text{Ο δορυφόρος έστειλε } 0) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{x^2}{2\sigma_0^2}}$$

$$p(x|\text{Ο δορυφόρος έστειλε } 1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-1)^2}{2\sigma_1^2}}$$

Το πρόβλημα που καλούμαστε να λύσουμε είναι το ακόλουθο: Ποιό είναι το βέλτιστο κατώφλι για το οποίο θα αποφασίζουμε ότι, αν το σήμα που λαμβάνουμε είναι μικρότερο από το κατώφλι τότε ο δορυφόρος έστειλε 0 διαφορετικά ο δορυφόρος έστειλε 1.

Αν ορίσουμε το κατώφλι που θέλουμε να βρούμε σαν T, τότε η απόφαση μας θα είναι

$$x \leq T \Rightarrow x \in \omega_0$$

$$x > T \Rightarrow x \in \omega_1$$

όπου ω_0 είναι το γεγονός ότι ο δορυφόρος έστειλε 0 και ω_1 είναι το γεγονός ότι ο δορυφόρος έστειλε 1.

Η στοχαστική απόφαση θα πρέπει να ληφθεί υπολογίζοντας τις πιθανότητες $p(\omega_0|x)$ και $p(\omega_1|x)$. Τα μεγέθη αυτά μετρούν την πιθανότητα ο δορυφόρος να έστειλε 0 ή 1 αντίστοιχα δεδομένου ότι έχουμε λάβει την μέτρηση του συγκεκριμένου πρότυπου x.

Η βέλτιστη απόφαση θα πρέπει να ληφθεί εξετάζοντας τις αντίστοιχες αριθμητικές τιμές των πιθανοτήτων, αν $p(\omega_0|x) > p(\omega_1|x)$ τότε "ο δορυφόρος έστειλε 0". Συνεπώς το κατώφλι T βρίσκεται από την τιμή του x που αποτελεί και λύση της εξίσωσης:

$$p(\omega_0|x) = p(\omega_1|x)$$

Εφαρμόζοντας τον κανόνα του Bayes και στα δύο μέλη της εξίσωσης έχουμε:

$$\frac{p(x|\omega_0)p(\omega_0)}{p(x)} = \frac{p(x|\omega_1)p(\omega_1)}{p(x)} \Leftrightarrow p(x|\omega_0)p(\omega_0) = p(x|\omega_1)p(\omega_1)$$

Αντικαθιστώντας τις αντίστοιχες πυκνότητες πιθανότητας έχουμε μια εξίσωση με άγνωστο το x. Η λύση της εξίσωσης μας δίνει το στοχαστικά βέλτιστο κατώφλι απόφασης T.

$$\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{x^2}{2\sigma_0^2}} p(\omega_0) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-1)^2}{2\sigma_1^2}} p(\omega_1) \Leftrightarrow$$

$$\frac{(x-1)^2}{2\sigma_1^2} - \frac{x^2}{2\sigma_0^2} = ln(\frac{p(\omega_1)\sigma_0}{p(\omega_0)\sigma_1})$$

Οι λύσεις που προκύπτουν είναι δύο:

$$T_1 = \frac{\sigma_0^2}{\sigma_0^2 - \sigma_1^2} - \frac{\sigma_0\sigma_1\sqrt{1 - 2(\sigma_0^2 - \sigma_1^2)ln(\frac{p(\omega_0)\sigma_1}{p(\omega_1)\sigma_0})}}{\sigma_0^2 - \sigma_1^2}$$

$$T_2 = \frac{\sigma_0^2}{\sigma_0^2 - \sigma_1^2} + \frac{\sigma_0\sigma_1\sqrt{1 - 2(\sigma_0^2 - \sigma_1^2)ln(\frac{p(\omega_0)\sigma_1}{p(\omega_1)\sigma_0})}}{\sigma_0^2 - \sigma_1^2}$$

Στην περίπτωση που οι διασπορές του θορύβου είναι ίδιες $\sigma_0 = \sigma_1 = \sigma$, τότε η λύση που προκύπτει είναι:

$$T = \frac{1}{2} + \sigma^2 ln(\frac{p(\omega_0)}{p(\omega_1)})$$

Για να δούμε μερικά χρήσιμα συμπεράσματα από την αλγεβρική λύση του κατωφλίου στην περιπτωση που ο προσθετικός θόρυβος έχει ίδια διασπορά στο σήμα που μεταφέρει το 0 και στο σήμα που μεταφέρει το 1.

Αν ο δορυφόρος στέλνει ισοπίθανα τα 0 και 1, τότε ανεξάρτητα από την ποσότητα του θορύβου που προστίθεται στο σήμα, το βέλτιστο κατώφλι απόφασης βρίσκεται στην τιμή 0.5.

Αν η διασπορά έχει αριθμητική τιμή 0.1 και ο δορυφόρος μεταδίδει διπλάσια 0 από 1, τότε το βέλτιστο κατώφλι έχει τιμή $T = 0.5 + 0.1 \times ln(2) = 0.5693$.

Αν το επίπεδο θορύβου αυξηθεί τόσο ώστε η διασπορά να πάρει την τιμή 0.5 και ο δορυφόρος μεταδίδει διπλάσια 0 από 1, τότε το βέλτιστο κατώφλι έχει τιμή $T = 0.5 + 0.5 \times ln(2) = 0.8466$.

Αν το επίπεδο θορύβου αυξηθεί σημαντικά έτσι ώστε η διασπορά να πάρει την τιμή 1.0 και ο δορυφόρος συνεχίσει να μεταδίδει διπλάσια 0 από 1, τότε το βέλτιστο κατώφλι έχει τιμή $T = 0.5 + 1 \times ln(2) = 1.19347$. Σε αυτή την περίπτωση παρουσιάζεται το φαινομενικά παράδοξο γεγονός να αποφασίζουμε ότι ο δορυφόρος έστειλε "1" μόνο στην περίπτωση που το σήμα που λάβαμε έχει αριθμητική τιμή μεγαλύτερη του 1.19347. Αν το σήμα έχει αριθμητική τιμή μικρότερη του 1.19347 τότε αποφασίζουμε ότι ο δορυφόρος έστειλε 0. Η απόφαση αυτή εξασφαλίζει, όπως θα αποδειχθεί παρακάνω, την μικρότερη πιθανότητα λανθασμένης απόφασης.

3.2 Στοχαστικά κριτήρια ταξινόμησης

Τα στοχαστικά συστήματα ταξινόμησης προτύπων παρέχουν περισσότερες δυνατότητες απο τα δομικά συστήματα διότι με αυτά μπορούμε να εκμεταλευτούμε και στατιστικές πληροφορίες, οι οποίες έχουν να κάνουν είτε με την συχνότητα με την οποία εμφανίζονται τα πρότυπα των κατηγοριών είτε με την σημαντικότητα της απόφασης που παίρνεται για κάθε κατηγορία. Δυστυχώς τα δομικά συστήματα και τα συστήματα διαδυνδεδεμένων δικτύων δεν έχουν την δυνατότητα χρήσης στατιστικών πληροφοριών οι οποίες στις περισσότερες πρακτικές εφαρμογές είναι διαθέσιμες.

3.2.1 Το κριτήριο μεγαλύτερης πιθανότητας

Η πιο απλή στοχαστική μέθοδος ταξινόμησης πραγματοποιείται με την εύρεση της κατηγορίας εκείνης η οποία έχει και την μεγαλύτερη πιθανότητα να συμβεί, δεδομένης της πληροφορίας που είναι διαθέσιμης δηλαδή της μέτρησης του πρότυπου $\mathbf{x}$. Το αντίστοιχο κριτήριο ταξινόμησης δίνεται από την σχέση:

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad p(\omega_j|\mathbf{x}) > p(\omega_i|\mathbf{x}) \quad \forall i \neq j \tag{3.1}$$

Η πιθανότητα $p(\omega_i|\mathbf{x})$ δεν μπορεί να υπολογιστεί εύκολα απευθείας. Απο τα παραδείγματα μπορεί να υπολογιστεί κατά προσέγγιση η πιθανότητα $p(\mathbf{x}|\omega_i)$. Με τον κανόνα του Bayes το κριτήριο της μεγαλύτερης πιθανότητας τροποποιείται ως εξής:

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad \frac{p(\mathbf{x}|\omega_j)p(\omega_j)}{p(x)} > \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(x)} \quad \forall i \neq j \tag{3.2}$$

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad p(\mathbf{x}|\omega_j)p(\omega_j) > p(\mathbf{x}|\omega_i)p(\omega_i) \quad \forall i \neq j \tag{3.3}$$

Στο τελικό κριτήριο ταξινόμησης βλέπουμε ότι συμμετέχουν και οι πιθανότητες εμφάνισης πρότυπου των κατηγοριών $p(\omega_i)$. Η συμβολή των πιθανοτήτων αυτών στην αύξηση της αξιοπιστίας του συνολικού συστήματος ταξινόμησης είναι σημαντική ιδιαίτερα στις περιπτώσεις στις οποίες παρουσιάζονται μεγάλες διαφορές στις πιθανότητες εμφάνισης των κατηγοριών, όπως συνέβει στο προηγούμενο παράδειγμα.

Παράδειγμα 26 *Εστω σύστημα ταξινόμησης δι-διάστατων προτύπων τριών κατηγοριών με γνωστές τις ακόλουθες πυκνότητες πιθανότητας:*

$$p(\mathbf{x}|\omega_1) = \begin{cases} \frac{1}{4} & |\mathbf{x}| < 1 \\ 0 & |\mathbf{x}| \geq 1 \end{cases}$$

$$p(\mathbf{x}|\omega_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mathbf{x}-(2,2)^T)^T(\mathbf{x}-(2,2)^T)}$$

$$p(\mathbf{x}|\omega_3) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mathbf{x}-(2,0)^T)^T(\mathbf{x}-(2,0)^T)}$$

Μας πληροφορούν ότι τα πρότυπα της δεύτερης κατηγορίας εμφανίζονται με διπλάσια συχνότητα από τα πρότυπα των άλλων κατηγοριών οι οποίες είναι ισοπίθανες. Μετρούμε το πρότυπο $(0.9, 0.9)^T$.

1. Σε ποιά κατηγορία μπορούμε να το ταξινομήσουμε με το κριτήριο της μεγαλύτερης πιθανότητας;

2. Τι θα συμβεί αν τα πρότυπα της δεύτερης κατηγορίας εμφανίζονται με τριπλάσια συχνότητα από τα πρότυπα των άλλων κατηγοριών;

Υπολογίζουμε τις αριθμητικές τιμές των παραστάσεων του κριτηρίου ταξινόμησης της μέγαλύτερης πιθανότητας όπως αυτές δίνονται στην εξίσωση 3.3:

$$p((0.9, 0.9)^T|\omega_1)p(\omega_1) = 0.25 \times p(\omega_1)$$

$$p((0.9, 0.9)^T|\omega_2)p(\omega_2) = \frac{2p(\omega_1)}{\sqrt{2\pi}} e^{-\frac{1}{2}((0.9,0.9)^T-(2,2)^T)^T((0.9,0.9)^T-(2,2)^T)} = 0.2379 \times p(\omega_1)$$

$$p((0.9, 0.9)^T|\omega_3)p(\omega_3) = \frac{p(\omega_1)}{\sqrt{2\pi}} e^{-\frac{1}{2}((0.9,0.9)^T-(2,0)^T)^T((0.9,0.9)^T-(2,0)^T)} = 0.0726 \times p(\omega_1)$$

Επειδή η ποσότητα $p(\omega_1)$ είναι θετικός αριθμός η μέγιστη αριθμητική τιμή εμφανίζεται για την πρώτη κατηγορία συνεπώς το πρότυπο $(0.9, 0.9)^T$ προέρχεται απο την πρώτη κατηγορία.

Αν τα πρότυπα της δεύτερης κατηγορίας εμφανίζονται με τριπλάσια συχνότητα από τα πρότυπα των άλλων κατηγοριών τότε αλλάζει η αριθμητική τιμή του κριτηρίου για την δεύτερη κατηγορία μονάχα και γίνεται:

$$p((0.9, 0.9)^T|\omega_2)p(\omega_2) = 0.35685 \times p(\omega_2)$$

Το πρότυπο $(0.9, 0.9)^T$ σε αυτή την περίπτωση ταξινομείται στην δεύτερη κατηγορία.

3.2.2 Το κριτήριο του ελάχιστου κόστους

Ας φανταστούμε για παράδειγμα την περίπτωση μιας αυτόματης ιατρικής διαγνωστικής μηχανής η οποία εκτιμά με επεξεργασία εξετάσεων την πιθανότητα ο εξεταζόμενος να έχει μία θανατηφόρα ασθένεια. Το σύστημα παίρνει τριών ειδών αποφάσεις, ο εξεταζόμενος "ασθενεί", "είναι φορέας", "είναι υγιής". Θα θέλαμε το σύστημα διάγνωσης να ελαχιστοποιήσει την πιθανότητα ο ασθενής να "είναι φορέας" ή να "ασθενεί" και το σύστημα να δώσει σαν αποτέλεσμα των εξετάσεων "είναι υγιής". Ολες οι άλλες περιπτώσεις σφαλμάτων κρίνονται λιγότερο σοβαρές, διότι λάθη διάγνωσης στα οποία το μηχάνημα δίνει αποτέλεσμα "είναι ασθενής" ή "είναι φορέας" ενώ ο εξεταζόμενος στην πραγματικότητα είναι υγιής θα έχουν σαν αποτέλεσμα την παραπομπή του ασθενή για λεπτομερέστερες εξετάσεις οι οποίες όμως θα αποδείξουν και την πραγματική κατάσταση του εξεταζόμενου.

Βλέπουμε λοιπόν ότι σε πολλές περιπτώσεις το κόστος του σφάλματος ταξινόμησης για κάποιες κατηγορίες αντικειμένων είναι διαφορετικό. Ας δούμε πως μπορούμε να εισάγουμε την έννοια του κόστους μιάς απόφασης σε στοχαστικά συστήματα για την περίπτωση ταξινόμησης πρότυπου σε δύο κατηγορίες αντικειμένων. Με την επαγωγική μέθοδο μπορούμε εύκολα να γενικεύσουμε τα αποτελέσματα σε πρόβλημα ταξινόμησης σε N κατηγορίες αντικειμένων.

Εστω λοιπόν ότι το σύστημα αναγνωρίζει δύο κατηγορίες προτύπων, τις ω_1 και ω_2.

Ορίζουμε τέσσερα μεγέθη που ονομάζονται κόστος λήψης απόφασης:

c_{11}, είναι το κόστος που προκύπτει όταν το σύστημα αποφασίσει ότι $\mathbf{x} \in \omega_1$ ενώ στην πραγματικότητα $\mathbf{x} \in \omega_1$

c_{12}, είναι το κόστος που προκύπτει όταν το σύστημα αποφασίσει ότι $\mathbf{x} \in \omega_2$ ενώ στην πραγματικότητα $\mathbf{x} \in \omega_1$

c_{21}, είναι το κόστος που προκύπτει όταν το σύστημα αποφασίσει ότι $\mathbf{x} \in \omega_1$ ενώ στην πραγματικότητα $\mathbf{x} \in \omega_2$

c_{22}, είναι το κόστος που προκύπτει όταν το σύστημα αποφασίσει ότι $\mathbf{x} \in \omega_2$ ενώ στην πραγματικότητα $\mathbf{x} \in \omega_2$

Οπου $c_{11}, c_{12}, c_{21}, c_{22}$ είναι μη αρνητικοί πραγματικοί αριθμοί.

