Άσκηση 1(a): Έστω ένα πρόβλημα ταξινόμησης σε δύο ισοπίθανες κλάσεις με βάση διανύσματα χαρακτηριστικών $\mathbf{x}$ που ακολουθούν πολυδιάστατες κανονικές κατανομές για κάθε κλάση (class conditional pdfs) με κοινό πίνακα συνδιασποράς (covariance matrix) $\mathbf{S}$ και διαφορετικά διανύσματα μέσων τιμών $\mathbf{m}_1$ και $\mathbf{m}_2$. Αποδείξτε ότι η πιθανότητα λάθους του Bayesian ταξινομητή ελάχιστου σφάλματος δίνεται από την

$$P_e \;=\; \int_{D/2}^{+\infty} \frac{1}{\sqrt{2\,\pi}}\;\exp\left(-\,z^2/2\right) \mathrm{d}\,z \;\;,$$

όπου το D είναι η απόσταση Mahalanobis μεταξύ των μέσων τιμών $\mathbf{m}_1$ και $\mathbf{m}_2$.

Λύση:

Ο ταξινομητής κατά Bayes βασίζεται στο τεστ του λόγου της πιθανοφάνειας,

$$L_{12} = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} \;>(<)\;1$$

(λόγω του ότι οι κλάσεις είναι ισοπίθανες), ή ισοδύναμα, αν θεωρήσουμε την τυχαία μεταβλητή

$$u \;=\; \ln L_{12} \;=\; \ln p(\mathbf{x}|\omega_1) \;-\; \ln p(\mathbf{x}|\omega_2) \;\;,$$

η ταξινόμηση βασίζεται στο αν η u είναι θετική (κλάση ω_1) ή αρνητική (κλάση ω_2).
Από τα δεδομένα της άσκησης (και την (2.27) του βιβλίου) έχουμε:

$$\begin{aligned} u \;&=\; -\,\frac{1}{2}\,(\mathbf{x}-\mathbf{m}_1)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{x}-\mathbf{m}_1) \;\;+\;\; \frac{1}{2}\,(\mathbf{x}-\mathbf{m}_2)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{x}-\mathbf{m}_2) \\ &=\;\; \mathbf{x}^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2) \;\;-\;\; \frac{1}{2}\,(\mathbf{m}_1+\mathbf{m}_2)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2) \end{aligned}$$

(λόγω και της συμμετρίας του κοινού πίνακα συνδιασποράς $\mathbf{S}$). Η τυχαία μεταβλητή u είναι λοιπόν γραμμικός συνδυασμός των l τυχαίων μεταβλητών που αποτελούν τα στοιχεία του l-διάστατου διανύσματος $\mathbf{x}$, που είναι από κοινού Γκαουσιανές, κατά συνέπεια η u ακολουθεί την κανονική κατανομή. Η μέση τιμή και διασπορά της u φυσικά εξαρτάται από την κλάση (ω_1 ή ω_2) από την οποία προέρχεται το $\mathbf{x}$. Στην περίπτωση που πρόκειται για την κλάση ω_1, η μέση τιμή της u είναι:

$$\begin{aligned} \mu_{1,u} \;\;&=\;\; E_1[u] \;\;=\;\; \mathbf{m}_1^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2) \;\;-\;\; \frac{1}{2}\,(\mathbf{m}_1+\mathbf{m}_2)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2) \\ &=\;\; \frac{1}{2}\,(\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2) \;\;=\;\; \frac{1}{2}\,D^2 \;\;, \end{aligned}$$

όπου D είναι η απόσταση Mahalanobis μεταξύ των $\mathbf{m}_1$ και $\mathbf{m}_2$ (εξίσωση (2.49) του βιβλίου). Επίσης, η διασπορά της είναι ίση με

$$\sigma_{1,u}^2 \;\;=\;\; E_1[\,(u-E_1[u])^{\,2}\,]$$

$$=\;\; E_1[\;(\;\mathbf{x}^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{m}_1-\mathbf{m}_2) \;-\; \frac{1}{2}\,(\mathbf{m}_1+\mathbf{m}_2)^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{m}_1-\mathbf{m}_2) \;-\; \frac{1}{2}\,(\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{m}_1-\mathbf{m}_2)\;)^{\,2}\,]$$

$$= \; E_1[\, (\,(\mathbf{x}-\mathbf{m}_1)^{\mathrm{T}}\, \mathbf{S}^{-1}(\mathbf{m}_1-\mathbf{m}_2)\,)^2\,] \;=\; (\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\, \mathbf{S}^{-1}\, E_1[\, (\mathbf{x}-\mathbf{m}_1)(\mathbf{x}-\mathbf{m}_1)^{\mathrm{T}}\,]\, \mathbf{S}^{-1}\, (\mathbf{m}_1-\mathbf{m}_2)$$

$$= \; (\mathbf{m}_1-\mathbf{m}_2)^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_1-\mathbf{m}_2) \;=\; D^2 \;.$$

Παρόμοια με τα παραπάνω, στην περίπτωση που πρόκειται για την κλάση ω_2, έχουμε:

$$\mu_{2,u} \;=\; -\,\frac{1}{2}\, D^2 \;\;, \qquad \sigma^2_{1,u} \;=\; D^2 \;\;.$$

Μπορούμε τώρα, με βάση τα παραπάνω, να εκφράσουμε περαιτέρω αναλυτικά την πιθανότητα λάθους του Bayesian ταξινομητή, που από την (2.7) του βιβλίου ισούται με:

$$P_e \;=\; \frac{1}{2}\, P(u<0, \mathbf{x}\in\omega_1) \;+\; \frac{1}{2}\, P(u>0, \mathbf{x}\in\omega_2) \;\;,$$

ως (βλέπε επίσης και την (2.24))

$$P_e \;=\; \frac{1}{2}\int_{-\infty}^{0} \frac{1}{\sqrt{2\,\pi}\; D}\; \exp\,(-\,\frac{(u\,-\,D^2/2)^2}{2\,D^2})\,\mathrm{d}u \;+\; \frac{1}{2}\int_{0}^{+\infty} \frac{1}{\sqrt{2\,\pi}\; D}\; \exp\,(-\,\frac{(u\,+\,D^2/2)^2}{2\,D^2})\,\mathrm{d}u$$

$$= \; \int_{0}^{+\infty} \frac{1}{\sqrt{2\,\pi}\; D}\; \exp\,(\,-\,\frac{(u\,+\,D^2/2)^2}{2\,D^2})\,\mathrm{d}u \;\;=\;\; \int_{D/2}^{+\infty} \frac{1}{\sqrt{2\,\pi}}\; \exp\,(-\,z^2/2)\,\mathrm{d}z \;\;,$$

όπου η προτελευταία ισότητα προέκυψε αντικαθιστώντας στο αριστερό ολοκλήρωμα $u \leftarrow -u$, και η τελευταία ισότητα προέκυψε με την αντικατάσταση $z \leftarrow (\,(u/D) + (D/2)\,)$.

Άσκηση 1(b): Δείξτε ότι οι εκτιμητές μέγιστης πιθανοφάνειας (maximum likelihood estimators) των άγνωστων παραμέτρων $\mathbf{m}$ και $\mathbf{S}$ της πολυδιάστατης Gaussian συνάρτησης πυκνότητας πιθανότητας

$$p(\mathbf{x}) \;=\; \frac{1}{(2\,\pi)^{l/2}\,(\det \mathbf{S})^{1/2}}\;\exp\left(-\,\frac{1}{2}\,(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\,\mathbf{S}^{-1}\,(\mathbf{x}-\mathbf{m})\right)$$

με βάση N ανεξάρτητες παρατηρήσεις / δείγματά της $\mathbf{x}_1\,,\mathbf{x}_2\,,\ldots,\mathbf{x}_N$, δίνονται από τις

$$\hat{\mathbf{m}}_{ML} \;=\; \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \quad , \quad \hat{\mathbf{S}}_{ML} \;=\; \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\hat{\mathbf{m}}_{ML})\,(\mathbf{x}_n-\hat{\mathbf{m}}_{ML})^{\mathrm{T}} \;\; .$$

Ερευνήστε αν οι παραπάνω εκτιμητές είναι αμερόληπτοι (unbiased), και, αν όχι, διορθώστε τους κατάλληλα ώστε να είναι.

