Ασκήσεις για το μάθημα «Αναγνώριση Προτύπων» - 4

1. Δίνονται τα σημεία $\boldsymbol{x}_1 = [-1,\ 1]^T$, $\boldsymbol{x}_2 = [-1,\ -1]^T$, $\boldsymbol{x}_3 = [1,\ -1]^T$, $\boldsymbol{x}_4 = [1,\ 1]^T$, εκ των οποίων τα δύο πρώτα (δύο τελευταία) ανήκουν στην κατηγορία ω_1 (ω_2), η οποία σημειώνεται με $+1$ (-1). Έτσι, οι ετικέτες των κλάσεων για τα παραπάνω διανύσματα είναι, αντίστοιχα, $y_1 = 1$, $y_2 = 1$, $y_3 = -1$, $y_4 = -1$. Είναι σαφές ότι το πρόβλημα είναι γραμμικώς διαχωρίσιμο. Προσδιορίστε τη γραμμή εκείνη που αφήνει το μέγιστο περιθώριο από τα διανύσματα των δύο κλάσεων, διατυπώνοντας και λύνοντας το αντίστοιχο (γραμμικό) πρόβλημα μηχανών διανυσματικής στήριξης (support vector machines - SVM).

   <u>Υπόδειξη</u>: Προσδιορίστε τους πολλαπλασιαστές Lagrange για τα παραπάνω σημεία λύνοντας το δυϊκό πρόβλημα

   $$\max_{\lambda_i}\left(\sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j\right)$$

   υπό τις προϋποθέσεις

   $$\sum_{i=1}^{N}\lambda_i y_i = 0, \qquad \lambda_i \geq 0, \quad i = 1, \ldots, N$$

   Στη συνέχεια προσδιορίστε το διάνυσμα $\boldsymbol{w}$ με βάση τον τύπο

   $$\boldsymbol{w} = \sum_{i=1}^{N}\lambda_i y_i \boldsymbol{x}_i$$

   και το w_0 χρησιμοποιώντας κάθε μία από τις σχέσεις $\lambda_i[y_i(\boldsymbol{w}^T\boldsymbol{x}_i + w_0) - 1] = 0$, $i = 1, 2, \ldots, N$ και παίρνοντας το μέσο όρο των τιμών που προκύπτουν.

2. Δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2 ως εξής

   ω_1: $[0.1, -0.2]^T, [0.2, 0.1]^T, [-0.15, 0.2]^T, [1.1, 0.8]^T, [1.2, 1.1]^T$

   ω_2: $[1.1, -0.1]^T, [1.25, 0.15]^T, [0.9, 0.1]^T, [0.1, 1.2]^T, [0.2, 0.9]^T$

   Ελέγξτε αν αυτά είναι γραμμικά διαχωρίσιμα και, αν όχι, σχεδιάστε ένα κατάλληλο perceptron πολλών επιπέδων με κόμβους που έχουν ως συνάρτηση ενεργοποίησης τη βηματική συνάρτηση, προκειμένου να ταξινομήσετε τα διανύσματα σε δύο κλάσεις.

3. Σχεδιάστε τις τρεις γραμμές στο δισδιάστατο χώρο

   $$x_1 + x_2 = 0$$

   $$x_2 = \frac{1}{4}$$

   $$x_1 - x_2 = 0$$

   Για κάθε ένα από τα πολύεδρα που σχηματίζονται από τις τομές τους, καθορίστε τις κορυφές του κύβου στον οποίο θα αντιστοιχηθούν από το πρώτο επίπεδο ενός perceptron πολλών επιπέδων, υλοποιώντας τις προηγούμενες γραμμές. Συνδυάστε τις περιοχές σε δύο κλάσεις έτσι ώστε (α) να επαρκεί ένα δίκτυο δύο επιπέδων για να τις ταξινομήσει και (β) να χρειάζεται ένα δίκτυο τριών επιπέδων. Και για τις δύο περιπτώσεις υπολογίστε αναλυτικά τα αντίστοιχα συναπτικά βάρη.