Ασφαλώς το κόστος της λανθασμένης απόφασης είναι σημαντικότερο (μεγαλύτερο αριθμητικά) από το κόστος της ορθής απόφασης (στις περισσότερες πρακτικές εφαρμογές το κόστος τίθεται στο μηδέν $c_{11} = c_{22} = 0$):

$$c_{12} > c_{11} \qquad \textit{και} \qquad c_{21} > c_{22} \tag{3.4}$$

Αν Ω_1 και Ω_2 είναι οι δύο περιοχές του πεδίου ορισμού του παραμετρικού διανύσματος των προτύπων για τις οποίες αποφασίζουμε ότι, αν $\mathbf{x} \in \Omega_1 \Rightarrow \mathbf{x} \in \omega_1$ και αν $\mathbf{x} \in \Omega_2 \Rightarrow \mathbf{x} \in \omega_2$, τότε η βέλτιστη επιλογή των δύο περιοχών μπορεί να υπολογιστεί με κριτήριο την ελαχιστοποίηση του *αναμενόμενου κόστους*:

$$C = E[\textit{κόστος}] =$$

$$\int_{\Omega_1} c_{11} p(\omega_1|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + \int_{\Omega_1} c_{21} p(\omega_2|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + \int_{\Omega_2} c_{12} p(\omega_1|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} + \int_{\Omega_2} c_{22} p(\omega_2|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} =$$

$$\int_{\Omega_1} (c_{11} p(\omega_1) p(\mathbf{x}|\omega_1) + c_{21} p(\omega_2) p(\mathbf{x}|\omega_2)) d\mathbf{x} + \int_{\Omega_2} (c_{12} p(\omega_1) p(\mathbf{x}|\omega_1) + c_{22} p(\omega_2) p(\mathbf{x}|\omega_2)) d\mathbf{x}$$

Εφόσον οι περιοχές Ω_1, Ω_2 δεν υπερκαλύπτονται και περιέχουν όλο τον χώρο των μετρήσεων ισχύει:

$$\int_{\Omega_1} p(\mathbf{x}|\omega_i) d\mathbf{x} + \int_{\Omega_2} p(\mathbf{x}|\omega_i) d\mathbf{x} = 1, \qquad i = 1, 2$$

Αντικαθιστώντας την παραπάνω εξίσωση στην συνάρτηση του αναμενόμενου κόστους έχουμε:

$$C = c_{12}p(\omega_1) + c_{22}p(\omega_2) +$$

$$\int_{\Omega_1} (-(c_{12} - c_{11})p(\omega_1)p(\mathbf{x}|\omega_1) + (c_{21} - c_{22})p(\omega_2)p(\mathbf{x}|\omega_2))d\mathbf{x}$$

Η λύση του προβλήματος δίνεται με την επιλογή της περιοχής Ω_1 για την οποία η ολοκληρωτική παράσταση είναι ελάχιστη. Η ολοκληρωτική παράσταση αποτελείται απο έναν θετικό και έναν αρνητικό όρο διότι οι πολλαπλασιαστικοί όροι είναι όλοι θετικοί (βλέπε εξίσωση 3.4).

Αν ο αρνητικός όρος είναι μεγαλύτερος του θετικού για όλα τα πρότυπα που ανήκουν στην περιοχή Ω_1 τότε η ολοκληρωτική παράσταση έχει αρνητικές τιμές πάντα και το συνολικό ολοκλήρωμα δίνει την ελάχιστη αρνητική της τιμή.

Η συνθήκη που πρέπει να ισχύει έτσι ώστε να αποφασίζουμε με το ελάχιστο κόστος ότι το άγνωστο πρότυπο $\mathbf{x}$ ανήκει στην περιοχή Ω_1 είναι:

$$(c_{12} - c_{11})p(\omega_1)p(\mathbf{x}|\omega_1) > (c_{21} - c_{22})p(\omega_2)p(\mathbf{x}|\omega_2) \quad \Rightarrow \quad \mathbf{x} \in \Omega_1 \tag{3.5}$$

Η εξίσωση 3.5 αποτελεί το κριτήριο ταξινόμησης του ελάχιστου κόστους.

Παράδειγμα 27 *Ασφαλιστική εταιρεία ήρθε να εκτιμήσει το σύστημα ασφαλείας που φτιάξαμε. Εξηγήσαμε στον υπεύθυνο της ασφαλιστικής εταιρείας ότι το σύστημα επιβεβαιώνει την ταυτότητα του ανθρώπου που ζητά πρόσβαση συγκρίνοντας την προφορά μιας λέξης από την ομιλία του με μία αντίστοιχη πρωτότυπη λέξη που έχει αποθηκευτεί και είναι μοναδική για κάθε ομιλητή. Εχουμε υπολογίσει ότι το μέτρο απόστασης που προκύπτει από την σύγκριση του πρότυπου της ομιλίας με το πρωτότυπο του ελεγχόμενου ομιλητή είναι μια στοχαστική μεταβλητή με πυκνότητα πιθανότητας την κανονική κατανομή και στην περίπτωση κατά την οποία ο ομιλητής δίνει την αληθινή του ταυτότητα και όταν δώσει ψεύτικη ταυτότητα.*

Η επιβεβαίωση της ταυτότητας του ομιλητή πραγματοποιείται με κατώφλι απόφασης. Αν η υπολογιζόμενη απόσταση του πρότυπου ξεπερνά το κατώφλι τότε απαγορεύεται η πρόσβαση του ομιλητή στο σύστημα ασφαλείας. Το κατώφλι υπολογίστηκε με έτσι ώστε να ελαχιστοποιείται η πιθανότητα λανθασμένης απόφασης.

Ο ασφαλιστής μας άκουσε προσεκτικά και μας ρώτησε στο τέλος: "Αν και δεν κατάλαβα τις τεχνικές λεπτομέρειες, το σύστημα που φτιάξατε δεν είναι κατάλληλο διότι για μας είναι πολύ σημαντικό να ελαχιστοποιήσουμε την πιθανότητα πρόσβασης στο σύστημα ασφαλείας ενός ομιλητή ο οποίος επικαλείται ξένη ταυτότητα. Για να μπορέσω να ασφαλίσω τα περιουσιακά στοιχεία που καλύπτονται από το παρόν σύστημα ασφαλείας θα πρέπει να μου εξασφαλίσετε ότι το σύστημα θα λειτουργεί 100 φορές πιό αυστηρά για να επιτρέψει σε κάποιον που επικαλείται ψεύτικη ταυτότητα να εισέλθει στο σύστημα από την περίπτωση που λανθασμένα απορρίπτει κάποιον ομιλητή. Στο κάτω-κάτω της γραφής αν είναι βραχνιασμένος ας ξαναπροσπαθήσει".

Πώς μπορείτε να ικανοποιήσετε αυτή την απαίτηση του ασφαλιστή;

Για να ικανοποιήσουμε την απαίτηση του ασφαλιστή είναι προφανές ότι θα πρέπει να τροποποιήσουμε το κριτήριο με το οποίο υπολογίζουμε το κατώφλι απόφασης. Αν ορίσουμε ότι το κόστος της λανθασμένης αποδοχής (c_1) είναι a φορές μεγαλύτερο απο το κόστος της λανθασμένης απόρριψης του ομιλητή (c_2), και επίσης αν ορίσουμε σαν ω_1 την κατηγορία γεγονότων η "ταυτότητα του ομιλητή είναι σωστή" και ω_2 την κατηγορία γεγονότων η "ταυτότητα του ομιλητή είναι λανθασμένη", τότε το κριτήριο λήψης απόφασης με το ελάχιστο κόστος γίνεται:

$$ac_2p(\omega_1)p(d(\mathbf{x},\mathbf{r})|\omega_1) > c_2p(\omega_2)p(d(\mathbf{x},\mathbf{r})|\omega_2) \quad \Rightarrow \quad \text{"Η ταυτότητα του ομιλητή είναι σωστή"}$$

Το βέλτιστο κατώφλι προκύπτει από την λύση της εξίσωσης:

$$ap(\omega_1)p(d(\mathbf{x},\mathbf{r})|\omega_1) = p(\omega_2)p(d(\mathbf{x},\mathbf{r})|\omega_2) \Leftrightarrow$$

$$ap(\omega_1)\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(Tr-m_1)^2}{2\sigma_1{}^2}} = p(\omega_2)\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(Tr-m_2)^2}{2\sigma_2{}^2}} \Leftrightarrow$$

$$Tr = \frac{m_1{\sigma_2}^2 - m_2{\sigma_1}^2 - \sigma_1\sigma_2\sqrt{2({\sigma_2}^2 - {\sigma_1}^2)log(\frac{ap(\omega_1)\sigma_2}{p(\omega_2)\sigma_1}) + (m_1 - m_2)^2}}{{\sigma_2}^2 - {\sigma_1}^2}$$

3.3 Σχέση στοχαστικών και δομικών κριτηρίων ταξινόμησης

Η εύρεση των σχέσεων που διέπουν τις διαφορετικές μεθόδους ταξινόμησης προτύπων βοηθά σημαντικά στην προσαρμογή λύσεων που έχουν βρεθεί ή προτείνονται γιά την μία κατηγορία μεθόδων να μεταφέρονται στην άλλη κατηγορία.

3.3.1 Συναρτήσεις διάκρισης και στοχαστικά κριτήρια ταξινόμησης

Είδαμε στην προηγούμενη παράγραφο δύο από τα σημαντικότερα στοχαστικά κριτήρια ταξινόμησης προτύπων, το κριτήριο μέγιστης πιθανότητας:

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad p(\mathbf{x}|\omega_j)p(\omega_j) > p(\mathbf{x}|\omega_i)p(\omega_i) \quad \forall i \neq j$$

και το κριτήριο του ελάχιστου κόστους:

$$(c_{12} - c_{11})p(\omega_1)p(\mathbf{x}|\omega_1) > (c_{21} - c_{22})p(\omega_2)p(\mathbf{x}|\omega_2) \quad \Rightarrow \quad \mathbf{x} \in \Omega_1$$

Και στις δύο περιπτώσεις υπολογίζουμε μία αριθμητική ποσότητα για κάθε κατηγορία και στην συνέχεια ταξινομούμε το άγνωστο πρότυπο στην κατηγορία που έχει την μεγαλύτερη αριθμητική τιμή, ότι δηλαδή κάναμε και για τα δομικά συστήματα ταξινόμησης όταν αυτά χρησιμοποιούν συναρτήσεις διάκρισης.

Συνεπώς αν ορίσω σαν συνάρτηση διάκρισης την

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

τότε το στοχαστικό κριτήριο ταξινόμησης της μέγιστης πιθανότητας γίνεται ισοδύναμο με την διαδικασία ταξινόμησης ενός δομικού συστήματος.

Με τον ίδιο τρόπο αν ορίσω σαν συνάρτηση διάκρισης:

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

τότε το στοχαστικό κριτήριο ταξινόμησης του ελάχιστου κόστους είναι ισοδύναμο με την κριτήριο ταξινόμησης ενός δομικού συστήματος με συναρτήσεις διάκρισης.

Το πιό σημαντικό πλεονέκτημα που προκύπτει από την επισήμανση αυτής της αντιστοιχίας είναι η δυνατότητα που έχουμε να υλοποιήσουμε τις μεθόδους εκπαίδευσης δομικών συστημάτων που χρησιμοποιούν συναρτήσεις διάκρισης στις αντίστοιχες στοχαστικές μεθόδους εκπαίδευσης.

Ο αλγόριθμος perceptron για τις συναρτήσεις διάκρισης, ο αλγόριθμος ελαχιστοποίησης του σφάλματος εκτίμησης, η προσέγγιση της συνάρτησης διάκρισης με συναρτήσεις δυναμικού ή με γενικευμένες ακτινικές συναρτήσεις διάκρισης μπορούν να επεκταθούν και να χρησιμοποιηθούν σε στοχαστικά συστήματα ταξινόμησης προτύπων.

Προσοχή θα πρέπει να δοθεί στις προσεγγιστικές μεθόδους υπολογισμού της συνάρτησης διάκρισης για τα στοχαστικά συστήματα διότι η τιμή της συνάρτησης διάκρισης πρέπει να βρίσκεται πάντα περιορισμένη σε κλειστό διάστημα τιμών ($[0,1]$ για τις συναρτήσεις διάκρισης που προκύπτουν από το κριτήριο μέγιστης πιθανότητας και στο διάστημα $[0, c_i]$ για τις συναρτήσεις διάκρισης που προκύπτουν από το κριτήριο του ελάχιστου κόστους).

Σε πολλές πρακτικές περιπτώσεις οι περιορισμοί αυτοί δυσκολεύουν σε πολλές περιπτώσεις την διαδικασία εκτίμησης των παραμέτρων του στοχαστικού μοντέλου από παραδείγματα.

3.3.2 Σχέση στοχαστικών και δομικών συστημάτων σύγκρισης προτύπων

Μια από τις σημαντικότερες σχέσεις ανάμεσα σε στοχαστικές και δομικές μεθόδους ταξινόμησης είναι εκείνη που συνδέει το κριτήριο ταξινόμησης της μικρότερης σταθμισμένης Ευκλείδιας απόστασης και το στοχαστικό κριτήριο ταξινόμησης στην κατηγορία με την μεγαλύτερη πιθανότητα, όταν η πυκνότητα πιθανότητας των προτύπων για κάθε κατηγορία ακολουθεί την κανονική κατανομή.

Είναι γνωστό ότι η πυκνότητα πιθανότητας των προτύπων γιά πολλές κατηγορίες φυσικών αντικειμένων μπορεί να προσεγγιστεί με σημαντική ακρίβεια με την πυκνότητα πιθανότητας της κανονικής κατανομής:

$$p(\mathbf{x},\mathbf{y},\mathbf{s}) = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{y})^T\mathbf{s}^{-1}(\mathbf{x}-\mathbf{y})} \tag{3.6}$$

Στην περίπτωση κατά την οποία υποθέτουμε ότι κάθε μία από τις παραμέτρους του πρότυπου είναι στοχαστικά ασυσχέτιστη από τις υπόλοιπες παραμέτρους, τότε η πυκνότητα πιθανότητας για την κανονική κατανομή απλοποιείται και γίνεται:

$$p(\mathbf{x},\mathbf{y},\mathbf{s}) = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-y_i)^2}{s_i}} \tag{3.7}$$

Οπου y_i και s_i είναι η μέση τιμή και η διασπορά της i παραμέτρου του διανύσματος των προτύπων.

Ας υποθέσουμε λοιπόν ότι έχουμε δύο κατηγορίες (ω_1, ω_2) τα πρότυπα των οποίων ακολουθούν την κανονική κατανομή (εξίσωση 3.7).

Η ταξινόμηση κάθε πρότυπου $\mathbf{x}$ μπορεί να πραγματοποιηθεί με έναν στοχαστικά βέλτιστο τρόπο, μετρώντας την πυκνότητα πιθανότητας κάθε μίας των κατηγοριών στην θέση $\mathbf{x}$ και ταξινομώντας το άγνωστο πρότυπο στην κατηγορία εκείνη στην οποία η πυκνότητα πιθανότητας μεγιστοποιείται.

Για σύστημα ταξινόμησης δύο κατηγοριών το πρότυπο $\mathbf{x}$ ταξινομείται στην κατηγορία ω_1 όταν ισχύει:

$$p(\mathbf{x},\mathbf{y}_1,\mathbf{s}_1) > p(\mathbf{x},\mathbf{y}_2,\mathbf{s}_2) \rightarrow x \in \omega_1 \Leftrightarrow$$

$$\frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}_1|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-y_i^{(1)})^2}{s_i^{(1)}}} > \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}_2|}} e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-y_i^{(2)})^2}{s_i^{(2)}}} \rightarrow x \in \omega_1$$

Αν $|\mathbf{s}_1| = |\mathbf{s}_2|$ τότε η προηγούμενη ανισότητα απλοποιείται σε:

$$e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-y_i^{(1)})^2}{s_i}} > e^{-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-y_i^{(2)})^2}{s_i}} \rightarrow x \in \omega_1 \Leftrightarrow$$

$$\sum_{i=1}^{N}\frac{(x_i-y_i^{(1)})^2}{s_i} < \sum_{i=1}^{N}\frac{(x_i-y_i^{(2)})^2}{s_i} \rightarrow x \in \omega_1 \Leftrightarrow$$

$$d_w(\mathbf{x},\mathbf{y}_1,\mathbf{m}) < d_w(\mathbf{x},\mathbf{y}_2,\mathbf{m}) \rightarrow x \in \omega_1$$

Οι παραπάνω συλογισμοί αποδεικνύουν ότι το στοχαστικό κριτήριο ταξινόμησης με την μεγαλύτερη πυκνότητα πιθανότητας είναι ισοδύναμο με το κριτήριο ταξινόμησης της μικρότερης σταθμισμένης Ευκλείδιας απόστασης όταν για όλες τις κατηγορίες ισχύει: $|\mathbf{s}_i| = |\mathbf{s}_j|, \forall i, j$.

