Κεφάλαιο 6

Εξαγωγή παραμέτρων

6.1 Εισαγωγή

Σε αυτό το κεφάλαιο περιγράφονται μέθοδοι προεπεξεργασίας των προτύπων που αποσκοπούν στην συμπίεση δεδομένων του διανύσματος του πρότυπου.

Οπως θα διαπιστώσατε και από την μελέτη των προγούμενων κεφαλαίων το είδος των παραμέτρων που λαμβάνονται από το φυσικό αντικείμενο και η τυχόν επεξεργασία τους επηρεάζει σημαντικά την αξιοπιστία του συστήματος ταξινόμησης προτύπων που θα κατασκευάσουμε. Μια από τις τεχνικές προεπεξεργασίας των προτύπων την μελετήσαμε ήδη σαν τμήμα ενός δομικού συστημάτων ταξινόμησης προτύπων (κεφάλαιο 2). Πιο συγκεκριμένα επιχειρήσαμε την αύξηση των διαστάσεων του διανύσματος του πρότυπου με μη γραμμικό συνδυασμό των διαθέσιμων μετρήσεων και είδαμε ότι η τεχνική αυτή βελτιώνει την αξιοπιστία της ταξινόμησης ιδιαίτερα στις περιπτώσεις που δεν είναι δυνατή η γραμμική διαχωρισιμότητα των προτύπων.

Στο παρόν κεφάλαιο θα μελετήσουμε μία διαφορετική κατηγορία τεχνικών προεπεξεργασίας των προτύπων που σκοπό έχουν την μείωση των διαστάσεων του διανύσματος του πρότυπου ή την επιλογή αυτών των παραμέτρων που συνισφέρουν την μεγαλύτερη πληροφορία διαχωρισιμότητας των κατηγοριών.

Τα προβλήματα που συναντάμε κατά την πρακτική εφαρμογή συστημάτων ταξινόμησης προτύπων αναφέρονται συνήθως σε αυτές τις κατηγορίες προβλημάτων διότι συνήθως μπορούμε να λάβουμε ένα μεγάλο πλήθος μετρήσεων που αναφέρονται σε διαφορετικά χαρακτηριστικά των φυσικών αντικειμένων.

Το πρώτο πρόβλημα που συναντάμε είναι να βρούμε ποιές από τις μετρήσεις που μπορούμε να λάβουμε συνισφέρουν πληροφορίες για τον διαχωρισμό των κατηγοριών των προτύπων. Συσχετιζόμενο είναι και τα προβλήματα οικονομικών περιορισμών που τίθενται ως εξής. Απο τις μετρήσεις των αντικειμένων που μπορούμε να λάβουμε θα θέλαμε να εντοπίσουμε εκείνες που μεγιστοποιούν την διαχωρισιμότητα των προτύπων των κατηγοριών και που ικανοποιούν τους περιορισμούς κόστους.

Το δεύτερο πρόβλημα συντάται στην απόκριση του συστήματος ταξινόμησης. Η απόκριση εξαρτάται σημαντικά από τον αριθμό των παραμέτρων του πρότυπου. Οσο περισσότερες είναι οι διαστάσεις του διανύσματος του πρότυπου τόσο μεγαλύτερος είναι και ο χρόνος απόκρισης του συστήματος ταξινόμησης προτύπων. Στις περιπτώσεις που αντιμετωπίζουμε περιορισμούς στον αριθμό των υπολογισμών που θα εκτελεί το σύστημα ταξινόμησης θα θέλουμε μαι βαθμίδα προεπεξεργασίας να συμπιέζει τα δεδομένα (να ελαττώνει τον αριθμό των διαστάσεων του πρότυπου) έτσι ώστε να ικανοποιηθούν τελικά και οι περιορισμοί σε υπολογιστική ισχύ του συστήματος. Μια καλή λύση θα ήταν ο μετασχηματισμός αυτός να ελαχιστοποιεί και την μείωση της πληροφορίας ως προς την διαχωρισιμότητα των κατηγοριών που επιφέρει η τυπική μείωση της πληροφοριακής ροής στην είσοδο του συστήματος ταξινόμησης προτύπων.

Η σωστή επιλογή των χαρακτηριστικών γνωρισμάτων από τα αντικείμενα που πρέπει να μετρήσουμε είναι κρίσιμη για την αξιοπιστία του συστήματος ταξινόμησης που θέλουμε να φτιάξουμε. Η επιλογή αυτή εξαρτάται σημαντικά και απο το είδος των κατηγοριών που θέλουμε να αναγνωρίσουμε, γεγονός που σημαίνει ότι πριν από την σχεδίαση κάθε συστήματος ταξινόμησης θα πρέπει να βρούμε πιό είναι το καταλληλότερο σύνολο παραμέτρων που πρέπει να προσδιοριστούν από τα πρότυπα έτσι ώστε να βελτιστοποιείται η αξιοπιστία της ταξινόμησης.

Η στενή σχέση ανάμεσα στο είδος των κατηγοριών που το σύστημα ταξινόμησης διαχωρίζει και το παραμετρικό διάνυσμα των προτύπων γίνεται φανερή από το το εξής απλό παράδειγμα: Εστω ότι μας έχει ζητηθεί να κατασκευάσουμε ένα σύστημα αυτόματης διαλογής ποιότητας πορτοκαλιών. Ο διαχωρισμός πρέπει να γίνει σε τέσσερις κατηγορίες:

α) Στα πορτοκάλια που πρέπει να καταστραφούν και είναι αυτά τα οποία έχουν σαπίσει κατά 50 % τουλάχιστον ανεξάρτητα μεγέθους,

β) Στα πορτοκάλια χυμοποίησης τα οποία πρέπει να έχουν διάμετρο μέχρι 7 εκατοστά και να μην παρουσιάζουν ανάπτυξη μικροοργανισμών μεγαλύτερη του 15 %,

γ) Στα ποτοκάλια εξαγωγής τα οποία δεν πρέπει να παρουσιάζουν ανάπτυξη μικροοργανισμών και να έχουν διάμετρο μεγαλύτερη των 10 εκατοστών,

δ) Οσα δεν εμπίπτουν στις παραπάνω κατηγορίες προορίζονται για την εγχώρια αγορά.

Οι αισθητήρες που είναι διαθέσιμοι μπορούν να μετρήσουν σε κάθε πορτοκάλι όγκο, χρώμα, βάρος, πυκνότητα, υγρασία και σκληρότητα της επιφάνειας.

Απο το είδος του διαχωρισμού που πρέπει να κάνουμε γίνεται φανερό ότι θα πρέπει να αναζητήσουμε τρόπους έτσι ώστε να διακρίνουμε τα αντικείμενα από τον όγκο, το χρώμα την υγρασία και την σκληρότητα της επιφάνειας.

6.2 Στατιστική προεπεξεργασία προτύπων

Μια απλή και πρακτική μέθοδος προεπεξεργασίας και επιλογής παραμέτρων μπορεί να επιτευχθεί με στατιστική επεξερργασία των μετρήσεων. Τα προτερήματα αυτής της μεθόδου είναι ότι δεν χρειάζεται να γνωρίζουμε πληροφορίες για την στοχαστική συμπεριφορά των προτύπων των κατηγοριών. Η στατιστική προεπεξεργασία αντλεί τα αποτελέσματά της με μεθόδους επεξεργασίας των παραδειγμάτων εκπαίδευσης.

Ορίζουμε σαν αναμενόμενη απόσταση των Μ προτύπων μιας κατηγορίας από τυχόν πρότυπο **x** το ακόλουθο μέγεθος:

$$D(\mathbf{x},\mathbf{X}) = \frac{1}{M}\sum_{j=1}^{M} d^2(\mathbf{x},\mathbf{x}_j) \tag{6.1}$$

όπου $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M)$.

