Κεφάλαιο 2

Δομικά συστήματα

2.1 Εισαγωγή

Τα δομικά συστήματα αποτελούν τις πρώτες χρονολογικά προσπάθειες κατασκευής αυτόματων μηχανών ταξινόμησης προτύπων. Σε αυτή την κατηγορία μεθόδων η διαδικασία ταξινόμησης βασίζεται στην ιδιότητα που έχουν τα πρότυπα της ίδιας κατηγορίας να έχουν ομοιάζουσα ή ιδιοματική παραμετρική περιγραφή σε σχέση με τα πρότυπα άλλων κατηγοριών.

Μία κατηγορία δομικών μεθόδων ταξινόμησης εκμεταλεύεται την στατιστική ιδιότητα που έχουν τα πρότυπα μιάς κατηγορίας να έχουν μικρότερη απόσταση από τα πρότυπα διαφορετικών κατηγοριών. Η ταξινόμηση πραγματοποιείται με σύγκριση προτύπων. Η σύγκριση γίνεται με μέτρηση της "απόστασης" του άγνωστου πρότυπου με πρωτότυπα που επιλέχτηκαν ή έχουν δημιουργηθεί τεχνητά έτσι ώστε να περιγράφουν τις κατηγορίες των αντικειμένων με τον καλύτερο δυνατό τρόπο.

Υπάρχει μία διαφορετική κατηγορία δομικών μεθόδων οι οποίες ταξινομούν πρότυπα μέσω της εύρεσης των χαρακτηριστικών εκείνων που περιγράφουν με κοινό και ίσως αποκλειστικό τρόπο τα πρότυπα μιάς κατηγορίας. Οταν τα χαρακτηριστικά γνωρίσματα είναι γνωστά, τότε χρησιμοποιούνται γραμμικές ή μη γραμμικές συναρτήσεις απόφασης γιά να ταξινομήσουν τα πρότυπα. Αν διαθέτουμε παραδείγματα προτύπων τότε η εύρεση των ιδιοματικών χαρακτηριστικών γίνεται με μεθόδους εκπαίδευσης. Στο δεύτερο μέρος αυτού του κεφαλαίου περιγράφονται οι μέθοδοι που ταξινομούν το άγνωστο πρότυπο με συναρτήσεις απόφασης. Οι τεχνικές αυτές χρησιμοποιούνται σε εφαρμογές στις οποίες το πλήθος των παραμέτρων του πρότυπου είναι πολύ μεγάλο και οι κατηγορίες είναι γραμμικά διαχωρισμένες. Στο τέλος του κεφάλαιου αυτουύ παρουσιάζονται μερικές από τις πλέον διαδεδομένες δομικές μέθοδους ταξινόμησης μη γραμμικά διαχωρίσιμων προτύπων.

2.2 Οι μέθοδοι σύγκρισης προτύπων

Τρία είναι τα κυριώτερα προβλήματα που πρέπει να λυθούν κατά την σχεδίαση συστημάτων ταξινόμησης προτύπων με σύγκριση.

Αρχικά πρέπει να προσδιοριστεί η συνάρτηση που θα χρησιμοποιηθεί για να μετρηθεί η απόσταση των προτύπων.

Στην συνέχεια πρέπει να βρεθεί μία μέθοδος υπολογισμού των πρωτότυπων που περιγράφουν με τον καλύτερο δυνατό τρόπο τα πρότυπα κάθε κατηγορίας. Επιθυμητή ιδιότητα της μεθόδου είναι η εκτέλεση των υπολογισμών να γίνεται με βέλτιστο τρόπο, δηλαδή να ελαχιστοποιείται το σφάλμα της ταξινόμησης των παραδειγμάτων εκπαίδευσης. Είναι αναμενόμενο η συνάρτηση απόστασης να επηρεάζει τον τρόπο υπολογισμού των πρωτότυπων και είναι γνωστό ότι σε πολύ σπάνιες περιπτώσεις είναι δυνατόν να επιτύχουμε τέτοιες βέλτιστες λύσεις. Στην πράξη υπάρχουν αλγόριθμοι οι οποίοι δίνουν μία προσέγγιση αυτών των λύσεων.

Το τρίτο πρόβλημα που πρέπει να αντιμετωπιστεί έχει να κάνει με τον αριθμό των πρωτότυπων που πρέπει να υπολογιστούν από τα παραδείγματα εκπαίδευσης έτσι ώστε το σύστημα ταξινόμησης να έχει ικανοποιητική αξιοπιστία.

Οι μέθοδοι ταξινόμησης με σύγκριση προτύπων έχουν μικρή δομική πολυπλοκότητα και μπορούν να υλοποιηθούν εύκολα. Δίνουν ικανοποιητική επιτυχία ταξινόμησης μόνο όταν τα πρότυπα των κατηγοριών είναι ικανοποιητικά διαχωρισμένα.

Η διαδικασία ταξινόμησης είναι απλή και πραγματοποιείται εκτελώντας δύο διαφορετικές λειτουργίες. Αρχικά υπολογίζονται όλες οι αποστάσεις μεταξύ του παραμετρικού διανύσματος του άγνωστου πρότυπου και των πρωτότυπων. Ακολουθεί η διαδικασία ταξινόμησης η οποία λαμβάνει υπόψη μόνο τα μεγέθη των αποστάσεων που υπολογίστηκαν στο προηγούμενο βήμα.

Αυξημένες υπολογιστικές απαιτήσεις έχουν οι μέθοδοι που χρησιμοποιούν πολλαπλά πρωτότυπα ή χρησιμοποιούν όλα τα παραδείγματα εκπαίδευσης κατά την διαδικασία της ταξινόμησης. Η υπολογιστική πολυπλοκότητα είναι ανάλογη του αριθμού των πρωτότυπων.

Η ταξινόμηση με σύγκριση προτύπων αποτελεί την πλέον διαδεδομένη και είναι γενικά μία αξιόπιστη ομάδα μεθόδων ταξινόμησης. Γιαυτό τον λόγο χρησιμοποιούνται πολύ συχνά σε πρακτικές εφαρμογές.

2.2.1 Οι συναρτήσεις απόστασης προτύπων

Δεν υπάρχει κοινά αποδεκτός ορισμός για το τι προυποθέσεις πρέπει να πληρεί μία συνάρτηση απόστασης. Θεωρητικά οποιαδήποτε συνάρτηση ικανοποιεί την ισχυρή συνθήκη που ακολουθεί και προαιρετικά τις ασθενείς συνθήκες μπορεί να χρησιμοποιηθεί για να δώσει ένα ποσοτικό μέγεθος της "απόστασης" προτύπων.

1. *Ισχυρή συνθήκη.* Σαν μέτρο απόστασης μπορεί να οριστεί κάθε βαθμωτή συνάρτηση $d(\mathbf{x}, \mathbf{y})$ τέτοια ώστε $d : S \times S \rightarrow \mathbb{R}$, όπου S είναι το πεδίο ορισμού των προτύπων ($S \subseteq \mathbb{R}^P$), και P είναι ο αριθμός των παραμέτρων του πρότυπου.

2. *Ασθενείς συνθήκες.* Οι συνθήκες αυτές έχουν σκοπό να περιγράψουν κάποιες επιθυμητές ιδιότητες που θα θέλαμε να έχει η συνάρτηση απόστασης. Οι σημαντικότερες ιδιότητες είναι οι ακόλουθες:

   (α) *Το πεδίο τιμών της συνάρτησης να μην περιέχει αρνητικούς αριθμούς.* Η απόσταση προτύπων είναι ένας μη αρνητικός πραγματικός αριθμός:

   $$d(\mathbf{x}, \mathbf{y}) \geq 0, \qquad \forall \mathbf{x}, \mathbf{y} \in S \tag{2.1}$$

   (β) *Κάθε πρότυπο έχει μηδενική απόσταση από τον εαυτό του:*

   $$d(\mathbf{x}, \mathbf{x}) = 0, \qquad \forall \mathbf{x} \in S \tag{2.2}$$

   (γ) *Συμμετρική ιδιότητα.* Η εναλλαγή της θέσης των προτύπων δεν επηρεάζει το μέγεθος της μετρούμενης απόστασης:

$$d(\mathbf{x},\mathbf{y}) = d(\mathbf{y},\mathbf{x}), \qquad \forall \mathbf{x},\mathbf{y} \in S \tag{2.3}$$

(δ) *Τριγωνική συνθήκη.* Η απόσταση δύο προτύπων είναι μικρότερη ή ίση από το άθροισμα των αποστάσεων τους από τρίτο πρότυπο:

$$d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},\mathbf{z}) + d(\mathbf{z},\mathbf{x}), \qquad \forall \mathbf{x},\mathbf{y},\mathbf{z} \in S \tag{2.4}$$

Η επιλογή της συνάρτησης απόστασης είναι μία δύσκολη εργασία διότι επηρεάζει καθοριστικά την αξιοπιστία του συστήματος ταξινόμησης. Η επιλογή της συνάρτησης απόστασης γίνεται συνήθως με μελέτη της κατανομής των προτύπων των κατηγοριών στο χώρο των μετρήσεων. Σε μερικές περιπτώσεις μπορούμε να προσδιορίσουμε εκ των προτέρων την συνάρτηση απόστασης εξαιτίας κάποιων χαρακτηριστικών ιδιοτήτων που έχουν οι παράμετροι των προτύπων. Παράδειγμα τέτοιου τύπου εφαρμογών είναι η ταξινόμηση προτύπων σε κατηγορίες που ακολουθούν κανονική πυκνότητα πιθανότητας. Στο επόμενο κεφάλαιο αποδεικνύεται ότι σε αυτό το είδος κατηγοριών η Ευκλείδια βεβαρυμένη απόσταση μεταξύ του άγνωστου πρότυπου και της αναμενόμενης τιμής του πρότυπου κάθε κατηγορίας είναι η συνάρτηση απόστασης που ελαχιστοποιεί την πιθανότητα λανθασμένης ταξινόμησης.

Στην πράξη δεν γνωρίζουμε συνήθως την κατανομή των προτύπων στο χώρο των μετρήσεων και γιαυτό τον λόγο δεν είμαστε σε θέση να κάνουμε τέτοιου είδους εκτιμήσεις. Βεβαίως πολλές φορές τέτοιου είδους υποθέσεις προσεγγίζουν αρκετά πιστά τα πραγματικά δεδομένα με άμεσο αποτέλεσμα την αποτελεσματική τους χρήση σε πραγματικά προβλήματα.

Επειδή δεν υπάρχει μία γενική μέθοδος επιλογής της καταλληλότερης συνάρτησης απόστασης γιαυτό τον λόγο καταφεύγουμε συνήθως σε πειραματικές μεθόδους. Φτιάχνουμε συστήματα ταξινόμησης τα οποία διαφέρουν μόνο ως προς τον τρόπο με τον οποίο μετρούν την απόσταση των προτύπων. Μετρούμε το σφάλμα των συστημάτων και στην συνέχεια διαλέγουμε σαν καλύτερη συνάρτηση απόστασης, εκείνη που δίνει το μικρότερο σφάλμα.

Για να εφαρμόσουμε αυτή την τεχνική είναι απαραίτητο να γνωρίζουμε έναν αριθμό συναρτήσεων απόστασης τις οποίες διαδοχικά ελέγχουμε ως προς την καταλληλότητά τους, όταν δεν μπορούμε να προσδιορίσουμε εκ των προτέρων την βέλτιστη συνάρτηση απόστασης.

Αν $\mathbf{x}$ είναι το άγνωστο πρότυπο και $\mathbf{y} \in S$ είναι το πρωτότυπο μιας κατηγορίας, τότε οι σημαντικότερες συναρτήσεις απόστασης ορίζονται από τις ακόλουθες αλγεβρικές εκφράσεις:

1. *Ευκλείδια απόσταση.* Είναι η πλέον διαδεδομένη συνάρτηση απόστασης, ικανοποιώντας ταυτόχρονα όλες τις ασθενείς συνθήκες. Η διαδοσή της οφείλεται κυρίως στην φυσική υπόσταση της Ευκλείδια απόστασης η οποία εκφράζει το μήκος της ευθείας που συνδέει το άγνωστο πρότυπο $\mathbf{x}$ και το πρωτότυπο της κατηγορίας $\mathbf{y}$ στον N-διάστατο Ευκλείδιο χώρο.

$$d_e(\mathbf{x},\mathbf{y}) = \left(\sum_{i=1}^{N} (x_i - y_i)^2\right)^{1/2} \tag{2.5}$$

όπου x_i, y_i είναι η i συνιστώσα του διανύσματος του άγνωστου πρότυπου και του αντίστοιχου πρωτότυπου.

2. *Γενικευμένη Ευκλείδια απόσταση.* Είναι λιγότερο γνωστή συνάρτηση απόστασης και αποτελεί γενίκευση της Ευκλείδιας απόσταση (s=2):

$$d_e(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{N} (x_i - y_i)^s \right)^{1/s} \qquad s = 2, 4, 6, ... \tag{2.6}$$

Η γενικευμένη Ευκλείδια απόσταση ικανοποιεί όλες τις ασθενείς συνθήκες.

3. *Σταθμισμένη Ευκλείδια απόσταση.* Η συνάρτηση αυτή είναι παραλλαγή της Ευκλείδιας απόστασης στην οποία υπισέρχονται και πολλαπλασιαστικοί όροι οι οποίοι είναι σταθερές (είναι μοναδικοί σε κάθε σύστημα ταξινόμησης) ή αποτελούν χαρακτηριστικό γνώρισμα κάθε μιάς των κατηγοριών (κάθε κατηγορία διαθέτει το δικό της σύνολο σταθερών συντελεστών). Η συνάρτηση σταθμισμένη Ευκλείδιας απόστασης ικανοποιεί επίσης όλες τις ασθενείς συνθήκες, εκτός από την τριγωνική ιδιότητα στην περίπτωση κατά την οποία κάθε κατηγορία διαθέτει το δικό της σύνολο συντελεστών.

$$d_w(\mathbf{x}, \mathbf{y}, \mathbf{m}) = \left(\sum_{i=1}^{N} m_i (x_i - y_i)^2 \right)^{1/2} \tag{2.7}$$

Απο την εξίσωση 2.7 γίνεται φανερό ότι για κάθε πρωτότυπο, εκτός από την παραμετρική του περιγραφή (το διάνυσμα $\mathbf{y}$), χρειάζεται και ο προσδιορισμός του διανύσματος βαρύτητας του $\mathbf{m}$ για κάθε μία από τις κατηγορίες (όταν αυτό μεταβάλλεται στις κατηγορίες). Αν παραμένει σταθερό θα πρέπει κατά την διάρκεια της εκπαίδευσης να υπολογιστεί ένα διάνυσμα βαρύτητας.

Ο υπολογισμός του διανύσματος βαρύτητας πρέπει να γίνει με τέτοιο τρόπο έτσι ώστε να περιγράφει χαρακτηριστικά γνωρίσματα της κατηγορίας του πρωτότυπου. Το ενδιαφέρον αυτό πρόβλημα δέχεται πολλές λύσεις. Μερικές από αυτές δίνονται σε αυτό το κεφάλαιο.

4. *Απόσταση Mahalanobis.* Στην συνάρτηση αυτή, που ικανοποιεί μέρος των ασθενών συνθηκών, δόθηκε το όνομα του Ινδού μαθηματικού Mahalanobis:

$$D(\mathbf{x}, \mathbf{y}, \mathbf{s}) = (\mathbf{x} - \mathbf{y})^T \mathbf{s}^{-1} (\mathbf{x} - \mathbf{y}) \tag{2.8}$$

όπου $\mathbf{s}$ είναι τετραγωνικός πίνακας $N \times N$.

Οι επόμενες δύο συναρτήσεις αποστάσεων δεν ικανοποιούν όλες τις ασθενείς συνθήκες.

5. *Απόσταση του αντίστροφου κανονικοποιημένου εσωτερικού γινομένου.*

$$d_k(\mathbf{x}, \mathbf{y}) = \frac{||\mathbf{x}||||\mathbf{y}||}{\mathbf{x}^T \mathbf{y}} \tag{2.9}$$

6. *Απόσταση Tanimoto.*

$$s(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{x} + \mathbf{y}^T \mathbf{y}}{\mathbf{x}^T \mathbf{y}} - 2 \tag{2.10}$$

Η επιλογή της συνάρτησης απόστασης, όπως θα φανεί στα παραδείγματα που ακολουθούν, επηρεάζει σημαντικά την αξιοπιστία του συστήματος ταξινόμησης και καθορίζει επίσης και τον τρόπο υπολογισμού των πρωτότυπων.

2.2.2 Τα κριτήρια ταξινόμησης

Μετά την επιλογή της συνάρτησης απόστασης πρέπει να περιγράψουμε τον μηχανισμό εκείνον ο οποίος εκτελεί την ταξινόμηση λαμβάνοντας υπόψη τον υπολογισμό των αποστάσεων του διανύσματος του άγνωστου πρότυπου από τα διανύσματα των πρωτότυπων των κατηγοριών. Σε αυτό το τμήμα του κεφαλαίου γίνεται μία λεπτομερής ανάλυση των τεχνικών ταξινόμησης προτύπων σε κατηγορίες με την μέθοδο της σύγκρισης.

Τα σημαντικότερα κριτήρια ταξινόμησης που χρησιμοποιούνται σε πρακτικές εφαρμογές είναι δύο, το κριτήριο της μικρότερης απόστασης των προτύπων και το κριτήριο ταξινόμησης των Κ-πλησιέστερων πρωτότυπων.

2.2.3 Η ταξινόμηση της μικρότερης απόστασης προτύπων

Το πλέον διαδεδομένο κριτήριο ταξινόμησης με σύγκριση πρότυπων είναι η εύρεση της κατηγορίας εκείνης το πρωτότυπο της οποίας έχει την μικρότερη απόσταση από το άγνωστο παραμετρικό διάνυσμα. Το κριτήριο ταξινόμησης ονομάζεται ταξινόμηση της μικρότερης απόστασης (minimum distance pattern classification):

$$\mathbf{x} \in \omega_j \qquad \alpha\nu\nu \qquad d(\mathbf{x}, \omega_j) < d(\mathbf{x}, \omega_i) \qquad \forall j = 1, N \tag{2.11}$$

όπου ανν είναι η συντομογραφία της έκφρασης "αν και μόνο αν", και N ο αριθμός των κατηγοριών που το σύστημα αναγνωρίζει.

2.2.4 Η ταξινόμηση των Κ-πλησιέστερων πρωτότυπων

Στην περίπτωση κατά την οποία η περιγραφή των κατηγοριών γίνεται με περισσότερα του ενός πρωτότυπα, τότε εκτός από το κριτήριο μικρότερης απόστασης έχουμε την δυνατότητα να χρησιμοποιήσουμε και το κριτήριο ταξινόμησης των Κ-πλησιέστερων πρωτότυπων (K-nearest neighbor classification). Το κριτήριο αυτό χρησιμοποιείται σπανιότερα, αλλά πολλές φορές δίνει μικρότερο σφάλμα από το κριτήριο ταξινόμησης της μικρότερης απόστασης.

Η ταξινόμηση πραγματοποιείται με την απομόνωση των Κ πρωτότυπων που παρουσιάζουν την μικρότερη απόσταση από το άγνωστο πρότυπο και ακολουθεί η ταξινόμηση στην κατηγορία εκείνη τα πρωτότυπα της οποίας υπερτερούν σε αριθμό στο σύνολο των Κ-πλησιέστερων πρωτότυπων.

Παράδειγμα 1 *Δίνεται σύστημα ταξινόμησης προτύπων δύο κατηγοριών και τα ακόλουθα παραδείγματα:*

$$\Omega = \{((1, 1.3), \omega_1), ((2.2, 2.5), \omega_1), ((3, 3.1), \omega_1), ((4, 3.5), \omega_1),$$

$$((1.1, 1.8), \omega_2), ((2.5, 4.3), \omega_2), ((3, 6.3), \omega_2), ((4, 7.6), \omega_2)\}.$$

Ο κατασκευαστής του συστήματος μας είπε ότι το σύστημα δουλεύει πολύ καλά με την αντίστροφη συνάρτηση του κανονικοποιημένου εσωτερικού γινομένου και μας έδωσε σαν πρωτότυπα των κατηγοριών, το $(1,1)$ για την ω_1 και το $(1,2)$ για την ω_2.

Είναι σωστοί οι ισχυρισμοί του;

Για να ελέγξουμε τους ισχυρισμούς του κατασκευαστή υπολογίζουμε την επιτυχία της ταξινόμησης απο τα παραδείγματα. Στον πίνακα που ακολουθεί δίνονται οι αποστάσεις των προτύπων από τα πρωτότυπα των κατηγοριών.