Παράδειγμα 28 *Στην περίπτωση κατά την οποία ισχύει $|\mathbf{s}_i| \neq |\mathbf{s}_j|$ ποιά είναι η σχέση που συνδέει το στοχαστικό κριτήριο μέγιστης πιθανότητας και το κριτήριο ταυξινόμησης της μικρότερης απόστασης όταν η πυκνότητα πιθανότητας των προτύπων των κατηγοριών ακολουθεί κανονική κατανομή;*

3.3.3 Σχέση κριτηρίου ταξινόμησης των k-πλησιέστερων προτύπων και του ταξινόμησης με την μεγαλύτερη πιθανότητα

Αν επιχειρήσουμε έναν στατιστικό υπολογισμό της συνάρτησης πυκνότητας πιθανότητας των προτύπων των κατηγοριών τότε μπορούμε να δούμε ότι:

$$P_j(|\mathbf{x}-\mathbf{y}|<\varepsilon) = \int_{|\mathbf{x}-\mathbf{y}|<\varepsilon} p_j(\mathbf{x})d\mathbf{x} \approx \frac{N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon)}{N_j}$$

όπου $N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon)$ είναι ο αριθμός των παραδειγμάτων της κατηγορίας ω_j τα οποία ικανοποιούν την σχέση $|\mathbf{x}_i - \mathbf{y}| < \varepsilon$, και ε μικρός θετικός πραγματικός αριθμός.

Αν θεωρήσουμε ότι στη περιοχή που ορίσαμε από τα σημεία $\mathbf{y}$ που ικανοποιούν την σχέση $|\mathbf{x}-\mathbf{y}| < \varepsilon$ βρίσκονται $N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon)$ παραδείγματα γιά κάθε μία από τις M κατηγορίες.

Το κριτήριο μέγιστης πιθανότητα σε αυτή την περίπτωση γίνεται:

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad p(\mathbf{x}|\omega_j)p(\omega_j) > p(\mathbf{x}|\omega_i)p(\omega_i) \quad \forall i \neq j \Leftrightarrow$$

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad \frac{N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon)}{N_j}p(\omega_j) > \frac{N_i(|\mathbf{x}-\mathbf{y}|<\varepsilon)}{N_i}p(\omega_i) \quad \forall i \neq j \Leftrightarrow$$

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad \frac{N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon)}{N_j}\frac{N_j}{\sum_{m=1}^{M} N_m} > \frac{N_i(|\mathbf{x}-\mathbf{y}|<\varepsilon)}{N_i}\frac{N_i}{\sum_{m=1}^{M} N_m} \quad \forall i \neq j \Leftrightarrow$$

$$\mathbf{x} \in \omega_j \quad \alpha\nu\nu \quad N_j(|\mathbf{x}-\mathbf{y}|<\varepsilon) > N_i(|\mathbf{x}-\mathbf{y}|<\varepsilon) \quad \forall i \neq j$$

Η τελευταία σχέση μας δείχνει ότι το κριτήριο ταξινόμησης της μεγαλύτερης πιθανότητας είναι ισοδύναμο με το κριτήριο ταξινόμησης στην κατηγορία εκείνη της οποίας τα περισσότερα παραδείγματα βρίσκονται στην γειτονιά του άγνωστου πρότυπου, πρόκειται δηλαδή για το κριτήριο ταξινόμησης των k-πλησιέστερων προτύπων που περιγράφηκε στο προηγούμενο κεφάλαιο.

3.4 Η συνάρτηση πυκνότητας πιθανότητας κατηγοριών

Απαραίτητη προυπόθεση για την κατασκευή στοχαστικών συστημάτων ταξινόμησης προτύπων είναι η εύρεση της συνάρτησης πυκνότητας πιθανότητας για κάθε κατηγορία προτύπων η οποία συνήθως υπολογίζεται από τα παραδείγματα και εκφράζεται με την ακόλουθη συνάρτηση: $p(\mathbf{x}|\omega_i)$.

Ο περιορισμός που πρέπει να ικανοποιεί μία συνάρτηση γιά να εκφράζει την πυκνότητα πιθανότητας κατηγορίας προτύπων είναι ο ακόλουθος:

$$\int_{\mathbf{x}\in\omega_i} p(\mathbf{x}|\omega_i)d\mathbf{x} = 1 \tag{3.8}$$

Μια σημαντική οικογένεια συναρτήσεων που εκφράζει πυκνότητες πιθανότητας δίνεται από την ακόλουθη αλγεβρική έκφραση:

$$p(\mathbf{x}|\omega_i) = C_n\sqrt{|\mathbf{w}_i|}f((\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}_i(\mathbf{x}-\mathbf{m}_i)) \tag{3.9}$$

όπου η σταθερά C_n τίθεται για να ικανοποιήσει τον περιορισμό των συναρτήσεων πυκνότητας πιθανότητας.

Ευκολα αποδεικνύεται ότι η πυκνότητα πιθανότητας της κανονικής κατανομής ανήκει σε αυτή την κατηγορία κατανομών.

Εκτός από την κανονική κατανομή υπάρχουν και οι ακόλουθες πυκνότητες πιθανότητας που χρησιμοποιούνται λιγότερο συχνά στις εφαρμογές.

1. Κατανομή Pearson τύπου-II (Pearson type-II):

$$p(\mathbf{x}|\omega_i) = \begin{cases} \frac{\Gamma(\frac{1}{2}n+K+1)}{\sqrt{\pi^n}\Gamma(k+1)}\sqrt{|\mathbf{w}_i|}(1-(\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}_i(\mathbf{x}-\mathbf{m}_i))^{-k}, & \mathbf{x}\in S \\ 0, & \mathbf{x}\notin S \end{cases} \tag{3.10}$$

   Η περιοχή S ορίσεται από την σχέση:

$$S = |1-(\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}_i(\mathbf{x}-\mathbf{m}_i))| < 1.$$

   Για $k = 0$ η κατανομή Pearson τύπου-II εκφυλίζεται σε ομοιόμορφη, και για $k = +\infty$ η κατανομή Pearson τύπου-II μετατρέπεται σε κανονική κατανομή.

2. Κατανομή Pearson τύπου-VII (Pearson type-VII):

$$p(\mathbf{x}|\omega_i) = \frac{\Gamma(K)}{\sqrt{\pi^n}\Gamma(k-\frac{1}{2}n)}\sqrt{|\mathbf{w}_i|}(1+(\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}_i(\mathbf{x}-\mathbf{m}_i))^{-k}, \quad k > \frac{1}{2}n \tag{3.11}$$

   Για $k = +\infty$ η κατανομή Pearson τύπου-VII μετατρέπεται σε κανονική κατανομή.

Παράδειγμα 29 *Εστω στοχαστικό σύστημα ταξινόμησης προτύπων δύο κατηγοριών που ακολουθούν την κατανομή Pearson τύπου-VII. Υπολόγισε το κατώφλι απόφασης με το οποίο κάθε πρότυπο ταξινομείται στην κατηγορία με την μεγαλύτερη πιθανότητα.*

Η πυκνότητα πιθανότητας της πρώτης κατηγορίας δίνεται από την σχέση:

$$p(\mathbf{x}|\omega_1) = \frac{\Gamma(K)}{\sqrt{\pi^n}\Gamma(k-\frac{1}{2}n)}\sqrt{|\mathbf{w}_1|}(1+(\mathbf{x}-\mathbf{m}_1)^T\mathbf{w}_1(\mathbf{x}-\mathbf{m}_1))^{-k}$$

Η πυκνότητα πιθανότητας της δεύτερης κατηγορίας δίνεται από την σχέση:

$$p(\mathbf{x}|\omega_2) = \frac{\Gamma(K)}{\sqrt{\pi^n}\Gamma(k-\frac{1}{2}n)}\sqrt{|\mathbf{w}_2|}(1+(\mathbf{x}-\mathbf{m}_2)^T\mathbf{w}_2(\mathbf{x}-\mathbf{m}_2))^{-k}$$

Το ζητούμενο κατώφλι απόφασης υπολογίζεται από την λύση της εξίσωσης:

$$p(\mathbf{x}|\omega_1)p(\omega_1) = p(\mathbf{x}|\omega_2)p(\omega_2)$$

Αντικαθιστώντας την αναλυτική έκφραση των πυκνοτήτων πιθανότητας έχουμε:

$$\sqrt{|\mathbf{w}_1|}(1+(\mathbf{x}-\mathbf{m}_1)^T\mathbf{w}_1(\mathbf{x}-\mathbf{m}_1))^{-k}p(\omega_1) = \sqrt{|\mathbf{w}_2|}(1+(\mathbf{x}-\mathbf{m}_2)^T\mathbf{w}_2(\mathbf{x}-\mathbf{m}_2))^{-k}p(\omega_2)$$

$$(|\mathbf{w}_2|p(\omega_2)^2)^{\frac{1}{2k}}(1+(\mathbf{x}-\mathbf{m}_1)^T\mathbf{w}_1(\mathbf{x}-\mathbf{m}_1)) = (|\mathbf{w}_1|p(\omega_1)^2)^{\frac{1}{2k}}(1+(\mathbf{x}-\mathbf{m}_2)^T\mathbf{w}_2(\mathbf{x}-\mathbf{m}_2))$$

Θέτοντας $C_1 = (|\mathbf{w}_1|p(\omega_1)^2)^{\frac{1}{2k}}$ και $C_2 = (|\mathbf{w}_2|p(\omega_2)^2)^{\frac{1}{2k}}$, έχουμε να επιλύσουμε την ακόλουθη δευτεροβάθμια εξίσωση για να βρούμε το κατώφλι απόφασης:

$$C_1(\mathbf{x}-\mathbf{m}_1)^T\mathbf{w}_1(\mathbf{x}-\mathbf{m}_1) - C_2(\mathbf{x}-\mathbf{m}_2)^T\mathbf{w}_2(\mathbf{x}-\mathbf{m}_2) + C_1 - C_2 = 0$$

3.5 Υπολογισμός της πυκνότητας πιθανότητας

Στην πλειονότητα των πρακτικών εφαρμογών δεν γνωρίζουμε τον αναλυτικό τύπο της συνάρτησης πυκνότητας πιθανότητας. Συνήθως αυτό που διαθέτουμε είναι ένας πεπερασμένος αριθμός παραδειγμάτων. Σε αυτές τις περιπτώσεις, που είναι και η πλειονότητα των πρακτικών εφαρμογών, ο υπολογισμός της συνάρτησης πυκνότητας πιθανότητας μπορεί να γίνει με διαφορετικές μεθόδους. Εμείς θα μελετήσουμε τις σημαντικότερες από αυτές.

3.5.1 Μεγιστοποίηση της εντροπίας

Είναι γνωστό ότι η εντροπία μας δίνει ένα ποσοτικό μέτρο της αταξίας ενός συστήματος. Η εντροπία της συνάρτησης πυκνότητας πιθανότητας μας δίνει την μέση τιμή της αβεβαιότητας ή ισοδύναμα της πληροφορίας που λαμβάνουμε από την μέτρηση της διανυσματικής παράστασης του πρότυπου $\mathbf{x}$:

$$E\nu\tau\rho o\pi\acute{\iota}\alpha = -\int_{\mathbf{x}} p(\mathbf{x})ln(p(\mathbf{x}))d\mathbf{x} \tag{3.12}$$

Μια μέθοδος υπολογισμού της συνάρτησης πυκνότητας πιθανότητας μπορεί να γίνει με την εύρεση της συνάρτησης εκείνης η οποία μεγιστοποιεί την εντροπία και ικανοποιεί ένα πλήθος περιορισμών που προκύπτουν από τα στατιστικά χαρακτηριστικά των παραδειγμάτων που διαθέτουμε. Η γενική περιγραφή του προβλήματος που θέλουμε να λύσουμε είναι η εξής: Εστω ότι έχουμε ένα σύνολο περιορισμών:

$$\int_{\mathbf{x}} p(\mathbf{x})d\mathbf{x} = 1 \tag{3.13}$$

και

$$\int_{\mathbf{x}} R_i(\mathbf{x})p(\mathbf{x})d\mathbf{x} = \rho_i, \qquad i = 1, I \tag{3.14}$$

Θέλουμε να βρούμε την συνάρτηση $p(\mathbf{x})$ η οποία ικανοποιεί τους παραπάνω περιορισμούς και μεγιστοποιεί την εντροπία της, δηλαδή την πληροφορία που λαμβάνουμε από τις μετρήσεις.

Οπως βλέπουμε θέλουμε να βρούμε τα ακρότατα μιας συνάρτησης (της εντροπίας) όταν αυτή ικανοποιεί ένα σύνολο περιορισμών. Μία γενική λύση του προβλήματος μπορεί να επιτευχθεί με την βοήθεια ενός θεωρήματος που απέδειξε ο Langrange.

Θεώρημα 3 *Εστω ότι θέλουμε να υπολογίσουμε τα ακρότατα της συνάρτησης $f(\mathbf{x})$ επάνω στην επιφάνεια που ορίζεται από ένα σύνολο περιορισμών:*

$$q_1(\mathbf{x}) = q_2(\mathbf{x}) = ... = q_q(\mathbf{x}) = 0$$

Σε κάθε ακρότατο της συνάρτησης ισχύει ότι:

$$\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) - \sum_{i=1}^{Q} \lambda_i \frac{\partial}{\partial \mathbf{x}} g_i(\mathbf{x}) = 0 \tag{3.15}$$

Οι συντελεστές λ_i ονομάζονται πολλαπλασιαστές του Langrange. Τα σημεία των ακρότατων της συνάρτησης βρίσκονται με την λύση των εξισώσεων των περιορισμών και της εξίσωσης των πολλαπλασιαστών του Langrange.

Θεώρημα 4 *Αν είναι γνωστό ότι το μονοδιάστατο πρότυπο κατηγορίας περιορίζεται στο διάστημα $x \in [a, b]$, τότε η συνάρτηση πυκνότητας πιθανότητας που μεγιστοποιεί την εντροπία της είναι:*

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

Ο μοναδικός περιορισμός του προβλήματος είναι:

$$\int_{a \le x \le b} p(x)dx = 1$$

Η εξίσωση των πολλαπλασιαστών του Langrange γίνεται:

$$\frac{\partial}{\partial p(x)} \int_{a \le x \le b} p(x)ln(p(x))dx + \lambda_0 \frac{\partial}{\partial p(x)} \int_{a \le x \le b} p(x)dx = 0 \Rightarrow$$

$$\int_{a \le x \le b} \frac{\partial}{\partial p(x)} p(x)ln(p(x))dx + \int_{a \le x \le b} \lambda_0 \frac{\partial}{\partial p(x)} p(x)dx = 0 \Rightarrow \int_{a \le x \le b} (ln(p(x)) + \lambda_0 + 1)dx = 0$$

Η προφανής λύση είναι:

$$p(x) = e^{\lambda_0+1} = c$$

Αντικαθιστώντας την ομοιόμορφη πυκνότητα πιθανότητας στον περιορισμό που έχουμε:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$$

Θεώρημα 5 *Αν είναι γνωστό ότι το μονοδιάστατο πρότυπο κατηγορίας περιορίζεται στο διάστημα $x \in \mathbb{R}^+$, και είναι γνωστή από στατιστικές μετρήσεις η μέση τιμή της συνάρτησης πυκνότητας πιθανότητας (μ) τότε η συνάρτηση που μεγιστοποιεί την εντροπία της είναι:*

$$p(x) = \begin{cases} \frac{1}{\mu}e^{-\frac{x}{\mu}}, & x \in \mathbb{R}^+ \\ 0, & x \notin \mathbb{R}^+ \end{cases}$$

Θεώρημα 6 *Αν είναι γνωστή η μέση τιμή (μ) και η διασπορά (σ) μονοδιάστατης τυχαίας μεταβλητής x τότε η συνάρτηση πυκνότητας πιθανότητας που μεγιστοποιεί την εντροπία της είναι η κανονική κατανομή.*