Λύση:

Γράφοντας την συνάρτηση πιθανοφάνειας σε λογαριθμική μορφή παίρνουμε (βλέπε (2.27) και (2.58))

$$L\,(\,\mathbf{m}\,,\mathbf{S}\,) \;=\; \sum_{n\,=\,1}^{N}\ln p\,(\,\mathbf{x}_n\,;\mathbf{m}\,,\mathbf{S}\,) \;=\; -\,\frac{N\,l}{2}\,\ln\,(2\,\pi) - \frac{N}{2}\,\ln|\,\mathbf{S}\,| - \frac{1}{2}\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,)\,\mathbf{S}^{\,-\,1}\,(\,\mathbf{x}_n-\,\mathbf{m}\,)^{\mathrm{T}}$$

$$=\; -\,\frac{N\,l}{2}\,\ln\,(2\,\pi)\;-\;\frac{N}{2}\,\ln|\,\mathbf{S}\,|\;-\;\frac{1}{2}\,\mathrm{tr}\left[\,\mathbf{S}^{\,-\,1}\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n\;-\;\mathbf{m}\,)\,(\,\mathbf{x}_n\;-\;\mathbf{m}\,)^{\mathrm{T}}\right] \;\; .$$

Παραγωγίζοντας παίρνουμε

$$\mathrm{d}\,L\,(\,\mathbf{m}\,,\mathbf{S}\,) \;=\; -\,\frac{N}{2}\,\mathrm{tr}\,[\,\mathbf{S}^{\,-\,1}\,\mathrm{d}\,\mathbf{S}\,]$$

$$-\,\frac{1}{2}\,\mathrm{tr}\left[\,\mathbf{S}^{\,-\,1}\,(\,\mathrm{d}\,\mathbf{S}\,)\,\mathbf{S}^{\,-\,1}\,\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,)\,(\,\mathbf{x}_n-\,\mathbf{m}\,)^{\mathrm{T}}\;\;-\;\;2\,\mathbf{S}^{\,-\,1}\,\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,)\,(\,\mathrm{d}\,\mathbf{m}\,)^{\mathrm{T}}\right] \;\; .$$

Εξισώνοντας το παραπάνω με το μηδέν, απαιτούμε για την επίτευξη μέγιστου οι όροι που πολλαπλασιάζουν τα $\mathrm{d}\,\mathbf{m}$ και $\mathrm{d}\,\mathbf{S}$ να είναι ίσοι με το μηδέν. Για τον πρώτο έχουμε εύκολα

$$\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,) \;=\; 0 \;\;\Leftrightarrow\;\; \hat{\mathbf{m}}_{ML} \;=\; \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n \;\; ,$$

ενώ για τον όρο που περιλαμβάνει το $\mathrm{d}\,\mathbf{S}$ έχουμε

$$\frac{1}{2}\,\mathrm{tr}\left(\mathbf{S}^{\,-\,1}\,(\,\mathrm{d}\,\mathbf{S}\,)\,\left[\,N\,\mathbf{I}\;-\;\mathbf{S}^{\,-\,1}\,\sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,)\,(\,\mathbf{x}_n-\,\mathbf{m}\,)^{\mathrm{T}}\right]\right) \;=\; 0 \;\;\Rightarrow$$

$$\Rightarrow\;\; N\,\mathbf{S} \;=\; \sum_{n\,=\,1}^{N}(\,\mathbf{x}_n-\,\mathbf{m}\,)\,(\,\mathbf{x}_n-\,\mathbf{m}\,)^{\mathrm{T}} \;\;\Rightarrow\;\; \hat{\mathbf{S}}_{ML} \;=\; \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\hat{\mathbf{m}}_{ML})\,(\mathbf{x}_n-\hat{\mathbf{m}}_{ML})^{\mathrm{T}} \;\; .$$

Εξετάζοντας τους παραπάνω εκτιμητές για αμεροληψία, έχουμε:

$$E\left[\hat{\mathbf{m}}_{ML}\right] \;=\; \frac{1}{N}\sum_{n=1}^{N} E\left[\mathbf{x}_n\right] \;=\; \frac{1}{N}\, N\,\mathbf{m} \;=\; \mathbf{m} \;\;,$$

δηλαδή ο εκτιμητής μέγιστης πιθανοφάνειας $\hat{\mathbf{m}}_{ML}$ του διανύσματος $\mathbf{m}$ είναι αμερόληπτος (unbiased). Για τον εκτιμητή $\hat{\mathbf{S}}_{ML}$ έχουμε (προσθέτοντας και αφαιρώντας το $\mathbf{m}$)

$$\begin{aligned}
E\left[\hat{\mathbf{S}}_{ML}\right] \;&=\; \frac{1}{N}\sum_{n=1}^{N} E\left[\,(\,(\,\mathbf{x}_n - \mathbf{m}\,) \,-\, (\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,)\,(\,(\,\mathbf{x}_n - \mathbf{m}\,) \,-\, (\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,)^{\mathrm{T}}\,\right] \\
&=\; \frac{1}{N}\sum_{n=1}^{N} E\left[\,(\,\mathbf{x}_n - \mathbf{m}\,)\,(\,\mathbf{x}_n - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;+\; \frac{1}{N}\sum_{n=1}^{N} E\left[\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)^{\mathrm{T}}\,\right] \\
&\quad -\frac{1}{N}\sum_{n=1}^{N} E\left[\,(\,\mathbf{x}_n - \mathbf{m}\,)\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;-\; \frac{1}{N}\sum_{n=1}^{N} E\left[\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,(\,\mathbf{x}_n - \mathbf{m}\,)^{\mathrm{T}}\,\right]
\end{aligned}$$

Στη συνέχεια, για κάθε έναν από τους τέσσερεις όρους του αθροίσματος έχουμε:

$$\begin{aligned}
&E\left[\,(\,\mathbf{x}_n - \mathbf{m}\,)\,(\,\mathbf{x}_n - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; \mathbf{S} \;\;, \\
&E\left[\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; E\left[\,\frac{1}{N}\sum_{i=1}^{N}(\,\mathbf{x}_i - \mathbf{m}\,)\,\frac{1}{N}\sum_{j=1}^{N}(\,\mathbf{x}_j - \mathbf{m}\,)^{\mathrm{T}}\,\right] \\
&\qquad\qquad\qquad\qquad =\; \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[(\,\mathbf{x}_i - \mathbf{m}\,)\,(\,\mathbf{x}_j - \mathbf{m}\,)^{\mathrm{T}}\right] \;=\; \frac{1}{N^2}\,N\mathbf{S} \;=\; \frac{1}{N}\,\mathbf{S} \;\;, \\
&E\left[\,(\,\mathbf{x}_n - \mathbf{m}\,)\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; E\left[\,(\,\mathbf{x}_n - \mathbf{m}\,)\,\frac{1}{N}\sum_{j=1}^{N}(\,\mathbf{x}_j - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; \frac{1}{N}\,\mathbf{S} \;\;, \\
&E\left[\,(\,\hat{\mathbf{m}}_{ML} - \mathbf{m}\,)\,(\,\mathbf{x}_n - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; E\left[\,\frac{1}{N}\sum_{i=1}^{N}(\,\mathbf{x}_i - \mathbf{m}\,)\,(\,\mathbf{x}_n - \mathbf{m}\,)^{\mathrm{T}}\,\right] \;=\; \frac{1}{N}\,\mathbf{S} \;\;.
\end{aligned}$$

Αντικαθιστώντας τα παραπάνω στην προηγούμενη εξίσωση, έχουμε:

$$E\left[\hat{\mathbf{S}}_{ML}\right] \;=\; \mathbf{S} \,+\, \frac{1}{N}\,\mathbf{S} \,-\, \frac{1}{N}\,\mathbf{S} \,-\, \frac{1}{N}\,\mathbf{S} \;=\; \frac{N-1}{N}\,\mathbf{S} \;\;.$$

Κατά συνέπεια ο εκτιμητής μέγιστης πιθανοφάνειας είναι μόνο ασυμπτωτικά αμερόληπτος (καθώς $N \to \infty$). Είναι εύκολο ωστόσο να διορθωθεί ο εκτιμητής, παίρνοντας τον

$$\hat{\mathbf{S}} \;=\; \frac{N}{N-1}\,\hat{\mathbf{S}}_{ML} \;=\; \frac{1}{N-1}\sum_{n=1}^{N}(\mathbf{x}_n - \hat{\mathbf{m}}_{ML})\,(\mathbf{x}_n - \hat{\mathbf{m}}_{ML})^{\mathrm{T}} \;\;.$$

Ο καινούργιος αυτός εκτιμητής είναι προφανώς αμερόληπτος.