Πίναχας 2.1: Αποστάσεις του αντίστροφου κανονικοποιημένου εσωτερικού γινομένου

Πρότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Προτότυπο (1,1)	1.0085	1.0062	1.0010	1.0022
Προτότυπο (1,2)	1.0187	1.0342	1.0485	1.0804
Πρότυπα της ω_2	(1.1,1.8)	(2.5,4.3)	(3,6.3)	(4,7.6)
Προτότυπο (1,1)	1.0400	1.0344	1.0611	1.0470
Προτότυπο (1,2)	1.0036	1.0019	1.0002	1.0002

Βλέπουμε ότι το σύστημα που ταξινομεί στην κατηγορία εκείνη το πρωτότυπο της οποίας έχει την μικρότερη απόσταση δίνει σωστή ταξινόμηση για όλα τα παραδείγματα Συνεπώς το σφάλμα του συστήματος ταξινόμησης που μας δόθηκε είναι 0%.

Ας δούμε τι επίδραση θα έχει στην αξιοπιστία του συστήματος η αλλαγή της συνάρτησης απόστασης. Ας δοκιμάσουμε την Ευκλείδια συνάρτηση απόστασης. Στον πίνακα 2.2 δίνονται οι Ευκλείδιες αποστάσεις των προτύπων των παραδειγμάτων απο τα πρωτότυπα των κατηγοριών.

Πίναχας 2.2: Ευκλείδιες αποστάσεις των προτύπων

Πρότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
ω_1	0.3000	1.9200	2.9000	3.9050
ω_2	0.7000	1.3000	2.2825	3.3541
Πρότυπα της ω_2	(1.1,1.8)	(2.5,4.3)	(3,6.3)	(4,7.6)
ω_1	0.8602	3.6249	5.6648	7.2498
ω_2	0.2236	2.7459	4.7423	6.3529

Απο τα αποτελέσματα του πίνακα βλέπουμε ότι αν προσπαθήσουμε να ταξινομήσουμε τα πρότυπα με βάση την μικρότερη απόστασή τους από τα αντίστοιχα πρωτότυπα των κατηγοριών έχουμε μόνο πέντε στις οκτώ επιτυχείς ταξινομήσεις, επιτυχία δηλαδή της τάξης του 62.5%.

Από τα αριθμητικά δεδομένα του παραδείγματος γίνεται φανερό ότι αν μεταβάλουμε την συνάρτηση απόστασης πρέπει ταυτόχρονα να μεταβάλλουμε και τα αντίστοιχα πρωτότυπα των κατηγοριών έτσι ώστε να μπορέσουμε να συγκρίνουμε με αντικειμενικό τρόπο τα συστήματα ταξινόμησης.

Παράδειγμα 2 *Για τα παραδείγματα του προηγούμενου παραδείγματος υπολογίστε την αξιοπιστία του συστήματος ταξινόμησης για την συνάρτηση απόστασης Tanimoto. Τι παρατηρείτε και πως μπορείτε να δικαιολογήσετε τις διαφορές των αποτελεσμάτων που παρουσιάζονται για τις τρείς διαφορετικές συναρτήσεις απόστασης.*

Παράδειγμα 3 *Σύστημα ταξινόμησης προτύπων τριών κατηγοριών έχει σαν πρωτότυπα τα διανύσματα που δίνονται στον πίνακα 2.3. Χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης βρείτε, με το κριτήριο ταξινόμησης της μικρότερης απόστασης και το κριτήριο ταξινόμησης των τριών κοντινότερων πρωτότυπων την κατηγορία των άγνωστων πρότυπων (0.0,0.0) και (2.0,2.0).*

Πίνακας 2.3: Πίνακας πρωτότυπων τριών κατηγοριών

Πρωτότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Πρωτότυπα της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
Πρωτότυπα της ω_3	(-1.1,0.4)	(-2.5,4.3)	(-3.0,3.3)	(-2.5,2.6)

Πίναχας 2.4: Πίναχας αποστάσεων προτύπων από τα πρωτότυπα των κατηγοριών

Πρωτότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
(0.0,0.0)	1.640	3.330	4.314	5.315
(2.0,2.0)	1.221	0.538	1.487	2.500
Πρωτότυπα της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
(0.0,0.0)	3.523	4.482	4.244	6.560
(2.0,2.0)	0.900	1.700	1.552	3.773
Πρωτότυπα της ω_3	(-1.1,0.4)	(-2.5,4.3)	(-3.0,3.3)	(-2.5,2.6)
(0.0,0.0)	1.077	4.974	4.460	3.607
(2.0,2.0)	3.488	5.053	5.166	4.540

Αρχικά κατασκευάζουμε τον πίνακα των αποστάσεων των προτύπων από τα πρωτότυπα των κατηγοριών (Πίνακας 2.4).

Το πρότυπο (0.0,0.0) ταξινομείται στην τρίτη κατηγορία με το κριτήριο μικρότερης απόστασης, διότι έχει απόσταση 1.077 από το πρώτο πρωτότυπο της τρίτης κατηγορίας. Με το κριτήριο των τριών κοντινότερων πρωτότυπων ταξινομείται στην πρώτη κατηγορία διότι στα τρία κοντινότερα πρωτότυπα δύο ανήκουν στην πρώτη κατηγορία (το πρώτο με 1.640 και το δεύτερο με 3.330) και ένα στην τρίτη κατηγορία (το κοντινότερο, με απόσταση 1.077).

Αντίστοιχα το δεύτερο πρότυπο (2.0,2.0) ταξινομείται στην πρώτη κατηγορία με το κριτήριο μικρότερης απόστασης (0.538). Με το κριτήριο των τριών κοντινότερων πρωτότυπων ταξινομείται πάλι στην πρώτη κατηγορία διότι στα τρία κοντινότερα πρωτότυπα δύο ανήκουν στην πρώτη κατηγορία (το πρώτο με απόσταση 1.221 και το δεύτερο με 0.538) και ένα στην δεύτερη κατηγορία με απόσταση 0.9.

2.2.5 Η διαδικασία της εκπαίδευσης

Η τελευταία εργασία που πρέπει να πραγματοποιήσουμε γιά να ολοκληρώσουμε την κατασκευή ενός συστήματος ταξινόμησης με σύγκριση προτύπων (μετά την επιλογή της συνάρτησης απόστασης και του κριτηρίου ταξινόμησης) είναι ο υπολογισμός του πρωτότυπου ή των πρωτότυπων κάθε κατηγορίας. Πρέπει δηλαδή να εκτελέσουμε την διαδικασία εκπαίδευσης με την οποία θα προσαρμόσουμε τις παραμέτρους του συστήματος έτσι ώστε να λειτουργεί με την μέγιστη δυνατή αξιοπιστία για το σύνολο των παραδειγμάτων που διαθέτουμε.

Ο τρόπος μέτρησης της απόστασης των προτύπων παίζει καθοριστικό ρόλο, όπως θα δούμε στις επόμενες παραγράφους, τόσο στον τρόπο υπολογισμού των πρωτότυπων των κατηγοριών, όσο και στην τελική αξιοπιστία του συστήματος ταξινόμησης.

Η διαδικασία εκπαίδευσης επηρεάζεται από το είδος του προβλήματος που αντιμετωπίζουμε, την επιθυμούμενη αξιοπιστία, τον αριθμό των παραδειγμάτων που διαθέτουμε κατά την διαδικασία εκπαίδευσης και τους τυχόν περιορισμούς στην υπολογιστική πολυπλοκότητα του συστήματος ταξινόμησης. Πέντε είναι οι σημαντικότεροι μέθοδοι που διαθέτουμε για την υλοποίηση μιάς διαδικασίας εκπαίδευσης:

1. Να χρησιμοποιήσουμε όλα τα πρότυπα των παραδειγμάτων σαν πρωτότυπα,
2. να χρησιμοποιήσουμε μόνο ένα απο αυτά,
3. να επιλέξουμε ένα πεπερασμένο αριθμό από τα πρότυπα των παραδειγμάτων σαν πρωτότυπα,
4. να κατασκευάσουμε ένα εικονικό (τεχνητό) πρωτότυπο για κάθε κατηγορία αντικειμένων, και

5. να κατασκευάσουμε πολλαπλά εικονικά πρωτότυπα για κάθε κατηγορία αντικειμένων βασιζόμενοι σε κάποιο κριτήριο βελτιστοποίησης.

Τα κριτήρια επιλογής των πρωτότυπων ως και η μέθοδος(οι) με τις οποίες θα το επιτύχουμε είναι το θέμα που διαπραγματεύονται οι επόμενες παραγράφοι.

2.2.6 Πρωτότυπα είναι όλα τα παραδείγματα

Σε αυτή την μέθοδο όλα τα πρότυπα των παραδειγμάτων που είναι διαθέσιμα χρησιμοποιούνται σαν πρωτότυπα των κατηγοριών. Οπως έχει ήδη προαναφερθεί η μεθοδολογία αυτή ακολουθείται όταν διαθέτουμε μικρό αριθμό παραδειγμάτων ή όταν θέλουμε να έχουμε πολύ μεγάλη αξιοπιστία ταξινόμησης χωρίς να μας ενδιαφέρει το υπολογιστικό κόστος.

Παράδειγμα 4 *Για να κατασκευάσουμε σύστημα ταξινόμησης προτύπων τριών κατηγοριών μας δίνονται τα παραδείγματα του πίνακα 2.5. Χρησιμοποιώντας την Ευκλείδια συνάρτηση απόστασης και το κριτήριο ταξινόμησης της μικρότερης απόστασης υπολογίστε το σφάλμα του συστήματος ταξινόμησης.*

Πίνακας 2.5: Πίνακας παραδειγμάτων για τρείς κατηγορίες αντικειμένων

Πρότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Πρότυπα της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
Πρότυπα της ω_3	(-1.1,0.4)	(-2.5,4.3)	(-3.0,3.3)	(-2.5,2.6)

Γνωρίζουμε ότι δεν μπορούμε να επιτύχουμε ακριβή υπολογισμό του σφάλματος σε συστήματα ταξινόμησης προτύπων όταν έχουμε στην διαθεσή μας πεπερασμένο αριθμό παραδειγμάτων (κεφάλαιο 1).

Με την μέθοδο C και U θα υπολογίσουμε το ελάχιστο και το μέγιστο σφάλμα του συστήματος ταξινόμησης.

Αν όλα τα παραδείγματα εκπαίδευσης χρησιμοποιηθούν σαν πρωτότυπα των κατηγοριών τότε, εφόσον δεν έχουμε ταυτόσημα πρότυπα σε περισσότερες των δύο κατηγορίες, το σφάλμα με την μέθοδο C είναι 0 % αφού όλα τα πρότυπα που θα χρησιμοποιηθούν για την μέτρηση της αξιοπιστίας του συστήματος θα είναι και πρωτότυπα των κατηγοριών.

Το μέγιστο σφάλμα θα το υπολογίσουμε με την μέθοδο του αχρησιμοποίητου παραδείγματος δηλαδή, κάθε φορά θα φτιάχνουμε από τα έντεκα παραδείγματα το σύστημα ταξινόμησης προτύπων και θα χρησιμοποιούμε το δωδέκατο για να υπολογίσουμε το σφάλμα του συστήματος. Συνεπώς βάσει της μεθόδου κατασκευής των πρωτότυπων, κάθε φορά όλα τα παραδείγματα θα είναι και πρωτότυπα των κατηγοριών εκτός από εκείνο το οποίο θα χρησιμοποιείται για την μέτρηση του σφάλματος.

Αρχικά κατασκευάζουμε τον πίνακα των αποστάσεων των προτύπων από τα πρωτότυπα των κατηγοριών (Πίνακας 2.6). Κάθε γραμμή του πίνακα αντιστοιχεί σε ένα σύστημα ταξινόμησης προτύπων που έχει σαν πρωτότυπα όλα τα παραδείγματα των εκτός από εκείνο το οποίο βρίσκεται στην πρώτη στήλη κάθε γραμμής. Το πρότυπο αυτό χρησιμοποιείται για να μετρηθεί το σφάλμα του συστήματος ταξινόμησης. Η μικρότερη απόστασή του από τα πρωτότυπα των κατηγοριών δίνονται στις επόμενες τρείς στήλες.

Τα πρότυπα που ταξινομούνται στις σωστές κατηγορίες είναι μόνο πέντε, δηλαδή το σφάλμα του συστήματος βρίσκεται στο διάστημα [0.00%, 58.33%].

2.2.7 Ενα παράδειγμα επιλέγεται σαν πρωτότυπο κατηγορίας

Η επιλογή του πλέον κατάλληλου πρότυπου πραγματοποιείται συνήθως όταν υπάρχουν σοβαροί περιορισμοί στην ταχύτητα απόκρισης του συστήματος ταξινόμησης ή τα πρότυπα των κατηγοριών είναι σαφώς διαχωρισμένα.

Πίνακας 2.6: Αποστάσεις προτύπων από τα πρωτότυπα των κατηγοριών

Πρότυπο ελέγχου	ω_1	ω_2	ω_3
(1.0,1.3) $\in \omega_1$	1.6970	1.8868	2.2847
(2.2,2.5) $\in \omega_1$	1.0000	0.4472	3.9115
(3.0,3.1) $\in \omega_1$	1.0000	0.4472	4.9091
(4.0,3.5) $\in \omega_1$	1.0770	1.2000	5.9680
(2.0,2.9) $\in \omega_2$	0.4472	1.0000	3.9826
(2.8,3.5) $\in \omega_2$	0.4472	1.0000	4.9819
(3.5,2.4) $\in \omega_2$	0.8602	1.3038	5.0159
(4.0,5.2) $\in \omega_2$	1.7000	2.0808	6.5620
(-1.1,0.4) $\in \omega_3$	2.2847	3.9824	2.6076
(-2.5,4.3) $\in \omega_3$	4.6097	4.7127	1.1180
(-3.0,3.3) $\in \omega_3$	4.4721	5.0159	0.8602
(-2.5,2.6) $\in \omega_3$	3.7336	4.5099	0.8602

Ο συνήθης τρόπος προσδιορισμού του πρωτότυπου πραγματοποιείται με την εύρεση του πρότυπου εκείνου που έχει την μικρότερη αθροιστικά απόσταση από τα παραδείγματα της κατηγορίας του:

$$\mathbf{y} = \underset{\mathbf{y}_j}{argmin} \sum_{i=1}^{M} d(\mathbf{y}_i, \mathbf{y}_j) \tag{2.12}$$

Παράδειγμα 5 *Μετρήστε το σφάλμα του συστήματος ταξινόμησης προτύπων του προηγούμενου παραδείγματος θεωρώντας σαν πρωτότυπο κάθε μιας των κατηγοριών εκείνο που ελαχιστοποιεί το άθροισμα των αποστάσεων του από τα πρότυπα των παραδειγμάτων της κατηγορίας του.*

Ακολουθώντας την ίδια μεθοδολογία λύσης προσπαθούμε να προσεγγίσουμε το σφάλμα του συστήματος ταξινόμησης με την μέθοδο C και U.

Η πρώτη διαφοροποίηση που συναντάμε σε σχέση με την λύση που δίνεται στο προηγούμενο παράδειγμα είναι ο τρόπος υπολογισμού του σφάλματος με την μέθοδο C. Αρχικά βρίσκουμε το πρότυπο εκείνο που ικανοποιεί την εξίσωση του κριτηρίου και κατόπιν ταξινομούμε όλα τα παραδείγματα στις κατηγορίες με το κριτήριο του κοντινότερου πρωτότυπου (σε αυτή την περίπτωση δεν μπορούμε να χρησιμοποιήσουμε άλλο κριτήριο ταξινόμησης) όπως αυτό δείχνεται στον πίνακα 2.7.

Στην πρώτη γραμμή του πίνακα δίπλα στο σύμβολο της κατηγορίας βρίσκεται το πρωτότυπο που έχει επιλεγεί από τα παραδείγματα με το κριτήριο της μικρότερης απόστασης.

Το σύστημα ταξινόμησης με την μέθοδο-C παρουσιάζει τρία σφάλματα σε δώδεκα πειράματα, συνεπώς το ελάχιστο σφάλμα του συστήματος είναι 25%.

Το μέγιστο σφάλμα θα το υπολογίσουμε με την μέθοδο του αχρησιμοποίητου παραδείγματος.

Ο πίνακας των αποστάσεων των προτύπων από τις κατηγορίες δίνεται στον πίνακα 2.8. Σε κάθε γραμμή του πίνακα υπολογίζουμε τα πρωτότυπα χρησιμοποιώντας όλα τα παραδείγματα εκτός από εκείνο που βρίσκεται στην πρώτη στήλη και στην συνέχεια υπολογίζουμε την απόσταση του αχρησιμοποίητου πρότυπου από το πρωτότυπο που έχει υπολογιστεί. Στην τελευταία στήλη δηλώνεται η κατηγορία στην οποία έχει ταξινομηθεί το αχρησιμοποίητο πρότυπο (το πρότυπο της πρώτης στήλης).

Τα πρότυπα που ταξινομούνται λανθασμένα είναι τέσσερα, συνεπώς το σφάλμα του συστήματος ταξινόμησης θα βρίσκεται στο διάστημα [25.00%, 33.33%].

Παρατηρούμε ότι σε σχέση με το σφάλμα του συστήματος ταξινόμησης που χρησιμοποιεί όλα τα παραδείγματα για πρωτότυπα το διάστημα αβεβαιότητας είναι σαφώς μικρότερο για την μέθοδο προσδιορισμού ενός πρωτότυπου από τα παραδείγματα.

Για την μέθοδο προσδιορισμού ενός πρωτότυπου από τα παραδείγματα βλέπουμε όμως ότι έχουμε φραγμό από ένα κατώτερο επιτρεπτό σφάλμα (25%) το οποίο είναι μηδενικό στην μέθοδο που χρησιμοποιεί όλα τα παραδείγματα σαν

Πίναχας 2.7: Πίναχας αποστάσεων προτύπων από τις κατηγορίες και ταξινόμηση

Πρότυπο	ω_1-(2.2,2.5)	ω_2-(2.8,3.5)	ω_3-(-2.5,2.6)	Ταξινόμηση
(1.0,1.3)	1.6970	2.8425	3.7336	ω_1
(2.2,2.5)	0.0000	1.1661	4.7010	ω_1
(3.0,3.1)	1.0000	0.4472	5.5226	ω_2
(4.0,3.5)	2.0591	1.2000	6.5620	ω_2
(2.0,2.9)	0.4472	1.0000	4.5099	ω_1
(2.8,3.5)	1.1661	0.0000	5.3758	ω_2
(3.5,2.4)	1.3038	1.3038	6.0033	ω_2
(4.0,5.2)	3.2449	2.0808	7.0007	ω_2
(-1.1,0.4)	3.9115	4.9819	2.6076	ω_3
(-2.5,4.3)	5.0328	5.3600	1.7000	ω_3
(-3.0,3.3)	5.2611	5.8034	0.8602	ω_3
(-2.5,2.6)	4.7010	5.3758	0.0000	ω_3

Πίναχας 2.8: Πίναχας αποστάσεων προτύπων από τις κατηγορίες και ταξινόμηση

Πρότυπο ελέγχου	ω_1	ω_2	ω_3	Ταξινόμηση
(1.0,1.3)	2.690725	2.842534	3.733631	ω_1
(2.2,2.5)	1.000000	1.166190	4.701064	ω_1
(3.0,3.1)	1.000000	0.447214	5.522681	ω_2
(4.0,3.5)	2.059126	1.200000	6.562012	ω_2
(2.0,2.9)	0.447214	1.000000	4.509989	ω_1
(2.8,3.5)	1.166190	1.303840	5.375872	ω_1
(3.5,2.4)	1.303840	1.303840	6.003332	ω_2
(4.0,5.2)	3.244996	2.080865	7.003571	ω_2
(-1.1,0.4)	3.911521	4.981967	2.607681	ω_3
(-2.5,4.3)	5.032892	5.360037	4.143670	ω_3
(-3.0,3.3)	5.261179	5.803447	3.466987	ω_3
(-2.5,2.6)	4.701064	5.375872	2.607681	ω_3

πρωτότυπα.

Παράδειγμα 6 *Μετρήστε το σφάλμα του συστήματος ταξινόμησης προτύπων του προηγούμενου παραδείγματος χρησιμοποιώντας σαν μέτρο απόστασης το αντίστροφο κανονικοποιημένο εσωτερικό γινόμενο διανυσμάτων και για τις δύο μεθόδους προσδιορισμού των πρωτότυπων των κατηγοριών. Τι διαφορές παρατηρείτε και πως μπορείτε να τις δικαιολογήσετε.*

2.2.8 Επιλογή k-πρότυπων από τα παραδείγματα

Η μέθοδος της επιλογής k-πρότυπων απο τα παραδείγματα είναι η πλέον διαδεδομένη στις εφαρμογές και αποτελεί συνήθως έναν συμβιβασμό ανάμεσα στην υπολογιστική πολυπλοκότητα του συστήματος ταξινόμησης που χρησιμοποιεί όλα τα διαθέσιμα παραδείγματα γιά πρωτότυπα (δίνοντας υψηλή αξιοπιστία ταξινόμησης) και του συστήματος ταξινόμησης που χρησιμοποιεί μονάχα ένα πρότυπο αναφοράς και παρουσιάζει χαμηλή υπολογιστική πολυπλοκότητα και συνήθως μικρότερη αξιοπιστία ταξινόμησης.

