ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

7ο εξάμηνο | Ακαδημαϊκό έτος 2024-2025

1η Σειρά Ασκήσεων

Ημερ. Παράδ.: 15.11.2024

Γενικές Οδηγίες: Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Εξηγήστε επαρκώς την εργασία σας. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των σημειώσεων για την λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων των αναλυτικών ασκήσεων της σειράς αυτής θα γίνει ηλεκτρονικά στην HELIOS ιστοσελίδα του μαθήματος και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: ML24_hwk1_AM_LastnameFirstname.pdf, όπου AM ειναι ο 8-ψήφιος αριθμός μητρώου σας. Σκαναριμένες χειρόγραφες λύσεις επιτρέπονται αρκεί να είναι καθαρογραμμένες και ευανάγνωστες. Επίσης στην 1η σελίδα των λύσεων θα αναγράφετε το ονοματεπώνυμο, Α.Μ., και email address σας. Συμπεριλάβετε και τον κώδικα προγραμμάτων, π.χ. Matlab ή Python, που χρησιμοποιήσατε για αριθμητική επίλυση. Να σημειωθεί ότι η καταληκτική ημερομηνία παράδοσης είναι τελική και δεν θα δοθεί παράταση.

Άσκηση 1.1 (Least squares estimation)

Θεωρούμε το παρακάτω γραμμικό μοντέλο γέννησης δεδομένων:

$$y_i = \mathrm{x}_i^T \mathrm{w} + \varepsilon_i, \;\; \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \;\; i = 1, 2, \ldots, N,$$

όπου $\mathrm{x}_i = [1, x_{i1}, \ldots, x_{ip}]^T$ και $\mathrm{w} = [w_0, w_1, \ldots, w_p]^T$ άγνωστο διάνυσμα παραμέτρων. Έστω επίσης X ο $N \times (p+1)$ πίνακας με $N > p$, του οποίου γραμμές είναι τα $\mathrm{x}_i^T, i = 1, 2, \ldots, N$ και για τον οποίο υποθέτουμε ότι $\mathrm{rank(X)} = p + 1$.

(α) Με βάση το παραπάνω μοντέλο, το $N \times 1$ διάνυσμα $\mathrm{y} = [y_1, y_2, \ldots, y_N]^T$ θα είναι τυχαίο διάνυσμα (τ.δ.). Προσδιορίστε τη συνάρτηση πυκνότητας πιθανότητας του τ.δ. αυτού.

(β) Αν θεωρήσουμε τα $\{\mathrm{x}_i, y_i\}, i = 1, 2, \ldots, N$ ως δεδομένα εκπαίδευσης, βρείτε τον εκτιμητή ελαχίστων τετραγώνων $\hat{\mathrm{w}}$ του w.

(γ) Δείξτε ότι ο εκτιμητής ελαχίστων τετραγώνων $\hat{\mathrm{w}}$ είναι αμερόληπτος, δηλαδή $\mathbb{E}[\hat{\mathrm{w}}] = \mathrm{w}$. Προσδιορίστε επίσης τον πίνακα συμμεταβλητότητας (covariance matrix) του $\hat{\mathrm{w}}$.

(δ) Έστω $\hat{y}_i = \mathrm{x}_i^T \hat{\mathrm{w}}$ και $\hat{\mathrm{y}} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N]^T$. Βρείτε την κατανομή του τ.δ. $\hat{\mathrm{y}}$ και δείξτε ότι το $\hat{\mathrm{y}}$ είναι η προβολή του y στον χώρο στηλών του X.

(ε) Έστω τώρα $\hat{\varepsilon}_i = y_i - \hat{y}_i$ και $\hat{\boldsymbol{\varepsilon}} = [\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_N]$, το διάνυσμα των "ελαχίστων σφαλμάτων". Βρείτε την κατανομή του τ.δ. $\hat{\boldsymbol{\varepsilon}}$. Στη συνέχεια αποδείξτε ότι το σφάλμα ελαχίστων τετραγώνων $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^N \hat{\varepsilon}_i^2$ μπορεί να εκφραστεί ως εξής: $\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}} = \mathrm{y}^T (\mathrm{I} - \mathrm{P}) \mathrm{y}$, όπου P είναι ο πίνακας προβολής στον χώρο στηλών του X.