Αν υπολογίσουμε την μέση τιμή της αναμενόμενης απόστασης των Μ προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά της θα έχουμε το ακόλουθο μέγεθος:

$$D(\mathbf{X}) = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{M-1}\sum_{i=1}^{M-1} d^2(\mathbf{x}_j,\mathbf{x}_i) \tag{6.2}$$

$$D(\mathbf{X}) = \frac{1}{M(M-1)}\sum_{j=1}^{M}\sum_{i=1}^{M-1} d^2(\mathbf{x}_j,\mathbf{x}_i) \tag{6.3}$$

Αν ορίσουμε σαν συνάρτηση απόστασης την Ευκλείδια (στον χώρο των $\mathbb{R}^N$) τότε μπορούμε να εκφράσουμε την απόσταση $D(\mathbf{X})$ σαν συνάρτηση της διασποράς κάθε μιας των παραμέτρων του διανύσματος του προτύπου, ως εξής:

$$D(\mathbf{X}) = \frac{M}{M-1}\sum_{n=1}^{N}\left(\frac{1}{M^2}\sum_{j=1}^{M}\sum_{i=1}^{M}(\mathbf{x_j}^{(n)} - \mathbf{x_i}^{(n)})^2\right) \tag{6.4}$$

$$D(\mathbf{X}) = \frac{M}{M-1}\sum_{n=1}^{N}\left(\frac{1}{M^2}\sum_{j=1}^{M}\sum_{i=1}^{M}(\mathbf{x_j}^{(n)})^2 - \frac{2}{M^2}\sum_{j=1}^{M}\sum_{i=1}^{M}\mathbf{x_j}^{(n)}\mathbf{x_i}^{(n)} + \frac{1}{M^2}\sum_{j=1}^{M}\sum_{i=1}^{M}(\mathbf{x_i}^{(n)})^2\right) \tag{6.5}$$

Βλέπουμε ότι μερικά αθροίσματα λειτουργούν ως πολλαπλασιαστές του σταθερού αριθμού M:

$$D(\mathbf{X}) = \frac{M}{M-1}\sum_{n=1}^{N}\left(\frac{M}{M^2}\sum_{j=1}^{M}M(\mathbf{x_j}^{(n)})^2 - 2\frac{1}{M}\sum_{j=1}^{M}\mathbf{x_j}^{(n)}\frac{1}{M}\sum_{i=1}^{M}\mathbf{x_i}^{(n)} + \frac{M}{M^2}\sum_{i=1}^{M}(\mathbf{x_i}^{(n)})^2\right) \tag{6.6}$$

$$D(\mathbf{X}) = \frac{M}{M-1}\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{j=1}^{M}(\mathbf{x_j}^{(n)})^2 - 2\frac{1}{M}\sum_{j=1}^{M}\mathbf{x_j}^{(n)}\frac{1}{M}\sum_{i=1}^{M}\mathbf{x_i}^{(n)} + \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x_i}^{(n)})^2\right) \tag{6.7}$$

$$D(\mathbf{X}) = \frac{2M}{M-1}\sum_{n=1}^{N}\left(\frac{1}{M}\sum_{j=1}^{M}(\mathbf{x_j}^{(n)})^2 - \left(\frac{1}{M}\sum_{i=1}^{M}(\mathbf{x_i}^{(n)})\right)\right) \tag{6.8}$$

$$D(\mathbf{X}) = \frac{2M}{M-1}\sum_{n=1}^{N}{\sigma_n}^2 \tag{6.9}$$

όπου ${\sigma_n}^2$ είναι η διασπορά των μετρήσεων της n παραμέτρου του διανύσματος του προτύπου.

Γνωρίζουμε ότι η πραγματική στατιστική διασπορά συνδέεται με την τυπική διασπορά με την σχέση:

$${s_n}^2 = \frac{M}{M-1}{\sigma_n}^2 \tag{6.10}$$

Αντικαθιστώντας στην προηγούμενη σχέση, έχω:

$$D(\mathbf{X}) = 2\sum_{n=1}^{N}{s_n}^2 \tag{6.11}$$

Αποδείξαμε λοιπόν ότι στην περίπτωση που συνάρτηση απόστασης είναι η Ευκλείδια τότε η μέση τιμή της αναμενόμενης απόστασης των M προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά της είναι το διπλάσιο του αθροίσματος των επιμέρους πραγματικών διασπορών του παραμέτρων του διανύσματος του πρότυπου για τα παραδείγματα εκπαίδευσης.

Ας δούμε λοιπόν τι θα συμβεί αν υποθέσουμε ότι μετασχηματίζουμε γραμμικά τα πρότυπα και μάλιστα χρησιμοποιούμε έναν σταθερό πολλαπλασιαστικό παράγοντα για κάθε παράμετρο του προτύπου.

Η Ευκλείδια απόσταση των μετασχηματισμένων προτύπων θα είναι:

$$d(\mathbf{x}^*, \mathbf{y}^*) = \sum_{n=1}^{N} (c_n x_n - c_n y_n)^2 \qquad \Leftrightarrow$$

$$d(\mathbf{x}^*, \mathbf{y}^*) = \sum_{n=1}^{N} w_n (x_n - y_n)^2 \tag{6.12}$$

Ευχολα αποδεικνύεται ότι η αναμενόμενη απόσταση των M προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά της δίνεται από την σχέση:

$$D(\mathbf{X}) = 2 \sum_{n=1}^{N} (w_n s_n)^2 \tag{6.13}$$

Σε αυτό το σημείο μπορούμε να θέσουμε ένα κριτήριο για τον βέλτιστο υπολογισμό των συντελεστών του διανύσματος $\mathbf{w} = (w_1, w_2, ..., w_N)$.

Η αναμενόμενη απόσταση των προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά της μας προσφέρει ένα ποσοτικό μέγεθος της διασποράς των προτύπων των παραδειγμάτων στο χώρο των μετρήσεων. Βολικό θα ήταν λοιπόν να ορίσουμε με τέτοιο τρόπο το διάνυσμα $\mathbf{w}$ έτσι ώστε μετά τον γραμμικό μετασχηματισμό των μετρήσεων η αναμενόμενη απόσταση των προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά να είναι ελάχιστη. Με αυτό τον τρόπο μπορούμε να εξασφαλίσουμε ότι τουλάχιστον τα διαθέσιμα παραδείγματα θα συσσωρευτούν σε μικρότερο χώρο. Αυτό το γεγονός συνισφέρει συνήθως στην διαχωρισιμότητα των προτύπων των κατηγοριών, αλλά δεν υπάρχει απόδειξη που να το εξασφαλίζει.

Η ελαχιστοποίηση της αναμενόμενης απόστασης των προτύπων της κατηγορίας από τα αντίστοιχα πρότυπά της μπορεί να πραγματοποιηθεί με την προσθήχη μία επιπλέον συνθήκης διότι η προφανής λύση $\mathbf{w} = \mathbf{0}$ η οποία προχαλεί μία σημειακή συσσώρευση των παραδειγμάτων δεν προσφέρεται για πρακτική εκμετάλλευση.

Δύο είναι οι σημαντικότεροι και απλούστεροι περιορισμοί που οδηγούν σε αναλυτική λύση.