3.5.2 Προσέγγιση της πυκνότητας πιθανότητας με ορθοκανονικές συναρτήσεις

Ας υποθέσουμε ότι μπορούμε να προσεγγίσουμε την πυκνότητα πιθανότητας των προτύπων κάποιας κατηγορίας προτύπων με ένα σύνολο Q συναρτήσεων:

$$\hat{p}(\mathbf{x}) = \sum_{i=1}^{Q} c_i \varphi_i(\mathbf{x}) \tag{3.16}$$

Θα θέλαμε με την βοήθεια κάποιου κριτηρίου βελτιστοποίησης να βρούμε τους συντελεστές $c_i, i = 1, Q$ έτσι ώστε η προσέγγιση να δίνει ικανοποιητική αξιοπιστία. Ενα καλό κριτήριο είναι η ελαχιστοποίηση του σταθμισμένου σφάλματος της προσέγγισης:

$$Er = \int_{\mathbf{x}} w(\mathbf{x})(p(x) - \hat{p}(\mathbf{x}))^2 d\mathbf{x} \tag{3.17}$$

όπου $w(\mathbf{x})$ είναι μία συνάρτηση που δίνει διαφορετική βαρύτητα στην προσέγγιση για τις διαφορετικές περιοχές του πεδίου τιμών των προτύπων. Ισοδύναμα η σχέση μπορεί να γίνει:

$$Er = \int_{\mathbf{x}} w(\mathbf{x})(p(x) - \sum_{i=1}^{Q} c_i \varphi_i(\mathbf{x}))^2 d\mathbf{x} \tag{3.18}$$

Οι συντελεστές $c_i, i = 1, Q$ προκύπτουν από ένα σύστημα Q γραμμικών εξισώσεων:

$$\frac{\partial}{\partial c_i} Er = 0, \qquad i = 1, Q \tag{3.19}$$

Υπολογίζοντας τις μερικές παραγώγους έχουμε:

$$\sum_{i=1}^{Q} c_i \int_{\mathbf{x}} w(\mathbf{x})\varphi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mathbf{x} = \int_{\mathbf{x}} w(\mathbf{x})\varphi_j(\mathbf{x})p(x)d\mathbf{x} \tag{3.20}$$

Ο όρος των εξισώσεων που βρίσκεται στο δεξί τμήμα βλέπουμε ότι περιγράφει την μέση τιμή της παράστασης $w(\mathbf{x})\varphi_j(\mathbf{x})$ συνεπώς αυτός ο όρος μπορεί να προσεγγιστεί με στατιστικές μεθόδους. Απο τα M παραδείγματα που διαθέτουμε υπολογίζουμε τις αριθμητικές τιμές των παραστάσεων $w(\mathbf{x}_i)\varphi_j(\mathbf{x}_i), i = 1, M$ και στην συνέχεια υποθέτουμε ότι ο αριθμητικός τους μέσος αποτελεί μία καλή προσέγγιση του αντίστοιχου στοχαστικού μέσου:

$$\int_{\mathbf{x}} w(\mathbf{x})\varphi_j(\mathbf{x})p(x)d\mathbf{x} \approx \frac{1}{M}\sum_{i=1}^{M} w(\mathbf{x}_i)\varphi_j(\mathbf{x}_i) \tag{3.21}$$

Αντικαθιστώντας την παραπάνω εξίσωση έχουμε:

$$\sum_{i=1}^{Q} c_i \int_{\mathbf{x}} w(\mathbf{x})\varphi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mathbf{x} \approx \frac{1}{M}\sum_{i=1}^{M} w(\mathbf{x}_i)\varphi_j(\mathbf{x}_i) \tag{3.22}$$

Αν οι συναρτήσεις $\varphi_i(\mathbf{x})$ έχουν εκλεγεί έτσι ώστε να είναι ορθοκανονικές ως προς την συνάρτηση βαρύτητας $w(\mathbf{x})$ τότε ισχύει ότι:

$$\int_{\mathbf{x}} w(\mathbf{x})\varphi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mathbf{x} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{3.23}$$

Λαμβάνοντας υπόψη τα παραπάνω ο τελικός υπολογισμός των συντελεστών γίνεται από τις σχέσεις:

$$c_i \approx \frac{1}{M}\sum_{j=1}^{M} w(\mathbf{x}_j)\varphi_i(\mathbf{x}_j) \qquad \forall i = 1, M \tag{3.24}$$

Ομάδες ορθοκανονικών συναρτήσεων δίνονται στο παράρτημα των σημειώσεων.

3.5.3 Τα πλαίσια Parzen

Μπορούμε να εκτιμήσουμε με διαφορετική μέθοδο την συνάρτηση πυκνότητας πιθανότητας από παραδείγματα, υποθέτοντας ότι το πρότυπο κάθε παραδείγματος αποτελεί από μόνο του μία κατηγορία αντικειμένων. Με την μέθοδο των πλαισίων Parzen κάθε παράδειγμα διαθέτει μία πυκνότητα πιθανότητας η οποία ονομάζεται και συνάρτηση πυρήνα (kernel) και η οποία είναι ίδιου τύπου γιά όλα τα παραδείγματα εκπαίδευσης.

Η συνολική πυκνότητα πιθανότητας της κατηγορίας υπολογίζεται από την σχέση:

$$p(x|\omega_i) \approx \frac{1}{M}\sum_{i=1}^{M} K(x - x_i) \tag{3.25}$$

Η συνάρτηση πυρήνα πρέπει να ικανοποιεί έναν αριθμό περιορισμών:

$$\int_{-\infty}^{+\infty} K(x)dx = 1 \tag{3.26}$$

$$\int_{-\infty}^{+\infty} |K(x)|dx = +\infty \tag{3.27}$$

$$\int_{x\to+\infty} |xK(x)| = 0 \tag{3.28}$$

Τυπικές συναρτήσεις πυρήνα δίνονται στον πίνακα που ακολουθεί:

Πίνακας 3.1: Τυπικές συναρτήσεις πυρήνα

Ονομασία συνάρτησης	πυρήνας
Ομοιόμορφος	$K_o(x) = \begin{cases} \frac{1}{2h}, & \lvert\frac{x}{h}\rvert \le 1 \\ 0, & \lvert\frac{x}{h}\rvert > 1 \end{cases}$
Τριγωνικός	$K_t(x) = \begin{cases} \frac{1}{h}(1-\lvert\frac{x}{h}\rvert), & \lvert\frac{x}{h}\rvert \le 1 \\ 0, & \lvert\frac{x}{h}\rvert > 1 \end{cases}$
Κανονικός	$K_n(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}$
Εκθετικός	$K_e(x) = \frac{1}{2h} e^{-\lvert\frac{x}{h}\rvert}$
Αντίστροφος	$K_i(x) = \frac{1}{h\pi}\frac{1}{1+(\frac{x}{h})^2}$
Ημιτόνου	$K_s(x) = \frac{1}{2h\pi}(\frac{\sin(\frac{x}{2h})}{\frac{x}{2h}})^2$

Παρατηρούμε ότι οι συναρτήσεις του πυρήνα εξαρτώνται από την μεταβλητή h η οποία καθορίζει και την διασπορά της πυκνότητα πιθανότητας του πυρήνα. Η τιμή της μεταβλητής h τίθεται εμπειρικά αλλά η γενική οδηγία που μπορούμε να δώσουμε είναι ότι όσο αυξάνει ο αριθμός των διαθέσιμων παραδειγμάτων, η τιμή αυτής της μεταβλητής πρέπει να μειώνεται.

Παράδειγμα 30 *Υπολογείστε το ελάχιστο σφάλμα στοχαστικού συστήματος ταξινόμησης προτύπων όταν η πυκνότητα πιθανότητας των προτύπων των κατηγοριών προσεγγίζεται με πλαίσια Parzen χρησιμοποιώντας συναρτήσεις εκθετικού πυρήνα με* $h = 1$.

Διαθέτουμε οκτώ παραδείγματα:

$$\Omega = \{((1,1),\omega_1),((2,2),\omega_1),((3.5,2.1),\omega_1),$$

$$((1.1,2.8),\omega_2),((2,5),\omega_2),((0,1.3),\omega_2),((1.5,3),\omega_2),((0.0,2),\omega_2)\}.$$

Κριτήριο ταξινόμησης είναι η κατηγορία με την μεγαλύτερη πιθανότητα.

Απο το σύνολο των οκτώ παραδειγμάτων υπολογίζω τις πιθανότητες εμφάνισης των προτύπων των κατηγοριών:

$$p(\omega_1) = \frac{3}{8} = 0.375$$

και

$$p(\omega_2) = \frac{5}{8} = 0.625.$$

Η πυκνότητα πιθανότητας της πρώτης κατηγορίας με τα πλαίσια Parzen δίνεται από την σχέση:

$$p(\mathbf{x}|w_1) = \frac{1}{6}(e^{-|x-(1,1)|} + e^{-|x-(2,2)|} + e^{-|x-(3.5,2.1)|})$$

Η πυκνότητα πιθανότητας της δεύτερης κατηγορίας δίνεται από την σχέση:

$$p(\mathbf{x}|w_2) = \frac{1}{10}(e^{-|x-(1.1,2.8)|} + e^{-|x-(2,5)|} + e^{-|x-(0,1.3)|} + e^{-|x-(1.5,3)|} + e^{-|x-(0,2)|})$$

Στον πίνακα που ακολουθεί δίνονται οι αριθμητικές τιμές του κριτηρίου ταξινόμησης της μεγαλύτερης πιθανότητας που υπολογίζεται από την σχέση $p(\omega_i)p(\mathbf{x}|\omega_i)$ για τις δύο κατηγορίες και για όλα τα παραδείγματα.

Πίνακας 3.2: Υπολογισμός του κριτηρίου ταξινόμησης της μεγαλύτερης πιθανότητας με συναρτήσεις εκθετικού πυρήνα

Κατηγορία	(1,1), ω_1	(2,2), ω_1	(3.5,2.1), ω_1	(1.1,2.8), ω_2
ω_1	0.06459	0.07429	0.06829	0.020961
ω_2	0.03259	0.03656	0.00851	0.100984
	(2,5), ω_2	(0,1.3), ω_2	(1.5,3), ω_2	(0.0,2), ω_2
ω_1	0.003822	0.019628	0.020013	0.016555
ω_2	0.063170	0.089726	0.097430	0.096387

Από τα αριθμητικά αποτελέσματα του πίνακα βλέπουμε ότι όλα τα παραδείγματα ταξινομούνται σωστά. Συνεπώς το ελάχιστο σφάλμα του συστήματος ταξινόμησης που κατασκευάσαμε από τα παραδείγματα είναι 0%.

3.6 Υπολογισμός των παραμέτρων της πυκνότητας πιθανότητας

Μερικές φορές είμαστε σε θέση να γνωρίζουμε την γενική μορφή της συνάρτησης που περιγράφει την πυκνότητα πιθανότητας αλλά χρειαζόμαστε τα παραδείγματα για να υπολογίσουμε τις σταθερές παραμέτρους της.

Σε αυτές τις περιπτώσεις εκμεταλευόμαστε μερικές από τις ιδιότητες της συνάρτησης πυκνότητας πιθανότητας που έχει να κάνει με τα αντίστοιχα στατιστικά της χαρακτηριστικά, δηλαδή:

1. Μέση τιμή:

$$\mathbf{M}r = \int_{\mathbf{x}} \mathbf{x}p(\mathbf{x}|\omega_i)d\mathbf{x} \approx \mathbf{M} = \frac{1}{Q}\sum_{j=1}^{Q}\mathbf{x}_j \tag{3.29}$$

Βλέπουμε ότι υπάρχει μία διαφορά ανάμεσα στον στοχαστικό και τον στατιστικό υπολογισμό του διανύσματος της τυχαίας μεταβλητής του πρότυπου. Η διαφορά αυτή δημιουργεί κάποια προβλήματα στον υπολογισμό της διασποράς της τυχαίας μεταβλητής, διότι στην πράξη δεν γνωρίζουμε την τιμή της πραγματικής μέσης τιμής ($\mathbf{M}r$) αλλά μιας προσέγγισης της, της στατιστικής μέσης τιμής ($\mathbf{M}$).

Η στατιστική διασπορά υπολογίζεται συνήθως από την ακόλουθη σχέση:

$$\mathbf{\Sigma} = \frac{1}{Q}\sum_{j=1}^{Q}(\mathbf{x}_j - \mathbf{M})^T(\mathbf{x}_j - \mathbf{M}) \tag{3.30}$$

Η ακριβέστερη εκτίμηση της στατιστικής διασποράς δίνεται από την σχέση:

$$\Sigma r = \frac{1}{Q}\sum_{j=1}^{Q}(\mathbf{x}_j - \mathbf{M}r)^T(\mathbf{x}_j - \mathbf{M}r) \tag{3.31}$$

Σε αυτή την περίπτωση βλέπουμε όμως ότι δεν γνωρίζουμε την πραγματική μέση τιμή $\mathbf{M}r$. Ας δούμε τι μπορούμε να κάνουμε γιαυτό το πρόβλημα.

Η εξίσωση 3.30 μπορεί να γραφτεί ισοδύναμα σαν:

$$\boldsymbol{\Sigma} = \frac{1}{Q}\sum_{j=1}^{Q}((\mathbf{x}_j - \mathbf{M}r) - (\mathbf{M} - \mathbf{M}r))^T((\mathbf{x}_j - \mathbf{M}r) - (\mathbf{M} - \mathbf{M}r)) \Leftrightarrow$$

$$\boldsymbol{\Sigma} = \frac{1}{Q}\sum_{j=1}^{Q}(\mathbf{x}_j - \mathbf{M}r)^T(\mathbf{x}_j - \mathbf{M}r) - (\mathbf{M} - \mathbf{M}r)^T(\mathbf{M} - \mathbf{M}r) \Leftrightarrow$$

$$\boldsymbol{\Sigma} r = \boldsymbol{\Sigma} + (\mathbf{M} - \mathbf{M}r)^T(\mathbf{M} - \mathbf{M}r) = \boldsymbol{\Sigma} + \frac{1}{Q}\boldsymbol{\Sigma} r \tag{3.32}$$

Απο αυτή την εξίσωση βλέπουμε ότι η πραγματική τιμή της στατιστικής διασποράς είναι στην πραγματικότητα μεγαλύτερη από αυτή που συνήθως υπολογίζουμε με την κλασσική σχέση της διασποράς (εξίσωση 3.30). Λύνοντας την εξίσωση ως προς $\boldsymbol{\Sigma} r$ έχουμε:

$$\boldsymbol{\Sigma} r = \frac{Q}{Q-1}\boldsymbol{\Sigma} = \frac{1}{Q-1}\sum_{j=1}^{Q}(\mathbf{x}_j - \mathbf{M})^T(\mathbf{x}_j - \mathbf{M}) \tag{3.33}$$

Οταν ο αριθμός των παραδειγμάτων αυξάνει τότε και η πραγματική στατιστική διασπορά προσεγγίζει την στατιστική διασπορά η οποία υπολογίζεται με την κλασσική σχέση της εξίσωσης 3.30.

Από τα παραπάνω γίνεται φανερό ότι η δεύτερη εξίσωση που μας προσφέρει δυνατότητες υπολογισμού των σταθερών παραμέτρων συνάρτησης πυκνότητας πιθανότητας είναι η ακόλουθη:

2. Διασπορά:

$$\int_{\mathbf{x}}(\mathbf{x} - \mathbf{M}r)^2 p(\mathbf{x}|\omega_i)d\mathbf{x} \approx \boldsymbol{\Sigma} r = \frac{1}{Q-1}\sum_{j=1}^{Q}(\mathbf{x}_j - \mathbf{M})^T(\mathbf{x}_j - \mathbf{M}) \tag{3.34}$$

Παράδειγμα 31 *Υπολογίστε τις παραμέτρους της συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής όταν γνωρίζετε πέντε πρότυπα μιάς μονοδιάστασης τυχαίας μεταβλητής, τα:* $\{2, 4, 5, 8, 6\}$.

Η μέση τιμή των μετρήσεων είναι $M = \frac{2+4+15+8+6}{5} = 7$.