4. Αποδείξτε ότι οποιαδήποτε Boolean συνάρτηση μπορεί να υλοποιηθεί από ένα δίκτυο perceptron δύο επιπέδων.

5. Δείξτε ότι αν τα $\boldsymbol{x}_1$ και $\boldsymbol{x}_2$ είναι δύο σημεία στον l-διάστατο χώρο, το υπερεπίπεδο που διχοτομεί το τμήμα με άκρα τα σημεία $\boldsymbol{x}_1, \boldsymbol{x}_2$, αφήνοντας το $\boldsymbol{x}_1$ στη θετική του πλευρά, δίνεται από
$$(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T \boldsymbol{x} - \frac{1}{2}\|\boldsymbol{x}_1\|^2 + \frac{1}{2}\|\boldsymbol{x}_2\|^2 = 0$$

6. Δείξτε ότι με την απεικόνιση
$$\boldsymbol{x} \equiv [x_1 \ x_2]^T \xrightarrow{\phi} [x_1^3 \ \sqrt{3}x_1^2 x_2 \ \sqrt{3}x_1 x_2^2 \ x_2^3]^T \equiv \boldsymbol{y}$$
από τον $\mathcal{R}^2$ στον $\mathcal{R}^4$, το εσωτερικό γινόμενο μεταξύ δύο διανυσμάτων στον χώρο $\mathcal{R}^2$ μπορεί να χρησιμοποιηθεί για τον υπολογισμό του εσωτερικού γινομένου των εικόνων των διανυσμάτων αυτών στο χώρο $\mathcal{R}^4$. Βρείτε την ακριβή σχέση μεταξύ των εσωτερικών γινομένων των δύο χώρων.

7. Έστω ένα σύνολο δεδομένων $X = \{(\boldsymbol{x}_i, y_i),\ \boldsymbol{x}_i \in \mathcal{R}^l,\ y_i \in \{0,1\},\ i = 1, \ldots, N\}$. Εξάγετε και διατυπώστε σε αλγοριθμική μορφή αλγόριθμο απότομης κατάδυσης για την εκπαίδευση ενός perceptron (ένας νευρώνας) με συνάρτηση εξόδου (ενεργοποίησης) την $f(z) = \frac{1}{1+e^{-az}}$, όπου a παράμετρος που καθορίζεται από το χρήστη, χρησιμοποιώντας ως συνάρτηση κόστους (α) την συνάρτηση του αθροίσματος των τετραγωνικών σφαλμάτων και (β) τη συνάρτηση δι-εντροπίας (cross entropy). Οι συναρτήσεις αυτές για την περίπτωσή μας γράφονται ως εξής:
$$J_1 = \frac{1}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$
και
$$J_2 = -\sum_{i=1}^{N}(y_i \ln \hat{y}_i + (1 - y_i)\ln(1 - \hat{y}_i))$$
όπου $\hat{y}_i = f(\boldsymbol{w}^T \boldsymbol{x}_i)$. Προκειμένου να κρατήσουμε το συμβολισμό όσο το δυνατόν πιο συμπαγή, αυξήσαμε τη διάσταση όλων των $\boldsymbol{x}_i$ κατά μία, θέτοντας την τελευταία συνιστώσα τους ίση με 1. Έτσι, το $\boldsymbol{w}$ γίνεται $\boldsymbol{w} = [w_1, w_2, \ldots, w_l, w_0]^T$. Συγκρίνετε τους αλγόριθμους που προκύπτουν.

<u>Υπόδειξη</u>: (i) Εκφράστε την παράγωγο της $f(z)$, $f'(z)$, σαν συνάρτηση της $f(z)$.
(ii) Για κάθε μία από τις παραπάνω περιπτώσεις υπολογίστε το $\frac{\partial J}{\partial \boldsymbol{w}}$ εκφράζοντάς το σε μορφή που να **μην** περιέχει την παράγωγο της $f(z)$. Στην συνέχεια χρησιμοποιήστε την εξίσωση του αλγορίθμου απότομης κατάδυσης και αντικαταστήστε το $\frac{\partial J}{\partial \boldsymbol{w}}$ με το ίσο του, που προέκυψε από τους υπολογισμούς σας.