1. *Ο αριθμητικός μέσος των συντελεστών του* $\mathbf{w}$ *είναι σταθερός.* Ας επιλύσουμε το πρόβλημα ελαχιστοποίησης για τον ακόλουθο περιορισμό:

$$\frac{1}{N} \sum_{n=1}^{N} w_n = \frac{1}{N} \tag{6.14}$$

Χρησιμποιώντας του τελεστές Lagrange το πρόβλημα ελαχιστοποίησης της ποσότητας $D(\mathbf{X})$ μετατρέπεται σε πρόβλημα ελαχιστοποίησης της ποσότητας:

$$D(\mathbf{X}) - \rho \left(\sum_{n=1}^{N} w_n - 1 \right) \tag{6.15}$$

ή ισοδύναμα

$$\sum_{n=1}^{N} (w_n s_n)^2 - \rho \left(\sum_{n=1}^{N} w_n - 1 \right) \tag{6.16}$$

Παραγωγίζοντας ως προς w_n και λύνοντας τις ακόλουθες N εξισώσεις έχουμε:

$$w_n = \frac{\rho}{4{s_n}^2} \tag{6.17}$$

Αθροίζοντας τις N εξισώσεις λαμβάνουμε:

$$\sum_{n=1}^{N} w_n = \rho \sum_{n=1}^{N} \frac{1}{4{s_n}^2} \tag{6.18}$$

Λύνοντας ως πρός ρ έχουμε την ακόλουθη τιμή για τον πολλαπλασιαστή του Lagrange:

$$\rho = \frac{4}{\sum\limits_{n=1}^{N} {s_n}^{-2}} \tag{6.19}$$

Αντικαθιστώντας στην λύση του προβλήματος έχουμε την ακόλουθη αναλυτική λύση:

$$w_n = \frac{\frac{1}{{s_n}^2}}{\sum\limits_{n=1}^{N} \frac{1}{{s_n}^2}} \tag{6.20}$$

Απο την αναλυτική μορφή της λύσης μπορούμε να δούμε ότι ο συντελεστής του γραμμικού μετασχηματισμού είναι αντιστρόφος ανάλογος της διασποράς της παραμέτρου στα παραδείγματα εκπαίδευσης. Οταν η διασπορά των μετρήσεων είναι μεγάλη ο συντελεστής γίνεται μικρός οπότε και συνισφορά αυτής της παραμέτρου στο συνολικό υπολογισμό της Ευκλείδιας απόστασης είναι μικρότερη από αυτήν που υπολογίζεται πριν τον γραμμικό μετασχηματισμό.

Τα συμπεράσματα αυτά είναι χρήσιμα και για την διαδικασία επιλογής μετρήσεων. Είναι φυσικό όταν τεθεί θέμα ελάττωσης των παραμέτρων του πρότυπου να κρατήσουμε μονάχα εκείνες που έχουν την μικρότερη διασπορά.

2. *Ο γεωμετρικός μέσος των συντελεστών του* $\mathbf{w}$ *είναι σταθερός.* Ας επιλύσουμε το ίδιο πρόβλημα ελαχιστοποίησης για τον ακόλουθο περιορισμό:

$$\prod_{n=1}^{N} w_n = 1 \tag{6.21}$$

Χρησιμποιώντας του τελεστές Lagrange το πρόβλημα ελαχιστοποίησης της ποσότητας $D(\mathbf{X})$ μετατρέπεται σε πρόβλημα ελαχιστοποίησης της ποσότητας:

$$\sum_{n=1}^{N} (w_n s_n)^2 - \rho \left(\prod_{n=1}^{N} w_n - 1 \right) \tag{6.22}$$

Παραγωγίζοντας ως προς w_n και λύνοντας τις ακόλουθες N εξισώσεις έχουμε:

$$w_n = \frac{\sqrt{\rho}}{2s_n} \tag{6.23}$$

Πολλαπλασιάζοντας τις N εξισώσεις λαμβάνουμε:

$$\prod_{n=1}^{N} w_n = \frac{\rho^N}{\prod_{n=1}^{N} 2s_n} \tag{6.24}$$

Λύνοντας ως πρός ρ έχουμε την ακόλουθη τιμή για τον πολλαπλασιαστή του Lagrange:

$$\rho = 4 \left(\prod_{n=1}^{N} s_n \right)^{2/N} \tag{6.25}$$

Αντικαθιστώντας στην λύση του προβλήματος έχουμε την ακόλουθη αναλυτική λύση:

$$w_n = \frac{1}{s_n} \left(\prod_{n=1}^{N} s_n \right) 1/N \tag{6.26}$$

Απο την αναλυτική μορφή της λύσης μπορούμε να δούμε ότι ο συντελεστής του γραμμικού μετασχηματισμού είναι αντιστρόφος ανάλογος της τυπικής απόκλισης της παραμέτρου στα παραδείγματα εκπαίδευσης. Συνεπώς ισχύουν τα ίδια συμπεράσματα που αναφέραμε στην προηγούμενη λύση.

Προσοχή θα πρεπει να δοθεί στο γεγονός ότι τα παραπάνω συμπεράσματα προκύπτουν όταν σαν συνάρτηση απόστασης χρησιμοποιείται η Ευκλείδια συνάρτηση. Συνεπώς στην περίπτωση που χρησιμοποιηθεί άλλη συνάρτηση απόστασης η ανάλυση που έγινε πρέπει να επαναληφθεί με την νέα συνάρτηση απόστασης.

Παράδειγμα 72 *Υποθέστε ότι έχουμε σύστημα ταξινόμησης T κατηγοριών και διαθέτουμε έναν αριθμό παραδειγμάτων για κάθε κατηγορία. Υπολογίστε τους συντελεστές γραμμικού μετασχηματισμού των προτύπων όταν χρησιμοποιούμε την Ευκλείδια συνάρτηση απόστασης με κριτήριο την ελαχιστοποίηση της αναμενόμενης απόστασης των προτύπων κάθε κατηγορίας από τα αντίστοιχα πρότυπά της.*

Η ποσότητα που θέλουμε να ελαχιστοποιήσουμε είναι η:

$$D = \sum_{t=1}^{T} D(\mathbf{X}_t)$$

Ακολουθούμε την ίδια πορεία για να επιτύχουμε την λύση του προβλήματος. Χρησιμποιώντας του τελεστές Lagrange το πρόβλημα ελαχιστοποίησης της ποσότητας $D(\mathbf{X})$ μετατρέπεται σε πρόβλημα ελαχιστοποίησης της ποσότητας:

$$\sum_{t=1}^{T} D(\mathbf{X}_t) - \rho \left(\sum_{n=1}^{N} w_n - 1 \right) \tag{6.27}$$

ή ισοδύναμα

$$\sum_{t=1}^{T}\sum_{n=1}^{N}(w_n s_n{}^{(t)})^2 - \rho\left(\sum_{n=1}^{N} w_n - 1\right) \tag{6.28}$$

Παραγωγίζοντας ως προς w_n και λύνοντας τις ακόλουθες N εξισώσεις έχουμε:

$$w_n = \frac{\rho}{4\sum_{t=1}^{T}(s_n{}^{(t)})^2} \tag{6.29}$$

Υπολογίζουμε τον πολλαπλασιαστή του Lagrange:

$$\rho = \frac{4}{\sum_{t=1}^{T}\sum_{n=1}^{N}(s_n{}^{(t)})^{-2}} \tag{6.30}$$

Αντικαθιστώντας στην λύση του προβλήματος έχουμε την ακόλουθη αναλυτική λύση:

$$w_n = \frac{\sum_{t=1}^{T}\frac{1}{s_n{}^2}}{\sum_{t=1}^{T}\sum_{n=1}^{N}\frac{1}{s_n{}^2}} \tag{6.31}$$

Παράδειγμα 73 *Θεωρώντας ότι ο γεωμετρικός μέσος των συντελεστών του* **w** *είναι σταθερός λύστε το προηγούμενο πρόβλημα.*

6.3 Επιλογή παραμέτρων με ελαχιστοποίηση της εντροπίας

Το πρόβλημα επιλογής του καταλληλότερου γνήσιου υποσυνόλου των μετρήσεων ενός συνόλου φυσικών αντικειμένων τα οποία θέλουμε με αυτόματο τρόπο να αναγνωρίζουμε είναι το πρώτο πρόβλημα που έχει να λύσει ο σχαδιαστής μιας μηχανής ταξινόμησης προτύπων.