Άσκηση 1(c): Βρείτε την εξίσωση που περιγράφει την επιφάνεια διαχωρισμού (και σχεδιάστε την) στο πρόβλημα Bayesian ταξινόμησης δύο ισοπίθανων κλάσεων με βάση 3-D διανύσματα χαρακτηριστικών $\mathbf{x}$ που ακολουθούν 3-D κανονικές κατανομές για κάθε κλάση (class conditional pdfs) με κοινό πίνακα συνδιασποράς (covariance matrix)

$$\mathbf{S} = \begin{bmatrix} 0.3 & 0.1 & 0.1 \\ 0.1 & 0.3 & -0.1 \\ 0.1 & -0.1 & 0.3 \end{bmatrix}$$

και διανύσματα μέσων τιμών $\mathbf{m}_1 = [0, 0, 0]^{\mathrm{T}}$ και $\mathbf{m}_2 = [0.5, 0.5, 0.5]^{\mathrm{T}}$.

Λύση:

Με βάση τα δεδομένα της άσκησης, το υπερεπίπεδο απόφασης (επιφάνεια διαχωρισμού) δίνεται από την εξίσωση (βλέπε και (2.35) στο βιβλίο):

$$-\frac{1}{2}(\mathbf{x}-\mathbf{m}_1)^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m}_1) + \ln\frac{P(\omega_1)}{(2\,\pi)^{l/2}\,|\mathbf{S}|^{1/2}} = -\frac{1}{2}(\mathbf{x}-\mathbf{m}_2)^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m}_2) + \ln\frac{P(\omega_2)}{(2\,\pi)^{l/2}\,|\mathbf{S}|^{1/2}} \ .$$

Καθώς οι δύο κλάσεις είναι ισοπίθανες ($P(\omega_1) = P(\omega_2) = 1/2$), και λόγω και του κοινού πίνακα συνδιασποράς $\mathbf{S}$, η εξίσωση διαχωρισμού των κλάσεων απλοποιείται στην:

$$\mathbf{x}^{\mathrm{T}}\,\mathbf{S}^{-1}\,\mathbf{m}_1 \;-\; \frac{1}{2}\,\mathbf{m}_1^{\mathrm{T}}\,\mathbf{S}^{-1}\,\mathbf{m}_1 \;=\; \mathbf{x}^{\mathrm{T}}\,\mathbf{S}^{-1}\,\mathbf{m}_2 \;-\; \frac{1}{2}\,\mathbf{m}_2^{\mathrm{T}}\,\mathbf{S}^{-1}\,\mathbf{m}_2 \ .$$

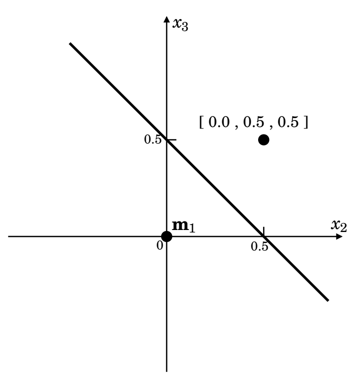
Αντικαθιστώντας τις τιμές των $\mathbf{m}_1$, $\mathbf{m}_2$, και $\mathbf{S}$, και βρίσκοντας εύκολα τον αντίστροφο πίνακα της συνδιασποράς,

$$\mathbf{S}^{-1} = \begin{bmatrix} 5.0 & -2.5 & -2.5 \\ -2.5 & 5.0 & 2.5 \\ -2.5 & 2.5 & 5.0 \end{bmatrix} \ ,$$

παίρνουμε την εξίσωση

$$2.5\,x_2 \;+\; 2.5\,x_3 \;-\; 1.25 \;=\; 0 \quad \Leftrightarrow \quad x_2 \;+\; x_3 \;-\; 0.5 \;=\; 0 \ .$$

Η επιφάνεια αυτή διαχωρισμού των κλάσεων είναι παράλληλη στον άξονα του x_1 (πρόκειται προφανώς για ειδική περίπτωση). Η τομή της με το κάθετο σε αυτήν επίπεδο των αξόνων (x_2 , x_3) σχεδιάζεται στο παρακάτω σχήμα.



Άσκηση 2: Έστω δύο κλάσεις ω_1 και ω_2 με τα ακόλουθα 2-D διανύσματα χαρακτηριστικών εκπαίδευσης (δύο για κάθε κλάση, γραμμένα σε μορφή $\mathbf{x} = [\,x_1\,, x_2\,]^{\mathrm{T}}$):

$$\omega_1 : \;\; [\,-1\,, 2\,]^{\mathrm{T}}\,,\;\; [\,0\,, 1\,]^{\mathrm{T}}$$

$$\omega_2 : \;\; [\,0\,, -1\,]^{\mathrm{T}}\,,\;\; [\,1\,, -2\,]^{\mathrm{T}}\;\; .$$

Έστω επίσης και ένα διάνυσμα δοκιμής $\mathbf{x}_{\text{test}} = [\,-3\,, 2\,]^{\mathrm{T}}$.

(a) Σχεδιάστε τα παραπάνω διανύσματα στο 2-D χώρο με άξονες τα x_1, x_2. Είναι οι κλάσεις γραμμικά διαχωρίσιμες (με βάση τα διανύσματα εκπαίδευσης); Αν ναι, δώστε μία συνάρτηση διάκρισης $g(\mathbf{x}) = w_1\,x_1 + w_2\,x_2 + w_0$ με διάνυσμα συντελεστών $\mathbf{w} = [\,\mathbf{w_1}\,, \mathbf{w_2}\,, \mathbf{w_0}\,]^{\mathrm{T}}$, που διαχωρίζει τις δύο κλάσεις με $g(\mathbf{x}) > 0$ για $\mathbf{x} \in \omega_1$ και $g(\mathbf{x}) < 0$ για $\mathbf{x} \in \omega_2$ (για τα διανύσματα εκπαίδευσης) και που ταξινομεί το διάνυσμα δοκιμής στην κλάση ω_1. Δώστε επίσης μία συνάρτηση διάκρισης της ίδιας μορφής που να ταξινομεί το διάνυσμα δοκιμής στην κλάση ω_2.

(b) Βρείτε τη γραμμική συνάρτηση διάκρισης των δύο κλάσεων με βάση τον αλγόριθμο perceptron (σε μορφή batch), αρχικοποιώντας τον αλγόριθμο με τιμές $\mathbf{w}(0) = [\,0\,, 0\,, 0\,]^{\mathrm{T}}$ και χρησιμοποιώντας βήμα $\rho = 1.0$. Σχεδιάστε την και ταξινομήστε το $\mathbf{x}_{\text{test}}$ με αυτήν.

(c) Βρείτε τη γραμμική συνάρτηση διάκρισης των δύο κλάσεων με βάση την ελαχιστοποίηση του αθροίσματος τετραγωνικών σφαλμάτων (minimum least squares error). Πόσο είναι το ελάχιστο σφάλμα εκπαίδευσης που πετυχαίνει ο αλγόριθμος; Σχεδιάστε την συνάρτηση διάκρισης και ταξινομήστε το $\mathbf{x}_{\text{test}}$ με αυτήν.

(d) Βρείτε τη γραμμική συνάρτηση διάκρισης των δύο κλάσεων με βάση την μέθοδο των μηχανών διανυσματικής στήριξης (support vector machines). Ποια είναι τα διανύσματα στήριξης (support vectors); Ποιο είναι το περιθώριο (margin) μεταξύ των κλάσεων που επιτυγχάνει ο αλγόριθμος; Σχεδιάστε την συνάρτηση διάκρισης και ταξινομήστε το $\mathbf{x}_{\text{test}}$ με αυτήν.

Λύση:

(a) Στο σχήμα της επόμενης σελίδας έχουν σχεδιαστεί τα τέσσερα διανύσματα εκπαίδευσης και το διάνυσμα δοκιμής. Οι κλάσεις είναι προφανώς γραμμικά διαχωρίσιμες, όπως φαίνεται και από τις τρεις ευθείες (γραμμικές συναρτήσεις διάκρισης) που έχουν σχεδιαστεί [έχουν σχεδιαστεί και οι ευθείες που προκύπτουν από τα υπόλοιπα υπο-ερωτήματα της άσκησης].