Η επιλογή k-πρότυπων από τα παραδείγματα αποτελεί έναν συμβιβασμό στις λύσεις που προτείνουν οι προηγούμενες μέθοδοι οι οποίες μάλιστα μπορεί να θεωρηθούν ότι αποτελούν ειδικές περιπτώσεις για k=1 (επιλογή ενός πρότυπου από τα παραδείγματα) και k=M (όλα τα πρότυπα των παραδειγμάτων θεωρούνται και πρωτότυπα των κατηγοριών).

Το κριτήριο επιλογής ενός συνόλου k βέλτιστων πρότυπων από τα παραδείγματα είναι το ακόλουθο:

$$\{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\} = \underset{\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_K}{argmin} \sum_{i=1}^{M} \underset{\mathbf{z}_x,\mathbf{y}_i \neq \mathbf{z}_x}{min} \; d(\mathbf{y}_i, \mathbf{z}_x), \quad \mathbf{z}_x \in \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\} \quad \mathbf{y}_i \in \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_M\} \quad (2.13)$$

Ο υπολογισμός των πρωτότυπων δεν μπορεί να γίνει αναζητώντας το ελάχιστο της συνάρτησης βελτιστοποίησης διότι η συνάρτηση min δεν είναι παραγωγίσιμη. Η αδυναμία αυτή δεν αποτελεί τελικά σημαντικό εμπόδιο για την λύση της εξίσωσης 2.13 διότι υπάρχει επαναληπτικός αλγόριθμος με τον οποίο μπορούμε να επιτύχουμε μία διαδοχική προσέγγιση της βέλτισης λύσης.

Αλγόριθμος 1

1. Αρχικές συνθήκες *Οπως έχει ήδη αναφερθεί στο κεφάλαιο 1 περιγράφοντας την στρατηγική που ακολουθούμε για να υλοποιήσουμε μία διαδικασία αυτο-εκπαίδευσης, ξεκινάμε με μία τυχαία επιλογή των πρωτότυπων* $\mathbf{z}_1, ..., \mathbf{z}_k$ *από τα πρότυπα των παραδειγμάτων και επαναληπτικά επαναπροσδιορίζουμε τις τιμές των.*

2. Ταξινόμηση προτύπων. *Βάσει του κριτηρίου ταξινόμησης τα πρότυπα της κατηγορίας κατατάσσονται σε μία από τις K εικονικές κατηγορίες που ορίζονται απο τα πρωτότυπα.*

3. Υπολογισμός πρωτότυπων. *Σε κάθε εικονική κατηγορία ορίζουμε το πρωτότυπο σαν το πρότυπο του παραδείγματος που ανήκει στην εικονική κατηγορία και ελαχιστοποιεί το άθροισμα των αποστάσεων του από τα υπόλοιπα πρότυπα της κατηγορίας.*

4. Ελεγχος του κριτηρίου τερματισμού των επαναλήψεων. *Αν κατά την διαδοχική εκτίμηση των πρωτότυπων αυτά παρέμειναν ίδια, τότε ο αλγόριθμος τερματίζει, διαφορετικά επαναλαμβάνεται απο το δεύτερο βήμα.*

Οι μέθοδοι υπολογισμού των πρωτότυπων με επανάληψη των τριών βημάτων που περιγράφησαν ονομάζονται αλγόριθμοι του τύπου k-μέσων (k-means). Οι αλγόριθμοι αυτοί παρουσιάζουν ευρεία διάδοση σε πρακτικά συστήματα διότι έχει αποδειχθεί ότι κάθε αλγόριθμος του τύπου K-μέσων συγκλίνει πάντα. Μειονέκτημα του αλγόριθμου είναι ότι η σύγκλιση επιτυγχάνεται σε ένα **τοπικό ελάχιστο** του αθροίσματος των αποστάσεων των προτύπων από τα πρωτότυπα.

Παράδειγμα 7 *Τα διαθέσιμα πρότυπα μιας κατηγορίας δίνονται από το σύνολο:*

$$\Omega = \{(1,-1), (2,7), (4,4), (4,2), (4,1), (2,2), (-4,1)\}$$

Χρησιμοποιώντας το τετράγωνο της Ευκλείδιας απόστασης υπολόγισε το βέλτιστο ζευγάρι πρότυπων που ελαχιστοποιεί το κριτήριο της εξίσωσης 2.13 με την μέθοδο των 2-μέσων.

Στον πίναχα που αχολουθεί δίνουμε τα πρωτότυπα, την ταξινόμηση προτύπων των παραδειγμάτων στις εικονικές κατηγορίες, και την συνολική απόσταση τους απο τα πρωτότυπα που κάθε φορά υπολογίζονται. Ξεκινάμε τον αλγόριθμο επιλέγοντας τυχαία σαν πρωτότυπα των εικονικών κατηγοριών τα δύο πρώτα πρότυπα (1,-1) και (2,7).

2.2.9 Υπολογισμός εικονικού πρωτότυπου

Μελετώντας τις προηγούμενες μεθόδους υπολογισμού των προτύπων αναφοράς μπορεί κάποιος να σκεφτεί εύλογα μήπως αντί της επιλογής κάποιου παραδείγματος, θα μπορούσε να προσδιορίσει στον

Πίναχας 2.9: Επαναλήψεις του αλγόριθμου 2-μέσων για την επιλογή δύο πρωτύπων

Επανάληψη	Αρχικές τιμές	Επανάληψη 1	Επανάληψη 2
1ο Πρωτότυπο	(1,-1)	(2,2)	(2,2)
Πρότυπα 1ης κατηγορίας	1,4,5,6,7	1,3,4,5,6,7	1,3,4,5,6,7
Απόσταση 1ης κατηγορίας	70	64	64
2ο Πρωτότυπο	(2,7)	(2,7)	(2,7)
Πρότυπα 2ης κατηγορίας	2,3	2	2
Απόσταση 2ης κατηγορίας	13	0.00000	0.00000
Συνολική απόσταση	83	64	64

χώρο των προτύπων το εικονικό εκείνο πρότυπο το οποίο θα ελαχιστοποιούσε την απόσταση του από τα πρότυπα των παραδειγμάτων κάθε κατηγορίας. Στα πλαίσια αυτού του συλλογισμού σε αυτή την παράγραφο θα αναζητηθεί το διάνυσμα $\mathbf{y}_o$, $\mathbf{y}_o \in S$ το οποίο ελαχιστοποιεί την ακόλουθη συνάρτηση:

$$\mathbf{y}_o = \underset{\mathbf{y}}{argmin} \sum_{i=1}^{M} d(\mathbf{y}_i, \mathbf{y}) \tag{2.14}$$

Οπως είναι γνωστό τα τοπικά ελάχιστα μιας συνεχούς συνάρτησης βρίσκονται στις ρίζες της πρώτης παραγώγου, όταν η δεύτερη παράγωγος έχει θετική τιμή. Συνεπώς οι βέλτιστες τιμές του διανύσματος $\mathbf{y}_o$ βρίσκονται απο την λύση της εξίσωσης:

$$\sum_{i=1}^{M} \frac{\partial}{\partial \mathbf{y}} d(\mathbf{y}_i, \mathbf{y}) = 0 \tag{2.15}$$

Δυστυχώς η επίλυση της εξίσωσης δεν είναι εφικτή για όλες τις συναρτήσεις απόστασης.

Σε αυτό το σημείο είναι σημαντικό επίσης να επισημάνουμε ότι ο τρόπος μέτρησης της απόστασης των προτύπων καθορίζει και τον τρόπο υπολογισμού του βέλτιστου πρωτότυπου από τα πρότυπα των παραδειγμάτων.

Συχνά στις εφαρμογές και ιδιαίτερα στις περιπτώσεις στις οποίες δεν μπορούμε να βρούμε αναλυτική η αλγοριθμική λύση καταφεύγουμε στην λύση της αναμενόμενης (μέσης) τιμής, η οποία δίνει συνήθως ένα πρωτότυπο που είναι κοντά στο πραγματικό βέλτιστο. Ο κανόνας αυτός δεν είναι απόλυτος και θα πρέπει να δοθεί ιδιαίτερη προσοχή αν χρησιμοποιούμε κάποια συνάρτηση απόστασης που δεν ικανοποιεί τις ασθενείς συνθήκες.

Με την μέθοδο της αναμενόμενης τιμής των προτύπων εκπαίδευσης το πρωτότυπο μπορεί να υπολογιστεί από την ακόλουθη σχέση:

$$\mathbf{y} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{y}_i$$

Παράδειγμα 8 *Αν υπάρχουν διαθέσιμα M παραδείγματα βρείτε το βέλτιστο εικονικό πρωτότυπο της κατηγορίας όταν σαν συνάρτηση απόστασης χρησιμοποιείται το τετράγωνο της Ευκλείδιας απόστασης.*

Το πρωτότυπο $\mathbf{y}$ θα προκύψει από την λύση της ακόλουθης εξίσωσης:

$$\sum_{i=1}^{M} \frac{\partial}{\partial \mathbf{y}} (\mathbf{y}_i - \mathbf{y})^T (\mathbf{y}_i - \mathbf{y}) = 0 \Rightarrow$$

$$\sum_{i=1}^{M} \frac{\partial}{\partial \mathbf{y}} (\mathbf{y}_i{}^T \mathbf{y}_i - \mathbf{y}_i{}^T \mathbf{y} - \mathbf{y}^T \mathbf{y}_i + \mathbf{y}^T \mathbf{y}) = 0 \Rightarrow$$

$$-2 \sum_{i=1}^{M} \mathbf{y}_i + 2 \sum_{i=1}^{M} \mathbf{y} = 0 \Rightarrow \quad M\mathbf{y} = \sum_{i=1}^{M} \mathbf{y}_i \Rightarrow \quad \mathbf{y} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{y}_i$$

Τι ωραία έκπληξη. Εκεί που λέγαμε ότι συνήθως μία καλή λύση είναι να υπολογίζουμε σαν πρωτότυπο την μέση τιμή των παραδειγμάτων κάθε κατηγορίας αποδείξαμε ότι, στην περίπτωση που χρησιμοποιούμε σαν συνάρτηση απόστασης το τετράγωνο της Ευκλείδιας απόστασης, το βέλτιστο εικονικό πρωτότυπο (εκείνο που ελαχιστοποιεί το άθροισμα των αποστάσεων του από τα παραδείγματα) δίνεται από την αντίστοιχη αναμενόμενη τιμή τους.

Παράδειγμα 9 *Τα διαθέσιμα πρότυπα μιας κατηγορίας δίνονται από το σύνολο:*

$$\Omega = \{(1, 1.5), (2, 1.8), (3, 3.1), (5, 6.4)\}$$

Υπολογίστε ένα εικονικό πρότυπο με την μέθοδο της αναμενόμενης τιμής και ένα πρότυπο με την μέθοδο επιλογής παραδείγματος.

Τι παρατηρείτε για το συνολικό άθροισμα των αποστάσεων του πρωτότυπου από τα πρότυπα των παραδειγμάτων και πως δικαιολογούνται τα αποτελέσματα για τις ακόλουθες συναρτήσεις απόστασης:

1. *την συνάρτηση απόστασης Tanimoto,*
2. *την Ευκλείδια απόσταση.*
3. *την σταθμισμένη Ευκλείδια απόσταση.*

Πρόγραμμα 1 *Πρόγραμμα υπολογισμού του ελάχιστου και μέγιστου σφάλματος συστήματος ταξινόμησης προτύπων με την βοήθεια του τετραγώνου της Ευκλείδιας απόστασης, το κριτήριο της μικρότερης απόστασης και χρησιμοποιώντας ένα πρότυπο αναφοράς ανά κατηγορία.*

```
function [Rc,Ru,Rep] = ClassMinDistEuclOne(x,c)
#
# [Rc,Ru,Rep] = ClassMinDistEuclOne(x,c)
# Pattern Recognition:
#     Distance measure:      Euclidian
#     Prototypes:            One Center
#     Classification rule:   Minimum Distance
#
# Input
#     x: Pattern Vectors
#     c: Classes
# Output
#     Rc: Correct classification rate using the C-method
#     Ru: Correct classification rate using the U-method
#     Rep: Pattern vectors on each class
#

NumOfClass = max(c) ;
NumOfPatterns = columns(x) ;
Mean = zeros(rows(x),NumOfClass) ;
Rep = zeros(NumOfClass,1) ;

#
```

```
#  C-Error
#
Rc = zeros(NumOfClass,1) ;
for j = 1:NumOfPatterns
    k = c(j) ;
    Rep(k,1) = Rep(k,1) + 1 ;
    Mean(:,k) = Mean(:,k) + x(:,j) ;
endfor
for j=1:NumOfClass
    Mean(:,j) = Mean(:,j) / Rep(j,1) ;
endfor
for i = 1:NumOfPatterns
    for j = 1:NumOfClass
        Dist(j) = (x(:,i) - Mean(:,j))' * ( x(:,i) - Mean(:,j) ) ;
    endfor
    Rec = ArgMin(Dist) ;
    if (Rec == c(i))
       Rc(Rec) = Rc(Rec) + 1 ;
    endif
endfor

#
#  U-Error
#

Ru = zeros(NumOfClass,1) ;
for j = 1:NumOfPatterns
    k = c(j) ;
    Mt = Mean(:,k) ;
    Mean(:,k) = Rep(k,1) / (Rep(k,1)-1) * ( Mean(:,k) - x(:,j) / Rep(k,1)) ;
    for i = 1:NumOfClass
        Dist(i) = (x(:,j) - Mean(:,i))' * ( x(:,j) - Mean(:,i) ) ;
    endfor
    Rec = ArgMin(Dist) ;
    if (Rec == k)
       Ru(Rec) = Ru(Rec) + 1 ;
    endif
    Mean(:,k) = Mt ;
endfor
endfunction


function [Res] = ArgMin(x)
#
#  Res = ArgMin(x)
#  Return the POSITION of the minimum value
#
Num = length(x) ;
Res = 1 ;
MinNum = x(1) ;
for i = 1:Num
    if (x(i) < MinNum)
       MinNum = x(i) ;
       Res = i ;
    endif
endfor
endfunction
```

2.2.10 Εύρεση k-εικονικών πρωτότυπων

Η εύρεση ενός βέλτιστου αριθμού K-εικονικών πρωτότυπων είναι η πλέον αποδοτική μέθοδος αυτόματου υπολογισμού των πρωτότυπων κατηγοριών. Αποτελεί δε την πλέον χρησιμοποιούμενη μέθοδο στις εφαρμογές. Το κριτήριο ελαχιστοποίησης διαφοροποιείται σε μικρό βαθμό από το αντίστοιχο κριτήριο της επιλογής k-προτύπων από τα παραδείγματα.

$$\mathbf{y} = \underset{\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_k}{argmin} \sum_{i=1}^{M} \underset{\mathbf{z}_x}{min}\, d(\mathbf{y}_i, \mathbf{z}_x), \quad \mathbf{z}_x \in \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\} \quad \mathbf{y}_i \in S \tag{2.16}$$

Οπως έχει ήδη προαναφερθεί ο υπολογισμός των πρωτότυπων με βάση την συνάρτηση βελτιστοποίησης δεν είναι δυνατός με αναλυτικό τρόπο διότι η συνάρτηση min δεν είναι παραγωγίσιμη. Γιαυτό τον λόγο καταφεύγουμε πάλι στην χρήση του επαναληπτικού αλγόριθμου των k-μέσων.

Αλγόριθμος 2

1. Αρχικές συνθήκες. *Ξεκινώντας από τυχαίες αρχικές τιμές για τα k πρωτότυπα εκτελούμε το επαναληπτικό τμήμα του αλγόριθμου.*

2. Ταυτοποίηση προτύπων. *Βάσει του κριτηρίου ταξινόμησης τα πρότυπα της κατηγορίας κατατάσονται σε μία από τις K εικονικές κατηγορίες που ορίζονται απο τα πρωτότυπα.*

3. Υπολογισμός πρωτότυπων. *Σε κάθε εικονική κατηγορία ορίζουμε σαν πρωτότυπο εκείνο το διάνυσμα που ελαχιστοποιεί το άθροισμα των αποστάσεων του απο τα πρότυπα της κατηγορίας.*

4. Ελεγχος του κριτηρίου τερματισμού των επαναλήψεων. *Αν κατά την διαδοχική εκτίμηση των πρωτότυπων αυτά παρέμειναν ίδια, τότε ο αλγόριθμος τερματίζει, διαφορετικά επαναλαμβάνεται απο το δεύτερο βήμα.*

Πρόγραμμα 2 *Υπολογισμός k-εικονικών προτύπων από παραδείγματα με συνάρτηση απόστασης το τετράγωνο της Ευκλείδιας απόστασης.*

```
function [Alphabet] = KMeans(Vector,KM,Threshold)
#
#  [Alphabet] = KMeans(Vectors,KM,Threshord)
#
#  Input
#      Vector: Pattern Vectors
#      KM: Size of the alphabet
#      Threshold: Covergence threshold
#  Output
#      Alphabet: Vector alphabet
#

N = rows(Vector) ;
NumOfVect = columns(Vector) ;
if ( NumOfVect <= KM )
printf( "The size of the alphabet is greater than the number of the Vectors\n" ) ;
return ;
endif
Cluster = zeros(NumOfVect,1) ;
Alphabet = zeros(N,KM) ;
Error = 1.0E28 ;
prError = 1.0E30 ;
for i=1:KM
while( 1 )
```

```
Alphabet(:,i) = Vector(:,floor(NumOfVect*rand())+1) ;
Same = 0 ;
for j=1:i-1
if ( Alphabet(:,j) == Alphabet(:,i) )
Same = 1 ;
break ;
endif
endfor
if ( Same == 0 )
break ;
endif
endwhile
endfor
while( (prError-Error)/Error > Threshold )
prError = Error ;
Error = 0.0 ;
NewAlphabet = zeros(N,KM) ;
Rep = zeros(KM,1) ;
for i = 1:NumOfVect
for j = 1:KM
Dist(j) = (Vector(:,i) - Alphabet(:,j))' * ( Vector(:,i) - Alphabet(:,j) ) ;
endfor
Rec = ArgMin(Dist) ;
Error = Error + min(Dist) ;
Cluster(i) = Rec ;
Rep(Rec) = Rep(Rec)+1 ;
NewAlphabet(:,Rec) = NewAlphabet(:,Rec) + Vector(:,i) ;
endfor
for i = 1:KM
Alphabet(:,i) = NewAlphabet(:,i) / Rep(i) ;
endfor
endwhile
endfunction
```

Παράδειγμα 10 *Τα διαθέσιμα πρότυπα μιας κατηγορίας δίνονται από το σύνολο:*

$$\Omega = \{(1,-1),(2,7),(4,4),(4,2),(4,1),(2,2),(-4,1)\}$$

Χρησιμοποιώντας το τετράγωνο της Ευκλείδιας συνάρτησης απόστασης υπολόγισε δύο πρωτότυπα με το κριτήριο των K-εικονικών πρωτότυπων. Σύγκρινε τα αποτελέσματα με το αντίστοιχο παράδειγμα στο οποίο ο υπολογισμός των πρωτότυπων γίνεται με επιλογή k-παραδειγμάτων.

Η συνάρτηση του τετραγώνου της Ευκλείδιας απόστασης μας διευκολύνει στους υπολογισμούς μας διότι κάθε φορά το πρωτότυπο της εικονικής κατηγορίας υπολογίζεται από τις αντίστοιχες μέσες τιμές των προτύπων που έχουν ταξινομηθεί στην κατηγορία.

Στον πίνακα που ακολουθεί δίνουμε τα πρωτότυπα, την ταξινόμηση των παραδειγμάτων στις εικονικές κατηγορίες, και την συνολική απόσταση των προτύπων κάθε κατηγορίας από το αντίστοιχο πρωτότυπο που υπολογίζεται σε κάθε βήμα του αλγόριθμου. Ξεκινάμε του υπολογισμούς μας επιλέγοντας τυχαία σαν πρωτότυπα των εικονικών κατηγοριών τα δύο πρώτα παραδείγματα.

Τα δύο πρωτότυπα που υπολογίσαμε είναι τα $\{(1.4,1),(3,5.5)\}$ που δίνουν συνολικό άθροισμα αποστάσεων 55.7. Το αντίστοιχο συνολικό άθροισμα των αποστάσεων με το κριτήριο επιλογής των k-πρότυπων είναι 64 για τα ίδια παραδείγματα. Παρατηρούμε επίσης ότι και οι δύο παραλλαγές του αλγόριθμου k-μέσων συγκλίνουν μετά από δύο επαναλήψεις.