Εχουν προταθεί πολλές λύσεις στο πρόβλημα της βέλτισης επιλογής των παραμέτρων. Στην συνέχεια αυτού του κεφαλαίου θα περιγράψουμε τις σημαντικότερες από αυτές.

Το κριτήριο επιλογής παραμέτρων με ελαχιστοποίηση της εντροπίας των προτύπων που προκύπτουν είναι ένα από τα αποτελεσματικότερα και συχνότερα χρησιμοποιούμενα κριτήρια συμπίεσης των διαστάσεων του πρότυπου. Το κριτήριο αυτό υπολογίζει τους συντελεστές γραμμικού μετασχηματισμού του πρότυπου με τέτοιο τρόπο έτσι ώστε να ελαχιστοποιείται η αβεβαιότητα ως προς την διαχωρισιμότητα των προτύπων των κατηγοριών.

Η εντροπία των παραδιεγμάτων της κατηγορίας ω_i δίνεται από την σχέση:

$$H_i = -\int_{\mathbf{x}} p(\mathbf{x}|\omega_i)log(p(\mathbf{x}|\omega_i))d\mathbf{x} \tag{6.32}$$

Αν μετασχηματίσουμε γραμμικά το πρότυπο $\mathbf{x}$:

$$\mathbf{y} = \mathbf{Q}\mathbf{x} \tag{6.33}$$

τότε η αναμενόμενη τιμή του διανύσματος $\mathbf{y}$ γίνεται:

$$E(\mathbf{y}) = E(\mathbf{Qx}) = \mathbf{Q}E(\mathbf{x}) \quad \Rightarrow \quad \mathbf{m}_y = \mathbf{Qm}_x \tag{6.34}$$

Η διασπορά υπολογίζεται ως εξής:

$$s(\mathbf{y}) = E((\mathbf{y}-\mathbf{m}_y)(\mathbf{y}-\mathbf{m}_y)^T) = \mathbf{Q}E((\mathbf{x}-\mathbf{m}_x)(\mathbf{x}-\mathbf{m}_x)^T)\mathbf{Q}^T = \mathbf{Qs}_x\mathbf{Q}^T \tag{6.35}$$

Αν υποθέσουμε ότι η πυκνότητα πιθανότητας του πρότυπου $\mathbf{x}$ είναι κανονική κατανομή:

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}_x|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_x)^T \mathbf{s}_\mathbf{x}^{-1}(\mathbf{x}-\mathbf{m}_x)} \tag{6.36}$$

τότε η πυκνότητα πιθανότητας της διανυσματικής μεταβλητής $\mathbf{y}$ θα είναι επίσης κανονική κατανομή με αναμενόμενη τιμή $\mathbf{m}_y$ και διασπορά $\mathbf{s}_y$:

$$p(\mathbf{y}|\omega_i) = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{s}_y|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{m}_y)^T (\mathbf{s}_\mathbf{y})^{-1}(\mathbf{y}-\mathbf{m}_y)} \tag{6.37}$$

Αντικαθιστώντας τις αντίστοιχες τιμές της μέσης τιμής και του πίνακα της διασποράς έχουμε την ακόλουθη κατανομή της $\mathbf{y}$ σαν συνάρτηση του πίνακα γραμμικού μετασχηματισμού $\mathbf{Q}$:

$$p(\mathbf{y}|\omega_i) = \frac{1}{(2\pi)^{N/2}\sqrt{|\mathbf{Qs}_x\mathbf{Q}^T|}} e^{-\frac{1}{2}\mathbf{Q}(\mathbf{x}-\mathbf{m}_x)^T (\mathbf{Qs}_x\mathbf{Q}^T)^{-1}\mathbf{Q}(\mathbf{x}-\mathbf{m}_x)} \tag{6.38}$$

Η εντροπία του νέου πρότυπου $\mathbf{y}$ είναι:

$$H_i{}^* = -\int_{\mathbf{y}} p(\mathbf{y}|\omega_i)log(p(\mathbf{y}|\omega_i))d\mathbf{y} \tag{6.39}$$

Ο βέλτιστος πίνακας $\mathbf{Q}$ μπορεί να υπολογιστεί σαν το ελάχιστο της συνάρτησης της εντροπίας της διανυσματικής μεταβλητής $\mathbf{y}$. Η επίλυση της ακόλουθης διανυσματικής εξίσωσης οδηγεί στον βέλτιστο πίνακα $\mathbf{Q}$:

$$\frac{\partial}{\partial \mathbf{Q}} H_i{}^* = \mathbf{0} \tag{6.40}$$

Η επίλυση της εξίσωσης είναι δυνατή αλλά παραλείπεται διότι καλύπτει μεγάλο αριθμό σελίδων. Τα συμπεράσματα της επίλυσης είναι τα ακόλουθα:

Ο πίνακας γραμμικού μετασχηματισμού $\mathbf{Q}$ των μετρήσεων προκύπτει με την τοποθέτηση των k κανονικοποιημένων ιδιοδιανυσμάτων του πίνακα διασποράς $\mathbf{s}_x$ που έχουν τις αντίστοιχες μικρότερες ιδιοτιμές.

Συνεπώς αν θέλουμε να επιλέξουμε τις βέλτιστες παραμέτρους των προτύπων με το κριτήριο ελαχιστοποίησης της εντροπίας πρέπει να υπολογίσουμε τις ιδιοτιμές του πίνακα των διασπορών των παραμέτρων του πρότυπου και κατόπιν να επιλέξουμε εκείνες τις παραμέτρους που έχουν τις μικρότερες ιδιοτιμές.

Αν θέλουμε να συμπιέσουμε το διάνυσμα του πρότυπου τότε πρέπει να φτιάξουμε έναν πίνακα $\mathbf{Q}$ ο οποίος θα αποτελείται από τα κανονικοποιημένα ιδιοδιανύσματα των μικρότερων ιδιοτιμών.

Παράδειγμα 74 *Εστω ότι θέλουμε να φτιάξουμε σύστημα ταξινόμησης προτύπων δύο κατηγοριών και έχουμε στην διάθεσή μας τα εξής πρότυπα:*

$$\Omega_1 = \{(1,1,0)^T, (1,0,0)^T, (1,0,1)^T, (0,0,0)^T\}$$

$$\Omega_2 = \{(0,0,1)^T, (0,1,1)^T, (0,1,0)^T, (1,1,1)^T\}$$

Μετασχηματίστε τα πρότυπα των παραδειγμάτων σε διανύσματα δύο διαστάσεων με κριτήριο την ελαχιστοποίηση της εντροπίας.