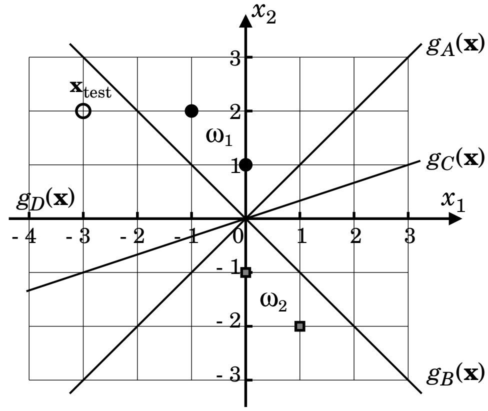
Δύο γραμμικές συναρτησεις διάκρισης που ικανοποιούν το πρώτο υπο-ερώτημα είναι οι

$$g_A(\mathbf{x}) \;=\; -x_1 + x_2 \;=\; 0\,,\;\; \text{δηλαδή}\;\; \mathbf{w}_A = [\,-1\,, 1\,, 0\,]^{\mathrm{T}}\;\; ,$$

και

$$g_B(\mathbf{x}) \;=\; x_1 + x_2 \;=\; 0\,,\qquad \text{δηλαδή}\;\; \mathbf{w}_B = [\,1\,, 1\,, 0\,]^{\mathrm{T}}\;\; .$$

Είναι εύκολο να δούμε ότι $g_A(\mathbf{x}) > 0$ για $\mathbf{x} = [\,-1\,, 2\,, 1\,]^{\mathrm{T}}$ και για $\mathbf{x} = [\,0\,, 1\,, 1\,]^{\mathrm{T}}$, όπως και ότι $g_A(\mathbf{x}) < 0$ για $\mathbf{x} = [\,0\,, -1\,, 1\,]^{\mathrm{T}}$ και για $\mathbf{x} = [\,1\,, -2\,, 1\,]^{\mathrm{T}}$, δηλαδή τα διανύσματα εκπαίδευσης ταξινομούνται σωστά (έχουμε επαυξήσει όλα τα διανύσματα με την σταθερά 1). Εύκολα βλέπουμε επίσης ότι $g_A(\mathbf{x}_{\text{test}}) = \mathbf{w}_A^{\mathrm{T}}\,\mathbf{x} = [\,-1\,, 1\,, 0\,]\,.\,[\,-3\,, 2\,, 1\,]^{\mathrm{T}} = 5 > 0$, κατά συνέπεια το διάνυσμα δοκιμής $\mathbf{x}_{\text{test}}$ ταξινομείται στην κλάση ω_1.



Επαναλαμβάνοντας τα παραπάνω για την ευθεία $g_B(\mathbf{x})$, παρατηρούμε ότι η ταξινόμηση των διανυσμάτων εκπαίδευσης γίνεται σωστά, και ότι $g_B(\mathbf{x}_{\text{test}}) = [\,1\,,1\,,0\,]\,.\,[\,-3\,,2\,,1\,]^{\mathrm{T}} = -1 < 0$, κατά συνέπεια το διάνυσμα δοκιμής $\mathbf{x}_{\text{test}}$ ταξινομείται τώρα στην κλάση ω_2.

(b) Είναι προφανές ότι η αρχική τιμή του διανύσματος $\mathbf{w}(0)$ δεν ταξινομεί σωστά κανένα από τα διανύσματα εκπαίδευσης. Κατά συνέπεια, η πρώτη επανάληψη του αλγορίθμου perceptron σε μορφή batch (εξίσωση (3.9) του βιβλίου) θα δώσει:

$$\mathbf{w}(1) = \mathbf{w}(0) - \sum_{n=1}^{4} \delta_{\mathbf{x}_n} \mathbf{x}_n = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} -1.0 \\ 2.0 \\ 1.0 \end{bmatrix} + \begin{bmatrix} 0.0 \\ 1.0 \\ 1.0 \end{bmatrix} - \begin{bmatrix} 0.0 \\ -1.0 \\ 1.0 \end{bmatrix} - \begin{bmatrix} 1.0 \\ -2.0 \\ 1.0 \end{bmatrix} = \begin{bmatrix} -2.0 \\ 6.0 \\ 0.0 \end{bmatrix} .$$

Παρατηρούμε εύκολα ότι η γραμμική συνάρτηση διάκρισης που προκύπτει είναι η

$$g_C(\mathbf{x}) \;=\; -\,2\,x_1 + 6\,x_2 \;=\; 0\,, \;\; \text{δηλαδή με } \; \mathbf{w}_C = [\,-2\,,6\,,0\,]^{\mathrm{T}} \;,$$

και ταξινομεί σωστά και τα τέσσερα διανύσματα εκπαίδευσης. Κατά συνέπεια ο αλγόριθμος έχει συγκλίνει. Η συνάρτηση διάκρισης έχει σχεδιαστεί στο παραπάνω σχήμα. Εύκολα επίσης βλέπουμε ότι $g_C(\mathbf{x}_{\text{test}}) = \mathbf{w}_C^{\mathrm{T}}\mathbf{x} = [\,-2\,,6\,,0\,]\,.\,[\,-3\,,2\,,1\,]^{\mathrm{T}} = 18 > 0$, κατά συνέπεια το διάνυσμα δοκιμής $\mathbf{x}_{\text{test}}$ ταξινομείται στην κλάση ω_1.

(c) Ακολουθώντας τη μεθοδολογία των εξισώσεων (3.44) και (3.45) του βιβλίου, σχηματίζουμε τον πίνακα $\mathbf{X}$ (χρησιμοποιώντας τα επαυξημένα διανύσματα δεδομένων εκπαίδευσης) και το διάνυσμα $\mathbf{y}$ ως

$$\mathbf{X} = \begin{bmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & -2 & 1 \end{bmatrix} \;,\quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} .$$

Στη συνέχεια σχηματίζουμε τους πίνακες

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 2 & -4 & 0 \\ -4 & 10 & 0 \\ 0 & 0 & 4 \end{bmatrix} \Rightarrow (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 2.50 & 1.00 & 0.00 \\ 1.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.25 \end{bmatrix} \;,\; \text{και } \mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} -2 \\ 6 \\ 0 \end{bmatrix} \;,$$

και εύκολα παίρνουμε το ζητούμενο διάνυσμα

$$\mathbf{w} \;=\; (\mathbf{X}^{\mathrm{T}}\,\mathbf{X})^{-1}\,\mathbf{X}^{\mathrm{T}}\,\mathbf{y} \;=\; \begin{bmatrix} 2.50 & 1.00 & 0.00 \\ 1.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.25 \end{bmatrix} . \begin{bmatrix} -2 \\ 6 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad .$$

Η γραμμική συνάρτηση διάκρισης που προκύπτει λοιπόν είναι η

$$g_B(\mathbf{x}) \;=\; x_1 + x_2 \;=\; 0 \;\;,$$

η οποία είχε οριστεί στην απάντησή μας στο υπο-ερώτημα 2(a). Εκεί είχαμε επίσης διαπιστώσει ότι $g_B(\mathbf{x}_{\mathrm{test}}) = -1 < 0$, δηλαδή ότι το διάνυσμα δοκιμής ταξινομείται στην κλάση ω_2. Παρατηρούμε επίσης ότι το άθροισμα των τετραγωνικών σφαλμάτων (εξίσωση (3.42) του βιβλίου) που πετυχαίνει η συγκεκριμένη συνάρτηση διάκρισης είναι μηδενικό, καθώς και για τα τέσσερα διανύσματα εκπαίδευσης τυχαίνει να ισχύει $y_i = \mathbf{w}_B^{\mathrm{T}}\,\mathbf{x}_i$.

