

Όνομα/νυμο:

Υπογραφή:

ΑΜ:

Εξάμηνο:

Αριθμός διφύλλων:

ΠΑΡΑΤΗΡΗΣΕΙΣ: Ανοιχτό βιβλίο μαθήματος & σημειώσεις μαθήματος. Κλειστά κινητά.

Θέμα 1: (25%) Τα (a) και (b) είναι ανεξάρτητα ερωτήματα. Απαντήστε αναλυτικά.

(a) Βρείτε την εξίσωση που περιγράφει την επιφάνεια διαχωρισμού (και σχεδιάστε την) στο πρόβλημα Bayesian ταξινόμησης δύο ισοπίθανων κλάσεων με βάση 2-D διανύσματα χαρακτηριστικών \mathbf{x} που ακολουθούν 2-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 $\mathbf{m}_1 = [0, 0]^T$ και $\mathbf{m}_2 = [1, 1]^T$.

(b) Βρείτε τον εκτιμητή μέγιστης πιθανοφάνειας (maximum likelihood estimator) της παραμέτρου $\theta > 0$ της συνάρτησης πυκνότητας πιθανότητας του Maxwell,

$$p(x|\theta) = \begin{cases} \frac{4}{\sqrt{\pi}} \theta^{\frac{3}{2}} x^2 e^{-\theta x^2}, & x \geq 0 \\ 0, & \text{αλλού} \end{cases},$$

με βάση N στατιστικά ανεξάρτητα δείγματά της x_n , $n = 1, 2, \dots, N$.

Θέμα 2: (25%) Τα (a) και (b) είναι ανεξάρτητα ερωτήματα. Απαντήστε αναλυτικά.

(a) Υπολογίστε με δυναμική χρονική στρέβλωση (dynamic time warping) το κόστος σύγκρισης μεταξύ της ακολουθίας τριών 2-D χαρακτηριστικών (πρότυπο αναφοράς)

$$X = \begin{bmatrix} 1.0 & 0.0 & 2.0 \\ 1.0 & 0.0 & 2.0 \end{bmatrix}$$

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

$$Y = \begin{bmatrix} 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 2.0 & 3.0 & 4.0 \end{bmatrix}$$

με κόστος σύγκρισης διανυσμάτων την απόσταση L_1 (Manhattan distance), και υπό δύο ειδών τοπικούς περιορισμούς μετάβασης: (i) Itakura (σχήμα 7.10 του βιβλίου) και (ii) Sakoe-Chiba (σχήμα 7.11.(α) του βιβλίου). Σχεδιάστε τα σχετικά διαγράμματα.

(b) Έστω τα διανύσματα

$$\mathbf{x}_1 = [1, 0]^T, \quad \mathbf{x}_2 = [-1, 0]^T, \quad \mathbf{x}_3 = [3, 1]^T, \quad \mathbf{x}_4 = [-3, 1]^T.$$

Ομαδοποιήστε τα διανύσματα ιεραρχικά, χρησιμοποιώντας τον συσσωρευτικό αλγόριθμο ομαδοποίησης (agglomerative clustering) με βάση την απόσταση L_1 (Manhattan distance). Χρησιμοποιείστε για ενημέρωση των αποστάσεων του πίνακα ανομοιότητας τον αλγόριθμο απλού δεσμού (single link). Σχεδιάστε το δεντρογράμμα ανομοιότητας (dendrogram) που προκύπτει.

Θέμα 3: (50%) Έστω δύο κλάσεις ω_1 και ω_2 με τα ακόλουθα 2-D διανύσματα χαρακτηριστικών εκπαίδευσης (τρία για κάθε κλάση, γραμμένα σε μορφή $\mathbf{x} = [x_1, x_2]^T$):

$$\omega_1 : [0, 1]^T, [5, 4]^T, [-5, 4]^T$$

$$\omega_2 : [0, -1]^T, [5, -4]^T, [-5, -4]^T .$$

Έστω επίσης και ένα διάνυσμα δοκιμής $\mathbf{x}_{\text{test}} = [5, 1]^T$.

- 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 $\mathbf{w} = [w_1, w_2, w_0]^T$, 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 .
- Βρείτε τη γραμμική συνάρτηση διάκρισης των δύο κλάσεων με βάση τον αλγόριθμο perceptron (σε μορφή batch), αρχικοποιώντας τον αλγόριθμο με τιμές $\mathbf{w}(0) = [0, 0, 0]^T$ και χρησιμοποιώντας βήμα $\rho = 1/6$. Σχεδιάστε την και ταξινομήστε το \mathbf{x}_{test} με αυτήν.
- 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}_{test} με αυτήν.
- 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.
- Τέλος, μειώστε την διάσταση των δεδομένων από δύο σε μία, αλλά χρησιμοποιώντας ανάλυση γραμμικής διάκρισης αυτήν τη φορά (Linear Discriminant Analysis – LDA). Όπως και πριν, 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.