Αρχικά υπολογίζουμε την αναμενόμενη τιμή και την διασπορά των προτύπων κάθε κατηγορίας ξεχωριστά:

$$\mathbf{m}_1 = 0.25(3,1,1)^T \qquad \mathbf{m}_2 = 0.25(1,3,3)^T$$

Η διασπορά των προτύπων τυχαίνει να είναι η ίδια και για τις δύο κατηγορίες:

$$\mathbf{s} = \mathbf{s}_1 = \mathbf{s}_2 = 0.0625 \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 3 & -1 & 3 \end{pmatrix}$$

Οι ιδιοτιμές του διανύσματος $\mathbf{s}$ προκύπτουν από την λύση της ακόλουθης εξίσωσης:

$$|\mathbf{s} - \lambda \mathbf{I}| = 0$$

Οι λύσεις της τριτοβάθμιας εξίσωσης που προκύπτει είναι:

$$\lambda_1 = \lambda_2 = 0.25, \qquad \lambda_3 = 0.0625$$

Τα ιδιοδιανύσματα $\mathbf{x}$ προκύπτουν από την λύση της ακόλουθης εξίσωσης:

$$(\mathbf{s} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

$$x_1 = (0.816497, 0.408248, 0.408248)^T$$

$$x_2 = (-0.816494, -0.410015, -0.406479)^T$$

$$x_3 = (0.57735, -0.57735, -0.57735)^T$$

Διαλέγουμε τα δύο ιδιοδιανύσματα που έχουν τις μικρότερες ιδιοτιμές. Το πρώτο επιλέγεται εύκολα, είναι το x_3 διότι αντιστοιχεί στο ιδιοδιάνυσμα με την μικρότερη αριθμητική τιμή ($\lambda_3 = 0.0625$). Το δεύτερο επιλέγετει τυχαία διότι οι αμέσως μεγαλύτερες ιδιοτιμές είναι ίσες.

Αν ορίσουμε τον πίνακα μετασχηματισμού $\mathbf{Q}$ σαν:

$$\mathbf{Q} = \begin{pmatrix} -0.816494 & -0.410015 & -0.406479 \\ 0.57735 & -0.57735 & -0.57735 \end{pmatrix}$$

τότε τα νέα διανύσματα δύο διαστάσεων των προτύπων προκύπτουν από την σχέση $\mathbf{y} = \mathbf{Q}\mathbf{x}$:

$$\Omega_1{}^* = \{(-1.622, 0)^T, (-0.816, 0.577)^T, (-1.222, 0)^T, (0,0)^T\}$$

$$\Omega_2{}^* = \{(-0.406, -0.577)^T, (-0.816, -1.154)^T, (-0.41, -0.577)^T, (-1.632, -0.577)^T\}$$

6.4 Διανυσματική επέκταση των Karhunen-Leone

Με την μέθοδο διανυσματικής επέκτασης των Karhunen-Leone προσπαθούμε να επιλύσουμε το αντίστροφο πρόβλημα. Θεωρούμε ότι τα πρότυπα $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_N)$ που είναι διαθέσιμα από τα παραδείγματα προκύπτουν από εναν γραμμικό μετασχηματισμό:

$$\mathbf{X} = \mathbf{\Phi C} = \sum_{j=1}^{N} \mathbf{c}_j \varphi_j \tag{6.41}$$

όπου φ_j είναι ορθοκανονικά διανύσματα. Λύνοντας ως προς την μεταβλητή $\mathbf{c}$:

$$\mathbf{c} = \mathbf{\Phi}^T \mathbf{x} \tag{6.42}$$

Απο την συνθήκη ορθοκανονικότητας γνωρίζουμε ότι:

$$\mathbf{c}_i = \mathbf{\Phi_i}^T \mathbf{X} \tag{6.43}$$

Υποθέτουμε ότι είμαστε σε θέση να υπολογίσουμε μόνο τα Μ (N¿M) διανύσματα $\mathbf{c}_i, i = 1, M$ αλλά παρόλα αυτά ζητάμε να έχουμε μία καλή εκτίμηση του διανύσματος $\mathbf{X}$. Υποθέτουμε τότε ότι μία εκτίμηση του $\mathbf{X}$ δίνεται από την σχέση:

$$\hat{\mathbf{X}} = \sum_{j=1}^{M} \mathbf{c}_j \varphi_j + \sum_{j=M+1}^{N} \mathbf{o}_j \varphi_j \tag{6.44}$$

Το σφάλμα της προσέγγισης είναι:

$$\mathbf{X} - \hat{\mathbf{X}} = + \sum_{j=M+1}^{N} (\mathbf{c}_j - \mathbf{o}_j) \varphi_j \tag{6.45}$$

Ο στατιστική αναμενόμενη τιμή του τετραγώνου του σφάλματος είναι:

$$E^2 = E(|\mathbf{X} - \hat{\mathbf{X}}|^2) = E(\sum_{j=M+1}^{N} \sum_{i=M+1}^{N} (\mathbf{c}_j - \mathbf{o}_j)(\mathbf{c}_i - \mathbf{o}_i) \varphi_j \varphi_i) \tag{6.46}$$

Λόγω της ορθοκανονικότητας των διανυσμάτων φ_j η παραπάνω σχέση απλοποιείται και γίνεται ισοδύναμα:

$$E^2 = \sum_{j=M+1}^{N} E((\mathbf{c}_j - \mathbf{o}_j)^2) \tag{6.47}$$

Το πρόβλημα που καλούμαστε να λύσουμε σε αυτή την περίπτωση είναι να βρούμε τα διανύσματα $\mathbf{o}_j$ τα οποία ελαχιστοποιούν το σφάλμα της προσέγγισης. Με την συνιθισμένη μας πια τακτική υπολογίζουμε το ελάχιστο της παράστασης του σφάλματος με την εύρεση των ριζών της πρώτης παραγώγου:

$$\frac{\partial}{\partial \mathbf{c}_j} E((\mathbf{c}_j - \mathbf{o}_j)^2) = -2E(\mathbf{c}_j) - \mathbf{o}_j = 0 \quad \Rightarrow \tag{6.48}$$

$$\mathbf{o}_j = E(\mathbf{c}_j) = \mathbf{\Phi_i}^T E(\mathbf{X}) \tag{6.49}$$

Αντικαθιστώντας την λύση που βρήκαμε στην εξίσωση του σφάλματος έχουμε:

$$E^2 = \sum_{j=M+1}^{N} E((\mathbf{c}_j - E(\mathbf{c}_j))^2) = \sum_{j=M+1}^{N} \mathbf{\Phi_i}^T E((\mathbf{X} - E(\mathbf{X}))E((\mathbf{X} - E(\mathbf{X})))^T \mathbf{\Phi}_i \tag{6.50}$$

$$E^2 = \sum_{j=M+1}^{N} \mathbf{\Phi_i}^T \mathbf{s}_X \mathbf{\Phi}_i \tag{6.51}$$

όπου $\mathbf{s}_X$ είναι ο πίνακας διασπορών του διανύσματος $\mathbf{X}$.

Αν λ_i είναι τα ιδιοδιανύσματα του πίνακα των διασπορών τότε γνωρίζουμε ότι ισχύει:

$$\mathbf{s}_X \mathbf{\Phi}_i = \lambda_i \mathbf{\Phi}_i \tag{6.52}$$

Αντικαθιστώντας στην εξίσωση του σφάλματος έχουμε:

$$E^2 = \sum_{j=M+1}^{N} \mathbf{\Phi_i}^T \lambda_i \mathbf{\Phi}_i = \sum_{j=M+1}^{N} \lambda_i \mathbf{\Phi}_i^T \mathbf{\Phi}_i = \sum_{j=M+1}^{N} \lambda_i \tag{6.53}$$

Η τελευταία ισότητα ισχύει διότι γνωρίζουμε ότι τα διανύσματα $\mathbf{\Phi}_i$ είναι ορθοκανονικά.

Απο τα αποτελέσματα αυτά βλέπουμε λοιπόν ότι τα σφάλμα της προσέγγισης εξαρτάται απο το άθροισμα των ιδιοτιμών των παραμέτρων που αφαιρούνται από την προσέγγιση. Συνεπώς το σφάλμα της προσέγγισης ελαχιστοποιείται όταν από τον πίνακα διασπορών του $\mathbf{X}$ αφαιρέσουμε εκείνες τις παραμέτρους που έχουν τις μικρότερες ιδιοτιμές.

Τα συμπεράσματα που βγάλαμε είναι αντίθετα από την προσέγγιση που επιχειρήσαμε στο προηγούμενο τμήμα με την προσπάθεια ελαχιστοποίσης της εντροπίας των νέων προτύπων και για την περίπτωση κατά την οποία τα πρότυπα ακολουθούν την κανονική κατανομή.