(d) Το ερώτημα λύνεται εύκολα (εποπτικά) αν διαπιστώσουμε ότι το μέγιστο περιθώριο (margin) μεταξύ των διανυσμάτων εκπαίδευσης των δύο κλάσεων είναι 2, και επιτυγχάνεται από τη γραμμική συνάρτηση διάκρισης

$$g_D(\mathbf{x}) \;=\; x_2 \;=\; 0\,, \;\; \text{δηλαδή} \;\; \mathbf{w}_D = [\,0\,,1\,,0\,]^{\mathrm{T}} \;\; .$$

Τα διανύσματα εκπαίδευσης που ικανοποιούν την $w_1x_1+w_2x_2+w_0 = \pm\,1$ είναι τα $[\,0\,,1\,]^{\mathrm{T}}$ και $[\,0\,,-1\,]^{\mathrm{T}}$ (γραμμένα στη μη επαυξημένη τους μορφή). Αυτά αποτελούν κατά συνέπεια τα διανύσματα υποστήριξης (support vectors). Παρατηρούμε όντως ότι η εξίσωση (3.79) του βιβλίου ισχύει, καθώς

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{2}\,(+\,1) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \;+\; \frac{1}{2}\,(-\,1) \begin{bmatrix} 0 \\ -1 \end{bmatrix} \;=\; \sum_i \lambda_i\, y_i\, \mathbf{x}_i \;\;,$$

δηλαδή έχουμε δύο μηδενικούς και δύο θετικούς πολλαπλασιαστές Lagrange (οι συγκεκριμένοι φυσικά αποτελούν μία από τις άπειρες επιλογές τέτοιων συντελεστών). Επίσης η εξίσωση (3.80) του βιβλίου ικανοποιείται για $w_0 = 0$. Τέλος, το διάνυσμα δοκιμής $\mathbf{x}_{\mathrm{test}}$ ταξινομείται στην κλάση ω_1, καθώς $g_D(\mathbf{x}_{\mathrm{test}}) = [\,0\,,1\,,0\,]\,.\,[\,-3\,,2\,,1\,]^{\mathrm{T}} = 2 > 0$.

Εναλλακτικά, στα παραπάνω φυσικά καταλήγουμε αν ακολουθήσουμε τη μεθοδολογία του Παραδείγματος 3.5 του βιβλίου.

Άσκηση 3: Έστω δύο κλάσεις ω_1 και ω_2 με τα ακόλουθα 2-D διανύσματα χαρακτηριστικών εκπαίδευσης (γραμμένα σε μορφή $\mathbf{x} = [\,x_1\,,x_2\,]^{\mathrm{T}}$):

$$\omega_1 : \; [\,0\,,0\,]^{\mathrm{T}}\,, \; [\,0\,,-1\,]^{\mathrm{T}}\,, \; [\,0\,,1\,]^{\mathrm{T}}\,, \; [\,-1\,,0\,]^{\mathrm{T}}\,, \; [\,1\,,0\,]^{\mathrm{T}}$$

$$\omega_2 : \; [\,2\,,0\,]^{\mathrm{T}}\,, \; [\,0\,,-2\,]^{\mathrm{T}}\,, \; [\,2\,,-2\,]^{\mathrm{T}}\,, \; [\,-2\,,0\,]^{\mathrm{T}}\,, \; [\,-2\,,2\,]^{\mathrm{T}}\,, \; [\,0\,,2\,]^{\mathrm{T}}\;.$$

Έστω επίσης και ένα διάνυσμα δοκιμής $\mathbf{x}_{\text{test}} = [\,2\,,2\,]^{\mathrm{T}}$.

(a) Σχεδιάστε τα παραπάνω διανύσματα στο 2-D χώρο με άξονες τα x_1, x_2. Είναι οι κλάσεις γραμμικά διαχωρίσιμες (με βάση τα διανύσματα εκπαίδευσης);

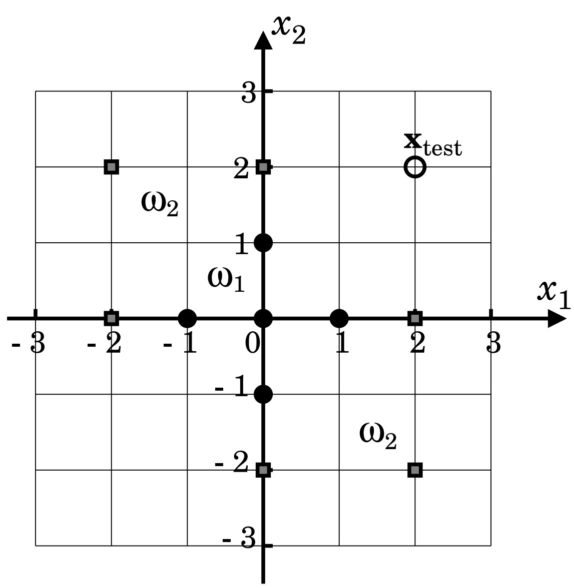
(b) Κατασκευάστε ένα perceptron δύο επιπέδων που να ταξινομεί τα διανύσματα εκπαίδευσης σωστά στις δύο κλάσεις. Σε ποια κλάση ταξινομείται το διάνυσμα δοκιμής $\mathbf{x}_{\text{test}}$;

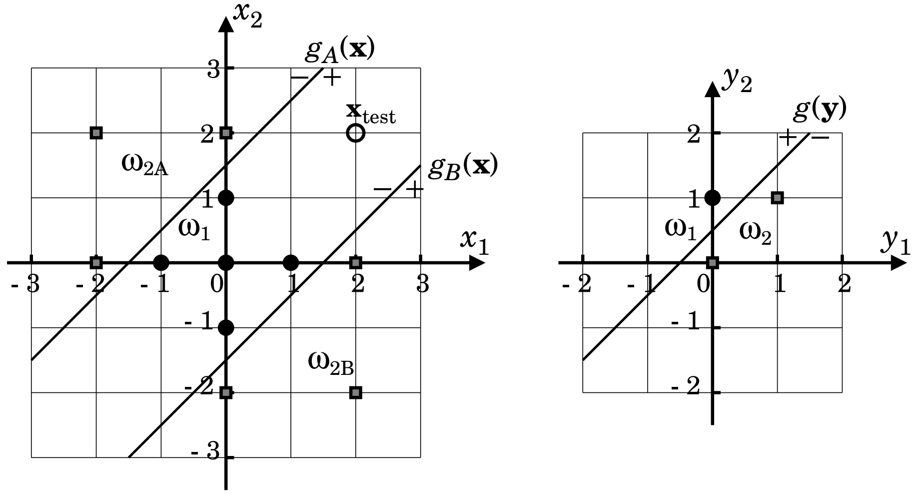
(c) Λύστε το παραπάνω πρόβλημα με δίκτυο ακτινωτής συνάρτησης βάσης (RBF). Σε ποια κλάση ταξινομείται το διάνυσμα δοκιμής;

(d) Λύστε το πρόβλημα χρησιμοποιώντας πολυωνυμικό ταξινομητή και ταξινομήστε το διάνυσμα δοκιμής με αυτόν.

Λύση:

(a) Τα δοθέντα διανύσματα εκπαίδευσης και το διάνυσμα δοκιμής έχουν σχεδιαστεί στο παρακάτω σχήμα. Είναι προφανές ότι δεν υπάρχει ευθεία που να διαχωρίζει τα διανύσματα εκπαίδευσης σωστά στις δύο κλάσεις ω_1 και ω_2, και κατά συνέπεια προχωρούμε στην σχεδίαση μη γραμμικών ταξινομητών στα παρακάτω υπο-ερωτήματα.





(b) Ακολουθούμε την μεθοδολογία της Ενότητας 4.3 του βιβλίου. Θεωρούμε τις ευθείες διάκρισης (βλέπε παραπάνω σχήμα)

$$g_A(\mathbf{x}) \;=\; x_1 - x_2 + 3/2 \;=\; 0\,, \;\; \text{δηλαδή} \;\; \mathbf{w}_A = [\,1\,,-1\,,3/2\,]^{\mathrm{T}} \;\;,$$

και

$$g_B(\mathbf{x}) \;=\; x_1 - x_2 - 3/2 \;=\; 0\,, \;\; \text{δηλαδή} \;\; \mathbf{w}_B = [\,1\,,-1\,,-3/2\,]^{\mathrm{T}} \;\;,$$

που ορίζουν αντίστοιχα perceptrons (ενός επιπέδου). Ορίζοντας ως τις εξόδους τους τα

$$y_1 = \mathrm{perc}_B(\mathbf{x}) \quad \text{και} \quad y_2 = \mathrm{perc}_A(\mathbf{x}) \;\;,$$