Άσκηση 1.2 (Multivariate Gaussian distribution)

Έστω x, y, τυχαίες μεταβλητές που ακολουθούν την κανονική κατανομή με μέσες τιμές $\mu_x = 5, \mu_y = 10$, τυπικές αποκλίσεις $\sigma_x = 3, \sigma_y = 4$ και συντελεστή συσχέτισης $\rho = 0.6$.

(α) Να δοθεί η από κοινού συνάρτηση πυκνότητας (σ.π.π.) των x, y.

(β) Προσδιορίστε την υπό συνθήκη σ.π.π. $p(y|x = 6)$ και υπολογίστε την πιθανότητα $p(y > 10|x = 6)$.

(γ) Επαναλάβετε το ερώτημα (β) για την σ.π.π. $p(x|y = 9)$ και την πιθανότητα $p(x > 5|y = 9)$

(δ) Βρείτε την εξίσωση της ισοσταθμικής καμπύλης για την οποία η από κοινού σ.π.π. των x, y έχει την τιμή που παίρνει για $(x, y) = (6, 12)$ και σχεδιάστε την.

Άσκηση 1.3 (Bayes classifier - 3 classes)

Θεωρούμε το πρόβλημα ταξινόμησης σε τρεις κλάσεις στο $\mathbb{R}^2$, με prior πιθανότητες $P(\omega_1) = 0.2, P(\omega_2) = 0.5, P(\omega_3) = 0.3$, αντίστοιχα. Επιπλέον υποθέτουμε ότι τα σημεία των κλάσεων προέρχονται από Gaussian κατανομές, δηλαδή $\mathrm{x}|\omega_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma), \mathrm{x}|\omega_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma), \mathrm{x}|\omega_3 \sim \mathcal{N}(\boldsymbol{\mu}_3, \Sigma)$ με

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \;\; \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \;\; \boldsymbol{\mu}_3 = \begin{bmatrix} 6 \\ 7 \end{bmatrix}, \;\; \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

(α) Αν δοθεί μια νέα παρατήρηση $\mathrm{x} = (3, 5)$, υπολογίστε τις posterior πιθανότητες $p(\omega_1|\mathrm{x})$, $p(\omega_2|\mathrm{x})$ και $p(\omega_3|\mathrm{x})$.
(β) Ταξινομήστε το πρότυπο $\mathrm{x} = (3, 5)$ σε μία από τις κλάσεις με βάση τον κανόνα ταξινόμησης κατά Bayes.
(γ) Βρείτε τις καμπύλες απόφασης μεταξύ κάθε ζεύγους κλάσεων (ω_1 και ω_2, ω_1 και ω_3, ω_2 και ω_3).
(δ) Να παραχθούν 500 σημεία από κάθε μία από τις παραπάνω κανονικές κατανομές και να δοθεί ένα σχήμα στο οποία θα φαίνονται όλα τα σημεία, τα κέντρα των κλάσεων και οι βέλτιστες καμπύλες απόφασης που βρήκατε στο (γ). Σχολιάστε.
(ε) Δίνεται ότι μια παρατήρηση x ανήκει στην κλάση ω_2. Υπολογίστε με αριθμητικές μεθόδους σε Python ή Matlab την πιθανότητα λανθασμένης ταξινόμησης του x, αν πραγματοποιούμε ταξινόμηση κατά Bayes.

Άσκηση 1.4 (Perceptron - MultiLayer Perceptron)

Θεωρήστε επέκταση του προβλήματος XOR, στην οποία η έξοδος είναι:

$$y = (x_1 + x_2) \mod (3) \tag{1}$$

ενώ οι είσοδοι x_1, x_2 παίρνουν τιμές 0, 1 και 2.
(α) Προσδιορίστε αναλυτικά (και όχι εκτελώντας αλγόριθμο ανανέωσης βαρών) έναν ταξινομητή τύπου MLP χρησιμοποιώντας μια μη γραμμική συνάρτηση ενεργοποίησης της επιλογής σας. Περιγράψτε αναλυτικά την αρχιτεκτονική και καταγράψτε τα βάρη και τις πολώσεις των νευρώνων. Η τελική αρχιτεκτονική του νευρωνικού δικτύου πρέπει να έχει τον ελάχιστο δυνατό αριθμό νευρώνων.
(β) Δείξτε σε διάγραμμα τα σημεία με τις εξόδους τους και σχεδιάστε τα όρια απόφασης.
(γ) Επιβεβαιώστε ότι το δίκτυο που προτείνατε υπολογίζει σωστά τις εξόδους για κάθε δυνατό συνδυασμό εισόδων.