Γενικά μπορούμε να πούμε ότι η σύγκλιση των αλγορίθμων k-μέσων επηρεάζεται σημαντικά:

1. από το πλήθος των παραδειγμάτων (αυξανομένου του αριθμού των παραδειγμάτων αυξάνει και ο αριθμός των επαναλήψεων του αλγόριθμου μέχρι να επιτευχθεί η σύγκλιση) και,

Πίνακας 2.10: Πίνακας επαναλήψεων του αλγόριθμου εύρεσης k-εικονικών πρωτότυπων

Επανάληψη	Αρχικές τιμές	Επαν. 1	Επαν. 2
1ο Πρωτότυπο	(1,-1)	(1.4,1.0)	(1.4,1.0)
Πρότυπα 1ης κατ.	1,4,5,6,7	1,4,5,6,7	1,4,5,6,7
Απόσταση 1ης κατ.	70	49.2	49.2
2ο Πρωτότυπο	(2,7)	(3,5.5)	(3,5.5)
Πρότυπα 2ης κατ.	2,3	2,3	2,3
Απόσταση 2ης κατ.	13	6.5	6.5
Συνολική απόσταση	83	55.7	55.7

2. την κατανομή τους στον χώρο των μετρήσεων (όσο πιο ομοιόμορφα κατανέμονται τα πρότυπα τόσο αυξάνουν και τα βήματα του αλγόριθμου).

Παράδειγμα 11 *Μετρήστε το ελάχιστο δυνατό σφάλμα συστήματος ταξινόμησης προτύπων τριών κατηγοριών από τα παραδείγματα που είναι διαθέσιμα και δίνονται στον πίνακα 2.11, χρησιμοποιώντας:*

1. Την συνάρτηση απόστασης Tanimoto,

2. το κριτήριο ταξινόμησης της μικρότερης απόστασης και

3. την μέθοδο υπολογισμού του πρωτότυπου κάθε κατηγορίας με την οποία επιλέγεται εκείνο το πρότυπο που ελαχιστοποιεί το άθροισμα των αποστάσεων από τα πρότυπα της κατηγορίας.

Πίνακας 2.11: Πίνακας προτύπων με παραδείγματα τριών κατηγοριών

Πρότυπα της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Πρότυπα της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
Πρότυπα της ω_3	(-1.1,0.4)	(-2.5,4.3)	(-3.0,3.3)	(-2.5,2.6)

Για να μετρήσουμε το σφάλμα του συστήματος ταξινόμησης προτύπων πρέπει αρχικά να υπολογίσουμε τα πρωτότυπα των κατηγοριών. Με βάση τις οδηγίες που έχουμε λάβει υπολογίζουμε αρχικά το άθροισμα των αποστάσεων κάθε πρότυπου από τα υπόλοιπα πρότυπα της κατηγορίας του. Απο τον ακόλουθο πίνακα (2.12) στον οποίο δίνονται οι αθροιστικές τιμές των αποστάσεων προκύπτει ότι, για τις τρείς κατηγορίες τα αντίστοιχα πρωτότυπα είναι το (2.2,2.5) για την ω_1, το (2.8,3.5) για την ω_2 και το (1.5,2.6) για την ω_3.

Πίνακας 2.12: Αθροιστική απόσταση των προτύπων από τα παραδείγματα των κατηγοριών

Πρότυπο της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Συνολική απόσταση	3.177	0.769	1.149	1.841
Πρότυπο της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
Συνολική απόσταση	0.645	0.304	0.578	0.855
Πρότυπο της ω_3	(-1.1,0.4)	(-2.5,4.3)	(-3.0,3.3)	(-2.5,2.6)
Συνολική απόσταση	9.898	4.855	3.632	2.147

Το ελάχιστο σφάλμα του συστήματος ταξινόμησης υπολογίζεται αν μετρήσουμε τις λανθασμένες αναγνωρίσεις στο σύνολο των παραδειγμάτων εκπαίδευσης (μέθοδος-C).

Στον πίνακα που ακολουθεί δίνεται η κατηγορία ταξινόμησης για όλα τα παραδείγματα εκπαίδευσης.

Τα σφάλματα είναι έξι σε ένα σύνολο δώδεκα παραδειγμάτων. Συνεπώς το ελάχιστο σφάλμα του συστήματος ταξινόμησης είναι 50%.

Πίναχας 2.13: Ταξινόμηση των παραδειγμάτων σε κατηγορίες με το κριτήριο της μικρότερης απόστασης

Πρότυπο της ω_1	(1.0,1.3)	(2.2,2.5)	(3.0,3.1)	(4.0,3.5)
Ταυτοποίηση	ω_3	ω_1	ω_2	ω_2
Πρότυπο της ω_2	(2.0,2.9)	(2.8,3.5)	(3.5,2.4)	(4.0,5.2)
Ταυτοποίηση	ω_1	ω_2	ω_2	ω_2
Πρότυπο της ω_3	(1.1,0.4)	(1.5,4.3)	(1.0,3.3)	(1.5,2.6)
Ταυτοποίηση	ω_1	ω_2	ω_3	ω_3

Παράδειγμα 12 *Μετρήστε το ελάχιστο σφάλμα του συστήματος ταξινόμησης προτύπων τριών κατηγοριών με τα παραδείγματα που δίνονται στον πίνακα 2.11. Χρησιμοποιείστε σαν συνάρτηση απόστασης το τετράγωνο της Ευκλείδιας απόστασης, κριτήριο ταξινόμησης στην κατηγορία της μικρότερης απόστασης και μέθοδο υπολογισμού του πρωτότυπου κάθε κατηγορίας την επιλογή εκείνου του παραδείγματος που ελαχιστοποιεί το άθροισμα των αποστάσων από τα πρότυπα των παραδειγμάτων.*

1. Ποιά είναι η διαφοροποίση των αποτελεσμάτων σε σχέση με το σύστημα του προηγούμενου παραδείγματος.

2. Εξηγήστε ποιούς επιπρόσθετους υπολογισμούς θα εκτελούσατε αν σας καλούσαν να αξιολογήσετε τα δύο συστήματα ταξινόμησης.

3. Αξιολογήστε τα συστήματα και αιτιολογείστε τους ισχυρισμούς σας.

Παράδειγμα 13 *Η εκπαίδευση συστήματος ταξινόμησης προτύπων με σύγκριση προτύπων έχει ήδη πραγματοποιηθεί με την μέθοδο υπολογισμού του εικονικού πρωτότυπου που ελαχιστοποιεί την απόσταση του από τα πρότυπα των παραδειγμάτων. Σε κάθε κατηγορία έχουν χρησιμοποιηθεί M παραδείγματα εκπαίδευσης.*

Λαμβάνοντας υπόψη ότι χρησιμοποιείται σαν συνάρτηση απόστασης το τετράγωνο της Ευκλείδιας απόστασης, μπορείτε να επαναπροσδιορίσετε το βέλτιστο πρωτότυπο της κατηγορίας αν σας γνωστοποιήσουν την ύπαρξη ενός νέου παραδείγματος;

Το βέλτιστο εικονικό πρωτότυπο, όταν χρησιμοποιείται το τετράγωνο της Ευκλείδιας συνάρτησης απόστασης, υπολογίζεται από M πρότυπα παραδειγμάτων με την βοήθεια της ακόλουθης σχέσης:

$$\mathbf{y} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i$$

Οταν έχουμε στην διάθεσή μας M+1 πρότυπα παραδειγμάτων, το νέο εικονικό πρωτότυπο δίνεται από την ακόλουθη σχέση:

$$\mathbf{y}' = \frac{1}{M+1}\sum_{i=1}^{M+1}\mathbf{y}_i \Rightarrow \mathbf{y}' = \frac{1}{M+1}\mathbf{y}_{M+1} + \frac{1}{M+1}\sum_{i=1}^{M}\mathbf{y}_i \Rightarrow \mathbf{y}' = \frac{1}{M+1}\mathbf{y}_{M+1} + \frac{M}{M+1}\left(\frac{1}{M}\sum_{i=1}^{M}\mathbf{y}_i\right) \Rightarrow$$

$$\mathbf{y}' = \frac{1}{M+1}\mathbf{y}_{M+1} + \mathbf{y} \tag{2.17}$$

2.3 Συναρτήσεις απόφασης

Μια διαφορετική κατηγορία συστημάτων ταξινόμησης προτύπων μπορεί να υλοποιηθεί με τις λεγόμενες συναρτήσεις απόφασης οι οποίες εκτελούν μία μονοσήμαντη απεικόνιση του χώρου των μετρήσεων των προτύπων στο σύνολο των πραγματικών αριθμών. Πιο συγκεκριμένα για ένα σύστημα ταξινόμησης προτύπων Ν κατηγοριών ορίζονται $N(N-1)$ συναρτήσεις απόφασης κάθε μία των οποίων χωρίζει τον χώρο των μετρήσεων σε δύο περιοχές, τον χώρο που καταλαμβάνουν τα πρότυπα της ω_i και τον χώρο που καταλαμβάνουν τα πρότυπα της ω_j, με ω_i, ω_j συμβολιζουμε δύο τυχαίες κατηγορίες του συστήματος ταξινόμησης. Αν $\mathbf{x}$ είναι τυχαίο πρότυπο τότε οι αντίστοιχες συναρτήσεις απόφασης δίνονται από την σχέσεις $g_{ij}(\mathbf{x})$ με $i,j=1,N$, $j \neq i$ και $g_{ij}(\mathbf{x}) : S \to \mathbb{R}$.

Το πρότυπο $\mathbf{x}$ ταξινομείται στην κατηγορία ω_i όταν:

$$g_{ij}(\mathbf{x}) > 0, \qquad \forall j \neq i \quad j=1,N \tag{2.18}$$

Ο αριθμός των συναρτήσεων απόφασης αυξάνει σημαντικά με τον αριθμό των κατηγοριών του συστήματος ταξινόμησης. Για παράδειγμα όταν το σύστημα αναγνωρίζει δέκα κατηγορίες αντικειμένων τότε απαιτείται η εύρεση $10 \times 9 = 90$ συναρτήσεων απόφασης. Ο μεγάλος αριθμός των συναρτήσεων απόφασης έχει σαν άμεση συνέπεια, την αύξηση της υπολογιστικής πολυπλοκότητας κατά την διαδικασία ταξινόμησης, έχει υψηλές απαιτήσεις μνήμης και απαιτείται μεγάλος αριθμός παραδειγμάτων για την εκπαίδευση.

Υπάρχουν δύο τεχνικές ελάττωσης του αριθμού των συναρτήσεων απόφασης:

1. *Ορισμός της αντίθετης συνάρτησης απόφασης.* Μπορούμε να υποθέσουμε ότι $g_{ji}(\mathbf{x}) = -g_{ij}(\mathbf{x})$ χωρίς να βλάψουμε την γενικότητα των μέχρι τώρα υποθέσεων (ότι μπορούν να υπάρξουν τέτοιες συναρτήσεις απόφασης). Είναι προφανές ότι αν χρησιμοποιήσουμε τον παραπάνω ορισμό ελαττώνουμε τον αριθμό των συναρτήσεων στο μισό, δηλαδή απαιτούνται $N(N-1)/2$ συναρτήσεις απόφασης για σύστημα ταξινόμησης Ν κατηγοριών.

2. *Διάσπαση της συνάρτησης απόφασης.* Σε εφαρμογές στις οποίες ο αριθμός των κατηγοριών είναι μεγάλος, ο αριθμός των συναρτήσεων απόφασης δεν μπορεί να περιοριστεί σημαντικά ακόμα και όταν χρησιμοποιήσουμε τον ορισμό της αντίθετης συνάρτησης απόφασης. Αν έχουμε σύστημα 100 κατηγοριών από το αρχικό σύνολο των 9900 συναρτήσεων μπορούμε, με τον ορισμό της αντίθετης συνάρτησης απόφασης, να ελαττώσουμε τις συναρτήσεις απόφασης σε 4950. Παρόλα αυτά ο αριθμός αυτός εξακολουθεί να είναι πρακτικά απαγορευτικός διότι παρουσιάζει μεγάλη υπολογιστική πολυπλοκότητα και έχει μεγάλες απαιτήσεις μνήμης 4950 * (P+1) δεκαδικών αριθμών στην περίπτωση που χρησιμοποιηθεί η απλούστερη των συναρτήσεων απόφασης, η γραμμική συνάρτηση. Με P συμβολίζουμε την διάσταση του παραμετρικού διανύσματος των προτύπων.

   Οταν έχουμε μεγάλο αριθμό κατηγοριών μπορούμε να πετύχουμε μία δραστική μείωση του αριθμού των συναρτήσεων απόφασης υποθέτοντας ότι μπορούμε να διασπάσουμε κάθε συνάρτηση σε δύο τμήματα. Στο τμήμα εκείνο που αποδίδει την επίδραση του διανύσματος $\mathbf{x}$ στα χαρακτηριστικά γνωρίσματα της κατηγορίας ω_i, συμβολίζεται σαν $r_i(\mathbf{x})$ και ονομάζεται συνάρτηση διάκρισης, και αντίστοιχα το τμήμα εκείνο που αποδίδει την επίδραση του διανύσματος $\mathbf{x}$ στην συνάρτηση απόφασης όσον αφορά την κατηγορία ω_j και συμβολίζεται σαν $r_j(\mathbf{x})$. Υποθέτοντας ότι η συνάρτηση απόφασης μπορεί να αναλυθεί σαν μία διαφορά των συναρτήσεων διάκρισης (discriminant functions) ο αριθμός των συναρτήσεων διάκρισης που απαιτούνται για την πλήρη περιγραφή του συστήματος ταξινόμησης ισούται με τον αριθμό των κατηγοριών:

$$g_{ij}(\mathbf{x}) = r_i(\mathbf{x}) - r_j(\mathbf{x}) = -g_{ji}(\mathbf{x}) \tag{2.19}$$

   Αν υποθέσουμε ότι οι συναρτήσεις απόφασης συστήματος ταξινόμησης 100 κατηγοριών μπορούν πράγματι να διασπαστούν σε διαφορές συναρτήσεων διάκρισης, τότε κατά την εκπαίδευση του συστήματος απαιτείται ο υπολογισμός εκατό μόνον συναρτήσεων.

Το κριτήριο ταξινόμησης για τις συναρτήσεις διάκρισης προκύπτει ισοδύναμα από το κριτήριο των συναρτήσεων απόφασης και είναι το εξής:

$$\mathbf{x} \in \omega_i \quad \alpha\nu\nu \quad r_i(\mathbf{x}) > r_j(\mathbf{x}) \quad \forall j \neq i \tag{2.20}$$

Σε αντίθεση με τις μεθόδους ταξινόμησης με σύγκριση προτύπων στις οποίες για κάθε πρότυπο μπορούμε να αποφασίσουμε σε ποιά κατηγορία αυτό ανήκει, όταν κατασκευάζουμε σύστημα ταξινόμησης προτύπων με συναρτήσεις απόφασης δεν είναι πάντα εφικτή η δυνατότητα λήψης απόφασης. Υπάρχουν πρότυπα για τα οποία δεν υπάρχει κατηγορία τέτοια ώστε όλες οι συναρτήσεις απόφασης να είναι θετικές, δηλαδή δεν μπορεί να ικανοποιηθεί το κριτήριο ταξινόμησης της εξίσωσης 2.18 για καμμία κατηγορία αντικειμένων. Σε αυτή την περίπτωση το σύστημα ταξινόμησης δεν μπορεί να αποφασίσει για την προέλευση του πρότυπου.

Η μετάπτωση του συστήματος ταξινόμησης στην κατάσταση αδυναμίας λήψης κάποιας απόφασης σε μερικές περιπτώσεις είναι επιθυμητή, π.χ. σε συστήματα ασφαλείας, ταξινόμησης δακτυλικών αποτυπωμάτων, ιατρική διάγνωση κ.ο.κ. διότι έχει φυσικό νόημα η απάντηση του συστήματος "Δεν μπορώ να αποφασίσω". Σε άλλες εφαρμογές όπως π.χ. συστήματα αυτόματης πλοήγησης, συστήματα αποφυγής εμποδίων, συστήματα ταξινόμησης ομιλίας, αυτόματης διόρθωσης σφαλμάτων, σε σύστημα οπτικής ταξινόμησης γραμμάτων ή λέξεων κ.ο.κ., η μη δυνατότητα λήψης απόφασης αποτελεί σαφές μειονέκτημα του συστήματος ταξινόμησης.

Θεώρημα 1 *Αν η συνάρτηση απόφασης μπορεί να διασπαστεί σε διαφορά συναρτήσεων διάκρισης τότε δεν υπάρχει πρότυπο το οποίο να μην ταξινομείται σε κάποια κατηγορία.*

Η απόδειξη είναι απλή. Για κάθε διάνυσμα $\mathbf{x}$ οι συναρτήσεις διάκρισης μπορούν να τοποθετηθούν κατά φθίνουσα αριθμητική σειρά, έστω: $r_i(\mathbf{x}) > r_j(\mathbf{x}) ... > r_m(\mathbf{x})$. Αφού η συνάρτηση $r_i(\mathbf{x})$ έχει την μεγαλύτερη αριθμητική τιμή, τότε ικανοποιείται το κριτήριο ταξινόμησης της εξίσωσης 2.20 για την κατηγορία ω_i.

Αναζητώντας κάποιο συνδετικό κρίκο ανάμεσα στις μεθόδους ταξινόμησης προτύπων με σύγκριση και με συναρτήσεις απόφασης μπορούμε να δούμε ότι, αν ορίσουμε σαν απόσταση προτύπων το αντίστροφο της συνάρτησης διάκρισης κάθε κατηγορίας τότε το κριτήριο ταξινόμησης με συναρτήσεις διάκρισης (εξίσωση 2.20) είναι ισοδύναμο με το κριτήριο ταξινόμησης προτύπων με την μικρότερη απόσταση (εξίσωση 2.11):

$$\mathbf{x} \in \omega_i \quad \alpha\nu\nu \quad r_i(\mathbf{x}) > r_j(\mathbf{x}), \quad \forall j \neq i \Leftrightarrow$$

$$\mathbf{x} \in \omega_i \quad \alpha\nu\nu \quad \frac{1}{d_i(\mathbf{x})} > \frac{1}{d_j(\mathbf{x})}, \quad \forall j \neq i \Leftrightarrow$$

$$\mathbf{x} \in \omega_i \quad \alpha\nu\nu \quad d_i(\mathbf{x}) < d_j(\mathbf{x}), \quad \forall j \neq i$$

Η ταξινόμηση προτύπων με συναρτήσεις απόφασης μπορεί να διακριθεί σε εκείνες τις μεθόδους που χρησιμοποιούν γραμμικές συναρτήσεις απόφασης και τις μεθόδους που χρησιμοποιούν μη-γραμμικές συναρτήσεις απόφασης, όπως τα συστήματα ταξινόμησης προτύπων που χρησιμοποιούν νευρωνικά δίκτυα. Τα συστήματα αυτά θα μελετηθούν σε επόμενο κεφάλαιο.

2.3.1 Γραμμικές συναρτήσεις απόφασης

Η γραμμική συνάρτηση απόφασης δίνεται από την σχέση $g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_{N+1}$, όπου $\mathbf{x}$ είναι η διανυσματική παράσταση των μετρήσεων του άγνωστου πρότυπου, $\mathbf{w}$ και w_{N+1} είναι σταθερές παράμετροι. Η επιφάνεια δύο κατηγοριών που ορίζεται από το σύνολο των σημείων που αποτελούν λύση της εξίσωσης $g(\mathbf{x}) = 0$ χωρίζει τον χώρο των μετρήσεων σε δύο περιοχές, των θετικών και των αρνητικών τιμών της.

Θεώρημα 2 *Η διαχωριστική επιφάνεια δύο κατηγοριών είναι κάθετη στο διάνυσμα* $\mathbf{w}$.

Θεωρούμε δύο σημεία επάνω στην διαχωριστική επιφάνεια το $\mathbf{x}_1$ και το $\mathbf{x}_2$. Η ευθεία που ενώνει τα δύο σημεία είναι το διάνυσμα $\mathbf{x}_1 - \mathbf{x}_2$. Για να δούμε αν η διαχωριστική επιφάνεια είναι κάθετη στο διάνυσμα $\mathbf{w}$ αρκεί να αποδείξουμε ότι το διάνυσμα $\mathbf{x}_1 - \mathbf{x}_2$ που ανήκει στην διαχωριστική επιφάνεια είναι κάθετο στο $\mathbf{w}$. Συνεπώς αρκεί να αποδείξουμε ότι:

$$(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{w} = 0$$

Η απόδειξη είναι απλή:

$$(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{w} = \mathbf{x}_1\mathbf{w} + x_{N+1} - (\mathbf{x}_2\mathbf{w} + x_{N+1}) = g(\mathbf{x}_1) - g(\mathbf{x}_2) = 0$$

Μπορούμε επίσης να παρατηρήσουμε ότι μόνο στην περίπτωση κατά την οποία κατασκευάζουμε σύστημα ταξινόμησης δύο κατηγοριών ο αριθμός των συντελεστών των γραμμικών συναρτήσεων απόφασης είναι μικρότερος από τον αντίστοιχο αριθμό των συντελεστών των συναρτήσεων διάκρισης.