παρατηρούμε ότι έχει επιτευχθεί η εξής αντιστοίχηση των περιοχών $\omega_1\,,\omega_{2A}\,,\omega_{2B}$ του επιπέδου εισόδου $\mathbf{x}$ (βλέπε παραπάνω σχήμα αριστερά) σε τρία σημεία του καινούργιου διδιάστατου χώρου $\mathbf{y} = [\,y_1\,,y_2\,]^{\mathrm{T}}$ (βλέπε παραπάνω σχήμα δεξιά):

$$\mathbf{x} \in \omega_1 \;\;\Rightarrow\;\; y_1 = 0\,,\; y_2 = 1 \;\;,$$

$$\mathbf{x} \in \omega_{2A} \;\;\Rightarrow\;\; y_1 = 0\,,\; y_2 = 0 \;\;,$$

$$\mathbf{x} \in \omega_{2B} \;\;\Rightarrow\;\; y_1 = 1\,,\; y_2 = 1 \;\;.$$

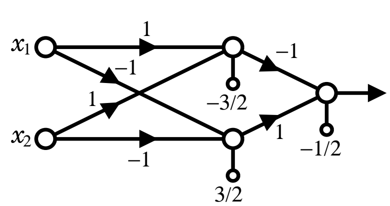
Οι δύο κλάσεις του προβλήματος αποτελούνται από συνολικά τρία πλέον σημεία

$$\omega_1 \;:\;\; [\,0\,,1\,]^{\mathrm{T}}\,, \qquad \omega_2 \;:\;\; [\,0\,,0\,]^{\mathrm{T}}\,,\;\; [\,1\,,1\,]^{\mathrm{T}}\,,$$

στον χώρο των $\mathbf{y}$, και είναι φυσικά γραμμικά διαχωρίσιμες. Μία ευθεία διάκρισης είναι η

$$g(\mathbf{y}) \;=\; -y_1 + y_2 - 1/2 \;=\; 0\,, \;\; \text{δηλαδή} \;\; \mathbf{w} = [\,-1\,,1\,,-1/2\,]^{\mathrm{T}} \;\;,$$

η οποία έχει σχεδιαστεί στο παραπάνω σχήμα, δεξιά, και η οποία δίνει το δεύτερο επίπεδο του ζητούμενου perceptron. Το συνολικό σύστημα ταξινόμησης δίνεται στο παρακάτω σχήμα. Είναι προφανές επίσης ότι το διάνυσμα δοκιμής ταξινομείται στην κλάση ω_1.



(c) Το πρόβλημα λύνεται εύκολα αν παρατηρήσουμε ότι οποιοσδήποτε κύκλος με κέντρο την αρχή των αξόνων και ακτίνα μεταξύ των τιμών 1 και 2 διαχωρίζει τις δύο κλάσεις. Κατά συνέπεια, μία μόνο συνάρτηση RBF αρκεί για να αντιστοιχίσει τον διδιάστατο χώρο σε ένα μονοδιάστατο χώρο όπου οι ω_1 και ω_2 είναι γραμμικά διαχωρίσιμες. Πράγματι, έστω

$$y = f(\mathbf{x}) = \exp\left(-\|\mathbf{x}\|^2\right) \ ,$$

δηλαδή μία συνάρτηση ακτινωτής βάσης της μορφής της εξίσωσης (4.53) του βιβλίου με κέντρο $\mathbf{c} = [\,0\,,0\,]^{\mathrm{T}}$ και $\sigma = 1/\sqrt{2}$. Παρατηρούμε τις ακόλουθες αντιστοιχίσεις:

$$\mathbf{x} = [\,0\,,0\,]^{\mathrm{T}} \ \Rightarrow \ y = \exp(0) = 1 \ ,$$

$$\mathbf{x} = [\,0\,,-1\,]^{\mathrm{T}},\ [\,0\,,1\,]^{\mathrm{T}},\ [\,-1\,,0\,]^{\mathrm{T}},\ [\,1\,,0\,]^{\mathrm{T}} \ \Rightarrow \ y = \exp(-1) = 0.3679 \ ,$$

$$\mathbf{x} = [\,2\,,0\,]^{\mathrm{T}},\ [\,0\,,-2\,]^{\mathrm{T}},\ [\,-2\,,0\,]^{\mathrm{T}},\ [\,0\,,2\,]^{\mathrm{T}} \ \Rightarrow \ y = \exp(-4) = 0.0183 \ ,$$

$$\mathbf{x} = [\,2\,,-2\,]^{\mathrm{T}},\ [\,-2\,,2\,]^{\mathrm{T}} \ \Rightarrow \ y = \exp(-8) = 0.0003 \ ,$$

δηλαδή τα δεδομένα εκπαίδευσης στον καινούργιο μας χώρο είναι τα

$$\omega_1 \ : \ 1\,,\ 0.3679 \ ,$$

και

$$\omega_2 \ : \ 0.0183\,,\ 0.0003 \ .$$

Αυτά πλέον είναι γραμμικά διαχωρίσιμα, για παράδειγμα από τη συνάρτηση διάκρισης

$g(y) = y - 0.2231$, ή, ισοδύναμα (στον αρχικό χώρο), $g(\mathbf{x}) = \exp\left(-\|\mathbf{x}\|^2\right) - 0.2231$.

Παρατηρώντας ότι $\exp(-1.5) = 0.2231$, είναι προφανές ότι η παραπάνω συνάρτηση διάκρισης αντιστοιχεί σε επιφάνεια διαχωρισμού που είναι κύκλος με κέντρο την αρχή των αξόνων και ακτίνα 1.5. Τέλος, έχουμε $g(\mathbf{x}_{\text{test}}) = \exp(-8) - \exp(-1.5) = -0.2228$, κατά συνέπεια το διάνυσμα δοκιμής ταξινομείται στην κλάση ω_2.

(d) Μία απλή αντιστοίχηση του αρχικού χώρου σε χώρο όπου το πρόβλημα είναι γραμμικά διαχωρίσιμο με χρήση πολυωνυμικών συναρτήσεων είναι η

$$y_1 = x_1^2 \ , \ y_2 = x_2^2 \ .$$

Παρατηρούμε τότε ότι

$$\begin{aligned}
\mathbf{x} &= [\,0\,,0\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,0\,,0\,]^{\mathrm{T}} \ ,\\
\mathbf{x} &= [\,0\,,-1\,]^{\mathrm{T}},\ [\,0\,,1\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,0\,,1\,]^{\mathrm{T}} \ ,\\
\mathbf{x} &= [\,-1\,,0\,]^{\mathrm{T}},\ [\,1\,,0\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,1\,,0\,]^{\mathrm{T}} \ ,\\
\mathbf{x} &= [\,2\,,0\,]^{\mathrm{T}},\ [\,-2\,,0\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,4\,,0\,]^{\mathrm{T}} \ ,\\
\mathbf{x} &= [\,0\,,-2\,]^{\mathrm{T}},\ [\,0\,,2\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,0\,,4\,]^{\mathrm{T}} \ ,\\
\mathbf{x} &= [\,2\,,-2\,]^{\mathrm{T}},\ [\,-2\,,2\,]^{\mathrm{T}} && \Rightarrow \ \mathbf{y} = [\,4\,,4\,]^{\mathrm{T}} \ .
\end{aligned}$$

Κατά συνέπεια, τα δεδομένα εκπαίδευσης στον καινούργιο χώρο είναι τα:

$$\omega_1 \ : \ [\,0\,,1\,]^{\mathrm{T}},\ [\,1\,,0\,]^{\mathrm{T}},\ [\,0\,,0\,]^{\mathrm{T}}\,.$$

και

$$\omega_2 : \; [\,0\,,4\,]^{\mathrm{T}}\,, \;\; [\,4\,,0\,]^{\mathrm{T}}\,, \;\; [\,4\,,4\,]^{\mathrm{T}}\,.$$

Αυτά πλέον είναι γραμμικά διαχωρίσιμα (φαίνεται εύκολα αν σχεδιαστούν σε σχήμα), για παράδειγμα από τη συνάρτηση διάκρισης

$$g(y) = -\,y_1 \,-\, y_2 \,+\, 2 \;\;, \;\; \text{ή, ισοδύναμα (στον αρχικό χώρο),} \;\; g(\mathbf{x}) = -\,x_1^2 \,-\, x_2^2 \,+\, 2 \;\;.$$

Παρατηρούμε ότι η παραπάνω συνάρτηση διάκρισης αντιστοιχεί σε επιφάνεια διαχωρισμού στον αρχικό χώρο που είναι κύκλος με κέντρο την αρχή των αξόνων και ακτίνα $\sqrt{2}$. Τέλος, έχουμε $g(\mathbf{x}_{\mathrm{test}}) = -4 - 4 + 2 = -6$, κατά συνέπεια το διάνυσμα δοκιμής ταξινομείται στην κλάση ω_2.