Η στατιστική διασπορά με τον κλασικό τρόπο υπολογισμού είναι:

$$\Sigma = \frac{(2-7)^2 + (4-7)^2 + (15-7)^2 + (8-7)^2 + (6-7)^2}{5} = 20.$$

Η πραγματική στατιστική διασπορά είναι: $\Sigma = \frac{(2-7)^2+(4-7)^2+(15-7)^2+(8-7)^2+(6-7)^2}{4} = 25.$

Η συνάρτηση πυκνότητας πιθανότητας για την κατηγορία των προτύπων δίνεται από την ακόλουθη σχέση:

$$p(x, \omega) = \frac{1}{5\sqrt{2\pi}}e^{-\frac{(x-7)^2}{50}} \tag{3.35}$$

3.7 Η επιλογή της συνάρτησης πυκνότητας πιθανότητας

Απο την παραπάνω ανάλυση προκύπτει εύλογα το ερώτημα: "Βεβαίως και έχουμε τον τρόπο να υπολογίσουμε από τα στατιστικά χαρακτηριστικά των προτύπων των παραδειγμάτων τις παραμέτρους της συνάρτησης πυκνότητας πιθανότητας, αλλά με ποιό τρόπο θα διαλέξω την πλέον κατάλληλη από αυτές;"

Αρχικά ας δούμε ποιές είναι οι πλέον συνήθεις συναρτήσεις πυκνότητας πιθανότητας. Στον πίνακα που ακολουθεί δίνουμε τις σημαντικότερες συναρτήσεις πυκνότητας πιθανότητας που χρησιμοποιούνται στις εφαρμογές.

Γνωρίζοντας ένα πλήθος συναρτήσεων καλούμαστε να λύσουμε το πρόβλημα της επιλογής της πλέον κατάλληλης από αυτές. Πρέπει να βρούμε ένα ποσοτικό μέγεθος με το οποίο να υπολογίσουμε κατά πόσο η συνάρτηση πυκνότητας πιθανότητας που χρησιμοποιούμε "ταιριάζει" με τα πραγματικά δεδομένα που δεν είναι τίποτα άλλο από τα πρότυπα των παραδειγμάτων.

Πίνακας 3.3: Συναρτήσεις πυκνότητας πιθανότητας και τα στατιστικά χαρακτηριστικά των

Ονομασία	πυκνότητα πιθανότητας	Μέση τιμή	Διασπορά
Arcsine, $a > 0$	$\frac{rect(\frac{x}{2a})}{\pi\sqrt{a^2-x^2}}$	$M = 0$	$\sigma^2 = \frac{a^2}{2}$
Beta,	$\frac{\Gamma(a+b)(u(x)-u(x-1))x^{a-1}(1-x)^{b-1}}{\Gamma(a)\Gamma(b)}$	$M = \frac{a}{a+b}, a > 0, b > 0$	$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)}$
Chi-Square	$\frac{x^{N/2-1}}{2^{N/2}\Gamma(N/2)}e^{-x/2}u(x)$	$M = N, N = 1, ...$	$\sigma^2 = 2N$
Erlang, $a > 0$	$\frac{a^N x^{N-1} e^{-ax}}{(N-1)!}u(x)$	$M = \frac{N}{a}, N = 1, ...$	$\sigma^2 = \frac{N}{a^2}$
Exponential	$ae^{-ax}u(x)$	$M = \frac{1}{a}, a > 0$	$\sigma^2 = \frac{1}{a^2}$
Gamma	$\frac{a^b x^{b-1} e^{-ax}}{\Gamma(b)}u(x)$	$M = \frac{b}{a}, a > 0, b > 0$	$\sigma^2 = \frac{b}{a^2}$
Gaussian	$\frac{e^{-(x-a)^2/b}}{\sqrt{\pi b}}$	$M = a, b > 0$	$\sigma^2 = \frac{b}{2}$
Laplace	$\frac{b}{2}e^{-b\|x-a\|}$	$M = a, b > 0$	$\sigma^2 = \frac{2}{b^2}$
Log-Normal	$\frac{u(x-b)e^{-(log(x-b)-a^2)/2\sigma^2}}{\sqrt{2\pi}(\chi-b)\sigma}$	$M = b + e^{a+\sigma^2/2}, \sigma > 0$	$\sigma^2 = (e^{\sigma^2} - 1)e^{2a+\sigma^2}$
Rayleigh	$\frac{2}{b}(x-a)e^{-(x-a)^2)/b}u(x-a)$	$M = a + \sqrt{\frac{\pi b}{4}}, b > 0$	$\sigma^2 = \frac{b(4-\pi)}{4}$
Uniform	$\frac{u(x-a)-u(x-b)}{b-a}$	$M = \frac{a+b}{2}, b > a$	$\sigma^2 = \frac{(b-a)^2}{12}$
Weibull	$abx^{b-1}e^{-ax^b}u(x)$	$M = \frac{\Gamma(1+b^{-1})}{\alpha^{1/b}}, a > 0, b > 0$	$\sigma^2 = \frac{\Gamma(1+2b^{-1})-\Gamma(1+b^{-1})^2}{\alpha^{2/b}}$

Η πλέον διαδεδομένη μέθοδος ελέγχου της ομοιότητας μιας συνεχούς συνάρτησης πυκνότητας πιθανότητας και των αντίστοιχων στατιστικών κατανομών που υπολογίζονται από τα πρότυπα παραδειγμάτων είναι ο έλεγχος-x^2 (x^2-test). Ο έλεγχος αυτός πραγματοποιείται εκτελώντας τα ακόλουθα βήματα:

1. *Ορισμός περιοχών ελέγχου.* Χωρίζουμε τον χώρο των προτύπων σε έναν αριθμό περιοχών έτσι ώστε κάθε περιοχή να περιέχει έναν ελάχιστο αριθμό προτύπων απο τα παραδείγματα. Εστω ότι το πεδίο ορισμού των προτύπων χωρίστηκε σε M περιοχές τις $S_1, S_2, ..., S_M$.

2. *Υπολογισμός των πιθανοτήτων των περιοχών ελέγχου.* Εστω ότι ελέγχουμε την συνεχή συνάρτηση πυκνότητα πιθανότητας $p(\mathbf{x})$. Για κάθε μία από αυτές τις περιοχές υπολογίζουμε την πιθανότητα το πρότυπο της κατηγορίας να ανήκει στην περιοχή. Η αριθμητική τιμή της πιθανότητας υπολογίζεται από την ακόλουθη σχέση:

$$Pt_i = \int_{\mathbf{x} \in S_i} p(\mathbf{x}) d\mathbf{x} \tag{3.36}$$

3. *Στατιστικός υπολογισμός των πιθανοτήτων των περιοχών ελέγχου.* Για κάθε μία από τις περιοχές εκτελούμε τον στατιστικό υπολογισμό της πιθανότητας το πρότυπο να ανήκει στην περιοχή που ελέγχουμε. Αν στην περιοχή S_i βρίσκονται t_i πρότυπα σε ένα σύνολο T παραδειγμάτων, τότε η στατιστική πιθανότητα υπολογίζεται από την σχέση:

$$Ps_i = \frac{t_i}{T} \tag{3.37}$$

4. *Υπολογισμός της απόκλισης.* Υπολογίζουμε την απόκλιση των πιθανοτήτων για όλες τις περιοχές:

$$Απόκλιση = \sum_{i=1}^{M} \frac{(Ps_i - Pt_i)^2}{Pt_i} \tag{3.38}$$

5. *Σύγκριση.* Για κάθε κατανομή που ελέγχουμε βρίσκουμε το αντίστοιχο ποσοτικό μέγεθος της απόκλισης κρατώντας σταθερές τις περιοχές ελέγχου. Η στοχαστική κατανομή που παρουσιάζει την μικρότερη απόσταση από την στατιστική κατανομή επιλέγεται σαν η πλέον αντιπροσωπευτική.

Ιδιαίτερη προσοχή πρέπει να δοθεί σε δύο σημαντικά γεγονότα που αποτελούν και τα αδύνατα σημεία του ελέγχου-x^2:

1. Οταν ο αριθμός των παραδειγμάτων που διαθέτουμε είναι μικρός τότε το τελικό μέγεθος της μετρούμενης απόκλισης δεν αποτελεί αξιόπιστη μέτρηση της διαφοράς της θεωρητικής από την στατιστική κατανομή. Ο μικρός αριθμός των παραδειγμάτων κάνει ανασφαλή την μέτρηση της στατιστικής πιθανότητας στις περιοχές που ορίζουμε.

2. Ο ορισμός των περιοχών ελέγχου παίζει πολύ σημαντικό ρόλο στον τελικό υπολογισμό της απόκλισης των κατανομών ιδιαίτερα όταν ο αριθμός των παραδειγμάτων δεν είναι μεγάλος ή όταν οι διαφορές αποκλίσεων των κατανομών που ελέγχονται είναι μικρές. Πολλές φορές μικρή διαφοροποίηση των περιοχών μπορεί να επιφέρει σημαντική μεταβολή στο μέτρο των διαφορών των κατανομών.

Παράδειγμα 32 *Εστω ότι έχουμε τα ακόλουθα παραδείγματα μονοδιάστατων προτύπων.*

$$\Omega = \{2.2, 3, 4, 4.4, 4.8, 5.5, 6.1, 6.6, 6.9, 7.3, 8, 10, 12\}$$

Βρείτε ποιά από τις ακόλουθες κατανομές πυκνότητας πιθανότητας μοιάζει περισσότερο με την στατιστική πυκνότητας πιθανότητας:

1. *Gaussian*
2. *Laplace*
3. *Uniform*

Επειδή δεν γνωρίζουμε τις σταθερές παραμέτρους των κατανομών τις υπολογίζουμε απο τα πρότυπα των παραδειγμάτων. Ο στατιστικός μέσος όρος είναι 6.21538, ενώ η στατιστική διασπορά βρέθηκε να είναι 7.49641. Οι πυκνότητες πιθανότητας που αντιστοιχούν σε αυτά τα στατιστικά χαρακτηριστικά είναι οι ακόλουθες:

1. Συνάρτηση Gauss

$$p_G(x) = 0.1457e^{-(x-6.21538)^2/14.9928}$$

2. Laplace

$$p_L(x) = 0.25826e^{-0.51652|x-6.21538|}$$

3. Uniform

$$p_U(x) = \frac{u(x-1.47310)-u(x-10.95766)}{9.48456}$$

Βάσει των βημάτων του ελέγχου-χ^2 εκτελούμε τις ακόλουθες εργασίες:

1. *Ορισμός περιοχών ελέγχου.* Χωρίζουμε τον χώρο των προτύπων σε τρεις περιοχές έτσι ώστε κάθε περιοχή να περιέχει τουλάχιστον τέσσερα πρότυπα. Το πεδίο ορισμού των προτύπων είναι το σύνολο των πραγματικών αριθμών και χωρίστηκε στις εξής κατηγορίες:

$$S_1 = (-\infty, 4.5] \quad S_2 = (4.5, 6.7] \quad S_3 = (6.7, +\infty)$$

2. *Υπολογισμός των πιθανοτήτων των περιοχών ελέγχου.* Για κάθε μία από τις ελεγχόμενες συναρτήσεις υπολογίζω την πιθανότητα εύρεσης πρότυπου στις περιοχές που πραγματοποιείται ο έλεγχος:

   1. Συνάρτηση Gauss

$$Pg_1 = \int_{-\infty}^{4.5} 0.1457e^{-(x-6.21538)^2/14.9928}dx = 0.265474$$

$$Pg_2 = \int_{4.5}^{6.7} 0.1457e^{-(x-6.21538)^2/14.9928}dx = 0.304743$$

$$Pg_3 = \int_{6.7}^{+\infty} 0.1457e^{-(x-6.21538)^2/14.9928}dx = 0.42973$$

   2. Laplace

$$Pl_1 = \int_{-\infty}^{4.5} 0.25826e^{-0.51652|x-6.21538|}dx = 0.206145$$

$$Pl_2 = \int_{4.5}^{6.7} 0.25826e^{-0.51652|x-6.21538|}dx = 0.404578$$

$$Pl_3 = \int_{6.7}^{+\infty} 0.25826e^{-0.51652|x-6.21538|}dx = 0.389277$$

   3. Uniform

$$Pu_1 = \int_{-\infty}^{4.5} \frac{u(x-1.47310)-u(x-10.95766)}{9.48456}dx = 0.31913$$

$$Pu_2 = \int_{4.5}^{6.7} \frac{u(x-1.47310)-u(x-10.95766)}{9.48456}dx = 0.23195$$

$$Pu_3 = \int_{6.7}^{+\infty} \frac{u(x-1.47310)-u(x-10.95766)}{9.48456}dx = 0.44890$$

3. *Στατιστικός υπολογισμός των πιθανοτήτων των περιοχών ελέγχου.* Η στατιστική πιθανότητα για κάθε μία από τις περιοχές είναι:

$$Ps_1 = \frac{4}{13} = 0.30769$$

$$Ps_2 = \frac{4}{13} = 0.30769$$

$$Ps_3 = \frac{5}{13} = 0.38461$$

4. *Υπολογισμός της απόκλισης.* Υπολογίζουμε την απόκλιση των πιθανοτήτων για όλες τις περιοχές:

$$\textit{ΑπόκλισηGaussian} = 0.0424$$

$$\textit{ΑπόκλισηLaplace} = 0.090322$$

$$\textit{ΑπόκλισηUniform} = 0.069562$$

5. *Σύγκριση.* Η κατανομή που προσεγγίζει περισσότερο την στατιστική κατανομή είναι η κανονική κατανομή διότι έχει την μικρότερη απόκλιση. Αμέσως καλύτερη είναι η ομοιόμορφη.

3.8 Στοχαστικός υπολογισμός πυκνότητας πιθανότητας

Σε πολλά προβλήματα οι μετρήσεις που παίρνουμε περιέχουν "θόρυβο", αποτελούν δηλαδή ποσέγγιση των πραγματικών τιμών εξαιτίας της πεπερασμένης ακρίβειας του μετρητικού οργάνου ή της παρουσίας θορύβου που παρεμβάλλεται.

Οι συναρτήσεις παλινδρόμησης (regression functions) σε αντίθεση με τις κοινές συναρτήσεις είναι στατιστικές συναρτήσεις. Χαρακτηριστικό γνώρισμα των συναρτήσεων παλινδρόμησης είναι το γεγονός ότι δίνουν κάθε φορά που ζητάμε την τιμή της συνάρτησης σε ένα σημείο διαφορετική αριθμητική τιμή. Οι μέθοδοι που θα παρουσιάσουμε είναι σε θέση να εκτιμήσουν ρίζες και ακρότατα μονοδιάστατων και πολυδιάστατων συναρτήσεων παλινδρόμησης. Οι μέθοδοι αυτοί χρησιμοποιούνται ευρύτατα σε προβλήματα αναγνώρισης προτύπων και κυρίως γιά την εκτίμηση της πυκνότητας πιθανότητας προτύπων.

3.8.1 Ο αλγόριθμος των Robbins-Monro

Εστω συνάρτηση $f(x)$ η οποία έχει μία μοναδική ρίζα $f(\hat{x}) = 0$. Υποθέτουμε ότι οι τιμές της συνάρτησης είναι αρνητικές στο διάστημα των μικρότερων του $\hat{x}$ τιμών και θετικές στο διάστημα των μαγαλύτερων του $\hat{x}$ τιμών. Η υπόθεση αυτή δεν βλάπτει την γενικότητα διότι αν συμβαίνει το αντίθετο τότε κάνουμε τους υπολογισμούς μας για την συνάρτηση $-f(x)$ η οποία ικανοποιεί τον περιορισμό που θέσαμε.

Υποθέτουμε ότι το μετρητικό όργανο που έχουμε για να μας μετρά τις τιμές της συνάρτησης για τις διάφορες τιμές του x παρενοχλείται από θόρυβο ο οποίος αλλοιώνει τις μετρήσεις. Γιαυτό τον λόγο δεν μπορούμε να υπολογίσουμε απευθείας την ρίζα της συνάρτησης χρησιμοποιώντας καποια γνωστή αριθμητική μέθοδο. Ετσι αντί της πραγματικής τιμής $f(x)$ λαμβάνουμε την τιμή της συνάρτησης παλινδρόμησης, την $g(x)$. Η διασπορά του σφάλματος εκτίμησης της τιμής της συνάρτησης δίνεται από την μέση τιμή της ποσότητας $|f(x) - g(x)|$.