Παράδειγμα 14 *Δίνεται σύστημα ταξινόμησης προτύπων τριών κατηγοριών το οποίο αναγνωρίζει πρότυπα στον χώρο δύο διαστάσεων με γραμμικές συναρτήσεις απόφασης. Στον πίνακα που ακολουθεί δίνονται οι συντελεστές για κάθε μία από τις συναρτήσεις των κατηγοριών.*

Πίνακας 2.14: Πίνακας συντελεστών των συναρτήσεων απόφασης

Κατηγορία ω_1	$\mathbf{w}_{12}$ - (1,2,3)	$\mathbf{w}_{13}$ - (1,3,1)
Κατηγορία ω_2	$\mathbf{w}_{21}$ - (-1,-2,-3)	$\mathbf{w}_{23}$ - (-1,2,2)
Κατηγορία ω_3	$\mathbf{w}_{31}$ - (-1,-3,-1)	$\mathbf{w}_{32}$ - (1,-2,-2)

Βάσει των παραδειγμάτων που ακολουθούν μετρήστε την αξιοπιστία του συστήματος ταξινόμησης για κάθε μία από τις κατηγορίες που το σύστημα αναγνωρίζει.

$$\Omega = \{((1,1),\omega_2),((2,1),\omega_1),((2,2),\omega_1),((-2,-1),\omega_2),((2,-1),\omega_1),$$

$$((1,-1),\omega_3),((-1,1),\omega_2),((3,1),\omega_1),((2,-2),\omega_3),((-2,2),\omega_1),$$

$$((1,-2),\omega_3),((-1,-2),\omega_2),((-4,-2),\omega_2)\}$$

Αρχικά μπορούμε να παρατηρήσουμε ότι υπάρχει μία ενδογενής αδυναμία στο σύστημα. Εκ των προτέρων γνωρίζουμε ότι σε κάθε σύστημα ταξινόμησης προτύπων το οποίο χρησιμοποιεί συναρτήσεις απόφασης όταν αυτές δεν μπορούν να αναλυθούν σε συναρτήσεις διάκρισης, τότε είναι δυνατόν να υπάρξουν πρότυπα για τα οποία το σύστημα ταξινόμησης δεν μπορεί να αποφασίσει σε ποιά κατηγορία ανήκουν.

Πίνακας 2.15: Ταξινόμηση των παραδειγμάτων με συναρτήσεις απόφασης

??

	1	2	3	4	5	6	7	8	9	10	11	12	13
Αναγνώριση	ω_1	ω_1	ω_1	ω_2	--	ω_3	ω_1	ω_1	ω_3	ω_1	ω_3	ω_3	ω_2
Κατηγορία	ω_2	ω_1	ω_1	ω_2	ω_1	ω_3	ω_2	ω_1	ω_3	ω_1	ω_3	ω_2	ω_2

Αρχικά υπολογίζουμε το εσωτερικό γινόμενο των διανυσμάτων των συντελεστών των συναρτήσεων απόφασης με το διάνυσμα του άγνωστου πρότυπου $(\mathbf{x}, 1)$. Το πρότυπο ταξινομείται στην κατηγορία εκείνη στην οποία το εσωτερικό γινόμενο όλων των διανυσμάτων των συναρτήσεων απόφασης της κατηγορίας έχουν θετική τιμή. Αν δεν υπάρχει κατηγορία γιά την οποία όλες οι συναρτήσεις απόφασης της κατηγορίας να έχουν θετική τιμή τότε δεν μπορούμε να ταξινομήσουμε το άγνωστο πρότυπο. Κατά την διαδικασία της ταξινόμησης υπάρχει και το ενδεχόμενο να βρούμε περισσότερες της μιάς κατηγορίες οι οποίες να παρουσιάζουν θετική τιμή γιά όλες τις συναρτήσεις απόφασης. Σε αυτή την περίπτωση το άγνωστο πρότυπο ταυτοποιείται σε όλες τις κατηγορίες που ικανοποιούν το κριτήριο ταξινόμησης.

Ακολουθώντας την διαδικασία ταξινόμησης βλέπουμε στον πίνακα τις ταξινομήσεις και την πραγματική κατηγορία κάθε πρότυπου. Με τα στοιχεία του πίνακα υπολογίζουμε το ποσοστό επιτυχών αναγνωρίσεων για κάθε μία κατηγορία ξεχωριστά ως και την συνολική αξιοπιστία της ταξινόμησης.

Επιτυχία ταξινόμησης της $\omega_1 = \frac{4}{5} = 80\%$.

Επιτυχία ταξινόμησης της $\omega_2 = \frac{2}{5} = 40\%$.

Επιτυχία ταξινόμησης της $\omega_3 = \frac{3}{3} = 100\%$.

Συνολική επιτυχία ταξινόμησης $\frac{9}{13} = 69.23\%$.

Παρατηρούμε επίσης ότι το έκτο παράδειγμα δεν μπορούμε να το ταξινομήσουμε απόφασης.

2.3.2 Διαδικασία εκπαίδευσης

Ενας από τους σημαντικότερους λόγους που συνισφέρουν στην διάδοση κάποιας μέθοδου ταξινόμησης προτύπων είναι και η ύπαρξη ενός πετυχημένου αλγόριθμου εκπαίδευσης από παραδείγματα.

Για τις συνάρτησεις απόφασης και διάκρισης υπάρχουν επαναληπτικοί αλγόριθμοι εκπαίδευσης οι οποίοι συγκλίνουν όταν τα παραδείγματα που είναι διαθέσιμα είναι γραμμικά διαχωρίσιμα.

Ορισμός 2 *Τα παραδείγματα λέγονται γραμμικά διαχωρίσιμα όταν για κάθε ζευγάρι κατηγοριών ω_i και ω_j υπάρχει $\mathbf{w} \in \mathbb{R}^{P+1}$ τέτοιο ώστε γιά την συνάρτηση $g_{ij}(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T\mathbf{x} + w_{P+1}$ να ισχύει $g_{ij}(\mathbf{x}) > 0$ για κάθε $\mathbf{x} \in \omega_i$ και $g_{ij}(\mathbf{x}) < 0$ για κάθε $\mathbf{x} \in \omega_j$.*

2.3.3 Ο αλγόριθμος perceptron

Εστω σύστημα ταξινόμησης N κατηγοριών για το οποίο υπάρχουν M_i παραδείγματα σε κάθε μία από τις κατηγορίες του. Αν χρησιμοποιήσουμε γραμμικές συναρτήσεις απόφασης για τις οποίες ισχύει ότι $g_{ij}(\mathbf{x}) = -g_{ji}(\mathbf{x})$ τότε ο ακόλουθος επαναληπτικός αλγόριθμος συγκλίνει αν τα παραδείγματα είναι γραμμικά διαχωρίσιμα.

Αλγόριθμος 3 *1.* Αρχικές τιμές. *Τοποθετούμε τυχαίες τιμές στους συντελεστές των $N(N-1)/2$ συναρτήσεων απόφασης.*

2. Αναγνώριση. *Επιλέγουμε τυχαία μία κατηγορία ω_i και από αυτήν το πρότυπο της $\mathbf{x}_m$. Υπολογίζουμε όλες τις συναρτήσεις απόφασης $g_{ij}(\mathbf{x}_m)$. Για όσες από αυτές η συνάρτηση απόφασης δεν είναι θετική επαναπροσδιορίζουμε του αντίστοιχους συντελεστές ως εξής:*

$$\mathbf{w}_{ij} = \mathbf{w}_{ij} + a(\mathbf{x}_m, 1) \tag{2.21}$$

όπου a είναι θετικός πραγματικός αριθμός.

3. Ελεγχος σύγκλισης. *Αν με τον έλεγχο όλων των παραδειγμάτων εκπαίδευσης δεν πραγματοποιηθεί επαναπροσδιορισμός των συντελεστών των συναρτήσεων απόφασης ο αλγόριθμος τερματίζει. Διαφορετικά επαναλαμβάνεται το προηγούμενο βήμα.*

Ο αλγόριθμος perceptron υπολογίζει τις συναρτήσεις απόφασης για σύστημα ταξινόμησης N κατηγοριών. Ο υπολογισμός των συναρτήσεων ανάμεσα σε δύο κατηγορίες εξαρτάται μόνο από τα αντίστοιχα παραδείγματα των κατηγοριών. Συνεπώς μπορούμε κάθε φορά να υπολογίσουμε ανεξάρτητα τους συντελεστές γιά κάθε ζευγάρι κατηγοριών. Υπολογίζουμε δηλαδή τους συντελεστές της συνάρτησης $g_{ij}(\mathbf{x})$ από τα αντίστοιχα παραδείγματα των κατηγοριών ω_i και ω_j.

Για να βελτιώσουμε την ταχύτητα σύγκλισης του αλγόριθμου, αρχικά τοποθετούμε μία σχετικά μεγάλη αρχική τιμή στον συντελεστή εκπαίδευσης, π.χ. $\alpha = 1$. Στην συνέχεια επιλέγοντας παραδείγματα εκτελούμε το επαναληπτικό τμήμα του αλγόριθμου perceptron. Οταν ολοκληρωθεί ένας ολόκληρος κύκλος ελέγχου των παραδειγμάτων και δεν πραγματοποιηθεί καμμία μεταβολή των συντελεστών της συνάρτησης απόφασης ο αλγόριθμος τερματίζει. Αν το πρότυπο κάποιου παραδείγματος δεν ταξινομηθεί σωστά, τότε το διάνυσμα των συντελεστών βαρύτητας που έχουν αρνητική τιμή επαναπροσδιορίζονται και επαναλαμβάνεται ο κύκλος ελέγχου των παραδειγμάτων. Αν μετά από κάποιες επαναλήψεις όλα τα πρότυπα ταξινομηθούν σωστά τότε το ελάχιστο σφάλμα του συστήματος ταξινόμησης είναι 0% και γνωρίζουμε επίσης ότι τα πρότυπα είναι γραμμικά διαχωρίσιμα. Αν το επαναληπτικό τμήμα του αλγόριθμου δεν τερματίζει τότε ελαττώνουμε τον συντελεστή εκπαίδευσης α και επαναλαμβάνουμε τον αλγόριθμο. Η διαδικασία μείωσης του συντελεστή εκπαίδευσης επαναλαμβάνεται ξανά αν και πάλι ο αλγόριθμος σεν συγκλίνει. Αν η αριθμητική τιμή του συντελεστή εκπαίδευσης μειωθεί σημαντικά χωρίς να πετύχουμε σύγκλιση, τότε μπορούμε να ισχυριστούμε με αρκετά μεγάλη βεβαιότητα ότι τα πρότυπα δεν είναι γραμμικά διαχωρίσιμα.

Πρόγραμμα 3 *Εκπαίδευση και υπολγοσμός του ελάχιστου σφάλματος συστήματος ταξινόμησης προτύπων με συναρτήσεις απόφασης γιά δύο κατηγορίες προτύπων.*

```
function [Rc,Rep] = Perceptron(x1,x2,Lr,MaxRep)
#
#  [Rc,Rep] = Perceptron(x,c,Lr)
#
#  Input
#      x1: Pattern Vectors for the first class
#      x2: Pattern Vectors for the second class
#      Lr: Learning rate
#      MaxRep: Maximum repetitions
#  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

#
#  C-Error
#
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 + Lr * Pat ;
endif
else
      k = floor(rand() * NumOfP2 + 1) ;
      Pat = [x2(:,k);1] ;
      Score = Weights' * Pat ;
      if ( Score >= 0 )
         Weights = Weights - Lr * 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 15 *Εστω ότι δίνονται τα ακόλουθα παραδείγματα.*

$$\Omega = \{((1,2),\omega_1),((-2,1),\omega_2),((2,1),\omega_3),((-0.5,-1),\omega_1),((-1,-1),\omega_2),((-0.5,1),\omega_3)\}$$

Κατασκευάστε σύστημα ταξινόμησης προτύπων τριών κατηγοριών με συναρτήσεις απόφασης.

Χρησιμοποιώντας την υπόθεση ότι $g_{ij}(x) = -g_{ji}(x)$ υπολογίζουμε τον αριθμό των συναρτήσεων σε: $3*(3-1)/2 = 3$. Εστω ότι οι συναρτήσεις απόφασης δίνονται από τις σχέσεις:

$$g_{12}(\mathbf{x}) = -g_{21}(\mathbf{x}) = \mathbf{w}_{12} * (\mathbf{x}, 1)$$

$$g_{13}(\mathbf{x}) = -g_{31}(\mathbf{x}) = \mathbf{w}_{13} * (\mathbf{x}, 1)$$

$$g_{23}(\mathbf{x}) = -g_{32}(\mathbf{x}) = \mathbf{w}_{23} * (\mathbf{x}, 1)$$

Αρχικά τοποθετώ τυχαίες τιμές στους συντελεστές της συνάρτησης απόφασης. Εστω ότι $w_{12} = -w_{21} = (1,0,0)$, $w_{13} = -w_{31} = (0,1,0)$, $w_{23} = -w_{32} = (0,0,1)$.

Διαλέγω κυκλικά ένα προς ένα τα πρότυπα των παραδειγμάτων και εφαρμόζω τον αλγόριθμο perceptron όπως αυτός δίνεται στον πίνακα που ακολουθεί (με συντελεστή εκπαίδευσης $a = 0.5$).

Μετά την σύγκλιση του αλγορίθμου (γραμμικά διαχωρίσιμα παραδείγματα ή ισοδύναμα ελάχιστο σφάλμα 0%), οι συναρτήσεις απόφασης κάθε κατηγορίας είναι οι ακόλουθες:

$$g_{12}(\mathbf{x}) = -g_{21}(\mathbf{x}) = (1,0,0) * (\mathbf{x},1)$$

$$g_{13}(\mathbf{x}) = -g_{31}(\mathbf{x}) = (-1.75,0,2) * (\mathbf{x},1)$$

$$g_{23}(\mathbf{x}) = -g_{32}(\mathbf{x}) = (1,0.5,0.5) * (\mathbf{x},1)$$

Παράδειγμα 16 *Εστω ότι δίνονται τα ακόλουθα παραδείγματα για την κατασκευή συστήματος ταξινόμησης προτύπων με συναρτήσεις απόφασης.*

$$\Omega = \{((1,1),\omega_1),((2,1),\omega_2),((2,2),\omega_1),((-2,-1),\omega_2),((1,2),\omega_1)\}$$

Υπολογίστε τους συντελεστές των συναρτήσεων απόφασης, τοποθετώντας αρχικές τιμές $\mathbf{w} = (1,1,1)$ *και τοποθετείστε τον συντελεστή επαναπροσδιορισμού του αλγόριθμου perceptron στην τιμή a=1. Τι παρατηρείτε κατά την εξέλιξη του αλγόριθμου και ποιές ενέργειες μπορείτε να κάνετε για να βελτιώσετε τα αποτελέσματα.*

2.3.4 Ο αλγόριθμος perceptron για τις συναρτήσεις διάκρισης

Αν οι συναρτήσεις απόφασης μπορούν να εκφραστούν σαν διαφορά γραμμικών συναρτήσεων διάκρισης $g_{ij}(\mathbf{x}) = -g_{ji}(\mathbf{x}) = r_i(\mathbf{x},\mathbf{w}_i) - r_j(\mathbf{x},\mathbf{w}_j)$ τότε ο επαναληπτικός αλγόριθμος perceptron μπορεί να τροποποιηθεί ως εξής:

Αλγόριθμος 4 *1.* Αρχικές τιμές. *Τοποθετούμε τυχαίες τιμές στους συντελεστές των N συναρτήσεων διάκρισης.*

2. Ταξινόμηση. *Επιλέγουμε τυχαία μία κατηγορία ω_i και από αυτήν το πρότυπο $\mathbf{x}_m$. Υπολογίζουμε την τιμή των N συναρτήσεων διάκρισης. Αν υπάρχουν κάποιες κατηγορίες ω_j για τις οποίες ισχύει $r_i(\mathbf{x}_m,\mathbf{w}_i) \le r_j(\mathbf{x}_m,\mathbf{w}_j)$ τότε επαναπροσδιορίζουμε του συντελεστές βαρύτητας των αντίστοιχων κατηγοριών:*

$$\mathbf{w}_i = \mathbf{w}_i + a(\mathbf{x}_m,1) \tag{2.22}$$

$$\mathbf{w}_j = \mathbf{w}_j - a(\mathbf{x}_m,1) \tag{2.23}$$

όπου a είναι θετικός πραγματικός αριθμός.

3. Ελεγχος σύγκλισης. *Αν με τον έλεγχο όλων των παραδειγμάτων εκπαίδευσης δεν πραγματοποιηθεί επαναπροσδιορισμός των συντελεστών των συναρτήσεων διάκρισης ο αλγόριθμος τερματίζει. Διαφορετικά επαναλαμβάνεται το προηγούμενο βήμα.*

Παράδειγμα 17 *Εστω ότι δίνονται τα ακόλουθα παραδείγματα για την κατασκευή συστήματος ταξινόμησης προτύπων δύο κατηγοριών με συναρτήσεις διάκρισης.*

$$\Omega = \{((1,1),\omega_1),((-3,0),\omega_2),((2,2),\omega_1),((-2,-2),\omega_2),((1,2),\omega_1)\}$$

Υπολογίστε τους συντελεστές των συναρτήσεων.

Επειδή το σύστημα αναγνωρίζει δύο κατηγορίες, θα έχουμε δύο γραμμικές συναρτήσεις διάκρισης

$$r_1(\mathbf{x}, \mathbf{w}_1) = w_{11}x_1 + w_{12}x_2 + w_{13}$$

$$r_2(\mathbf{x}, \mathbf{w}_2) = w_{21}x_1 + w_{22}x_2 + w_{23}$$

όπου $\mathbf{x} = (x_1, x_2)$.

Τοποθετώ αρχικές τιμές στους συντελεστές της συνάρτησης απόφασης. Εστω ότι $w_{11} = w_{12} = -1, w_{13} = 1$ και $w_{21} = w_{22} = 1, w_{23} = -1$.

Διαλέγω κυκλικά ένα προς ένα τα πρότυπα των παραδειγμάτων και εφαρμόζω τον αλγόριθμο perceptron με συντελεστή $a = 0.5$. Τα αποτελέσματα του αλγόριθμου σε κάθε επαναληπτικό βήμα του δίνεται στον πίνακα που ακολουθεί.

Οταν ολοκληρωθεί ένας ολόκληρος κύκλος ελέγχου των παραδειγμάτων και δεν πραγματοποιηθεί καμμία μεταβολή των συντελεστών της συνάρτησης απόφασης ο αλγόριθμος τερματίζει (Βήμα 6–10).

Οι συναρτήσεις διάκρισης για κάθε μία από τις κατηγορίες είναι:

$$r_1(\mathbf{x}) = 2x_1 + 0.5x_2 + 1.5$$

$$r_2(\mathbf{x}) = -2x_1 - 0.5x_1 - 1.5$$

Η διαφορά της ταχύτητας σύγκλισης των αλγορίθμων perceptron που χρησιμοποιούνται γιά τον υπολογισμό των συναρτήσεων απόφασης και διάκρισης οφείλεται σε δύο παράγοντες:

Ο πρώτος παράγοντας οφείλεται στην ισχυρή επίδραση που έχουν οι αρχικές συνθήκες στην σύγκλιση του αλγόριθμου. Κακή επιλογή, που συνήθως σημαίνει ότι οι αρχικές τιμές διαφέρουν σημαντικά από τις τελικές τιμές της σύγκλισης, αυξάνει σημαντικά τον αριθμό των επαναλήψεων του αλγόριθμου.

Ο δεύτερος παράγοντας έχει να κάνει με τον αριθμό των άγνωστων παραμέτρων σε σχέση με τον αριθμό των παραδειγμάτων εκπαίδευσης. Γνωρίζουμε ότι, κατά την προσεγγιστική επίλυση πολυδιάστατων προβλημάτων βελτιστοποίησης ο αριθμός των επαναλήψεων του αλγόριθμου έχει άμεση σχέση με τον αριθμό των παραμέτρων που θέλουμε να υπολογίσουμε. Ο αριθμός των επαναλήψεων αυξάνει με την αύξηση των παραμέτρων που θέλουμε να εκτιμήσουμε και μάλιστα με εκθετική τάξη στις περισσότερες των περιπτώσεων. Επίσης όταν αυξάνεται ο αριθμός των παραδειγμάτων εκπαίδευσης συνήθως μειώνεται και το σύνολο των αποδεκτών λύσεων με συνέπεια να αυξάνουν τα βήματα του αλγόριθμου διότι αυξάνει η πιθανότητα η αρχική τιμή που τίθεται στους συντελεστές βαρύτητας να βρίσκεται μακρυά από την κοντινότερη αποδεκτή λύση.

Παράδειγμα 18 *Εστω ότι δίνονται τα ακόλουθα παραδείγματα για την κατασκευή συστήματος ταξινόμησης προτύπων τριών κατηγοριών.*

$$\Omega = \{((-2,0),\omega_1),((2,2),\omega_2),((-1,-4),\omega_3),((-3,1),\omega_1),((3,-1),\omega_2),((1,-3),\omega_3)\}$$

Υπολογίστε τους συντελεστές των γραμμικών συναρτήσεων απόφασης και του συντελεστές των γραμμικών συναρτήσεων διάκρισης και ελέγξετε αν οι ισχυρισμοί που διατυπώσαμε σαν συμπεράσματα της προηγούμενης άσκησης ευσταθούν.

2.3.5 Ο αλγόριθμος ελαχιστοποίησης του σφάλματος εκτίμησης

Ο αλγόριθμος perceptron έχει ένα ισχυρό μειονέκτημα. Συγκλίνει μόνο όταν τα παραδείγματα είναι γραμμικά διαχωρισμένα, διαφορετικά ο επαναπροσδιορισμός των συντελεστων των συναρτήσεων απόφασης ή των συντελεστών των συναρτήσεων διάκρισης καταλήγει σε μία ατέρμονη διαδικασία χωρίς καμμία πιθανότητα σύγκλισης. Σε αυτές τις περιπτώσεις είναι δύσκολο να κρίνουμε αν η αδυναμία σύγκλισης οφείλεται στην μεγάλη τιμή του συντελεστή εκπαίδευσης ή στην παρουσία μη γραμμικών διαχωρίσιμων παραδειγμάτων.

Ο αλγόριθμος ελαχιστοποίησης του σφάλματος εκτίμησης είναι ένας επαναληπτικός αλγόριθμος ο οποίος συγκλίνει ακόμα και στις περιπτώσεις στις οποίες τα πρότυπα των παραδειγμάτων δεν είναι γραμμικά διαχωρισμένα.

Ας δούμε όμως πως προκύπτει ο αλγόριθμος. Οπως έχει προαναφερθεί μελετάμε το πρόβλημα ορισμού της συνάρτησης απόφασης δύο κατηγοριών μία και ο υπολογισμός των συντελεστών εξαρτάται μονάχα από τα παραδείγματα δύο κατηγοριών κάθε φορά. Εστω $g_{12}(\mathbf{x}) = -g_{21}(\mathbf{x})$ και N_1, N_2 είναι ο αριθμός των παραδειγμάτων για τις δύο κατηγορίες αντίστοιχα.

Ορίζω σαν $\mathbf{x}$ την επέκταση του διανύσματος κάθε πρότυπου που προκύπτει με την προσθήκη μιάς νέας διάστασης με τιμή 1, $\mathbf{x} = (\mathbf{x}', 1)$ όπου $\mathbf{x}'$ είναι το διάνυσμα του πρότυπου. Επίσης για κάθε ένα από τα διανύσματα εκπαίδευσης της κατηγορίας ω_2 αντιστρέφω το πρόσημο των συντελεστών $\mathbf{x} = -\mathbf{x}$. Η μετατροπή αυτή μετασχηματίζει το πρόβλημα εκπαίδευσης του συστήματος στην εύρεση των συντελεστών $\mathbf{w}$ οι οποίοι να ικανοποιούν τις ακόλουθες ανισότητες:

$$\mathbf{x}_i\mathbf{w} > 0, \qquad \forall i = 1, ..., M \quad \textit{και} \quad M = N_1 + N_2 \tag{2.24}$$

Ολες τις συναρτήσεις μπορούμε να τις ομαδοποιήσουμε σε:

$$\mathbf{Xw} > \mathbf{0}, \qquad \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M) \tag{2.25}$$

Το πρόβλημα επίλυσης της ανίσωσης μπορεί να αντικατασταθεί από ένα πρόβλημα επίλυσης της εξίσωσης:

$$\mathbf{Xw} = \mathbf{b}, \qquad \mathbf{b} \in (\mathbb{R}^+)^M \tag{2.26}$$

Αναλυτική λύση της 2.26 δεν μπορεί να επιτευχθεί, γιαυτό τον λόγο προσπαθούμε να επιτύχουμε μία επαναληπτική λύση η οποία να βελτιστοποναί διαδοχικά μία συνάρτηση κριτηρίου. Η συνάρτηση αυτή θα μας δείχνει ταυτόχρονα πόσο έχουμε πλησιάσει την επιθυμούμενη λύση του προβλήματος σε κάθε επανάληψη του αλγόριθμου. Η συνάρτηση σφάλματος δίνεται από την εξίσωση:

$$Er(\mathbf{X}, \mathbf{w}, \mathbf{b}) = \frac{1}{2}\sum_{i=1}^{M}(\mathbf{w}^T\mathbf{x}_i - b_i)^2 = \frac{1}{2}||\mathbf{Xw} - \mathbf{b}||^2 \tag{2.27}$$

Η βέλτιστη τιμή του $\mathbf{w}$ προκύπτει από την ελαχιστοποίηση της συνάρτησης του σφάλματος.

Επειδή μπορούμε να ελαχιστοποιήσουμε την συνάρτηση μεταβάλλοντας δύο ανεξάρτητες μεταβλητές τις $\mathbf{w}$, και $\mathbf{b}$ υπολογίζουμε την παράγωγο της συνάρτησης σφάλματος ως προς αυτές τις μεταβλητές.

$$\frac{\partial}{\partial \mathbf{w}}Er(\mathbf{X}, \mathbf{w}, \mathbf{b}) = \mathbf{X}^T(\mathbf{Xw} - \mathbf{b}) \tag{2.28}$$

$$\frac{\partial}{\partial \mathbf{b}}Er(\mathbf{X}, \mathbf{w}, \mathbf{b}) = -(\mathbf{Xw} - \mathbf{b}) \tag{2.29}$$

Με γνωστή την τιμή $\mathbf{b}$, η ελάχιστη τιμή της συνάρτησης βρίσκεται στο σημείο στο οποίο ισχύει:

$$\frac{\partial}{\partial \mathbf{w}}Er(\mathbf{X}, \mathbf{w}, \mathbf{b}) = 0 \Rightarrow \mathbf{X}^T(\mathbf{Xw} - b) = 0 \Rightarrow \mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{b}$$

Η έκφραση $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ ονομάζεται γενικευμένος αντίστροφος πίνακας του $\mathbf{X}$.

Ο συντελεστής $\mathbf{b}$ επαναπροσδιορίζεται βάσει της μεθόδου βηματικής σύγκλισης:

$$\mathbf{b} = \mathbf{b} - c\frac{\partial}{\partial \mathbf{b}}Er(\mathbf{x}, \mathbf{w}, \mathbf{b}) \tag{2.30}$$

Για να έχουμε αποδεκτή λύση δεν πραγματοποιούμε την απλή αντικατάσταση της παραγώγου που έχουμε βρεί διότι γνωρίζουμε ότι οι συνιστώσες του διανύσματος $\mathbf{b}$ πρέπει να είναι πάντα θετικοί αριθμοί. Με την αναδρομική σχέση που δίνουμε στην επόμενη εξίσωση επαναπροσδιορίζουμε τους συντελεστές του διανύσματος $\mathbf{b}$ μόνο στην περίπτωση κατά την οποία αυτοί θα αυξηθούν αριθμητικά σε κάποιες από τις συνιστώσες τους.

$$\mathbf{b} = \mathbf{b} + \frac{c}{2}(\mathbf{Xw} - \mathbf{b} + |\mathbf{Xw} - \mathbf{b}|) \tag{2.31}$$

Με την τεχνική αυτή προσεγγίζουμε την επίλυση της ανίσωσης που επιθυμούμε, ικανοποιώντας ταυτόχρονα και τους περιορισμούς που θέσαμε.

Με βάση τα παραπάνω παραθέτουμε τον αλγόριθμο υπολογισμού των συναρτήσεων απόφασης από παραδείγματα που βασίζεται στην ελαχιστοποίηση του σφάλματος εκτίμησης. Συνήθης επιλογή του συντελεστή c είναι κάποιος μικρός θετικός πραγματικός αριθμός $0 < c \leq 1$. Για αυτές τις τιμές ο αλγόριθμος, που είναι γνωστός και με το όνομα Ho-Kashyap, συγκλίνει ακόμα και αν τα διανύσματα δεν είναι γραμμικά διαχωρίσιμα.

Αλγόριθμος 5 *1.* Αρχικές τιμές. *Τοποθετούμε τυχαίους μικρούς θετικούς αριθμούς στους συντελεστές του διανύσματος* $\mathbf{b}$.

2. Υπολογισμός των συντελεστών της συνάρτησης απόφασης. *Με την σχέση που ακολουθεί πραγματοποιούμε μία εκτίμηση του* $\mathbf{w}$.

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{b} \tag{2.32}$$

3. Ελεγχος σύγκλισης. *Αν* $\mathbf{Xw} > \mathbf{0}$ *τότε τα παραδείγματα είναι γραμμικά διαχωρίσιμα οπότε και ο αλγόριθμος τερματίζει. Αν η παραπάνω συνθήκη δεν ικανοποιείται και οι μεταβολές του* $\mathbf{w}$ *κατά δυο διαδοχικές προσεγγίσεις έχουν μειωθεί σημαντικά τότε έχουμε κατάσταση σύγκλισης του αλγόριθμου. Σε αυτή την περίπτωση γνωρίζουμε ότι τα παραδείγματα δεν είναι γραμμικά διαχωρίσιμα.*

4. Επαναπροσδιορισμός του $\mathbf{b}$. *Η νέα τιμή του διανύσματος* $\mathbf{b}$ *υπολογίζεται από την ακόλουθη αναδρομική σχέση:*

$$\mathbf{b} = \mathbf{b} + \frac{c}{2}(\mathbf{Xw} - \mathbf{b} + |\mathbf{Xw} - \mathbf{b}|) \tag{2.33}$$

5. Επανάληψη του αλγόριθμου. *Το επαναληπτικό τμήμα του αλγόριθμου εκτελείται από το δεύτερο βήμα.*

Πρόγραμμα 4 *Υπολογισμός των συντελεστών βαρύτητας γραμμικού ταξινομητή με την βοήθεια του αλγόριθμου Ho-Kasyap*