Άσκηση 4(a): Αποδείξτε ότι η απόκλιση (Kullback-Leibler divergence) $d_{ij}(\mathbf{x})$ ενός διανύσματος χαρακτηριστικών $\mathbf{x} = [\,x_1\,,\, x_2\,, \ldots,\, x_L\,]^{\mathrm{T}}$ ως προς κλάσεις ω_i, ω_j (εξίσωση (5.21) του βιβλίου) ικανοποιεί τη σχέση $d_{ij}(\mathbf{x}) \geq 0$.

Λύση:

Χρησιμοποιώντας τη γνωστή ιδιότητα $\ln x \leq x - 1$, παίρνουμε (βλέπε επίσης εξίσωση (5.19) του βιβλίου):

$$D_{ij} \;=\; -\int p\,(\mathbf{x}|\omega_i)\,\ln\frac{p\,(\mathbf{x}|\omega_j)}{p\,(\mathbf{x}|\omega_i)}\,\mathrm{d}\mathbf{x} \;\geq\; -\int p\,(\mathbf{x}|\omega_i)\left(\frac{p\,(\mathbf{x}|\omega_j)}{p\,(\mathbf{x}|\omega_i)} \;-\; 1\right)\mathrm{d}\mathbf{x}$$

$$= \int p\,(\mathbf{x}|\omega_j)\,\mathrm{d}\mathbf{x} \;-\; \int p\,(\mathbf{x}|\omega_i)\,\mathrm{d}\mathbf{x} \;=\; 1 \;-\; 1 \;=\; 0 \;\;.$$

Παρόμοια, παίρνουμε $D_{ji} \geq 0$ και, κατά συνέπεια, $d_{ij} = D_{ij} + D_{ji} \geq 0$.

Άσκηση 4(b): Αποδείξτε ότι $d_{ij}(\mathbf{x}) \;=\; \sum_{l=1}^{L} d_{ij}(x_l)$, για στατιστικά ανεξάρτητες μεταβλητές $x_1\,,\, x_2\,, \ldots,\, x_L$.

Λύση:

Θα αποδείξουμε το παραπάνω για $L = 2$. Η γενίκευση για $L > 2$ είναι προφανής. Έχουμε λοιπόν,

$$\begin{aligned}
d_{ij}(x_1, x_2) \;&=\; \int_{x_1}\int_{x_2}(\,p_i(x_1,x_2) \;-\; p_j(x_1,x_2)\,)\;\ln\frac{p_i(x_1,x_2)}{p_j(x_1,x_2)}\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2 \\
&= \int_{x_1}\int_{x_2}(\,p_i(x_1)\,p_i(x_2) \;-\; p_j(x_1)\,p_j(x_2)\,)\;(\,\ln\frac{p_i(x_1)}{p_j(x_1)} \;+\; \ln\frac{p_i(x_2)}{p_j(x_2)}\,)\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2 \\
&= \int_{x_1}\int_{x_2}(\,p_i(x_1)\,p_i(x_2) \;-\; p_j(x_1)\,p_j(x_2)\,)\;\ln\frac{p_i(x_1)}{p_j(x_1)}\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2 \\
&+ \int_{x_1}\int_{x_2}(\,p_i(x_1)\,p_i(x_2) \;-\; p_j(x_1)\,p_j(x_2)\,)\;\ln\frac{p_i(x_2)}{p_j(x_2)}\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2 \\
&= \int_{x_1}(\,p_i(x_1) - p_j(x_1)\,)\;\ln\frac{p_i(x_1)}{p_j(x_1)}\,\mathrm{d}\,x_1 \;+\; \int_{x_2}(\,p_i(x_2) - p_j(x_2)\,)\;\ln\frac{p_i(x_2)}{p_j(x_2)}\,\mathrm{d}\,x_2 \\
&= d_{ij}(x_1) \;+\; d_{ij}(x_2) \;\;.
\end{aligned}$$

Άσκηση 4(c): Αποδείξτε ότι για πολυδιάστατες κανονικές κατανομές για κάθε κλάση (multivariate Gaussian class conditional pdfs) με κοινό πίνακα συνδιασποράς $\mathbf{S}$ και διανύσματα μέσων τιμών $\mathbf{m}_i$ (για την κλάση ω_i), η απόκλιση ισούται με την απόσταση Mahalanobis μεταξύ των αντίστοιχων διανυσμάτων μέσων τιμών, δηλ. $d_{ij}(\mathbf{x}) = (\mathbf{m}_i - \mathbf{m}_j)^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_i - \mathbf{m}_j)$.

Λύση:

Όπως έχουμε ήδη δείξει στις πρώτες εξισώσεις της Άσκησης 1(a), ο λογάριθμος του λόγου των συναρτήσεων πυκνότητας πιθανότητας δύο πολυδιάστατων κανονικών κατανομών με κοινό πίνακα συνδιασποράς είναι

$$\ln p(\mathbf{x}|\omega_i) \;-\; \ln p(\mathbf{x}|\omega_j) \;=\; \mathbf{x}^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_i - \mathbf{m}_j) \;\;-\;\; \frac{1}{2}\, (\mathbf{m}_i + \mathbf{m}_j)^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_i - \mathbf{m}_j) \;\; .$$

Κατά συνέπεια,

$$D_{ij} \;=\; \int p\,(\mathbf{x}|\omega_i)\, \ln \frac{p\,(\mathbf{x}|\omega_i)}{p\,(\mathbf{x}|\omega_j)}\, \mathrm{d}\mathbf{x} \;=\; \mathbf{m}_i^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_i - \mathbf{m}_j) \;\; ,$$

αφού ο δεύτερος όρος είναι μία σταθερά και η μέση τιμή του διανύσματος $\mathbf{x}$ που ακολουθεί την $p\,(\mathbf{x}|\omega_i)$ είναι η $\mathbf{m}_i$. Αντίστοιχα, έχουμε

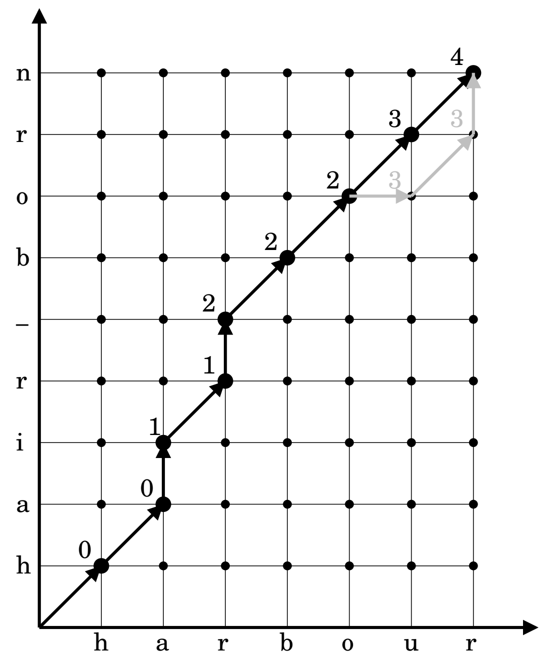
$$D_{ji} \;=\; \int p\,(\mathbf{x}|\omega_j)\, \ln \frac{p\,(\mathbf{x}|\omega_j)}{p\,(\mathbf{x}|\omega_i)}\, \mathrm{d}\mathbf{x} \;=\; \mathbf{m}_j^{\mathrm{T}}\, \mathbf{S}^{-1}\, (\mathbf{m}_j - \mathbf{m}_i) \;\; .$$

Αθροίζοντας τα D_{ij} και D_{ji}, παίρνουμε το ζητούμενο.

Άσκηση 5(a): Υπολογίστε με διαδικασία δυναμικής χρονικής στρέβλωσης (dynamic time warping) την απόσταση edit μεταξύ των ακολουθιών συμβόλων "harbour" (πρότυπο αναφοράς) και "hair born" (υπό εξέταση πρότυπο) και σχεδιάστε το σχετικό διάγραμμα δυναμικού προγραμματισμού.

Λύση:

Η λύση του προβλήματος φαίνεται στο παρακάτω σχήμα, ακολουθώντας την μεθοδολογία της ενότητας 7.2.2 του βιβλίου (σύμφωνα επίσης με τα σχήματα 7.4). Είναι προφανές ότι η ζητούμενη απόσταση ισούται με 4 (Παρατήρηση: Στο βέλτιστο μονοπάτι έχουμε σημειώσει το **συνολικό** κόστος του σε κάθε σημείο). Παρατηρούμε επίσης ότι το βέλτιστο μονοπάτι δεν είναι μοναδικό. Βέλτιστη επίσης είναι και η παραλλαγή του που έχει σημειωθεί με γκρι χρώμα.