Το πλεονέκτημα της στατιστικής προσέγγισης που δίνεται με την διανυσματική επέκταση των Karhunen-Leone είναι ότι η μέθοδος αυτή έχει γενικότερη εφαρμογή διότι δεν θέτει καμμία προυπόθεση για την στατιστική συμπεριφορά των προτύπων.

Συνοψίζοντας τα αποτελέσματα της μεθόδου διανυσματικής επέκτασης των Karhunen-Leone δίνουμε μους υπολογισμούς που πρέπει να εκτελέσουμε έτσι ώστε να υπολογίσουμε τα νέα μειωμένων παραμέτρων πρότυπα.

1. *Υπολογισμός του πίνακα διασπορών.* Στα διαθέσιμα πρότυπα υπολογίζουμε τον πίνακα διασπορών.

$$\mathbf{s}_X = E((\mathbf{X} - E(\mathbf{X}))E((\mathbf{X} - E(\mathbf{X})))^T \tag{6.54}$$

2. *Υπολογισμός των ιδιοτιμών του πίνακα διασπορών.* Μετά τον υπολογισμό των ιδιοτιμών του πίνακα διασπορών υπολογίζουμε τα ιδιοδιανύσματα που αντιστοιχούν στις μεγαλύτερες ιδιοτιμές και με αυτά κατασκευάζουμε τον πίνακα $\mathbf{\Phi}$:

$$|\mathbf{s}_X - \lambda_i \mathbf{I}| = 0 \tag{6.55}$$

$$\lambda_1 > \lambda_2 > ... > \lambda_N \tag{6.56}$$

$$(\mathbf{s}_X - \lambda_i \mathbf{I})\varphi_i = 0, \qquad i = 1, M \qquad M < N \tag{6.57}$$

$$\mathbf{\Phi} = (\varphi_1, \varphi_1, ..., \varphi_M) \tag{6.58}$$

3. *Αναστρέφουμε τον πίνακα* $\mathbf{\Phi}$.

$$\mathbf{\Psi} = \mathbf{\Phi}^T \tag{6.59}$$

4. *Υπολογίζουμε τα νέα πρότυπα.*

$$\mathbf{c} = \mathbf{\Psi}\mathbf{x} \tag{6.60}$$

Παράδειγμα 75 *Κατασκευάστε σύστημα ταξινόμησης προτύπων με την μέθοδο της ταξινόμησης στο κοντινότερο εικονικό πρότυπο και χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης και υπολογίστε το ελάχιστο σφάλμα ταξινόμησης του συστήματος που κατασκευάσατε με πρότυπα ίδια με αυτά που δίνονται στο προηγούμενο παράδειγμα.*

Ποιά είναι τα συμπεράσματά σας από την σύγκριση των αντίστοιχων ελάχιστων σφαλμάτων, για την μέθοδο διανυσματικής επέκτασης των Karhunen-Leone

Σε πρακτικές εφαρμογές όπου η μέθοδος της διανυσματικής επέκτασης των Karhunen-Leone εφαρμόζεται για επιλογή παραμέτρων σε συστήματα ταξινόμησης προτύπων η αναμενόμενη τιμή και ο πίνακας διασπορών των προτύπων κάθε κατηγορίας δεν είναι ο ίδιος.

Σε αυτές τις περιπτώσεις θεωρούμε ότι τα πρότυπα όλων των κατηγοριών ανήκουν σε μία ευρύτερη ομάδα και υπολογίζουμε τον πίνακα διασπορών όλων των προτύπων των παραδειγμάτων που έχουμε στην διάθεσή μας. Ετσι η αναμενόμενη τιμή και ο πίνακας διασπορών των προτύπων των παραδειγμάτων είναι οι ακόλουθοι:

$$\mathbf{m}_1 = (\frac{1}{4}, -\frac{3}{8}, \frac{3}{4})$$

$$\mathbf{s}_x = \begin{pmatrix} \frac{7}{16} & -\frac{1}{32} & -\frac{1}{16} \\ -\frac{1}{32} & \frac{31}{64} & \frac{1}{32} \\ -\frac{1}{16} & \frac{1}{32} & \frac{3}{16} \end{pmatrix}$$

Οι ιδιοτιμές του πίνακα διασπορών είναι:

$$\lambda_1 = 0.510316, \qquad \lambda_2 = 0.428064, \qquad \lambda_3 = 0.170995$$

Επιλέγουμε το κανονικοποοιημένο ιδιοδιάνυσμα που έχει την μεγαλύτερη ιδιοτιμή

$$\mathbf{e}_1 = (0.514831, -0.83801, -0.180799)^T$$

Τα νέα διανύσματα των προτύπων προκύπτουν από την σχέση:

$$\mathbf{y} = \mathbf{e_1}^T\mathbf{x} = (0.514831, -0.83801, -0.180799)\mathbf{x}$$

$$\Omega_1{}^* = \{1.35284, -0.695631, -0.503977, 0.\}$$

$$\Omega_2{}^* = \{-0.180799, 0.657211, 0.657211, 1.17204\}$$

Τα εικονικά πρωτότυπα των κατηγοριών προκύπτουν από τις αναμενόμενες τιμές των παραδειγμάτων:

$$\mu_1 = 0.0383, \qquad \mu_2 = 0.5764$$

Ταυτοποιώντας τα πρότυπα με την Ευκλείδια συνάρτηση απόστασης στο κοντινότερο εικονικό πρωτότυπο έχουμε ένα σφάλμα σε τέσσερεις ταξινομήσεις (το πρώτο πρότυπο) για την πρώτη κατηγορία και επίσης ένα σφάλμα σε τέσσερεις ταξινομήσεις (το πρώτο πρότυπο) για την δεύτερη κατηγορία.

Το συνολικό σφάλμα του συστήματος ταξινόμησης είναι 2 /8 * 100 = 25%.

6.5 Επιλογή παραμέτρων με στοχαστική προσέγγιση

Στα κεφάλαια στα οποία περιγράψαμε δομικές και στοχαστικές μεθόδους ταξινόμησης είδαμε ότι μπορούμε να προσεγγίσουμε την συνάρτηση διάκρισης με συνεχείς συναρτήσεις. Σε αυτό το τμήμα του κεφαλαίου θα περιγράψουμε μία τεχνική επιλογής παραμέτρων με την βοήθεια της μεθόδου προσέγγισης του ελάχιστου παλίνδρομης συνάρτηση με τον αλγόριθμο των Kiefer-Wolfowitz.

Εστω ότι η κατηγορία αντικειμένων ω χαρακτηρίζεται από την συνάρτηση διάκρισης $f(\mathbf{x})$.

Ας θεωρήσουμε μία προσέγγιση της $f(\mathbf{x})$ την $\hat{f}(\mathbf{x})$. Η αναμενόμενη τιμή του σφάλματος της προσέγγισης δίνεται από την σχέση:

$$e = \int_{\mathbf{x}\in\omega} (f(\mathbf{x}) - \hat{f}(\mathbf{x}))^2 p(\mathbf{x}|\omega) d\mathbf{x} \tag{6.61}$$

Υποθέτοντας ότι μπορούμε να εκφράσουμε την προσέγγιση σαν ένα γραμμικό συνδυασμό ορθογώνιων συναρτήσεων:

$$\hat{f}(\mathbf{x}) = \mathbf{c}\varphi(\mathbf{x}) \tag{6.62}$$

Συνεπώς το σφάλμα μπορεί να εκφραστεί σαν μία παλίνδρομη συνάρτηση:

$$e = \int_{\mathbf{x}\in\omega} (f(\mathbf{x}) - \mathbf{c}\varphi(\mathbf{x}))^2 p(\mathbf{x}|\omega) d\mathbf{x} \tag{6.63}$$

Το πρόβλημα που θέλουμε να λύσουμε είναι να βρούμε τους συντελεστές c_i για τους οποίους η στοχαστική συνάρτηση του σφάλματος ελαχιστοποιείται:

Γνωρίζουμε ότι μπορούμε να προσεγγίσουμε το ελάχιστο μιας στοχαστικής συνάρτησης με τον επαναληπτικό αλγόριθμο των Kiefer-Wolfowitz.