Άσκηση 1.5 (Support Vector Machines - Kernels)

(α) Οι συναρτήσεις πυρήνα (kernel functions) ορίζουν έμμεσα μια συνάρτηση απεικόνισης $\phi(\cdot)$ που μετασχηματίζει ένα δείγμα εισόδου $\mathrm{x} \in \mathbb{R}^d$ σε έναν χώρο υψηλής διάστασης Q, δίνοντας τη μορφή του εσωτερικού γινομένου στον Q: $K(\mathrm{x}_i, \mathrm{x}_j) \equiv \langle \phi(\mathrm{x}_i), \phi(\mathrm{x}_j) \rangle$.

(α1) Αποδείξτε εάν ο πυρήνας είναι συμμετρικός ή όχι, δηλαδή εάν ισχύει $K(\mathrm{x}_i, \mathrm{x}_j) = K(\mathrm{x}_j, \mathrm{x}_i)$.

(α2) Υποθέστε ότι χρησιμοποιούμε μια συνάρτηση πυρήνα RBF (radial basis function) $K(\mathrm{x}_i, \mathrm{x}_j) = exp(-\frac{1}{2}\|\mathrm{x}_i - \mathrm{x}_j\|^2)$. Επομένως, υπάρχει μια έμμεσα ορισμένη και άγνωστη συνάρτηση απεικόνισης $\phi(\mathrm{x})$. Αποδείξτε ότι για οποιεσδήποτε δύο εισόδους $\mathrm{x}_i, \mathrm{x}_j$, η τετραγωνική Ευκλείδεια απόσταση των αντίστοιχων σημείων τους στον χώρο χαρακτηριστικών Q είναι μικρότερη από 2, δηλαδή αποδείξτε ότι:

$$\|\phi(\mathrm{x}_i) - \phi(\mathrm{x}_j)\|^2 \leq 2 \tag{2}$$

(β) Με τη βοήθεια της συνάρτησης πυρήνα, ένα SVM προσπαθεί να κατασκευάσει έναν υπερεπίπεδο στον χώρο χαρακτηριστικών Q, ο οποίος μεγιστοποιεί το περιθώριο μεταξύ δύο κλάσεων. Η απόφαση ταξινόμησης για οποιοδήποτε x λαμβάνεται με βάση το πρόσημο της ποσότητας:

$$\langle \hat{\mathrm{w}}, \phi(\mathrm{x}) \rangle + \hat{w_0} = \sum_{t \in SV} y_i \alpha_i K(\mathrm{x_i}, \mathrm{x}) + \hat{w_0} = f(\mathrm{x}; \alpha, \hat{w_0}) \tag{3}$$

όπου οι $\hat{\mathrm{w}}$ και $\hat{w_0}$ είναι παράμετροι του υπερεπιπέδου ταξινόμησης στον χώρο χαρακτηριστικών Q, το SV είναι το σύνολο των διανυσμάτων στήριξης, και το α_i είναι ο συντελεστής για το i-στο διάνυσμα στήριξης. Θεωρήστε ότι χρησιμοποιούμε ξανά τη συνάρτηση πυρήνα RBF και ποθέστε ότι τα δείγματα εκπαίδευσης είναι γραμμικά διαχωρίσιμα στον χώρο χαρακτηριστικών Q και ότι το SVM βρίσκει ένα περιθώριο που διαχωρίζει τέλεια τα σημεία.

Αν επιλέξουμε ένα σημείο ελέγχου x_{far} που βρίσκεται μακριά από οποιοδήποτε δείγμα εκπαίδευσης x_i (η απόσταση εδώ μετράται στον αρχικό χώρο $\mathbb{R}^d$), αποδείξτε ότι:

$$f(x_{far}; \alpha, \hat{w_0}) \approx \hat{w_0}. \tag{4}$$