```
function w = HoKaOrg( z, ir )

#     [Rc,Rep] = HoKaOrg( x1,x2, ir )
#
# z: Pattern Array        ir: Algorithm Repetitions
```

```
#
# This function estimates the weights of a two classes linear classifier
# using the Ho Kasyap algorithm

Nv = rows(z) ;
Np = columns(z) ;
b = 0.1 * rand( Np, 1 ) ;
piz = pinv(z') ;
w = piz * b ;
b1 = z' * w ;
e = b1 - b ;
i = 0 ;
while( i < ir )
  printf( "Step %d\n", i ) ;
  printf( "Classification Score: %7.4f%%\n", ( 100 *
          NoGreatValMat( b1, 0.0 ) ) / ( rows(b1) * columns(b1)) ) ;
  if ( GreatValMat(b1,0.0) == 1 )
     printf( "Linear Separation of classes in %d repetitions\n", i ) ;
     return ;
  endif
  ea = AbsMat(e) ;
  b = b + 0.5 * ( e + ea ) ;
  w = piz * b ;
  b1 = z' * w ;
  e = b1 - b ;
  i++ ;
endwhile
printf( "Original Ho-Kasyap not convergence in %d repetitions\n", i ) ;
printf( "Classification Score: %7.4f%%\n", ( 100 *
        NoGreatValMat( b1, 0.0 ) ) / ( rows(b1) * columns(b1)) ) ;
endfunction

function a = AbsMat( b )
# usage: a = AbsMat( b )
#
# This function returns the absolute values of the array elements |b(i,j)|
#
r = rows(b) ;
c = columns(b) ;
for i=1:r
  for j = 1:c
     if ( b(i,j) < 0.0 )
        a(i,j) = - b(i,j) ;
     else
        a(i,j) = b(i,j) ;
     endif
  endfor
endfor
endfunction

function a = GreatValMat( b, v )
#usage: a = GreatValMat( b, v )
#
# This function returns 1 if one element of the matrix b
# is less than v otherwise returns 0
#
r = rows(b) ;
c = columns(b) ;
for i=1:r
```