Άσκηση 5(b): Υπολογίστε με διαδικασία δυναμικής χρονικής στρέβλωσης (dynamic time warping) το κόστος σύγκρισης μεταξύ της ακολουθίας τεσσάρων 2-D χαρακτηριστικών (πρότυπο αναφοράς)

$$X = \begin{bmatrix} 1.0 & -1.0 & -1.0 & 0.0 \\ 1.0 & 0.0 & -1.0 & -1.0 \end{bmatrix}$$

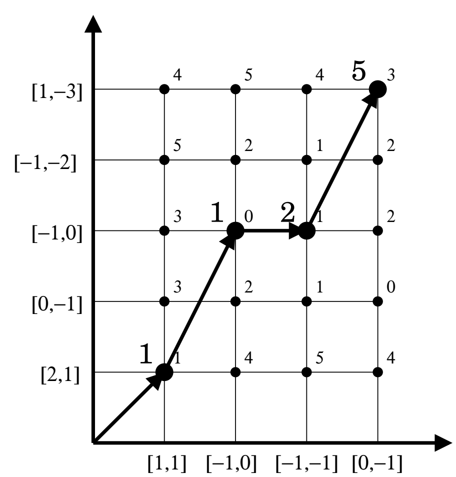
και της ακολουθίας πέντε 2-D χαρακτηριστικών (υπό εξέταση πρότυπο)

$$Y = \begin{bmatrix} 2.0 & 0.0 & -1.0 & -1.0 & 1.0 \\ 1.0 & -1.0 & 0.0 & -2.0 & -3.0 \end{bmatrix}$$

υπό τοπικούς περιορισμούς μετάβασης Itakura (σχήμα 7.10) του βιβλίου και κόστος σύγκρισης διανυσμάτων την απόσταση L_1 (Manhattan distance). Σχεδιάστε το σχετικό διάγραμμα δυναμικού προγραμματισμού.

Λύση:

Η λύση του προβλήματος φαίνεται στο παρακάτω σχήμα, ακολουθώντας την μεθοδολογία της ενότητας 7.2.3 του βιβλίου. Για διευκόλυνση, υπολογίζουμε πρώτα όλες τις αποστάσεις L_1 μεταξύ των διανυσμάτων (βλέπε αριθμούς σε μικρή γραμματοσειρά). Στη συνέχεια εφαρμόζουμε τον αλγόριθμο υπό τοπικούς περιορισμούς μετάβασης του σχήματος 7.10 του βιβλίου. Το βέλτιστο μονοπάτι έχει συνολικό κόστος 5. Ωστόσο, αν επιτρέπαμε χαλαρούς περιορισμούς άκρων, τότε το βέλτιστο μονοπάτι θα είχε κόστος 4, τερματίζοντας ένα σημείο πιο κάτω από το μονοπάτι που έχει σχεδιαστεί στο σχήμα. Αξίζει επίσης να παρατηρηθεί, ότι όπως και πριν, και στις δύο περιπτώσεις, το βέλτιστο μονοπάτι δεν είναι μοναδικό.



Άσκηση 6:

Θεωρήστε την μοντελοποίηση ενός πειράματος ρίψης τριών νομισμάτων με προκατάληψη, χρησιμοποιώντας ένα HMM με τρεις καταστάσεις.

- Η πρώτη κατάσταση μοντελοποιεί το πρώτο νόμισμα c_1 και έχει πιθανότητες παρατήρησης $P(H|c_1) = 0.50$, $P(T|c_1) = 0.50$ (για αποτέλεσμα ρίψης heads (H) ή tails (T)).
- Για το δεύτερο νόμισμα έχουμε $P(H|c_2) = 0.75$, $P(T|c_2) = 0.25$.
- Για το τρίτο νόμισμα έχουμε $P(H|c_3) = 0.25$, $P(T|c_3) = 0.75$.

Υποθέτουμε επίσης ότι η πιθανότητα μετάβασης μεταξύ νομισμάτων είναι ομοιόμορφη και ίση με $a_{ij} = 1/3$ (όπως επίσης και η αρχική πιθανότητα).

(a) Αν έχει παρατηρηθεί η ακολουθία αποτελεσμάτων ρίψης $\mathbf{O} = (HHTH)$, μπορείτε να δείτε εύκολα ποιά είναι η πιο πιθανή ακολουθία καταστάσεων $\hat{\mathbf{c}}$ (δηλαδή των νομισμάτων των τεσσάρων ρίψεων);

(b) Υπολογίστε την κοινή πιθανότητα $P[\mathbf{O}, \hat{\mathbf{c}}]$.

(c) Πόσες φορές πιο μικρή από την απάντηση στο (b) είναι η πιθανότητα ότι η ακολουθία αποτελεσμάτων ρίψης προήλθε από τέσσερεις ρίψεις του νομίσματος c_1;

(d) Υπολογίστε την ακολουθία $\hat{\mathbf{c}}$ του ερωτήματος (a) χρησιμοποιώντας τον αλγόριθμο Viterbi.

Λύση:

(a) Εφόσον όλες οι πιθανότητες μετάβασης και οι αρχικές πιθανότητες έχουν την ίδια τιμή, η πιο πιθανή ακολουθία καταστάσεων θα είναι αυτή που μεγιστοποιεί την πιθανότητα κάθε παρατήρησης, δηλαδή κατάσταση c_2 κάθε φορά που εμφανίζεται H, και κατάσταση c_3 κάθε φορά που εμφανίζεται T. Κατά συνέπεια, η ζητούμενη ακολουθία είναι η $\hat{\mathbf{c}} = (\, c_2\, , c_2\, , c_3\, , c_2\,)$.

(b) Είναι εύκολο να δούμε πως η ζητούμενη πιθανότητα ισούται με

$$P\,[\,\mathbf{O}, \hat{\mathbf{c}}\,] \;=\; (1/3)^4\,(0.75)^4 \;=\; 0.0039 \;\;.$$

(c) Παρόμοια, βρίσκουμε, συμβολίζοντας με $\mathbf{c} = (\, c_1\, , c_1\, , c_1\, , c_1\,)$, ότι

$$\frac{P\,[\,\mathbf{O}, \mathbf{c}\,]}{P\,[\,\mathbf{O}, \hat{\mathbf{c}}\,]} \;=\; \frac{(1/3)^4\,(0.50)^4}{(1/3)^4\,(0.75)^4} \;=\; 0.1975 \;\;,$$

δηλαδή η συγκεκριμένη ακολουθία καταστάσεων είναι περί τις 5 φορές λιγότερο πιθανή από την $\hat{\mathbf{c}}$.

(d) Εφαρμόζοντας τον αλγόριθμο Viterbi, έχουμε (απλοποιώντας το συμβολισμό γράφουμε i αντί για c_i, και ορίζουμε $\delta_t(i) = \max_{q_1, q_2, ..., q_{t-1}} P(o_1, o_2, ..., o_t, q_1, q_2, ..., q_{t-1}, q_t = i)$):

$$\delta_1(1) = (1/3)\,0.5\ , \quad \delta_1(2) = (1/3)\,0.75\ , \quad \delta_1(3) = (1/3)\,0.25\ ,$$

$$\delta_2(1) = (1/3)^2\,0.75\,.\,0.5\ , \quad \delta_2(2) = (1/3)^2\,(0.75)^2\ , \quad \delta_2(3) = (1/3)^2\,0.75\,.\,0.25$$

(με καλύτερη προηγούμενη κατάσταση την 2),

$$\delta_3(1) = (1/3)^3\,(0.75)^2\,0.5\ , \quad \delta_3(2) = (1/3)^3\,(0.75)^2\,0.25\ , \quad \delta_3(3) = (1/3)^3\,(0.75)^3\ ,$$

(με καλύτερη προηγούμενη κατάσταση την 2),

$$\delta_4(1) = (1/3)^4\,(0.75)^3\,0.5\ , \quad \delta_4(2) = (1/3)^4\,(0.75)^4\ , \quad \delta_4(3) = (1/3)^4\,(0.75)^3\,0.25\ ,$$

(με καλύτερη προηγούμενη κατάσταση την 3). Το μέγιστο από τα δ_4 επιτυγχάνεται από την κατάσταση 2, κατά συνέπεια με υπαναχώρηση (backtracking) παίρνουμε την ακολουθία (2,2,3,2).
