Η παράδοση των ασκήσεων θα γίνει στην τελική εξέταση (Ιουνίου) του μαθήματος. Επισημαίνεται ότι οι εργασίες είναι **ατομικές** και ότι συνεισφέρουν **8%** του τελικού βαθμού.

Άσκηση 1: Βρείτε τους εκτιμητές μέγιστης πιθανοφάνειας (maximum likelihood estimators) των παραμέτρων των παρακάτω κατανομών, δοθέντων N στατιστικά ανεξάρτητων δειγμάτων $x_n\,,\ n = 1\,,2\,,...,N$:

(a) Εκθετικής κατανομής, δηλ. της:

$$p\,(\,x\,;\,\theta\,) \;=\; \begin{cases} \dfrac{1}{\theta}\,\exp\Big(-\,\dfrac{x}{\theta}\Big) \;\;,\;\; x\,\ge\,0 \\ 0\;\;, \qquad\qquad\qquad \text{αλλού} \end{cases} \;.$$

(b) Συμμετρικής ομοιόμορφης κατανομής, δηλ. της:

$$p\,(\,x\,;\,\theta\,) \;=\; \begin{cases} 1/(2\,\theta)\;\;,\;\; \text{εάν} \;\; -\,\theta\,\le\,x\,\le\,\theta \\ 0\;\;, \quad\;\; \text{αλλιώς} \end{cases} \;.$$

Άσκηση 2: Βρείτε τις εξισώσεις που περιγράφουν τις επιφάνειες διαχωρισμού και σχεδιάστε τις περιοχές αποφάσεων στο πρόβλημα Bayesian ταξινόμησης τριών ισοπίθανων κλάσεων ω_1, ω_2, και ω_3, 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) τον μοναδιαίο διαγώνιο πίνακα $\mathbf{I}$ και διανύσματα μέσων τιμών $\mathbf{m}_1 = [\,1\,,1\,,-1\,]^{\mathrm{T}}$, $\mathbf{m}_2 = [\,-1\,,1\,,-1\,]^{\mathrm{T}}$, και $\mathbf{m}_3 = [\,0\,,-1\,,-1\,]^{\mathrm{T}}$.

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

$$\omega_1\;:\;\;[\,1\,,1\,]^{\mathrm{T}}\,,\;\;[\,1\,,2\,]^{\mathrm{T}}\,,\;\;[\,0\,,3\,]^{\mathrm{T}}\;\;,$$

$$\omega_2\;:\;\;[\,-1\,,-1\,]^{\mathrm{T}}\,,\;\;[\,-1\,,-2\,]^{\mathrm{T}}\,,\;\;[\,0\,,-3\,]^{\mathrm{T}}\;\;.$$

Έστω επίσης και ένα διάνυσμα δοκιμής $\mathbf{x}_{\mathrm{test}} = [\,0\,,1/2\,]^{\mathrm{T}}$.

(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 $\mathbf{w} = [\,w_1\,,w_2\,,w_0\,]^{\mathrm{T}}$, 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_2.

(b) Ταξινομήστε το διάνυσμα δοκιμής με βάση τον κανόνα του πλησιέστερου γείτονα (1-NN), χρησιμοποιώντας οποιοδήποτε μέτρο απόστασης μεταξύ διανυσμάτων στον δυδιάστατο χώρο επιθυμείτε.

(c) 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 ρ της αρεσκείας σας. Σχεδιάστε την και ταξινομήστε το $\mathbf{x}_{\text{test}}$ με αυτήν. Σημείωση: Ο συμβολισμός της $g(\mathbf{x})$ και τα πρόσημα επιλογής των κλάσεων είναι όπως στο υπο-ερώτημα (a).

(d) Βρείτε τη γραμμική συνάρτηση διάκρισης των δύο κλάσεων με βάση την μέθοδο των μηχανών διανυσματικής στήριξης (support vector machines), εκφράζοντας το διάνυσμα $[\,w_1\,,w_2\,]^{\mathrm{T}}$ ως γραμμικό συνδυασμό των διανυσμάτων στήριξης (support vectors) και στη συνέχεια βρίσκοντας και το w_0. 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.

(e) 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.

(f) Μειώστε την διάσταση των δεδομένων από δύο σε μία χρησιμοποιώντας ανάλυση κύριων συνιστωσών (Principal Component Analysis – PCA). Βρείτε τις προβολές των διανυσμάτων εκπαίδευσης στον μονοδιάστατο χώρο που προκύπτει, και διαπιστώστε αν οι κλάσεις είναι γραμμικά διαχωρίσιμες σε αυτόν τον χώρο ή όχι. Αν ναι, ταξινομήστε το παράδειγμα δοκιμής στην κατάλληλη κλάση, με βάση τον κανόνα του πλησιέστερου γείτονα (1-NN) στον μονοδιάστατο χώρο που προκύπτει.

(g) Επαναλάβετε τα βήματα του υπο-ερωτήματος (g), αλλά χρησιμοποιώντας αυτή τη φορά ανάλυση γραμμικής διάκρισης (Linear Discriminant Analysis – LDA).

Άσκηση 4: Έστω δύο κλάσεις ω_1 και ω_2 με τα ακόλουθα 2-D διανύσματα χαρακτηριστικών εκπαίδευσης (τρία για κάθε κλάση, γραμμένα σε μορφή $\mathbf{x} = [\,x_1\,,x_2\,]^{\mathrm{T}}$):

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

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

(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.