```
  for j = 1:c
     if ( b(i,j) < v )
        a = 0 ;
return ;
     endif
  endfor
endfor
a = 1 ;
endfunction
```

Παράδειγμα 19 *Απο τα ακόλουθα παραδείγματα κατασκευάστε σύστημα ταξινόμησης προτύπων δύο κατηγοριών με συναρτήσεις απόφασης.*

$$\Omega = \{((0,0),\omega_1),((0,1),\omega_1),((1,0),\omega_2),((1,1),\omega_2)\}$$

Οι συντελεστές των γραμμικών συναρτήσεων απόφασης θα υπολογιστούν με την μέθοδο Ho-Kashyap.

Πολλαπλασιάζουμε τα πρότυπα της ω_2 με τον συντελεστή -1, οπότε το διάνυσμα $\mathbf{X}$ θα είναι:

$$\mathbf{X} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

Ο γενικευμένος αντίστροφος πίνακας υπολογίζεται:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ 0.5 & 0.5 & -0.5 & 0.5 \end{pmatrix}$$

Ξεκινώντας τον αλγόριθμο (βήμα 1) με αρχική τιμή $\mathbf{b} = (1,1,1,1)^T$ και $c = 1$, η πρώτη εκτίμηση των συντελεστών της συνάρτησης απόφασης (βήμα 2) είναι:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{b} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

Υπολογίζοντας το κριτήριο τερματισμού του αλγόριθμου (βήμα 3) βρίσκουμε:

$$\mathbf{Xw} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} > \mathbf{0}$$

Τα πρότυπα των παραδειγμάτων είναι γραμμικά διαχωρίσιμα από το πρώτο βήμα του αλγόριθμου. Η ζητούμενη συνάρτηση απόφασης θα είναι: $g_{12}(\mathbf{x}) = (\ -2\ 0\ 1\ ) * (\ x_1\ x_2\ 1\ )^T = -2x_1 + 1$.

Παράδειγμα 20 *Χρησιμοποιώντας τα ακόλουθα παραδείγματα κατασκευάστε σύστημα ταξινόμησης προτύπων δύο κατηγοριών με συναρτήσεις απόφασης.*

$$\Omega = \{((-2,-2),\omega_1),((2,2),\omega_2),((3,-1),\omega_1),((1,-1),\omega_2),((1,1),\omega_1)\}$$

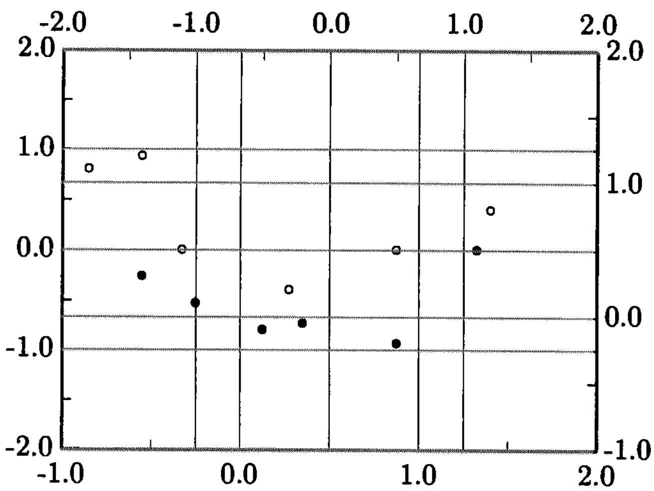
Υπολογίστε τους συντελεστές των γραμμικών συναρτήσεων απόφασης με την βοήθεια του αλγόριθμου Ho-Kashyap και αποδείξτε ότι τα πρότυπα των παραδειγμάτων είναι γραμμικά διαχωρίσιμα.

2.3.6 Μη-γραμμικές συναρτήσεις απόφασης

Οι μη-γραμμικές συναρτήσεις απόφασης μπορούν να θεωρηθούν ότι έχουν στενή συγγένεια με τα νευρωνικά συστήματα ταξινόμησης διότι αυτά περιέχουν μη-γραμμικούς τελεστές. Θεωρώντας όμως ότι χαρακτηριστικό γνώρισμα των νευρωνικών δικτύων είναι η διασύνδεση όμοιων μη-γραμμικών υπολογιστικών μονάδων θα αναφέρουμε σε αυτό το τμήμα των σημειώσεων εκείνες τις μη-γραμμικές μεθόδους που δεν μπορούν να αναλυθούν σαν ένα πολυεπίπεδο διασυνδεδεμένο σύστημα όμοιων μη-γραμμικών τελεστών.

2.3.7 Οι γενικευμένες συναρτήσεις απόφασης

Γραμμικά διαχωρίσιμες κατηγορίες είναι πολύ δύσκολο να βρεθούν σε πρακτικά προβλήματα. Η συνήθης κατάσταση είναι η ύπαρξη μη γραμμικών διαχωριστικών επιφανειών. Στο σχήμα 2.1 απεικονίζονται παραδείγματα προτύπων δύο κατηγοριών τα οποία δεν είναι γραμμικά διαχωρίσιμα.



Σχήμα 2.1: Πρότυπα δύο κατηγοριών τα οποία δεν μπορούν να διαχωριστούν με γραμμικές συναρτήσεις απόφασης

Γιά να κατασκευάσουμε σύστημα ταξινόμησης προτύπων με υψηλά αξιοπιστία σε αυτές τις περιπτώσεις καταφεύγουμε σε μία τεχνική που σαν σκοπό έχει τον μη-γραμμικό μετασχηματισμό του χώρου των προτύπων σε ένα νέο χώρο μετρήσεων στον οποίο τα παραδείγματα να είναι γραμμικά διαχωρίσιμα. Ο σκοπός αυτός επιτυγχάνεται με το ορισμό της ακόλουθης συνάρτησης διάκρισης:

$$d(\mathbf{x}) = \sum_{k=1}^{K+1} w_k f_k(\mathbf{x}) = w_1 f_1(\mathbf{x}) + w_2 f_2(\mathbf{x}) + \ldots + w_K f_K(\mathbf{x}) + w_{K+1} \quad (2.34)$$

όπου $f_k(\mathbf{x}), k = 1, \ldots, K$ είναι μονοδιάστατες συναρτήσεις του παραμετρικού διανύσματος του πρότυπου. Η σχέση των συντελεστών βαρύτητας και των τιμών των συναρτήσεων είναι γραμμική συνεπώς, η συνάρτηση απόστασης μπορεί να γραφεί ισοδύναμα σαν:

$$d(\mathbf{x}) = \mathbf{w}\mathbf{x}^* \quad (2.35)$$

όπου $\mathbf{x}^* = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_K(\mathbf{x}), 1)^T$

Στον μετασχηματισμένο χώρο των μετρήσεων ελέγχουμε (εαν αυτό είναι δυνατόν) την γραμμική διαχωρισιμότητα των παραδειγμάτων και εφαρμόζουμε τον αλγόριθμο perceptron για τις συναρτήσεις διάκρισης των κατηγοριών για τα μετασχηματισμένα διανύσματα $\mathbf{x}^*$.

Με την ίδια μέθοδο μπορούμε να μετασχηματίσουμε τον χώρο των μετρήσεων και να υπολογίσουμε αντί των συναρτήσεων διάκρισης, τις αντίστοιχες συναρτήσεις απόφασης ή με τον αλγόριθμο perceptron ή με την μέθοδο των Ho-Kashyap.

Γενική μέθοδος προσδιορισμού της διανυσματικής παράστασης των συναρτήσεων $f_k(\mathbf{x})$ σε σχέση με την διασπορά των προτύπων στο χώρο των μετρήσεων δυστυχώς δεν υπάρχει. Η αναζήτηση των καταλληλότερων συναρτήσεων αποτελεί θέμα εμπειρίας του σχεδιαστή του συστήματος. Συνηθίζεται να επιλέγονται συναρτήσεις πολυωνυμικής μορφής διότι είναι απλές στον υπολογισμό τους και διότι γνωρίζουμε ότι μπορούν να προσεγγίσουν οποιαδήποτε συνάρτηση άλλου τύπου (ανάπτυξη κατά Taylor).

Ενα δευτερεύον γεγονός που διευκολύνει την διαχωρισιμότητα των προτύπων στον νέο χώρο των μετρήσεων είναι η αύξηση των διαστάσεων του διανύσματος των προτύπων. Η τεχνική αυτή βοηθά την πιθανότητα εύρεσης ενός συντελεστή βαρύτητας $\mathbf{w}$ ο οποίος θα είναι σε θέση να διαχωρίσει τα πρότυπα των παραδειγμάτων, διότι με την αύξηση των διαστάσεων του διανύσματος του πρότυπου έχουμε αυτόματα και ίσου αριθμού αύξηση των διαστάσεων του $\mathbf{w}$, με άμεση συνέπεια την αύξηση των βαθμών ελευθερίας του αλγόριθμου εκπαίδευσης. Η αύξηση των βαθμών ελευθερίας αυξάνει την υπολογιστική πολυπλοκότητα του αλγόριθμου αλλά αυξάνει ταυτόχρονα και την διακριτική ικανότητα της συνάρτησης απόφασης.

Η τεχνητή αύξηση των διαστάσεων του διανύσματος των προτύπων είναι μια πολύ χρήσιμη τεχνική η οποία μπορεί να χρησιμοποιηθεί σε διαφορετικές μεθόδους ταξινόμησης προτύπων. Μια παρόμοια εφαρμογή θα δούμε στο κεφάλαιο στο οποίο περιγράφονται τα συστήματα ταξινόμησης προτύπων που περιέχουν νευρωνικά δίκτυα.

Παράδειγμα 21 *Για τα πρότυπα του σχήματος 2.1*

$$\Omega_1 = \{(-1.8, 1.1), (-1.1, 0.5), (-1.4, 1.2), (-0.3, 0.2), (0.5, 0.5), (1.2, 0.8)\}$$

$$\Omega_2 = \{(-1.4, 0.3), (-1, 0.1), (-0.5, -0.1), (-0.2, -0.05), (0.5, -0.2), (1.1, 0.5)\}$$

βρείτε μία συνάρτηση απόφασης που να διαχωρίζει τα πρότυπα των παραδειγμάτων.

Από την διάταξη των προτύπων στο επίπεδο βλέπουμε ότι δεν υπάρχει ευθεία η οποία να χωρίζει το επίπεδο σε δύο τμήματα κάθε ένα των οποίων να περιέχει όλα τα πρότυπα κάθε κατηγορίας. Μια δευτεροβάθμια εξίσωση της μορφής $ax_2 + bx_1{}^2 + cx_1 + d = 0$, με $\mathbf{x} = (x_1, x_2)^T$ ίσως να μπορούσε να ορίσει μία καμπύλη που να διαχωρίζει τα πρότυπα των κατηγοριών.

Επειδή δεν μπορούμε εύκολα να υπολογίσουμε τους συντελεστές της εξίσωσης, αφήνουμε την εργασία αυτή στον αλγόριθμο εκπαίδευσης, ορίζοντας το διάνυσμα μετασχηματισμού $\mathbf{x}^* = (x_1{}^2, x_2{}^2, x_1x_2, x_1, x_2, 1)^T$.

Ο πίνακας $\mathbf{X}$ μετά τον μετασχηματισμό των προτύπων και την αντιστροφή του προσήμου στα πρότυπα της κατηγορίας Ω_2 γίνεται:

$$\mathbf{X} = \begin{pmatrix} 3.24 & 1.21 & -1.98 & -1.8 & 1.1 & 1 \\ 1.21 & 0.25 & -0.55 & -1.1 & 0.5 & 1 \\ 1.96 & 1.44 & -1.68 & -1.4 & 1.2 & 1 \\ 0.09 & 0.04 & -0.06 & -0.3 & 0.2 & 1 \\ 0.25 & 0.25 & 0.25 & 0.5 & 0.5 & 1 \\ 1.44 & 0.64 & 0.96 & 1.2 & 0.8 & 1 \\ -1.96 & -0.09 & 0.42 & 1.4 & -0.3 & -1 \\ -1 & -0.01 & 0.1 & 1 & -0.1 & -1 \\ -0.25 & -0.01 & -0.05 & 0.5 & 0.1 & -1 \\ -0.04 & -0.0025 & -0.01 & 0.2 & 0.05 & -1 \\ -0.25 & -0.04 & 0.1 & -0.5 & 0.2 & -1 \\ -1.21 & -0.25 & -0.55 & -1.1 & -0.5 & -1 \end{pmatrix}$$

Ο γενικευμένος αντίστροφος πίνακας υπολογίζεται από την σχέση:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T =$$

$$\begin{pmatrix} 0.22 & -0.09 & -0.22 & -0.23 & -0.23 & 0.10 & -0.23 & -0.05 & 0.04 & 0.10 & -0.13 & -0.17 \\ -0.27 & -0.92 & 0.93 & -0.61 & -0.85 & 1.04 & 0.36 & -0.33 & -1.02 & -0.32 & -0.19 & 0.82 \\ -0.38 & -0.18 & 0.15 & -0.20 & -0.30 & 0.71 & -0.14 & -0.38 & -0.44 & -0.04 & 0.46 & 0.36 \\ 0.21 & -0.03 & -0.17 & 0.00 & 0.14 & -0.18 & 0.13 & 0.28 & 0.27 & 0.05 & -0.39 & -0.38 \\ -0.06 & 0.91 & -0.15 & 0.71 & 1.07 & -0.45 & 0.00 & 0.34 & 0.89 & 0.33 & 0.91 & -0.44 \\ -0.06 & 0.07 & 0.03 & 0.2 & 0.17 & -0.09 & 0.04 & -0.04 & -0.142 & -0.20 & -0.26 & -0.05 \end{pmatrix}$$

Ο υπολογισμός των αρχικών τιμών του αλγόριθμου των Ho-Kashyap ολοκληρώθηκε. Για το επαναληπτικό τμήμα του αλγόριθμου θέτω αρχικές τιμές για το διάνυσμα $\mathbf{b} = (111111111111)^T$ και θέτω την σταθερά c στην τιμή 2.

Στον πίνακα δίνονται οι αριθμητικές τιμές της συνάρτησης απόφασης για κάθε ένα απο τα παραδείγματα στα διαδοχικά βήματα του αλγόριθμου. Οπως βλέπουμε στο πρώτο βήμα του αλγόριθμου, όλα τα παραδείγματα της δεύτερης κατηγορίας αρχικά ταξινομούνται σαν πρότυπα της πρώτης (με την αλλαγή του προσήμου τα διανύσματα της δεύτερης κατηγορίας έχουν θετική τιμή). Συνεπώς το σφάλμα του συστήματος ταξινόμησης με την τυχαία αρχική τιμή είναι 50%.

Με το δεύτερο βήμα του αλγόριθμου μονάχα το τελευταίο πρότυπο ταξινομείται λανθασμένα στην πρώτη κατηγορία. Μετά από πέντε διαδοχικούς επαναπροσδιορισμούς των συντελεστών βαρύτητας βλέπουμε ότι οι συντελεστές βαρύτητας που υπολογίστηκαν μπορούν να διαχωρίσουν επιτυχώς όλα τα παραδείγματα της εκπαίδευσης.

Το παραμετρικό διάνυσμα του πρότυπου είναι το $\mathbf{x} = (x_1, x_2)^T$ και λαμβάνοντας υπόψιν τον αρχικό μετασχηματισμό των προτύπων, η τελική μη-γραμμική συνάρτηση απόφασης για τις δύο κατηγορίες είναι η ακόλουθη:

$$g_{12}(\mathbf{x}) = -4.349{x_1}^2 + 16.403{x_2}^2 + 9.402x_1x_2 - 5.726x_1 + 8.382x_2 - 2.333$$

Παράδειγμα 22 *Για τα πρότυπα του σχήματος 2.1 βρείτε μία συνάρτηση απόφασης που να διαχωρίζει τα πρότυπα των παραδειγμάτων χρησιμοποιώντας σαν διάνυσμα μετασχηματισμού το* $\mathbf{x}^* = ({x_1}^2, x_1, x_2, 1)^T$.

Τι παρατηρείτε σε σχέση με τα αποτελέσματα της λύσης που προτείνεται στο προηγούμενο παράδειγμα;

2.3.8 Οι συναρτήσεις δυναμικού

Χωρίς να βλάψουμε την γενικότητα της προσέγγισης μπορούμε να θεωρήσουμε ότι η συνάρτηση απόφασης των κατηγοριών ω_i και ω_j αποτελείται από ένα άθροισμα ορθοκανονικών συναρτήσεων:

$$g_{ij}(\mathbf{x}) = \sum_{k=1}^{\infty} c_i F_i(\mathbf{x}) \tag{2.36}$$

Οι ορθοκανονικές συναρτήσεις που θα επιλεγούν καθορίζονται συνήθως από την κατανομή των παραδειγμάτων και τους περιορισμούς στην υπολογιστική πολυπλοκότητα της διαδικασίας ταξινόμησης. Οι συντελεστές c_i υπολογίζονται κατά την διαδικασία της εκπαίδευσης από τα διαθέσιμα παραδείγματα.

Ορισμός 3 *Συνάρτηση δυναμικού (potential functions) του διανυσματικού χώρου των προτύπων λόγω της παρουσίας του πρότυπου $\mathbf{x}_k$ ορίζεται να είναι το ακόλουθο βαθμωτό μέγεθος:*

$$K(\mathbf{x}, \mathbf{x}_k) = \sum_{k=1}^{\infty} \lambda_i F_i(\mathbf{x}) F_i(\mathbf{x}_k) \tag{2.37}$$

όπου $\lambda_i, i = 1, ..., +\infty$ είναι πραγματικοί αριθμοί.

Το όνομά τους αυτές οι συναρτήσεις το οφείλουν στην φυσική υπόσταση που μπορούμε να δώσουμε σε αυτές. Συγκεκριμένα η συνάρτηση δυναμικού περιγράφει την επίδραση την οποία έχει το παράδειγμα $\mathbf{x}_k$ στον περιβάλλοντα χώρο. Η παρουσία πολλών παραδειγμάτων έχει σαν αποτέλεσμα την αθροιστική παρουσία των αντίστοιχων δυναμικών τους.

Η παρουσία του αθροίσματος απείρων όρων εμποδίζει την κατασκευή πρακτικών συστημάτων, γιαυτό τον λόγο καταφεύγουμε συνήθως στον υπολογισμό ενός πεπερασμένου αθροίσματος ορθοκανονικών συναρτήσεων. Οι συναρτήσεις αυτού του τύπου ονομάζονται ορθοκανονικές συναρτήσεις τύπου-1 (potential functions of Type 1):

$$K(\mathbf{x}, \mathbf{x}_k) = \sum_{k=1}^{Mm} F_i(\mathbf{x}) F_i(\mathbf{x}_k) \tag{2.38}$$

Ορθοκανονικές συναρτήσεις τύπου-2 (potential functions of Type 2), είναι ένας τύπος συμμετρικών συναρτήσεων οι οποίες μπορούν να αναλυθούν σε άθροισμα απείρων όρων ορθοκανονικών συναρτήσεων. Χρησιμοποιούνται συχνότερα στις εφαρμογές διότι παρουσιάζουν σημαντικά μικρότερη υπολογιστική πολυπλοκότητα.

Οι πλέον κοινές συναρτήσεις αυτού του τύπου είναι οι ακόλουθες:

$$K(\mathbf{x}, \mathbf{x}_k) = e^{-a|\mathbf{x}-\mathbf{x}_k|^2} \tag{2.39}$$

$$K(\mathbf{x}, \mathbf{x}_k) = \frac{1}{1 + a|\mathbf{x} - \mathbf{x}_k|^2} \tag{2.40}$$

$$K(\mathbf{x}, \mathbf{x}_k) = \frac{sin(a|\mathbf{x} - \mathbf{x}_k|^2)}{a|\mathbf{x} - \mathbf{x}_k|^2} \tag{2.41}$$

όπου a είναι θετικός πραγματικός αριθμός.

2.3.9 Διαδικασία εκπαίδευσης

Η συνάρτηση απόφασης μπορεί να θεωρηθεί ότι αποτελεί ένα άθροισμα συναρτήσεων δυναμικού. Η επίδραση του δυναμικού που παράγει κάθε παράδειγμα είναι ανεξάρτητη απο την επίδραση που ασχούν τα άλλα παραδείγματα, συνεπώς ξεκινώντας από ένα παράδειγμα μπορούμε διαδοχικά να υπολογίσουμε την συνάρτηση διάκρισης του διανυσματικού χώρου των προτύπων με βάση την ακόλουθη αναδρομική παράσταση:

$$g_{ij}(\mathbf{x}) = g_{ij}(\mathbf{x}) + r_k K(\mathbf{x}, \mathbf{x}_k) \tag{2.42}$$

Ο υπολογισμός του συντελεστή r_k είναι κρίσιμος για την απόδοση του συστήματος ταξινόμησης και υπολογίζεται έτσι ώστε να είναι μηδέν αν το παράδειγμα $\mathbf{x}_k$ ταξινομείται σωστά, διαφορετικά, σε περίπτωση λανθασμένης ταξινόμησης να αυξάνει ή να μειώνει την αριθμητική τιμή της συνάρτησης απόφασης στο σημείο $\mathbf{x}_k$. Η παραπάνω λειτουργία μπορεί να περιγραφεί από την ακόλουθη αλγεβρική σχέση:

$$r_k = \begin{cases} 0, & \alpha\nu \quad \mathbf{x}_k \in \omega_i \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) > 0 \\ 0, & \alpha\nu \quad \mathbf{x}_k \in \omega_j \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) < 0 \\ 1, & \alpha\nu \quad \mathbf{x}_k \in \omega_i \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) \leq 0 \\ -1, & \alpha\nu \quad \mathbf{x}_k \in \omega_j \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) \geq 0 \end{cases} \tag{2.43}$$

Η διαδικασία εκπαίδευσης του συστήματος, ανεξάρτητα από το είδος των συναρτήσεων δυναμικού που θα χρησιμοποιηθούν, επιτυγχάνεται με την εκτίμηση κάθε συνάρτησης απόφασης σε κάθε ζεύγος κατηγοριών ξεχωριστά. Είναι δε η ακόλουθη:

Αλγόριθμος 6 *1.* Αρχικές τιμές. *Επιλέγουμε τον τύπο των δυναμικών συναρτήσεων. Ορίζουμε μία τυχαία αρχική τιμή για την συνάρτηση απόφασης $g_{ij}(\mathbf{x})$.*

2. Επαναπροσδιορισμός της συνάρτησης απόφασης σε θέση τυχαίου παραδείγματος. *Επιλέγουμε τυχαία ένα παράδειγμα, έστω το $\mathbf{x}_k$. Υπολογίζουμε την τιμή της συνάρτησης απόφασης σε αυτό το σημείο βάσει της σχέσης:*

$$g_{ij}(\mathbf{x}) = g_{ij}(\mathbf{x}) + r_k K(\mathbf{x}, \mathbf{x}_k) \tag{2.44}$$

$$r_k = \begin{cases} 0, & \alpha\nu \quad \mathbf{x}_k \in \omega_i \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) > 0 \\ 0, & \alpha\nu \quad \mathbf{x}_k \in \omega_j \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) < 0 \\ 1, & \alpha\nu \quad \mathbf{x}_k \in \omega_i \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) \leq 0 \\ -1, & \alpha\nu \quad \mathbf{x}_k \in \omega_j \quad \kappa\alpha\iota \quad g_{ij}(\mathbf{x}_k) \geq 0 \end{cases} \tag{2.45}$$

3. Σύγκλιση του αλγόριθμου. *Αν ο συντελεστής r_k είναι μηδέν για όλα τα παραδείγματα (ισοδύναμα αν όλα τα παραδείγματα ταξινομηθούν σωστά), τότε ο αλγόριθμος τερματίζει. Διαφορετικά επαναλαμβάνεται το προηγούμενο βήμα.*

Παράδειγμα 23 *Φτιάξτε σύστημα ταξινόμησης δύο κατηγοριών προτύπων με συναρτήσεις απόφασης που αποτελούνται από δυναμικές συναρτήσεις του τύπου:*

$$K(\mathbf{x}, \mathbf{x}_k) = e^{-|\mathbf{x}-\mathbf{x}_k|^2}$$

όταν είναι γνωστά τέσσερα παραδείγματα (το πρόβλημα της πύλης XOR):

$$\Omega = \{((0,0),\omega_1), ((1,1),\omega_1), ((1,0),\omega_2), ((0,1),\omega_2)\}.$$

Το πρόβλημα προσδιορισμού της εξόδου της πύλης XOR είναι το πλέον απλό και διαδεδομένο πρόβλημα ταξινόμησης προτύπων με εκπαίδευση από παραδείγματα τα οποία δεν είναι γραμμικά διαχωρίσιμα. Γιαυτό το λόγο πολλές φορές το πρόβλημα της εκπαίδευσης συστήματος ταξινόμησης προτύπων στην πύλη XOR χρησιμοποιείται γιά να συγκρίνονται οι αποδόσεις των αλγόριθμων εκπαίδευσης.

1. Αρχικά ορίζουμε αρχική τιμή για την συνάρτηση απόφασης των δύο κατηγοριών:

$$g_{12}(\mathbf{x}) = 0$$

2. Κυκλικά επιλέγω τα παραδείγματα και επαναπροσδιορίζω την συνάρτηση απόφασης. Αν $\mathbf{x}_k = (0,0)^T$, τότε $g_{12}(\mathbf{x}_k) = 0$, οπότε και η συνάρτηση απόφασης επαναπροσδιορίζεται ως εξής:

$$r_k = 1 \quad \Rightarrow \quad g_{ij}(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}_k) = e^{-|\mathbf{x}-(0,0)^T|^2} = e^{-|\mathbf{x}|^2}$$

3. Αν $\mathbf{x}_k = (1,1)^T$, τότε $g_{12}(\mathbf{x}_k) = e^{-4} > 0$. Η συνάρτηση απόφασης παραμένει ως έχει, διότι $r_k = 0$:
4. Αν $\mathbf{x}_k = (0,1)^T$, τότε $g_{12}(\mathbf{x}_k) = e^{-1}$. Η συνάρτηση απόφασης με $r_k = -1$ επαναπροσδιορίζεται ως εξής:

$$r_k = -1 \quad \Rightarrow \quad g_{ij}(\mathbf{x}) = g_{ij}(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_k) = e^{-|\mathbf{x}|^2} - e^{-|\mathbf{x}-(0,1)^T|^2}$$

5. Αν $\mathbf{x}_k = (1,0)$, τότε $g_{12}(\mathbf{x}_k) = e^{-1} - e^{-4} > 0$, οπότε και η συνάρτηση απόφασης επαναπροσδιορίζεται διότι $r_k = -1$:

$$r_k = -1 \quad \Rightarrow \quad g_{ij}(\mathbf{x}) = g_{ij}(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_k) = e^{-|\mathbf{x}|^2} - e^{-|\mathbf{x}-(0,1)^T|^2} - e^{-|\mathbf{x}-(1,0)^T|^2}$$

6. Επιλέγω πάλι το πρώτο παράδειγμα $\mathbf{x}_k = (0,0)^T$, τότε $g_{12}(\mathbf{x}_k) = e^0 - 2e^{-1} > 0$. Η συνάρτηση απόφασης παραμένει ως έχει.
7. Αν $\mathbf{x}_k = (1,1)^T$, τότε $g_{12}(\mathbf{x}_k) = e^{-4} - 2e^{-1} < 0$. Η συνάρτηση απόφασης με $r_k = 1$ επαναπροσδιορίζεται ως εξής:

$$r_k = 1 \quad \Rightarrow \quad g_{ij}(\mathbf{x}) = g_{ij}(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_k) = e^{-|\mathbf{x}|^2} - e^{-|\mathbf{x}-(0,1)^T|^2} - e^{-|\mathbf{x}-(1,0)^T|^2} + e^{-|\mathbf{x}-(1,1)^T|^2}$$

8. Αν $\mathbf{x}_k = (0,1)$, τότε $g_{12}(\mathbf{x}_k) = 2e^{-1} - 1 - e^{-4} < 0$. Η συνάρτηση απόφασης παραμένει ως έχει.
9. Αν $\mathbf{x}_k = (1,0)$, τότε $g_{12}(\mathbf{x}_k) = 2e^{-1} - 1 - e^{-4} < 0$. Η συνάρτηση απόφασης παραμένει ως έχει.
10. Αν $\mathbf{x}_k = (1,1)$, τότε $g_{12}(\mathbf{x}_k) = 1 - 2e^{-1} + e^{-4} > 0$. Η συνάρτηση απόφασης παραμένει ως έχει.
11. Αν $\mathbf{x}_k = (0,0)$, τότε $g_{12}(\mathbf{x}_k) = e^{-4} - 2e^{-1} + 1 > 0$. Η συνάρτηση απόφασης παραμένει ως έχει.
12. Ο αλγόριθμος τερματίζει διότι στα τελευταία τέσσερα βήματα του αλγόριθμου, κατά τα οποία ελέγχθηκαν όλα τα πρότυπα των παραδειγμάτων, η συνάρτηση απόφασης δίνει θετική τιμή γιά τα πρότυπα που ανήκουν στην κατηγορία και αρνητική τιμή γιά τα πρότυπα που δεν ανήκουν σε αυτή. Συνεπώς οι συναρτήσεις απόφασης για το πρόβλημα XOR είναι οι ακόλουθες:

$$g_{ij}(\mathbf{x}) = -g_{ji}(\mathbf{x}) = e^{-|\mathbf{x}|^2} - e^{-|\mathbf{x}-(0,1)^T|^2} - e^{-|\mathbf{x}-(1,0)^T|^2} + e^{-|\mathbf{x}-(1,1)^T|^2}$$

2.3.10 Γενικευμένες ακτινικές συναρτήσεις διάκρισης

Οι γενικευμένες ακτινικές συναρτήσεις διάκρισης (Generalized Radial-Basis Functions Networks) πολλές φορές θεωρούνται ότι αποτελούν τμήμα νευρωνικών συστημάτων επεξεργασίας σήματος και ταξινόμησης προτύπων. Εμείς θα μελετήσουμε αυτή την μέθοδο ταξινόμησης σαν τμήμα των δομικών συστημάτων ταξινόμησης προτύπων με συναρτήσεις απόφασης διότι η δομή της μεθόδου, όπως θα δούμε σε επόμενες παραγράφους, μοιάζει με την μέθοδο των συναρτήσεων δυναμικού. Η ίδια λύση προκύπτει και με τα RBF νευρωνικά δίκτυα που περιγράφονται στο κεφάλαιο 4.

Αν θέλουμε να προσεγγίσουμε την συνάρτηση διάκρισης της κατηγορίας ω_i με έναν πεπερασμένο αριθμό συναρτήσεων, τότε μπορούμε να εκφράσουμε την συνάρτηση διάκρισης σαν το γραμμικό άθροισμα αυτών των συναρτήσεων:

$$g_i(\mathbf{x}) = \sum_{k=1}^{K} w_k \varphi_k(\mathbf{x}) \tag{2.46}$$

όπου $\varphi_k(\mathbf{x}), k = 1, K$ είναι K βαθμωτές συναρτήσεις.

Εστω ότι έχουμε στην διάθεσή μας M παραδείγματα. Η συνάρτηση διάκρισης για την κατηγορία ω_i θέλουμε να έχει κάποια ιδιαίτερα χαρακτηριστικά για τα σημεία του χώρου των προτύπων που καλύπτονται από τα διαθέσιμα παραδείγματα. Η συνάρτηση διάκρισης $g_i(\mathbf{x})$ για τα πρότυπα των παραδειγμάτων που ανήκουν στην κατηγορία ω_i θα θέλαμε να έχει μεγαλύτερη αριθμητική τιμή από την τιμή της συνάρτησης διάκρισης των άλλων κατηγοριών.

Συνεπώς το πρόβλημα ορισμού της κατάλληλης συνάρτησης διάκρισης έχει μετατραπεί σε πρόβλημα εύρεσης των K συντελεστών w_k της εξίσωσης 2.46 έτσι ώστε η συνάρτηση να έχει τις επιθυμούμενες τιμές στα σημεία του χώρου που καλύπτονται από τα παραδείγματα:

$$g_i(\mathbf{x}_k) \approx d_k \qquad k = 1, ..., M \tag{2.47}$$

όπου d_k είναι οι επιθυμούμενες τιμές.

Οταν ο αριθμός των παραδειγμάτων είναι μεγαλύτερος από τον αριθμό των συντελεστών ($M > K$) τότε μπορούμε να πετύχουμε μόνο μία προσέγγιση των επιθυμούμενων τιμών. Αν (M=K) τότε υπό κάποιες προυποθέσεις μπορούμε να βρούμε μία μόνο λύση ($g_i(\mathbf{x}_k) = d_k, k = 1, ..., M$).

Η τελευταία περίπτωση συναντάται συχνότερα σε προβλήματα σχεδίασης τροχιάς σε διαστημικά οχήματα στα οποία το πρόβλημα είναι να βρούμε την τροχιά του στο διαστημικό χώρο η οποία πρέπει να διέρχεται **ακριβώς** από κάποια σημεία του χώρου σε θέσεις κοντά ή επάνω σε πλανήτες).

Για τον υπολογισμό των συντελεστών της συνάρτησης θα χρησιμοποιήσουμε μια παραλλαγή της μεθόδου ελαχιστοποίησης του σφάλματος εκτίμησης της συνάρτησης διάκρισης από τις επιθυμούμενες τιμές που έχει ήδη αναφερθεί σε προηγούμενη παράγραφο.

Το σφάλμα εκτίμησης της συνάρτησης διάκρισης για τα M παραδείγματα είναι το ακόλουθο:

$$E = \sum_{j=1}^{M} \left(d_j - \sum_{k=1}^{K} w_k \varphi_k(\mathbf{x}_j) \right)^2 \tag{2.48}$$

Αν οι συναρτήσεις $\varphi_k(\mathbf{x})$ μπορούν να εκφραστούν μέσα από έναν ενιαίο τύπο της μορφής $\varphi_k(\mathbf{x}) = G(\mathbf{x}, \mathbf{s}_k) = G(|\mathbf{x} - \mathbf{s}_k|)$ τότε το σφάλμα εκτίμησης μπορεί να εκφραστεί με την ακόλουθη διανυσματική έκφραση:

$$E = |\mathbf{d} - \mathbf{G}\mathbf{w}| \tag{2.49}$$

όπου $\mathbf{d} = (d_1, d_2, ..., d_M)^T$, $\mathbf{w} = (w_1, w_2, ..., w_K)^T$, και

$$\mathbf{G} = \begin{pmatrix} G(\mathbf{x}_1, \mathbf{s}_1) & G(\mathbf{x}_1, \mathbf{s}_2) & ... & G(\mathbf{x}_1, \mathbf{s}_K) \\ G(\mathbf{x}_2, \mathbf{s}_1) & G(\mathbf{x}_2, \mathbf{s}_2) & ... & G(\mathbf{x}_2, \mathbf{s}_K) \\ ... & ... & ... & ... \\ G(\mathbf{x}_M, \mathbf{s}_1) & G(\mathbf{x}_M, \mathbf{s}_2) & ... & G(\mathbf{x}_M, \mathbf{s}_K) \end{pmatrix}$$

Παραγωγίζοντας την συνάρτηση σφάλματος ως προς $\mathbf{w}$ και υπολογίζοντας τις ρίζες της παράστασης που προκύπτει βρίσκουμε ότι οι συντελεστές δίνονται από την ακόλουθη αναλυτική σχέση:

$$\mathbf{w} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \tag{2.50}$$

Παράδειγμα 24 *Κατασκευάστε σύστημα ταξινόμησης προτύπων το οποίο να αναγνωρίζει τις κατηγορίες του προβλήματος της πύλης XOR. Οι δύο συναρτήσεις διάκρισης πρέπει να έχουν την ακόλουθη μορφή:*

Για την πρώτη κατηγορία,

$$g_1(\mathbf{x}) = w_{11}e^{-|\mathbf{x}-(0,0)^T|^2} + w_{12}e^{-|\mathbf{x}-(1,1)^T|^2}.$$

Για την δεύτερη κατηγορία,

$$g_2(\mathbf{x}) = w_{21}e^{-|\mathbf{x}-(0,1)^T|^2} + w_{22}e^{-|\mathbf{x}-(1,0)^T|^2}.$$

Επειδή ο προσδιορισμός των συντελεστών της συνάρτησης διάκρισης μιάς κατηγορίας είναι ανεξάρτητος των παραδειγμάτων των άλλων κατηγοριών, ο υπολογισμός θα πραγματοποιηθεί για κάθε μία από τις δύο συναρτήσεις ξεχωριστά.

1. Για την συνάρτηση διάκρισης της πρώτης κατηγορίας θέτουμε επιθυμούμενες τιμές $\mathbf{d} = (1,1,0,0)^T$. Αν $\mathbf{w} = (w_{11}, w_{12})^T$ είναι το διάνυσμα των συντελεστών βαρύτητας, τότε

$$\mathbf{G} = \begin{pmatrix} e^0 & e^{-4} \\ e^{-4} & e^0 \\ e^{-1} & e^{-1} \\ e^{-1} & e^{-1} \end{pmatrix}$$

Υπολογίζουμε τον γενικευμένο αντίστροφο πίνακα:

$$(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T = \begin{pmatrix} 0.831926 & -0.186731 & 0.233085 & 0.233085 \\ -0.186731 & 0.831926 & 0.233085 & 0.233085 \end{pmatrix}$$

Η λύση προκύπτει από την εξίσωση:

$$\mathbf{w} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} = (0.645195, 0.645195)^T$$

Συνεπώς η συνάρτηση απόφασης για την πρώτη κατηγορία είναι:

$$g_1(\mathbf{x}) = 0.645195e^{|\mathbf{x}-(0,0)^T|^2} + 0.645195e^{|\mathbf{x}-(1,1)^T|^2}$$

2. Ακολουθεί ο υπολογισμός της συνάρτησης διάκρισης για την δεύτερη κατηγορία η οποία εκτελείται με τον ίδιο τρόπο:

   Επιθυμούμενες τιμές της εξόδου της συνάρτησης διάκρισης $\mathbf{d} = (0,0,1,1)^T$, με $\mathbf{w} = (w_{21}, w_{22})^T$, και

$$\mathbf{G} = \begin{pmatrix} e^{-1} & e^{-1} \\ e^{-1} & e^{-1} \\ e^0 & e^{-4} \\ e^{-4} & e^0 \end{pmatrix}$$

Ο γενικευμένος αντίστροφος πίνακας είναι:

$$(\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T = \begin{pmatrix} 0.233085 & 0.233085 & 0.831926 & -0.186731 \\ 0.233085 & 0.233085 & -0.186731 & 0.831926 \end{pmatrix}$$

Η λύση προκύπτει από την εξίσωση:

$$\mathbf{w} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} = (0.466169, 0.466169)^T$$

Η συνάρτηση απόφασης για την πρώτη κατηγορία είναι:

$$g_1(\mathbf{x}) = 0.466169e^{|\mathbf{x}-(0,1)^T|^2} + 0.466169e^{|\mathbf{x}-(1,0)^T|^2}$$

Το σύστημα ταξινόμησης που κατασκευάσαμε θα ταξινομεί τα πρότυπα στην κατηγορία εκείνη της οποίας η συνάρτηση διάκρισης έχει την μεγαλύτερη αριθμητική τιμή.

2.4 Λυμένα Προβλήματα

Πρόβλημα 1 *Σε σύστημα ταξινόμησης προτύπων δύο κατηγοριών κατά την εκπαίδευση χρησιμοποιήθηκαν τα παραδείγματα που δίνονται στον πίνακα 2.19. Δεδομένου ότι το σύστημα ταξινόμησης προτύπων χρησιμοποιεί,*

1. *το τετράγωνο της Ευκλείδειας συνάρτησης απόστασης,*
2. *το κριτήριο ταξινόμησης της μικρότερης απόστασης και*
3. *σαν μέθοδο υπολογισμού του πρωτότυπου κάθε κατηγορίας, το εικονικό πρότυπο που ελαχιστοποιεί το άθροισμα των αποστασεών του από τα πρότυπα των παραδειγμάτων της κατηγορίας.*

Υπολογίστε το σφάλμα του συστήματος ταξινόμησης και συγκρίνετε το αποτέλεσμα που θα λάβετε με το αντίστοιχο σφάλμα του συστήματος που θα προκύψει αν τα παραδείγματα της εκπαίδευσης αυξηθούν σε έξι. Τα δύο παραδείγματα που προστίθενται είναι το $\{(12.5, 2.5), (5.0, 5.0)\}$ *για την* ω_1, *και το* $\{(3.0, 1.0), (1.0, 1.1)\}$ *για την* ω_2.

Το σφάλμα του συστήματος ταξινόμησης θα υπολογιστεί μέσα σε ένα διάστημα εμπιστοσύνης, ελάχιστο των οποίων είναι το σφάλμα που υπολογίζεται με την μέθοδο-C και μέγιστο αυτό που θα υπολογιστεί με την μέθοδο του αχρησιμοποίητου παραδείγματος.

Με την μέθοδο-C και τα τέσσερα πρότυπα του πίνακα υπολογίζουμε το πρωτότυπο κάθε κατηγορίας και ταξινομούμε τα οχτώ παραδείγματα με κριτήριο την ελάχιστη απόσταση. Το ελάχιστο σφάλμα του συστήματος είναι 75.00% για την ω_1 και 50.00% για την ω_2. Το συνολικό σφάλμα μετρημένο με την μέθοδο-C είναι 62.50%. Το ελάχιστο σφάλμα χρησιμοποιώντας και τα έξι πρότυπα για κάθε κατηγορία είναι 50.00% για την ω_1 και 16.66% για την ω_2. Το συνολικό σφάλμα είναι 33.33%.

Το μέγιστο σφάλμα το υπολογίζουμε με την μέθοδο του αχρησιμοποίητου παραδείγματος. Επιλέγουμε ένα πρότυπο και στην συνέχεια χρησιμοποιούμε τα υπόλοιπα για να εκπαιδεύσουμε το σύστημα. Με το αχρησιμοποίητο παράδειγμα μετράμε το σφάλμα του συστήματος.

Χρησιμοποιώντας τα τέσσερα παραδείγματα ανά κατηγορία, έχουμε σαν μέγιστο σφάλμα του συστήματος 100.00% για την ω_1 και 100.00% για την ω_2, δηλαδή μέγιστο σφάλμα μετρημένο με την μέθοδο-U 100.00%.

Χρησιμοποιώντας έξι παραδείγματα ανά κατηγορία, έχουμε σαν μέγιστο σφάλμα του συστήματος 50.00% για την ω_1 και 66.66% για την ω_2, δηλαδή μέγιστο σφάλμα μετρημένο με την μέθοδο-U 58.33%.

Αν διατάξουμε τα λάθη που υπολογίσαμε σε σχέση με τον αριθμό παραδειγμάτων που κάθε φορά είναι διαθέσιμα βλέπουμε ότι, με οχτώ παραδείγματα εκπαίδευσης το σφάλμα του συστήματος βρίσκεται μέσα στα όρια [62.5%,100.0%] και για δώδεκα παραδείγματα το σφάλμα βρίσκεται στα όρια [33.3%,58.3%].

Μπορούμε να παρατηρήσουμε ότι το διάστημα εμπιστοσύνης του σφάλματος στην πρώτη περίπτωση είναι 37.5%, ενώ στην δεύτερη περίπτωση στην οποία ο αριθμός των διαθέσιμων παραδειγμάτων αυξάνει, το διάστημα εμπιστοσύνης του σφάλματος ελαττώνεται στο 25%.

Γενικότερα μπορούμε να πούμε ότι αυξανομένου των αριθμού των διαθέσιμων παραδειγμάτων το σφάλμα με τις μεθόδους C και U υπολογίζεται με αυξανόμενη ακρίβεια διότι το διάστημα εμπιστοσύνης έχει την τάση να ελαττώνεται. Παρατηρείται επίσης και μία στατιστική τάση να ελαττώνεται και το συνολικό σφάλμα του συστήματος διότι αυξανομένου του αριθμού των παραδειγμάτων ο υπολογισμός της μνήμης του συστήματος τείνει να προσεγγίσει τις πραγματικές του τιμές.

2.5 Αλυτα Προβλήματα

Ασκηση 1 *Δώστε μία λύση για το πρόβλημα ταξινόμησης των εξόδων της λογικής πύλης XOR με γενικευμένες συναρτήσεις απόφασης.*

Ασκηση 2 *Κατασκευάστε σύστημα ταξινόμησης δύο κατηγοριών προτύπων με συναρτήσεις απόφασης που αποτελούνται από δυναμικές συναρτήσεις του τύπου:*

$$K(\mathbf{x}, \mathbf{x}_k) = \frac{1}{1 + a|\mathbf{x} - \mathbf{x}_k|^2}$$

για το πρόβλημα της πύλης XOR

Ασκηση 3 *Κατασκευάστε σύστημα ταξινόμησης προτύπων το οποίο να αναγνωρίζει δύο κατηγορίες δεδομένου ότι είναι γνωστά τα ακόλουθα παραδείγματα:*

$$\Omega = \{((1,1),\omega_1),((1,0),\omega_1),((0,0),\omega_2),((0,1),\omega_2)\}$$

Οι δύο συναρτήσεις διάκρισης πρέπει να έχουν την μορφή:

$$g_1(\mathbf{x}) = w_1 e^{|\mathbf{x}-(0,0)^T|^2}$$

$$g_2(\mathbf{x}) = w_2 e^{|\mathbf{x}-(0,1)^T|^2}$$

Ποιό είναι το ελάχιστο σφάλμα του συστήματος ταξινόμησης που φτιάξατε;

Ασκηση 4 *Εστω ότι στα παραδείγματα της προηγούμενης άσκησης προστίθεται και το παράδειγμα* $((0.5,0.5),\omega_1)$. *Με ίδια τα υπόλοιπα δεδομένα υπολογίστε το ελάχιστο σφάλμα του συστήματος ταξινόμησης.*

1. *Τι παρατηρείτε και πως μπορείτε να βελτιώσετε την αξιοπιστία του συστήματος;*
2. *Ποιό είναι το μέγιστο σφάλμα του συστήματος ταξινόμησης;*

Πίνακας 2.16: Βήματα επαναπροσδιορισμού των συντελεστών βαρύτητας των συναρτήσεων απόφασης

Πρότυπο	Συν.	Τιμή	Επαν.	w	Συν.	Τιμή	Επαν.	w
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	2	Οχι	--
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	-1	Ναι	0.25,0.5,0.5
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	1	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	1	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	-1.5	Ναι	-0.75,0,0	$g_{32}(\mathbf{x})$	-1	Ναι	-1,-0.5,0.5
(-0.5,1)	$g_{31}(\mathbf{x})$	0.375	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	-0.75	Ναι	-0.25.1.0.5
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	-0.625	Ναι	0,0.5,1
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	-1.5	Ναι	-1,0,0.5	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	1	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	-0.5	Ναι	-0.5,1,1
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	-0.25	Ναι	-0.25,0.5,1.5
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	-1.5	Ναι	-1.25,0,1	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	1.625	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	-0.25	Ναι	-0.75,1,1.5
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	0.125	Οχι	--
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	-1	Ναι	-1.75,0.5,1	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	2.375	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	0.25	Οχι	--
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	-0.375	Ναι	-1.5,0,1.5
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	1.5	Οχι	--	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	2.25	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	0	Ναι	-1,1,2
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	0.5	Οχι	--
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	-1	Ναι	-2,0.5,1.5	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	3	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	0.5	Οχι	--
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	2	Ναι	-1.75,0,2
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	1.5	Οχι	--	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	2.875	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--
(1,2)	$g_{12}(\mathbf{x})$	1	Οχι	--	$g_{13}(\mathbf{x})$	0.25	Οχι	--
(-0.5,-1)	$g_{12}(\mathbf{x})$	0.5	Οχι	--	$g_{13}(\mathbf{x})$	1.125	Οχι	--
(-2,1)	$g_{21}(\mathbf{x})$	2	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(-1,-1)	$g_{21}(\mathbf{x})$	1	Οχι	--	$g_{23}(\mathbf{x})$	2	Οχι	--
(2,1)	$g_{31}(\mathbf{x})$	1.5	Οχι	--	$g_{32}(\mathbf{x})$	2	Οχι	--
(-0.5,1)	$g_{31}(\mathbf{x})$	2.875	Οχι	--	$g_{32}(\mathbf{x})$	0.5	Οχι	--

Πίνακας 2.17: Βήματα επαναπροσδιορισμού των συντελεστών βαρύτητας της συνάρτησης διάκρισης

ΒΗΜΑ	πρότυπο	Νέο w_1	$r_1(x)$	Νέο w_2	$r_2(x)$	Επαν.
1	$(1,1) \in \omega_1$	-0.5,-0.5,1.5	-1.0	0.5,0.5,-1.5	1.0	Ναι
2	$(2,2) \in \omega_1$	0.5,0.5,2	-0.5	-0.5,-0.5,-2	0.5	Ναι
3	$(1,2) \in \omega_1$	--	3.5	--	-3.5	Οχι
4	$(-3,-0) \in \omega_2$	2,0.5,1.5	0.5	-2,0.5,1.5	-0.5	Ναι
5	$(-2,-2) \in \omega_2$	--	-3.5	--	-0.5	Οχι
6	$(1,1) \in \omega_1$	--	4	--	-4	Οχι
7	$(2,2) \in \omega_1$	--	6.5	--	-6.5	Οχι
8	$(1,2) \in \omega_1$	--	4.5	--	-4.5	Οχι
9	$(-3,-0) \in \omega_2$	--	-4.5	--	4.5	Οχι
10	$(-2,-2) \in \omega_2$	--	-3.5	--	3.5	Οχι

Πίνακας 2.18: Αριθμητική τιμή της συνάρτηση απόφασης κατά τα διαδοχικά βήματα επαναπροσδιορισμού των συντελεστών βαρύτητας

ΒΗΜΑ	1	2	3	4	5	6	7	8	9	10	11	12
1	2.77	1.31	2.52	0.97	2.75	6.04	-1.53	-1.01	-0.71	-0.8	-1.49	-4.61
2	2.63	0.70	6.38	0.19	3.09	8.73	0.58	0.21	0.31	0.64	2.38	-3.17
3	3.54	1.28	10.02	0.46	3.55	9.60	1.21	0.26	0.31	0.95	4.69	-1.7
4	3.9	1.56	12.12	0.60	3.86	10.11	1.75	0.41	0.41	1.2	6.19	-0.78
5	4.1	1.64	13.48	0.64	4.07	10.42	2.27	0.67	0.60	1.43	7.15	-0.21
6	4.21	1.74	14.41	0.71	4.23	10.63	2.59	0.80	0.68	1.55	7.8	0.18

Πίνακας 2.19: Πρότυπα παραδειγμάτων δύο κατηγοριών

Πρότυπα της ω_1	(4.0,4.3)	(-2.2,4.5)	(-3.0,1.5)	(1.0,3.5)
Πρότυπα της ω_2	(-1.3,5.9)	(4.0,1.0)	(-2.0,-0.5)	(0.5,0.5)