Η επαναληπτική σχέση που δίνει τους συντελεστές c_i είναι η ακόλουθη:

$$c_{k+1}{}^{(i)} = c_k{}^{(i)} + 2a_k(f(\mathbf{x}) - \mathbf{c}^{(i)}\varphi(\mathbf{x}))\varphi_i(\mathbf{x}) \tag{6.64}$$

Οπου a_k είναι η συμβολοσειρά θετικών πραγματικών αριθμών που πρέπει να ικανοποιεί τις γνωστές συνθήκες του αλγόριθμου των Kiefer-Wolfowitz.

Η μέθοδος επιλογής παραμέτρων με την μέθοδο στοχαστικής προσέγγισης ακολουθεί την εξής διαδικασία: Μειώνουμε τις παραμέτρους του διανύσματος του πρότυπου και υπολογίζουμε το αθροιστικό σφάλμα της προσέγγισης της συνάρτησης διάκρισης για όλες τις κατηγορίες (αφού προηγουμένως έχουμε υπολογίσει τους συντελεστές $\mathbf{c}$ για κάθε κατηγορία ξεχωριστά).

Οι παράμετροι του πρότυπου που χρησιμοποιήσαμε στο μικρότερο μετρούμενο αθροιστικό σφάλμα της προσέγγισης μας δίνει και τις καλύτερες παραμέτρους που πρέπει να απαρτίζουν το διάνυσμα του πρότυπου.

6.6 Επιλογή παραμέτρων και στοχαστική απόκλιση κατηγοριών

Η στοχαστική απόκλιση κατηγοριών είναι ένα μέγεθος μέτρησης της απόστασης ή ανομοιότητας μεταξύ δύο κατηγοριών.

Ορίζοντας σαν μέγεθος πληροφορίας διάκρισης της κατηγορίας ω_i από την κατηγορία ω_j το ακόλουθο μέγεθος:

$$U_{ij} = log(\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})}) \tag{6.65}$$

η αναμενόμενη τιμή της πληροφορίας διάκρισης της κατηγορίας ω_i από την κατηγορία ω_j δίνεται από το ακόλουθο μέγεθος:

$$I_{ij} = \int_{\mathbf{x}} p_i(\mathbf{x}) log(\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})}) d\mathbf{x} \tag{6.66}$$

Συμμετρικά η αναμενόμενη τιμή της πληροφορίας διάκρισης της κατηγορίας ω_j από την κατηγορία ω_i θα είναι:

$$I_{ji} = \int_{\mathbf{x}} p_j(\mathbf{x}) log(\frac{p_j(\mathbf{x})}{p_i(\mathbf{x})}) d\mathbf{x} \tag{6.67}$$

Ορίζουμε σαν στοχαστική απόκλιση κατηγοριών το άθροισμα των επιμέρους τιμών των αναμενόμενων τιμών της πληροφορίας διάκρισης:

$$Sd(i,j) = Sd(j,i) = I_{ij} + I_{ji} = \int_{\mathbf{x}} (p_i(\mathbf{x}) - p_j(\mathbf{x})) log(\frac{p_i(\mathbf{x})}{p_j(\mathbf{x})}) d\mathbf{x} \tag{6.68}$$

Οι σπουδαιότερες ιδιότητες της στοχαστικής απόκλισης κατηγοριών είναι οι ακόλουθες:

1. *Η στοχαστική απόκλιση είναι μη αρνητικός πραγματικός αριθμός.*

$$Sd(i,j) \geq 0 \tag{6.69}$$

2. *Η στοχαστική απόκλιση κάθε κατηγορίας από τον εαυτό της είναι πάντα μηδέν.*

$$Sd(i,i) = 0 \tag{6.70}$$

3. *Η στοχαστική απόκλιση είναι συμμετρική συνάρτηση.*

$$Sd(i,j) = Sd(j,i) \tag{6.71}$$

4. *Αν το διάνυσμα του πρότυπου αποτελείται από N στοχαστικά ανεξάρτητες παραμέτρους η στοχαστική απόκλιση ισούται με το άθροισμα των επιμέρους στοχαστικών αποκλίσεων των παραμέτρων.*

$$Sd(i,j,(x_1,x_2,...,x_N)^T) = \sum_{r=1}^{N} Sd(i,j,x_r) \tag{6.72}$$

5. *Η επέκταση του διανύσματος του πρότυπου με νέα παράμετρο αυξάνει την στοχαστική απόκλιση των κατηγοριών.*

$$Sd(i,j,(x_1,...,x_N)^T) \leq Sd(i,j,(x_1,...,x_N,x_{N+1})^T) \tag{6.73}$$

Με την υπόθεση ότι οι δύο κατηγορίες προτύπων ακολουθούν κανονική πυκνότητα πιθανότητας με διαφορετική μέση τιμή και διασπορά:

$$p_i(\mathbf{x}) = p(\mathbf{x}|\omega_i) = N(\mathbf{m}_i,\mathbf{s}_i), \qquad p_j(\mathbf{x}) = p(\mathbf{x}|\omega_j) = N(\mathbf{m}_j,\mathbf{s}_j) \tag{6.74}$$

Στην περίπτωση των κανονικών κατανομών πυκνότητας πιθανότητας έχουμε την ακόλουθη σχέση που μας δίνει την στοχαστική απόκλιση των κατηγοριών σαν συνάρτηση της μέσης τιμής και της διασποράς των:

$$I_{ij} = \frac{1}{2}log(\frac{\mathbf{s}_j}{\mathbf{s}_i}) + \frac{1}{2}Tr[\mathbf{s}_i(\mathbf{s}_j{}^{-1} - \mathbf{s}_i{}^{-1})] + \frac{1}{2}Tr[\mathbf{s_j}^{-1}(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T] \tag{6.75}$$

Η συνάρτηση $Tr[\mathbf{x}]$ μας δίνει το άθροισμα των στοιχείων της διαγωνίου ενός διαγώνιου πίνακα $\mathbf{x}$.

Η στοχαστική απόκλιση κατηγοριών θα είναι:

$$Sd(i,j) = \frac{1}{2}Tr[(\mathbf{s}_i - \mathbf{s}_j)(\mathbf{s}_j{}^{-1} - \mathbf{s}_i{}^{-1})] + \frac{1}{2}Tr[(\mathbf{s_i}^{-1} + \mathbf{s_j}^{-1})(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T] \tag{6.76}$$

Ας δούμε πως μπορούμε να χρησιμοποιήσουμε την ποσότητα της στοχαστικής απόκλιση των κατηγοριών στην διαδικασία συμπίεσης των διαστάσεων του διανύσματος του πρότυπου.