(c) Λύστε το παραπάνω πρόβλημα με δίκτυο ακτινωτής συνάρτησης βάσης (RBF).

(d) Λύστε το παραπάνω πρόβλημα με έναν πολυωνυμικό ταξινομητή.

(e) Σχεδιάστε ένα δέντρο αποφάσεων (decision tree) για το παραπάνω πρόβλημα.

Άσκηση 5: Έστω τα επτά παρακάτω διανύσματα εκπαίδευσης:

$\mathbf{x}_1 = [\,0\,,0\,]^{\mathrm{T}}$, $\mathbf{x}_2 = [\,2\,,0\,]^{\mathrm{T}}$, $\mathbf{x}_3 = [\,0\,,-1\,]^{\mathrm{T}}$, $\mathbf{x}_4 = [\,0\,,1\,]^{\mathrm{T}}$, $\mathbf{x}_5 = [\,2\,,1\,]^{\mathrm{T}}$, $\mathbf{x}_6 = [\,2\,,-1\,]^{\mathrm{T}}$, $\mathbf{x}_7 = [\,3\,,0\,]^{\mathrm{T}}$.

(a) Έστω ότι αυτά παρουσιάζονται στο βασικό ακολουθιακό αλγοριθμικό σχήμα ομαδοποίησης (BSAS) όπως παραπάνω (με σειρά από αριστερά προς τα δεξιά). Ομαδοποιήστε τα με τον αλγόριθμο BSAS, χρησιμοποιώντας την απόσταση L_1 (Manhattan distance), κατώφλι $\theta = 2.1$, και θεωρώντας ότι οι αντιπρόσωποι των ομάδων που προκύπτουν είναι η μέση τιμή των μελών της ομάδας (και χωρίς περιορισμό στον αριθμό ομάδων που προκύπτουν).

(b) Στη συνέχεια χρησιμοποιήστε τις ομάδες που προέκυψαν ως αρχικοποίηση του αλγορίθμου ομαδοποίησης των K-μέσων (K-means), με χρήση Ευκλείδειας απόστασης (L_2). Ποιες είναι οι τελικές ομάδες που προκύπτουν;

(c) Ποια είναι τα medoids των ομάδων που προκύπτουν στο (a) και ποια των ομάδων του (b);

(d) Ομαδοποιήστε ξανά τα επτά διανύσματα, αυτήν τη φορά ιεραρχικά χρησιμοποιώντας τον συσσωρευτικό αλγόριθμο ομαδοποίησης (agglomerative clustering) με βάση την απόσταση L_1 (Manhattan distance). Χρησιμοποιείστε δύο διαφοροποιήσεις του αλγορίθμου, όσον αφορά την ενημέρωση των αποστάσεων του πίνακα ανομοιότητας, τον πρώτο με βάση τον αλγόριθμο απλού δεσμού (single link), και τον δεύτερο με βάση τον αλγόριθμο πλήρους δεσμού (complete link). Και στις δύο περιπτώσεις σχεδιάστε τα δεντρογράμματα ανομοιότητας (dendrograms) που προκύπτουν.

Άσκηση 6: Υποθέστε δύο κλάσεις με ένα πρότυπο αναφοράς (template) η καθεμία:

$$\omega_1 : \ \mathbf{X}_1 = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 1 \end{bmatrix} \quad , \quad \omega_2 : \ \mathbf{X}_2 = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -3 \end{bmatrix} \ .$$

Έστω επίσης ένα πρότυπο δοκιμής (υπό εξέταση πρότυπο)

$$\mathbf{Y} = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & 0 \end{bmatrix} \ ,$$

το οποίο θέλουμε να ταξινομήσουμε στην ω_1 ή ω_2, με βάση τη μεθοδολογία δυναμικής χρονικής στρέβλωσης (dynamic time warping). Για τον σκοπό αυτό χρησιμοποιούμε ως κόστος σύγκρισης διανυσμάτων την απόσταση L_1 (Manhattan distance), τοπικούς περιορισμούς μετάβασης Sakoe-Chiba (δηλαδή στο δυδιάστατο πλέγμα της δυναμικής χρονικής στρέβλωσης επιτρέπονται μεταβάσεις σε κάθε καινούργιο κόμβο από τον αμέσως προηγούμενο κάτω του κόμβο, τον αμέσως προηγούμενο αριστερά του, και τον διαγώνιο κάτω αριστερά του), και περιορισμούς άκρων που απαιτούν το βέλτιστο μονοπάτι να διέρχεται από τον αρχικό και τον τελικό κόμβο που αποτελούνται από τα ζεύγη των πρώτων και τελευταίων διανυσμάτων των προτύπων αναφοράς και δοκιμής, αντίστοιχα. Βρείτε τις αποστάσεις $d_{\mathrm{DTW}}(\mathbf{X}_1\,,\mathbf{Y})$ και $d_{\mathrm{DTW}}(\mathbf{X}_2\,,\mathbf{Y})$, σχεδιάζοντας τα σχετικά διαγράμματα δυναμικής χρονικής στρέβλωσης. Σε ποια κλάση θα ταξινομηθεί το πρότυπο δοκιμής με βάση τις αποστάσεις αυτές;