Κάνουμε δύο υποθέσεις για τα χαρακτηριστικά του θορύβου:

1. *Ο θόρυβος είναι τυχαίος.* Αν ο θόρυβος που αλλοιώνει τις μετρήσεις είναι στοχαστικά ανεξάρτητος από το σήμα και έχει μηδενική μέση τιμή τότε ισχύει:

$$E\{g(x,t)\} = f(x) \tag{3.39}$$

2. *Ο θόρυβος είναι πεπερασμένος.* Αν η ενέργεια του θορύβου είναι πεπερασμένη τότε θα ισχύει ότι:

$$\sigma^2 = E[E\{(g(x,t) - f(x))^2\}] < T_h, \qquad T_h \in \mathbb{R}^+ \tag{3.40}$$

Ο περιορισμός αυτός τίθεται για να επισημάνουμε ότι η ενέργεια του θορύβου που προστίθεται είναι περιορισμένη. Οσο πιο μικρή είναι η ενέργεια του θορύβου τόσο καλύτερη θα είναι και η εκτίμηση της ρίζας της συνάρτησης.

Απο του Robbins και Monro αποδείχθηκε ότι αν ο θόρυβος ικανοποιεί τις παραπάνω συνθήκες τότε ο επαναληπτικός αλγόριθμος που ακολουθεί συγκλίνει πάντα σε σημείο που βρίσκεται στην γειτονιά της πραγματικής ρίζας.

1. *Αρχικές τιμές, k=0.* Επιλέγουμε μία τιμή για το x έστω x_0 τέτοια ώστε να είναι κοντά στην ρίζα της εξίσωσης. Αν αυτό δεν είναι δυνατόν να πραγματοποιηθεί διαλέγουμε μία οποιαδήποτε τυχαία τιμή x_0.

2. *Επαναπροσδιορίζουμε την τιμή της ρίζας της εξίσωσης.* Με την σχέση που ακολουθεί επαναπροσδιορίζουμε την εκτίμησή μας για την τιμή της ρίζας της εξίσωσης:

$$x_{k+1} = x_k - a(k)g(x_k) \tag{3.41}$$

3. *Ελεγχος τερματισμού των επαναλήψεων, k=k+1.* Επειδή το φαινόμενο είναι στοχαστικό δεν μπορούμε να θεωρήσουμε ότι τα συνήθη κριτήρια σύγκλισης είναι αξιόπιστα, όπως π.χ. το κριτήριο ελέγχου της απόλυτης μεταβολής της εκτίμησης της ρίζας της εξίσωσης. Το κριτήριο που χρησιμοποιούμε σε αυτές τις περιπτώσεις είναι ο αριθμός των επαναλήψεων του αλγόριθμου ή ισοδύναμα, ο ελέγχουμε την αριθμητική τιμή της σταθεράς $a(k)$.

Ο παράγοντας $a(k)$ είναι μία ακολουθία θετικών πραγματικών αριθμών η οποία πρέπει να ικανοποιεί τις ακόλουθες συνθήκες:

$$\lim_{k\to\infty} a(k) = 0 \tag{3.42}$$

$$\sum_{k=1}^{\infty} a(k) = +\infty \tag{3.43}$$

$$\sum_{k=1}^{\infty} a(k)^2 = W,, \qquad W \in \mathbb{R}^+ \tag{3.44}$$

Ο τελευταίος περιορισμός μας λέει ότι το άθροισμα των τετραγώνων της ακολουθίας θα πρέπει να είναι πεπερασμένος θετικός αριθμός.

Παράδειγμα ακολουθίας αριθμών που ικανοποιούν τους περιορισμούς και χρησιμοποιούνται πολύ συχνά στις εφαρμογές, λόγω της απλότητας των υπολογισμών, είναι η:

$$a(k) = \frac{1}{k}, \qquad k = 1, 2, ... \tag{3.45}$$

Λίγα χρόνια μετά την παρουσίαση του αλγόριθμου των Robbins-Monro αποδείχθηκε ότι ο αλγόριθμος συγκλίνει **πάντα** σε τιμή που βρίσκεται κοντά στην πραγματική ρίζα. Συγκεκριμένα ο Blum απέδειξε ότι:

$$Prob(\lim_{k\to\infty} x_k = \hat{x}) = 1 \tag{3.46}$$

Πολλές φορές ο αλγόριθμος παρουσιάζει πολύ αργούς ρυθμούς σύγκλισης γιαυτό τον λόγο καταφεύγουμε συχνά σε διάφορα τεχνάσματα γιά να αυξήσουμε την ταχύτητα σύγκλισης. Ενας αποτελεσματικός τρόπος για να επιταχύνουμε την σύγκλιση του αλγόριθμου είναι να μην αλλάζουμε τον συντελεστή $a(k)$ όταν το πρόσημο της συνάρτησης δεν έχει αλλάξει σε δύο διαφορετικά βήματα. Με την τεχνική αυτή διατηρούμε σταθερό τον συντελεστή επαναπροσδιορισμού της ρίζας όταν βρισκόμαστε μακρυά από την πραγματική τιμή της ρίζας.

3.8.2 Ο αλγόριθμος των Kiefer-Wolfowitz

Ο αλγόριθμος των Robbins-Monro μπορεί εύκολα να προσαρμοστεί έτσι ώστε να υπολογίζει τα ακρότατα μιας συνάρτησης παλινδρόμησης. Είναι γνωστό ότι τα ακρότατα της συνάρτησης $f(x)$ βρίσκονται στις ρίζες της συνάρτησης:

$$\frac{\partial}{\partial x} f(x) = 0$$

οπότε και το πρόβλημα εύρεσης των ακρότατων της συνάρτησης μπορεί εύκολα να μετατραπεί σε αντίστοιχο πρόβλημα εύρεσης των ριζών της πρώτης παραγώγου της. Επειδή η συνάρτηση παλινδρόμησης είναι στοχαστική, η παράγωγος σε οποιδήποτε σημείο μπορεί να υπολογιστεί μόνο με αριθμητική μέθοδο.

Γνωρίζουμε ότι η παράγωγος μιας συνάρτησης μπορεί να προσεγγιστεί από την ακόλουθη σχέση:

$$\frac{\partial}{\partial x} f(x) \approx \frac{f(x+h) - f(x-h)}{2h} \tag{3.47}$$

Από τον αλγόριθμο των Robbins-Monro, η αναδρομική σχέση που επαναπροσδιορίζει τις ρίζες της παραγώγου της συνάρτησης, που είναι ταυτόχρονα και τα αντίστοιχα τοπικά ελάχιστα, δίνεται από την σχέση:

$$x_{k+1} = x_k - a(k) \frac{g(x_k + c(k)) - g(x_k - c(k))}{2c(k)} \tag{3.48}$$

ενώ η αναδρομική σχέση που επαναπροσδιορίζει τις τιμές των μέγιστων της συνάρτησης παλινδρόμησης είναι αντίστοιχα:

$$x_{k+1} = x_k + a(k)\frac{g(x_k + c(k)) - g(x_k - c(k))}{2c(k)} \tag{3.49}$$

Η επαναληπτική εξίσωση ονομάζεται μέθοδος εύρεσης των ακρότατων παλίνδρομης συνάρτησης των Kiefer-Wolfowitz.

Ολες οι παρατηρήσεις και οι περιορισμοί που τέθηκαν κατά την περιγραφή της μεθόδου εύρεσης των ριζών παλίνδρομης συνάρτησης των Robbins-Monro ισχύουν και σε αυτή την μέθοδο. Επιπρόσθετα πρέπει να τονίσουμε ότι οι ακολουθίες των θετικών πραγματικών αριθμών $a(k), c(k)$ πρέπει να ικανοποιούν τις ακόλουθες συνθήκες:

$$\lim_{k\to\infty} a(k) = 0 \tag{3.50}$$

$$\lim_{k\to\infty} c(k) = 0 \tag{3.51}$$

$$\sum_{k=1}^{\infty} a(k) = +\infty \tag{3.52}$$

$$\sum_{k=1}^{\infty} \left(\frac{a(k)}{c(k)}\right)^2 = W, \qquad W \in \mathbb{R}^+ \tag{3.53}$$

Συνήθεις ακολουθίες θετικών πραγματικών αριθμών που χρησιμοποιούνται στις εφαρμογές λόγω της υπολογιστικής απλότητας είναι οι:

$$a(k) = \frac{1}{k}, \qquad k = 1, 2, ... \tag{3.54}$$

$$c(k) = \frac{1}{log(k) + 1}, \qquad k = 1, 2, ... \tag{3.55}$$

Παράδειγμα 33 *Τηλεκατευθυνόμενο βαθυσκάφος μελετά υποθαλάσσια ρεύματα. Σκοπός του είναι να μεταβεί στο κέντρο υποθαλάσσιου ρεύματος και να πάρει διάφορες μετρήσεις. Η ανίχνευση του ρεύματος πραγματοποιείται με μετρήσεις της θερμοκρασίας του νερού.*

Θεωρώντας ότι το κέντρο του υποθαλάσσιου ρεύματος βρίσκεται στο σημείο που μεγιστοποιείται η διαφορά θερμοκρασίας, από σημείο αναφοράς σταθερής θερμοκρασίας, μπορείτε να φτιάξετε έναν μηχανισμό που να κατευθύνει το βαθυσκάφος στο κέντρο του, γνωρίζοντας ότι οι μετρήσεις που παίρνετε από το θερμόμετρο του βαθυσκάφους υπόκεινται σε πολλαπλές επιδράσεις που αλλοιώνουν την πραγματική θερμοκρασία του νερού που περιβάλλει το βαθυσκάφος.

Εκτελέστε μία αριθμητική προσωμοίωση θεωρώντας ότι η μέτρηση της θερμοκρασίας που παίρνετε σαν συνάρτηση του βάθους (σε χιλιόμετρα) είναι μία παλίνδρομη συνάρτηση της μορφής: $\theta(x) = 5{+}10e^{|x-0.521|}{+}0.5rnd()$, *Κελσίου*

όπου η συνάρτηση $rnd()$ *είναι μία συνάρτηση που δίνει έναν τυχαίο αριθμό στο διάστημα* $[0, 1)$.

Είναι φανερό ότι για να λύσουμε το πρόβλημα πρέπει να βρούμε το μέγιστο της παλίνδρομης συνάρτησης $\theta(x)$.

Θα εφαρμόσουμε τον αλγόριθμο των Kiefer-Wolfowitz χρησιμοποιώντας τις ακολουθίες πραγματικών αριθμών

$$a(k) = \frac{1}{k}, \qquad k = 1, 2, ...$$

$$c(k) = \frac{1}{log(k) + 1}, \qquad k = 1, 2, ...$$

και αρχική τιμή βάθους 200 μέτρα.

Στον πίνακα που ακολουθεί θα δούμε τις διαφορετικές προσεγγίσεις του βάθους στο οποίο βρίσκεται το κέντρο του υποθαλάσσιου ρεύματος. Οι συντελεστές $a(k), c(k)$ δεν μεταβάλλονται όσο η ποσότητα $g(x_k + c(k)) - g(x_k - c(k))$ παραμένει ομόσημη στις διαδοχικές εκτιμήσεις του ακρότατου.

Πίνακας 3.4: Τα διαδοχικά βήματα προσέγγισης του βάθους στο οποίο βρίσκεται το κέντρο του υποθαλάσσιου ρεύματος

Βήμα	1	2	3	4	5	6	7	8	9
Βάθος	0.4680	0.5376	0.5360	0.5451	0.4829	0.5352	0.4669	0.5932	0.4348
$\|x_k - x_{k-1}\|$	0.2680	0.0696	0.0016	0.0091	0.0623	0.0523	0.0683	0.1263	0.1584
Βήμα	10	11	12	13	14	15	16	17	18
Βάθος	0.5636	0.4924	0.5222	0.5264	0.5247	0.5122	0.5270	0.5112	0.5102
$\|x_k - x_{k-1}\|$	0.1288	0.0712	0.0297	0.0042	0.0017	0.0126	0.0149	0.0158	0.0010

Μετά από δεκαοχτώ βήματα του αλγόριθμου η τελική θέση του κέντρου του υποθαλάσιου ρεύματος εκτιμήθηκε ότι πρέπει να είναι στα 510.2 μέτρα (το πραγματικό κέντρο βρίσκεται σε βάθος 521 μέτρων).

3.8.3 Αλγόριθμοι προσέγγισης των ακρότατων πολυδιάστατων συναρτήσεων παλινδρόμησης

Η επέκταση των αλγορίθμων των Robbins-Monro και των Kiefer-Wolfowitz σε πολυδιάστατες συναρτήσεις παλινδρόμησης είναι σχετικά απλή υπόθεση. Οι ενέργειες που πρέπει να κάνουμε και οι συνθήκες που πρέπει να πληρεί ο θόρυβος που παραμορφώνει το σήμα μας παραμένουν οι ίδιες. Η αναδρομική εκτίμηση της ρίζας της παλίνδρομης εξίσωσης γίνεται:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - a(k)g(\mathbf{x}_k) \tag{3.56}$$

Αν $\mathbf{x} \in \mathbb{R}^P$ είναι το παραμετρικό διάνυσμα, οι αναδρομικές σχέσεις που επαναπροσδιορίζει τις τιμές των τοπικά ελάχιστων της συνάρτησης $g(\mathbf{x}_k)$ των Kiefer-Wolfowitz γίνεται:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - a(k) \begin{pmatrix} \frac{\partial g(\mathbf{x}_k)}{\partial x_1} \\ \frac{\partial g(\mathbf{x}_k)}{\partial x_2} \\ \cdots \\ \frac{\partial g(\mathbf{x}_k)}{\partial x_P} \end{pmatrix} \tag{3.57}$$

με

$$\frac{\partial g(\mathbf{x}_k)}{\partial x_i} \approx \frac{g(\mathbf{x}_k + c(k)\mathbf{e}_i) - g(\mathbf{x}_k - c(k)\mathbf{e}_i)}{2c(k)} \tag{3.58}$$

Το διάνυσμα $\mathbf{e}_i$ έχει τιμή μηδέν σε όλες τις θέσεις του εκτός από την διάσταση i στην οποία έχει τιμή ένα: $\mathbf{e}_i = (0, ..., 0, 1, 0, ..., 0)^T$

Αντίστοιχα η αναδρομική σχέση που επαναπροσδιορίζει τις τιμές των μέγιστων είναι αντίστοιχα:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + a(k) \begin{pmatrix} \frac{\partial g(\mathbf{x}_k)}{\partial x_1} \\ \frac{\partial g(\mathbf{x}_k)}{\partial x_2} \\ \cdots \\ \frac{\partial g(\mathbf{x}_k)}{\partial x_P} \end{pmatrix} \tag{3.59}$$

3.8.4 Υπολογισμός κατωφλίων απόφασης με στοχαστικές μεθόδους

Είδαμε σε προηγούμενο τμήμα αυτού του κεφαλαίου ότι μπορούμε να προσεγγίσουμε την συνάρτηση πιθανότητας με γραμμικό τρόπο μετασχηματίζοντας τα διανύσματα των προτύπων των παραδειγμάτων. Η συνάρτηση διάκρισης μπορεί να εκφραστεί σαν:

$$g(\mathbf{x}) = p(\omega|\mathbf{x}) \approx \sum_{i=1}^{Q+1} c_i \varphi_i(\mathbf{x}) = \mathbf{c}^T \mathbf{x}_\varphi \tag{3.60}$$

όπου

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_Q \\ c_{Q+1} \end{pmatrix} \quad \text{και} \quad \mathbf{x}_\varphi = \begin{pmatrix} \varphi_1(\mathbf{x}) \\ \varphi_2(\mathbf{x}) \\ \cdots \\ \varphi_Q(\mathbf{x}) \\ 1 \end{pmatrix} \tag{3.61}$$

Μια απλή μέθοδος υπολογισμού της πιθανότητας $p(\omega|\mathbf{x})$ μπορεί να γίνει με την γνώση της τιμής της στα διάφορα σημεία του χώρου των μετρήσεων π.χ. στα σημεία που βρίσκονται τα πρότυπα των παραδειγμάτων.