Εστω ότι μετασχηματίζουμε γραμμικά το πρότυπο του διανύσματος $\mathbf{x}, \mathbf{x} \in \mathbb{R}^N$ με πίνακα $\mathbf{Q}$ διαστάσεων $M \times N, M < N$. Το αποτέλεσμα θα είναι ένα διάνυσμα $\mathbf{y}, \mathbf{y} \in \mathbb{R}^M$ μικρότερων διαστάσεων:

$$\mathbf{y} = \mathbf{Q}\mathbf{x} \tag{6.77}$$

Η στοχαστική απόκλιση των κατηγοριών της νέας μεταβλητής $\mathbf{y}$ γίνεται:

$$Sd(i,j,\mathbf{y}) = \int_{\mathbf{y}} (p_i(\mathbf{y}) - p_j(\mathbf{y}))log(\frac{p_i(\mathbf{y})}{p_j(\mathbf{y})})d\mathbf{y} \tag{6.78}$$

Αν τα πρότυπα των κατηγοριών της μεταβλητής $\mathbf{x}$ ακολουθούν την κανονική κατανομή, τότε και τα πρότυπα της νέας μεταβλητής $\mathbf{y}$ θα ακολουθούν επίσης την κανονική κατανομή με αναμενόμενη τιμή και διασπορά τις ακόλουθες:

$$\mathbf{m_i}^* = \mathbf{Q}\mathbf{m}_i, \qquad \mathbf{m_j}^* = \mathbf{Q}\mathbf{m}_j \tag{6.79}$$

$$\mathbf{s_i}^* = \mathbf{Q}\mathbf{s}_i\mathbf{Q}^T, \qquad \mathbf{s_j}^* = \mathbf{Q}\mathbf{s}_j\mathbf{Q}^T \tag{6.80}$$

Η στοχαστική απόκλιση των κατηγοριών της νέας μεταβλητής $\mathbf{y}$ θα γίνει:

$$Sd(i,j,\mathbf{y}) = \frac{1}{2}Tr[(\mathbf{s}_j^*)^{-1}\mathbf{s}_i^* + (\mathbf{s}_i^*)^{-1}\mathbf{s}_j^*] - M + \frac{1}{2}Tr[((\mathbf{s}_i^*)^{-1} + (\mathbf{s}_j^*)^{-1})(\mathbf{m_i}^* - \mathbf{m_j}^*)(\mathbf{m_i}^* - \mathbf{m_j}^*)^T] \tag{6.81}$$

Επειδή η συνάρτηση $Tr[\mathbf{x}]$ του πίνακα $\mathbf{x}$ μπορεί να υπολογιστεί ισοδύναμα σαν το άθροισμα των ιδιοτιμών της γιαυτό τον λόγο η στοχαστική απόκλιση των κατηγοριών μπορεί να γραφεί ισοδύναμα:

$$Sd(i,j,\mathbf{y}) = \frac{1}{2}\sum_{i=1}^{M}(\lambda_i + \frac{1}{\lambda_i}) - M + \frac{1}{2}\lambda \tag{6.82}$$

Οπου $\lambda_i, i = 1, M$ είναι οι ιδιοτιμές της $(\mathbf{s}_j^*)^{-1}\mathbf{s}_i^* + (\mathbf{s}_i^*)^{-1}\mathbf{s}_j^*$ και λ είναι η ιδιοτιμή του συντελεστή $((\mathbf{s}_i^*)^{-1} + (\mathbf{s}_j^*)^{-1})(\mathbf{m_i}^* - \mathbf{m_j}^*)(\mathbf{m_i}^* - \mathbf{m_j}^*)^T$.

Ενας βέλτιστος τρόπος υπολογισμού των στοιχείων του πίνακα $\mathbf{Q}$ είναι να βρούμε για ποιές τιμές του η στοχαστική απόκλιση των κατηγοριών μεγιστοποιείται.

Η ιδέα είναι καλή αλλά δυστυχώς δεν μπορούμε να επιτύχουμε μια αναλυτική λύση στο πρόβλημα αυτό και ιδιαίτερα στην γενική της περίπτωση κατά την οποία τα στατιστικά χαρακτηριστικά των κανονικών πυκνοτήτων πιθανότητας των κατηγοριών είναι διαφορετικά. Σε αυτή την περίπτωση χρησιμοποιούμε μεθόδους επαναληπτικής προσέγγισης ενός τοπικά βέλτιστου της συνάρτησης $Sd(i,j,\mathbf{y})$.

Επίσης πρέπει να προσέξουμε ότι οδηγηθήκαμε σε αυτά τα συμπεράσματα υποθέτοντας ότι τα πρότυπα των κατηγοριών ακολουθούν την κανονική πυκνότητα πιθανότητας. Συνεπώς οι σχέσεις που προαναφέραμε ισχύουν μόνο για αυτή την κατανομή.

6.7 Λυμένα Προβλήματα

Πρόβλημα 11 *Υπολογίστε τους συντελεστές γραμμικού μετασχηματισμού των προτύπων όταν χρησιμοποιούμε το αντίστροφο εσωτερικό γινόμενο σαν συνάρτηση απόστασης προτύπων με κριτήριο την ελαχιστοποίηση της αναμενόμενης απόστασης των προτύπων κάθε κατηγορίας από τα αντίστοιχα πρότυπά της.*

Πρόβλημα 12 *Υποθέστε ότι διαθέτουμε παραδείγματα από τρείς κατηγορίες προτύπων:*

$$\Omega_1 = \{(1,2,3),(3,3,1),(4,4,2),(-2,3,3),(0,2,-1)\}$$

$$\Omega_2 = \{(-1,1,1),(2,-3,-1),(-1,4,3),(-2,4,0)\}$$

$$\Omega_3 = \{(-1,-2,-1),(-3,1,-2),(2,2,0),(2,-6,-1),(0,0,-3)\}$$

Αφαιρέστε μία παράμετρο από τα πρότυπα, φτιάξτε το σύστημα ταξινόμησης προτύπων και υπολογίστε τον ελάχιστο σφάλμα του, χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης.

6.8 Αλυτα Προβλήματα

Ασκηση 27 *Υποθέστε ότι διαθέτουμε παραδείγματα δύο κατηγοριών προτύπων:*

$$\Omega_1 = \{(1,2,3),(3,3,1),(4,4,2),(-2,3,3),(0,2,-1)\}$$

$$\Omega_2 = \{(-1,1,1),(2,-3,-1),(-1,4,3),(-2,4,0)\}$$

Ελαττώστε τον αριθμό των παραμέτρων του πρότυπου σε δύο χρησιμοποιώντας τον μετασχηματισμό της ελάχιστης εντροπίας.

Ασκηση 28 *Ελαττώστε τον αριθμό των παραμέτρων του πρότυπου της προηγούμενης άσκησης σε μία χρησιμοποιώντας τον μετασχηματισμό της ελάχιστης εντροπίας.*

Ασκηση 29 *Αφαιρέστε μία παράμετρο από τα παραδείγματα της προηγούμενης άσκησης χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης και κριτήριο την ελαχιστοποίηση της αναμενόμενης απόστασης των προτύπων κάθε κατηγορίας από τα αντίστοιχα πρότυπά της.*

Ασκηση 30 *Για τα πρότυπα των παραδειγμάτων που προκύπτουν από τις προηγούμενες ασκήσεις φτιάξτε σύστημα ταξινόμησης προτύπων χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης και κριτήριο ταξινόμησης το κοντινότερο εικονικό πρωτότυπο που ελαχιστοποιεί την απόστασή του από τα πρότυπα της κατηγορίας.*

Υπολογίστε το ελάχιστο και το μέγιστο σφάλμα του συστήματος ταξινόμησης προτύπων και διατυπώστε τα συμεράσματά σας.

Ασκηση 31 *Αποδείξτε τις ιδιότητες της στοχαστικής απόκλισης κατηγοριών.*

Ασκηση 32 *Υποδείξτε μία γενική μέθοδο μείωσης των παραμέτρων του πρότυπου με γραμμικό μετασχηματισμό ο οποίος να μεγιστοποιεί την στοχαστική απόκλιση N κατηγοριών υποθέτοντας ότι τα πρότυπα των παραδειγμάτων ακολουθούν την κανονική κατανομή.*