Η πληροφορία που διαθέτουμε γιαυτά τα σημεία κατά την διάρκεια της εκπαίδευσης είναι η ταυτότητα του πρότυπου. Εκμεταλευόμενοι αυτή την πληροφορία μπορούμε να ορίσουμε την παλίνδρομη συνάρτηση $z(\mathbf{x})$ η οποία παίρνει τις ακόλουθες τιμές:

$$z(\mathbf{x}_\varphi) = \begin{cases} 1, & \mathbf{x}_\varphi \in \omega \\ 0, & \text{αλλού} \end{cases} \tag{3.62}$$

Υποθέτουμε ότι η συνάρτηση $z(\mathbf{x}_\varphi)$ αποτελεί μία προσέγγιση της πραγματικής συνάρτησης πυκνότητας πιθανότητας :

$$z(\mathbf{x}_\varphi) = p(\omega|\mathbf{x}_\varphi) + n \tag{3.63}$$

όπου n είναι θόρυβος ο οποίος υποθέτουμε ότι έχει μηδενική μέση τιμή και πεπερασμένη διασπορά. Η συνάρτηση:

$$R(z(\mathbf{x}_\varphi), \mathbf{c}) = \sum_{i=1}^{M} |z(\mathbf{x}_\varphi{}^{(i)}) - \mathbf{c}^T \mathbf{x}_\varphi{}^{(i)}| \tag{3.64}$$

είναι και αυτή μία συνάρτηση παλινδρόμησης που εξαρτάται από το διάνυσμα $\mathbf{c}^T$, και M είναι ο αριθμός των διαθέσιμων παραδειγμάτων. Αν χρησιμοποιήσουμε την μέθοδο των Kiefer-Wolfowitz για να βρούμε τις τιμές του διανύσματος $\mathbf{c}$ για το οποίο η συνάρτηση παλινδρόμησης γίνεται ελάχιστη τότε έχουμε την ακόλουθη αναδρομική εξίσωση που οδηγεί στην εκτίμηση του ελάχιστου της συνάρτησης:

$$\mathbf{c}_{k+1} = \mathbf{c}_k - a(k)\left(\frac{\partial R(z(\mathbf{x}_\varphi), \mathbf{c})}{\partial \mathbf{c}}\right)_{\mathbf{c}=\mathbf{c}_k} \tag{3.65}$$

Βρίσκοντας την παράγωγο της συνάρτησης ως προς την διανυσματική μεταβλητή $\mathbf{x}_\varphi$ έχουμε:

$$\frac{\partial R(z(\mathbf{x}_\varphi), \mathbf{c})}{\partial \mathbf{c}} = -\sum_{i=1}^{M} \mathbf{x}_\varphi sgn(z(\mathbf{x}_\varphi{}^{(i)}) - \mathbf{c}^T \mathbf{x}_\varphi{}^{(i)}) \tag{3.66}$$

όπου

$$sgn(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases} \tag{3.67}$$

Με βάσει τα παραπάνω συμπεράσματα η εξίσωση 3.65 τελικά έρχεται στην μορφή:

$$\mathbf{c}_{k+1} = \begin{cases} \mathbf{c}_k + a(k)\mathbf{x}_\varphi, & z(\mathbf{x}_\varphi) > \mathbf{c}^T \mathbf{x}_\varphi \\ \mathbf{c}_k - a(k)\mathbf{x}_\varphi, & z(\mathbf{x}_\varphi) \leq \mathbf{c}^T \mathbf{x}_\varphi \end{cases} \tag{3.68}$$

Από την μορφή της σχέσης μπορούμε να παρατηρήσουμε ότι ο στοχαστικός αλγόριθμος προσδιορισμού ενός βέλτιστου αριθμού συντελεστών της γραμμικής συνάρτησης απόφασης μοιάζει με τον αντίστοιχο αλγόριθμο perceptron που περιέχεται στο κεφάλαιο 2. Η μοναδική τους διαφορά βρίσκεται στον συντελεστή επαναπροσδιορισμού ο οποίος στην περίπτωση της στοχαστικής προσέγγισης συνεχώς ελαττώνεται ενώ στον αλγόριθμο perceptron παραμένει σταθερός.

Η σημαντικότερη διαφορά των δύο αλγορίθμων βρίσκεται στο γεγονός ότι ο στοχαστικός αλγόριθμος **συγκλίνει πάντα ανεξάρτητα από την κατανομή των προτύπων των παραδειγμάτων στο χώρο των μετρήσεων**, σε αντίθεση με τον αλγόριθμο perceptron ο οποίος θέτει σαν απαραίτητη προυπόθεση για την σύγκλιση την γραμμική διαχωρισιμότητα των προτύπων.

Πρόγραμμα 5 *Υπολογισμός των συντελεστών γραμμικού ταξινομητή με την βοήθεια του αλγόριθμου των Kiefer-Wolfowitz*

```
function [Rc,Rep] = KieWolLin(x1,x2,MaxRep,Const)
#
#  [Rc,Rep] = KieWolLin(x1,x2,MaxRep,Const)
#
#  Input
#      x1: Pattern Vectors for the first class
#      x2: Pattern Vectors for the second class
#      MaxRep: Maximum number of repetitions
#      Const: Constand multiplier factor
#  Output
#      Rc: Correct classification rate using the C-method
#      Rep: Pattern vectors on each class
#

NumOfP1 = columns(x1) ;
NumOfP2 = columns(x2) ;
Weights = 2 * rand(rows(x1)+1,1) - 1 ;
Rep = [NumOfP1,NumOfP2] ;
```

```
TotPat = sum(Rep) ;
Rc = zeros(2,1) ;
if ( rows(x1) != rows(x2) )
printf( "Error in vectors x1, x2\n" ) ;
endif

for j = 1:MaxRep
if ( rand() > 0.5 )
k = floor(rand() * NumOfP1 + 1) ;
Pat = [x1(:,k);1] ;
Score = Weights' * Pat ;
if ( Score < 0 )
Weights = Weights + (Const/j) * Pat ;
endif
else
      k = floor(rand() * NumOfP2 + 1) ;
      Pat = [x2(:,k);1] ;
      Score = Weights' * Pat ;
      if ( Score >= 0 )
         Weights = Weights - (Const/j) * Pat ;
      endif
endif
Rc = zeros(2,1) ;
for i=1:NumOfP1
if ( Weights' * [x1(:,i);1] >= 0 )
Rc(1) = Rc(1) + 1 ;
endif
endfor
   for i=1:NumOfP2
      if ( Weights' * [x2(:,i);1] < 0 )
         Rc(2) = Rc(2) + 1 ;
      endif
   endfor
printf( "%8d", sum(Rc) ) ;
fflush(stdout) ;
if ( sum(Rc) == TotPat )
break ;
endif
endfor
printf( "\n" ) ;
endfunction
```

Παράδειγμα 34 *Δίνονται τα πρότυπα των παραδειγμάτων δύο μη-γραμμικά διαχωρίσιμων κατηγοριών:*

$$\Omega_1 = \{(0,-3),(-0.5,2),(1,-1),(1,1),(2,-1)\}$$

$$\Omega_2 = \{(0.5,-1),(-0.5,-0.5),(-1,2.5),(-2,1),(-2,-1)$$

Βρείτε την συνάρτηση απόφασης για το στοχαστικό μοντέλο ταξινόμησης της μεγαλύτερης πιθανότητας με την μέθοδο των Kiefer-Wolfowitz.

Επεκτείνω την διάσταση των προτύπων $\mathbf{x} = (\mathbf{x}, 1)^T$, οπότε και τα πρότυπα των παραδειγμάτων γίνονται:

$$\Omega_1 = \{(0,-3,1),(-0.5,2,1),(1,-1,1),(1,1,1),(2,-1,1)\}$$

$$\Omega_2 = \{(0.5,-1,1),(-0.5,-0.5,1),(-1,2.5,1),(-2,1,1),(-2,-1,1)$$

Η συνάρτηση απόφασης δίνεται από την σχέση:

$$g(\mathbf{x}) = p(\omega|\mathbf{x}) \approx \sum_{i=1}^{Q} c_i \varphi_i(\mathbf{x}) = \mathbf{c}^T \mathbf{x}_\varphi$$

Εφόσον δεν πραγματοποιήσουμε κάποιον μη-γραμμικό μετασχηματισμό, η συνάρτηση απόφασης δίνεται από την σχέση:

$$g(\mathbf{x}) = p(\omega|\mathbf{x}) \approx \mathbf{c}^T \mathbf{x}$$

Φτιάχνω την συνάρτηση που δίνει την διαφορά των συναρτήσεων απόφασης:

$$d(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) = p(\omega_1|\mathbf{x}) - p(\omega_2|\mathbf{x}) \approx \mathbf{c_1}^T\mathbf{x} - \mathbf{c_2}^T\mathbf{x} = (\mathbf{c_1}^T - \mathbf{c_2}^T)\mathbf{x} = \mathbf{c}^T\mathbf{x}$$

Αντί λοιπόν για έξι συντελεστές που πρέπει να υπολογίσουμε, αν ορίσουμε δύο συναρτήσεις απόφασης, μπορούμε να χρησιμοποιήσουμε την συνάρτηση $d(\mathbf{x})$ και να ορίσουμε την εξής παλίνδρομη συνάρτηση:

$$z(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \omega_1 \\ -1 & \mathbf{x} \in \omega_2 \end{cases}$$

Ο στοχαστικός αλγόριθμος επαναπροσδιορισμού των συντελεστών βαρύτητας γίνεται:

$$\mathbf{c}_{k+1} = \begin{cases} \mathbf{c}_k + a(k)\mathbf{x}, & z(\mathbf{x}) > \mathbf{c}^T\mathbf{x} \\ \mathbf{c}_k - a(k)\mathbf{x}, & z(\mathbf{x}) \leq \mathbf{c}^T\mathbf{x} \end{cases}$$

Ξεκινώντας με αρχική τιμή των συντελεστών $c_0 = \{0, 0, 0\}$ στον πίνακα που ακολουθεί αναγράφουμε την τιμή $\mathbf{c}_k\mathbf{x}$ γιά κάθε ένα από τα πρότυπα των παραδειγμάτων που είναι διαθέσιμα, όταν συμπληρωθεί ένας κύκλος επαναπροσδιορισμού των συντελεστών με όλα τα διαθέσιμα παραδείγματα. Στην τελευταία στήλη αναγράφεται το αθροιστικό σφάλμα των διαφορών $|z(\mathbf{x}) - \mathbf{c}_k\mathbf{x}|$ για τα παραδείγματα της κατηγορίας.

Αν θα θέλαμε να μετρήσουμε το ελάχιστο σφάλμα του συστήματος που φτιάξαμε κατά την διάρκεια επαναπροσδιορισμού των συντελεστών της γραμμικής συνάρτησης απόφασης μπορούμε να δούμε ότι τα σφάλματα ήταν 5,5,5,4,3,3 σε δέκα πρότυπα.

Η τελική συνάρτηση απόφασης είναι:

$$d(\mathbf{x}) = d((x_1, x_2)) = 0.381183x_1 - 0.209777x_2 - 0.0711072$$

Η διαδικασία ταξινόμησης πραγματοποιείται με την εύρεση της τιμής της $d(\mathbf{x})$ για το πρότυπο που ενδιαφερόμαστε. Αν η τιμή της συνάρτησης είναι θετική το πρότυπο ταξινομείται στην κατηγορία ω_1 διαφορετικά ταξινομείται στην κατηγορία ω_2.

3.8.5 Στοχαστικός αλγόριθμος ελάχιστου τετραγωνικού σφάλματος

Αν ορίσουμε με διαφορετικό τρόπο την στοχαστική συνάρτηση του σφάλματος:

$$R(z(\mathbf{x}_\varphi), \mathbf{c}) = \sum_{i=1}^{M} (z(\mathbf{x}_\varphi{}^{(i)}) - \mathbf{c}^T\mathbf{x}_\varphi{}^{(i)})^2 \tag{3.69}$$

ακολουθώντας τον ίδιο τρόπο απόδειξης μπορούμε να δείξουμε ότι ο προσδιορισμός του ελάχιστου της συνάρτησης $R(z(\mathbf{x}_\varphi), \mathbf{c})$ μπορεί να επιτευχθεί με την εξής αναδρομική σχέση:

$$\mathbf{c}_{k+1} = \mathbf{c}_k + a(k)\mathbf{x}_\varphi(z(\mathbf{x}_\varphi) - \mathbf{c}^T\mathbf{x}_\varphi) \tag{3.70}$$

Πίνακας 3.5: Υπολογισμός γραμμικού κατωφλίου απόφασης με την μέθοδο των Kiefer-Wolfowitz

Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	4.477	-1.934	1.675	-0.929	1.477	11.3333
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	1.774	1.321	-2.486	-0.335	2.269	10.445
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	2.303	-1.123	0.789	-0.605	0.670	7.759
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	0.849	0.620	-1.41	-0.246	1.148	9.02397
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	1.452	-0.858	0.568	-0.349	0.601	5.01083
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	0.551	0.288	-1.104	-0.450	0.467	8.243
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	0.865	-0.693	0.418	-0.175	0.566	3.930
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	0.345	0.048	-0.916	-0.618	-0.024	7.701
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	0.904	-0.819	0.501	-0.140	0.740	3.611
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	0.382	-0.017	-1.097	-0.856	-0.214	7.670
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$c_k x$	0.710	-0.700	0.482	-0.026	0.762	3.575
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$c_k x$	0.342	-0.064	-0.967	-0.866	-0.357	7.833

Παράδειγμα 35 *Βρείτε την συνάρτηση κατωφλίου για το στοχαστικό μοντέλο ταξινόμησης με την μεγαλύτερη πιθανότητα με την μέθοδο των Kiefer-Wolfowitz, για τα πρότυπα του προηγούμενου παραδείγματος χρησιμοποιώντας την συνάρτηση παλινδρόμησης:*

$$R(z(\mathbf{x}_\varphi),\mathbf{c}) = \sum_{i=1}^{M} (z(\mathbf{x}\varphi^{(i)}) - \mathbf{c}^T\mathbf{x}\varphi^{(i)})^2$$

και την ίδια αρχική τιμή.

Επαναλαμβάνοντας την διαδικασία ορισμού της συνάρτησης κατωφλίου και τους ίδιους μετασχηματισμούς που έγιναν και στο προηγούμενο παράδειγμα, η αναδρομική συνάρτηση επαναπροσδιορισμού των συντελεστών βαρύτητας γίνεται:

$$\mathbf{c}_{k+1} = \mathbf{c}_k + a(k)\mathbf{x}(z(\mathbf{x}) - \mathbf{c}^T\mathbf{x})$$

Ξεκινώντας με αρχική τιμή συντελεστών την $c_0 = \{0,0,0\}$, στον πίνακα που ακολουθεί αναγράφουμε την τιμή $\mathbf{c}_k\mathbf{x}$ γιά κάθε ένα από τα πρότυπα των παραδειγμάτων που είναι διαθέσιμα, όταν συμπληρωθεί ένας κύκλος επαναπροσδιορισμού των συντελεστών με όλα τα διαθέσιμα παραδείγματα. Στην τελευταία στήλη αναγράφεται το αθροιστικό σφάλμα των διαφορών $|z(\mathbf{x}) - \mathbf{c}_k\mathbf{x}|$ για τα παραδείγματα της κατηγορίας.

Το ελάχιστο σφάλμα του συστήματος κατά την διάρκεια επαναπροσδιορισμού των συντελεστών της γραμμικής συνάρτησης απόφασης παρουσιάζει χαμηλότερο ρυθμό σφαλμάτων από την προηγούμενη προσέγγιση και προσεγγίζει την τελική τιμή ταχύτερα. Τα σφάλματα σε αυτή την λύση είναι 4,3,3,3,3,3 γιά δέκα πρότυπα.

Η τελική συνάρτηση απόφασης είναι επίσης ελαφρώς διαφοροποιημένη, σε:

$$d(\mathbf{x}) = d((x_1,x_2)) = 0.287111x_1 + 0.0237388x_2 - 0.109837$$

Πίναχας 3.6: Υπολογισμός γραμμιχού χατωφλίου απόφασης με την μέθοδο των Kiefer-Wolfowitz

Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	3.940	-1.744	2.521	0.390	3.234	13.5
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	2.165	0.920	-2.633	-1.747	0.384	10.036
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	-0.013	-0.052	0.467	0.534	0.881	4.821
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	0.260	-0.136	-0.242	-0.707	-0.774	8.253
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	-0.115	-0.101	0.306	0.381	0.653	3.13069
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	0.132	-0.195	-0.256	-0.659	-0.734	8.243
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	-0.145	-0.137	0.241	0.308	0.561	3.579
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	0.081	-0.221	-0.280	-0.651	-0.718	8.244
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	-0.159	-0.161	0.207	0.267	0.513	3.765
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	0.054	-0.237	-0.299	-0.651	-0.711	8.248
Πρότυπα της ω_1	(0,-3)	(-0.5,2)	(1 ,-1)	(1,1)	(2,-1)	
$\mathbf{c}_k\mathbf{x}$	-0.167	-0.177	0.186	0.241	0.484	3.875
Πρότυπα της ω_2	(0.5,-1)	(-0.5,-0.5)	(-1,2.5)	(-2,1)	(-2,-1)	
$\mathbf{c}_k\mathbf{x}$	0.0369	-0.247	-0.312	-0.653	-0.708	8.252

3.9 Λυμένα Προβλήματα

Πρόβλημα 2 *Εστω ότι η πυχνότητα πιθανότητας των προτύπων μιας χατηγορίας έχει προσεγγιστεί με Q ορθοχανονιχές συναρτήσεις χαι την βοήθεια M παραδειγμάτων.*

$$\hat{p}(\mathbf{x}) = \sum_{i=1}^{Q} c_i \varphi_i(\mathbf{x})$$

Μπορείτε να βρείτε μία νέα προσέγγιση της συνάρτησης πυχνότητας πιθανότητας δεδομένου ότι πληροφορηθήχατε την ύπαρξη ενός νέου παραδείγματος, ενώ δεν έχετε πλέον στην διάθεσή σας τα υπόλοιπα παραδείγματα;

Οι συντελεστές της συνάρτησης πυχνότητας πιθανότητας δίνονται από την σχέση:

$$c_i \approx \frac{1}{M} \sum_{j=1}^{M} w(\mathbf{x}_j)\varphi_i(\mathbf{x}_j)$$

Με την προσθήχη ενός νέου παραδείγματος οι νέοι συντελεστές δίνονται από την σχέση:

$$c_i{}' \approx \frac{1}{M+1} \sum_{j=1}^{M+1} w(\mathbf{x}_j)\varphi_i(\mathbf{x}_j) \Rightarrow$$

$$c_i{}' \approx \frac{M}{M+1}\frac{1}{M}\sum_{j=1}^{M} w(\mathbf{x}_j)\varphi_i(\mathbf{x}_j) + \frac{1}{M+1}w(\mathbf{x}_{M+1})\varphi_i(\mathbf{x}_{M+1}) \Rightarrow$$

$$c_i{}' \approx \frac{M}{M+1}c_i + \frac{1}{M+1}w(\mathbf{x}_{M+1})\varphi_i(\mathbf{x}_{M+1})$$

Η αναδρομικότητα στον υπολογισμό των συντελεστών των συναρτήσεων που προσεγγίζουν την πυκνότητα πιθανότητας μιάς κατηγορίας προτύπων μας δίνει ένα πολύ σημαντικό πλεονέκτημα. Με αυτές τις σχέσεις μπορούμε να επιτύχουμε μία αποτελεσματική και συνεχή εκπαίδευση του συστήματος ταξινόμησης.

Μπορούμε αρχικά να εκτιμήσουμε τους συντελεστές c_i με τα διαθέσιμα δείγματα, στην συνέχεια να τα διαγράψουμε και κάθε φορά που βρίσκουμε ένα νέο παράδειγμα εκπαίδευσης να επαναπροσδιορίζουμε τους συντελεστές. Η μόνη επιπρόσθετη πληροφορία που χρειάζεται να κρατάμε είναι ο συνολικός αριθμός των παραδειγμάτων τα οποία έχουν χρησιμοποιηθεί για να εκπαιδεύσουν το σύστημα.

Πρόβλημα 3 *Υπολογείστε το ελάχιστο σφάλμα στοχαστικού συστήματος ταξινόμησης προτύπων όταν η πυκνότητα πιθανότητας των προτύπων των κατηγοριών προσεγγίζεται με πέντε ορθοχανονικές συναρτήσεις και την βοήθεια των ακόλουθων παραδειγμάτων:*

$$\Omega = \{((1,1),\omega_1),((2.5,2.5),\omega_1),((3,1.1),\omega_1),((2,1.5),\omega_1),$$

$$((1,2.8),\omega_2),((2.5,5.3),\omega_2),((0,1),\omega_2),((1.5,3.5),\omega_2),((0.5,3),\omega_2)\}.$$

Αρχικά θα χρησιμοποιήσουμε μία οικογένεια ορθοχανονικών συναρτήσεων. Εστω ότι διαλέγουμε τα πολυώνυμα Hermite. Γνωρίζουμε ότι, με συνάρτηση βαρύτητας την $w(x) = e^{-x^2}$ στο διάστημα $-\infty < x < +\infty$ τα ακόλουθα πολυώνυμα είναι ορθογώνια.

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x \end{aligned}$$

Μετατρέπω τις ορθογώνιες συναρτήσεις σε ορθοχανονικές βάσει του μετασχηματισμού:

$$\varphi_i(x) = \sqrt{\frac{w(x)}{A_i}}H_i(x)$$

όπου

$$A_i = \int_{-\infty}^{\infty} e^{-x^2} H_i{}^2(x)dx$$

Τα αντίστοιχα ορθοχανονικά πολυώνυμα Hermite είναι:

$$\varphi_0(x) = \sqrt{\frac{e^{-x^2}}{\sqrt{\pi}}}$$

$$\varphi_1(x) = \sqrt{\frac{e^{-x^2}}{2\sqrt{\pi}}}2x$$

$$\varphi_2(x) = \sqrt{\frac{e^{-x^2}}{8\sqrt{\pi}}}(4x^2 - 2)$$

$$\varphi_3(x) = \sqrt{\frac{e^{-x^2}}{48\sqrt{\pi}}}(8x^3 - 12x)$$

Για τα διανύσματα δύο διαστάσεων ορίζω από τα ορθοκανονικά πολυώνυμα Hermite πέντε συναρτήσεις:

$$\Phi_1(\mathbf{x}) = \varphi_0(x_1)\varphi_0(x_2) = \sqrt{\frac{e^{-x_1^2-x_2^2}}{\pi}}$$

$$\Phi_2(\mathbf{x}) = \varphi_1(x_1)\varphi_0(x_2) = \sqrt{\frac{2e^{-x_1^2-x_2^2}}{\pi}}x_1$$

$$\Phi_3(\mathbf{x}) = \varphi_0(x_1)\varphi_1(x_2) = \sqrt{\frac{2e^{-x_1^2-x_2^2}}{\pi}}x_2$$

$$\Phi_4(\mathbf{x}) = \varphi_1(x_1)\varphi_1(x_2) = \sqrt{\frac{e^{-x_1^2-x_2^2}}{\pi}}x_1x_2$$

$$\Phi_5(\mathbf{x}) = \varphi_2(x_1)\varphi_0(x_2) = \sqrt{\frac{e^{-x_1^2-x_2^2}}{2\pi}}(2x_1^2-1)$$

Η συνάρτηση που προσεγγίζει την πυκνότητα πιθανότητας για κάθε μία από τις κατηγορίες δίνεται από τον γενικό τύπο:

$$\hat{p}(\mathbf{x}) = \sum_{i=1}^{5} c_i \Phi_i(\mathbf{x})$$

Οι συντελεστές c_i θα υπολογιστούν από τα M παραδείγματα από την γνωστή σχέση:

$$c_i \approx \frac{1}{M}\sum_{j=1}^{M} \Phi_i(\mathbf{x}_j)$$

Το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων θα υπολογιστεί χρησιμοποιώντας όλα τα παραδείγματα και για εκπαίδευση και για τον έλεγχο της απόδοσης.

Αρχικά υπολογίζουμε την πυκνότητα πιθανότητας για κάθε μία από τις κατηγορίες του συστήματος ταξινόμησης. Στον πίνακα που ακολουθεί δίνονται οι πέντε συντελεστές c_i για τις δύο κατηγορίες.

Πίνακας 3.7: Συντελεστές ορθοκανονικών συναρτήσεων

Κατηγορία	c_1	c_2	c_3	c_4	c_5
ω_1	0.0592136	0.0955025	0.0888213	0.07500540	0.0798648
ω_2	0.0709839	0.0028726	0.1072550	0.00588822	-0.0476267

Το κριτήριο ταξινόμησης με βάση την μεγαλύτερη πιθανότητα απαιτεί και την γνώση της πιθανότητας εμφάνισης των προτύπων κάθε κατηγορίας. Επειδή δεν έχουμε ένα τέτοιο δεδομένο κάνουμε μία στατιστική προσέγγιση με βάση τον αριθμό των παραδειγμάτων κάθε κατηγορίας που διαθέτουμε. Σε σύνολο εννέα παραδειγμάτων τα τέσσερα ανήκουν στην πρώτη κατηγορία και πέντε στην δεύτερη.

Συνεπώς $p(\omega_1) = \frac{4}{9}$ και $p(\omega_2) = \frac{5}{9}$.

Στον πίνακα που ακολουθεί δίνονται οι αριθμητικές τιμές του κριτηρίου ταξινόμησης της μεγαλύτερης πιθανότητας $(p(\omega_i)\hat{p}(\mathbf{x}|\omega_i))$ για τις δύο κατηγορίες και για όλα τα παραδείγματα.

Το ελάχιστο σφάλμα του συστήματος ταξινόμησης που φτιάξαμε από τα παραδείγματα είναι 3 / 9 * 100 = 33.33%.

Σε αυτό το σημείο πρέπει να δώσουμε δύο σημαντικές παρατηρήσεις:

1. Το κριτήριο ταξινόμησης της μεγαλύτερης πιθανότητας για το πρότυπο (3,1.1) και την δεύτερη κατηγορία έδωσε αρνητικό αριθμό, παρόλο που αυτό δεν επιτρέπεται μία και το κριτήριο αποτελείται από γινόμενο πιθανοτήτων.

Πίναχας 3.8: Αριθμητιχές τιμές του χριτηρίου ταξινόμησης της μεγαλύτερης πιθανότητας

Παραδείγματα της ω_1	(1,1)	(2.5,2.5)	(3,1.1)	(2,1.5)	
ω_1	0.04380	0.00089	0.002790	0.012808	
ω_2	0.02632	0.00008	-0.00052	0.001624	
Παραδείγματα της ω_2	(1,2.8)	(2.5,5.3)	(0,1)	(1.5,3.5)	(0.5,3)
ω_1	0.002525	2.39×10^{-8}	0.02325	0.00023	0.00150
ω_2	0.001930	6.62×10^{-9}	0.05432	0.00012	0.00179

Το γεγονός αυτό μας οδηγεί στο συμπέρασμα ότι η προσέγγιση που χάναμε δεν είναι αρχετά χαλή. Χρειάζονται περισσότερες ορθοχανονιχές συναρτήσεις για να υπολογίσουμε με μεγαλύτερη αχρίβεια τις πυχνότητες πιθανότητας των χατηγοριών.

2. Αν απειχονήσουμε τα παραδείγματα στο επίπεδο θα δούμε ότι τα πρότυπα είναι γραμμιχά διαχωρίσιμα χατά την θετιχή διαγώνιο των 45 μοιρών. Θα περιμέναμε το ελάχιστο σφάλμα του συστήματος ταξινόμησης να είναι 0%. Το γεγονός ότι δεν πετύχαμε το ελάχιστο σφάλμα ταξινόμησης οφείλεται πάλι στον μιχρό αριθμό των ορθοχανονιχών συναρτήσεων που χρησιμοποιήσαμε για την προσέγγιση.

3.10 Αλυτα Προβλήματα

Ασχηση 5 *Αποδείξτε το θεώρημα 5.*

Ασχηση 6 *Αποδείξτε το θεώρημα 6.*

Ασχηση 7 *Αν είναι γνωστό ότι το μονοδιάστατο πρότυπο χατηγορίας περιορίζεται στο διάστημα* $x \in [\alpha, \beta], \alpha, \beta \in \mathbb{R}^+$, *χαι είναι γνωστή από στατιστιχές μετρήσεις η μέση τιμή του πρότυπου βρείτε την συνάρτηση πυχνότητας πιθανότητας που μεγιστοποιεί την εντροπία της.*

Ασχηση 8 *Υπολογείστε το μέγιστο σφάλμα του στοχαστιχού συστήματος ταξινόμησης προτύπων του προβλήματος 3.*

Ασχηση 9 *Αυξήστε τα ορθοχανονιχά διανύσματα του λυμένου προβλήματος 3 σε δέχα χαι υπολογίστε ξανά το ελάχιστο σφάλμα με το χριτήριο της μεγαλύτερης πιθανότητας. Τι παρατηρείτε;*

Ασχηση 10 *Εστω ότι η πυχνότητα πιθανότητας δύο χατηγοριών μονοδιάστατων προτύπων αχολουθούν την χατανομή:*

$$p(x|\omega_1) = \begin{cases} -0.5x + 1, & 0 \leq x \leq 2 \\ 0, & x < 0 \quad \textit{χαι} \quad x > 2 \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 0.5x - 0.5, & 1 \leq x \leq 3 \\ 0, & x < 1 \quad \textit{χαι} \quad x > 3 \end{cases}$$

Υπολογίστε τα χατώφλια απόφασης του χριτηρίου της μεγαλύτερης πιθανότητας σαν συνάρτηση του λόγου των πιθανοτήτων εμφάνισης των προτύπων των χατηγοριών.

Ασκηση 11 *Γνωρίζουμε ότι η πυκνότητα πιθανότητας δύο κατηγοριών αντικειμένων ακολουθεί κανονική κατανομή. Δεδομένου ότι έχουμε στην διάθεσή μας τα ακόλουθα παραδείγματα:*

$$\Omega_1 = \{(0,4),(4,4),(0,0),(4,0)\}$$

$$\Omega_2 = \{(3,5),(3,3),(5,5),(5,3)\}$$

Υπολογίστε το κατώφλι απόφασης για το κριτήριο ταξινόμησης της μεγαλύτερης πιθανότητας, γνωρίζοντας ότι τα πρότυπα της δεύτερης κατηγορίας έχουν διπλάσια πιθανότητα εμφάνισης από τα πρότυπα της πρώτης κατηγορίας.

Ασκηση 12 *Εστω ότι η πυκνότητα πιθανότητας τριών κατηγοριών μονοδιάστατων προτύπων ακολουθούν την κατανομή του Reyleigh:*

$$p(x|\omega_i) = \begin{cases} \frac{x}{{\sigma_i}^2} e^{-\frac{x^2}{2{\sigma_i}^2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Υπολογίστε τα κατώφλια απόφασης του κριτηρίου της μεγαλύτερης πιθανότητας δεδομένου ότι τα πρότυπα των κατηγοριών έχουν ισοπίθανη πιθανότητα εμφάνισης.

Μπορείτε να υπολογίσετε την θεωρητική πιθανότητα σφάλματος ταξινόμησης;

Ασκηση 13 *Εστω ότι η πυκνότητα πιθανότητας δύο κατηγοριών μονοδιάστατων προτύπων ακολουθούν την κατανομή Pearson τύπου-2. Υπολογίστε το κατώφλι απόφασης του κριτηρίου της μεγαλύτερης πιθανότητας σαν συνάρτηση του λόγου πιθανοτήτων εμφάνισης των προτυπων των κατηγοριών.*

Ασκηση 14 *Εστω ότι έχουμε τα ακόλουθα παραδείγματα μονοδιάστατων προτύπων.*

$\Omega = \{1.0, 2.9, 3.8, 4, 4.5, 5.5, 6, 6.5, 6.9, 7, 8, 10, 12, 15, 20\}$

Γνωρίζοντας ότι το πρότυπο παίρνει μόνο θετικές τιμές, βρείτε ποιά από τις γνωστές κατανομές πυκνότητας πιθανότητας (Πίνακας 3.3) ταιριάζει περισσότερο στα παραδείγματα που διαθέτουμε;