Ασκήσεις για το μάθημα «Αναγνώριση Προτύπων» - 3

1. Θεωρείστε ένα πρόβλημα M κλάσεων στον l-διάστατο χώρο, όπου κάθε κλάση μοντελοποιείται από μία κανονική κατανομή. Κάτω από ποιες συνθήκες ο απλοϊκός ταξινομητής Bayes είναι ισοδύναμος με τον ταξινομητή Bayes; Μπορείτε για την περίπτωση αυτή να δώσετε μία ισοδύναμη μορφή του η οποία δεν απαιτεί υπολογισμό εκθετικών;

2. Αποδείξτε ότι η προβολή ενός σημείου $\boldsymbol{x}$ πάνω στο υπερεπίπεδο $g(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0 = 0$ είναι το σημείο

$$\boldsymbol{x}_p = \boldsymbol{x} - \frac{g(\boldsymbol{x})}{||\boldsymbol{w}||^2}\boldsymbol{w}$$

   Υπόδειξη: Ελαχιστοποιήστε τη συνάρτηση $||\boldsymbol{x} - \boldsymbol{x}_p||^2$ υπό τον περιορισμό ότι $g(\boldsymbol{x}_p) = 0$. Για το σκοπό αυτό ορίστε τη συνάρτηση Lagrange $\mathcal{L} = \frac{1}{2}||\boldsymbol{x} - \boldsymbol{x}_p||^2 - \lambda g(\boldsymbol{x}_p)$, όπου λ είναι ένας παράγοντας που ονομάζεται πολ/στής Lagrange. Θέσατε την πρώτη παράγωγο της $\mathcal{L}$ ίση με το $\mathbf{0}$ και χρησιμοποιήστε και την εξίσωση-περιορισμό $g(\boldsymbol{x}_p) = \boldsymbol{w}^T\boldsymbol{x}_p + w + 0 = 0$, προκειμένου να υπολογίσετε τα $\boldsymbol{x}_p$ και λ.

3. Η *κυρτή θήκη* (*convex hull*) ενός συνόλου διανυσμάτων $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ είναι το σύνολο όλων των διανυσμάτων που γράφονται στη μορφή

$$\boldsymbol{x} = \sum_{i=1}^{n} a_i\boldsymbol{x}_i, \qquad \mu\epsilon\ (i)\ a_i \geq 0,\ i = 1, \ldots, n,\ (ii)\ \sum_{i=1}^{n} a_i = 1$$

   Έστω δύο σύνολα δεδομένων, $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}\}$ και $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2}\}$, που περιέχουν στοιχεία από δύο κλάσεις, ω_1 και ω_2, αντίστοιχα. Δείξτε ότι: είτε τα X_1 και X_2 θα είναι γραμμικώς διαχωρίσιμα, είτε οι κυρτές θήκες τους θα τέμνονται.

   Υπόδειξη: Υποθέστε ότι οι δύο παραπάνω προτάσεις είναι ταυτόχρονα αληθείς και εξετάστε την ταξινόμηση ενός σημείου που βρίσκεται στην τομή των κυρτών θηκών.

4. Θεωρείστε ένα πρόβλημα M κλάσεων όπου η ταξινόμηση γίνεται με τη χρήση γραμμικών συναρτήσεων διάκρισης της μορφής $g_i(\boldsymbol{x}) = \boldsymbol{w}_i^T\boldsymbol{x} + w_{i0}$, $i = 1, \ldots, M$. Αποδείξτε ότι οι περιοχές απόφασης που αντιστοιχούν σε κάθε κλάση είναι κυρτές (convex), δείχνοντας ότι αν $\boldsymbol{x}_1 \in \mathcal{R}_i$ και $\boldsymbol{x}_2 \in \mathcal{R}_i$, τότε $\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \mathcal{R}_i$, για $0 \leq \lambda \leq 1$, όπου $\mathcal{R}_i$ η περιοχή απόφασης για την κλάση ω_i..

   Υπόδειξη: Θυμηθείτε ότι, στην περίπτωση αυτή, ένα σημείο $\boldsymbol{x}$ καταχωρείται στην κατηγορία ω_i αν $g_i(\boldsymbol{x}) > g_j(\boldsymbol{x})$, για $j \neq i$, $i = 1, \ldots, M$.

5. Θεωρείστε ένα πρόβλημα ταξινόμησης δύο κλάσεων όπου η κλάση ω_1 αποτελείται από τα διανύσματα $[0,\ 0]^T$ και $[0,\ 1]^T$, ενώ η κλάση ω_2 αποτελείται από τα διανύσματα $[1,\ 0]^T$ και $[1,\ 1]^T$. Χρησιμοποιήστε την μορφή «επιβράβευσης και τιμωρίας» του αλγορίθμου perceptron, με $\rho = 1$ και $\boldsymbol{w}(0) = [0,\ 0]^T$, προκειμένου να προσδιορίσετε μία λύση στο πρόβλημα. Στη συνέχεια, σχεδιάστε την αντίστοιχη ευθεία πάνω στο χώρο των δεδομένων.

6. Θεωρείστε την περίπτωση ενός προβλήματος ταξινόμησης δύο κλάσεων στον l-διάστατο χώρο, όπου οι συνιστώσες των διανυσμάτων μπορούν να παίρνουν τις τιμές 0 ή 1. Ας υποθέσουμε επίσης ότι ένα διάνυσμα $\boldsymbol{x}$ καταχωρείται στην κλάση ω_1 αν ο αριθμός των μη μηδενικών συνιστωσών του είναι περιττός, ενώ, διαφορετικά, καταχωρείται στην ω_2. Αποδείξτε ότι το πρόβλημα αυτό δεν είναι γραμμικά διαχωρίσιμο για $l > 1$.

7. Θεωρείστε ένα πρόβλημα δύο γραμμικώς διαχωρίσιμων κλάσεων, ω_1 και ω_2, για τις οποίες είναι διαθέσιμο ένα σύνολο δεδομένων διανυσμάτων $X = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N),\ \boldsymbol{x}_i \in \mathcal{R}^l,\ y_i \in \{-1,\ +1\},\ i = 1, \ldots, N\}$

   Αποδείξτε ότι ο αλγόριθμος perceptron συγκλίνει σε πεπερασμένο αριθμό βημάτων σε λύση. Θεωρείστε το ρ σταθερό και μικρότερο του 2.

   Υπόδειξη: Έστω $\boldsymbol{w}^*$ μία λύση. Αν a είναι μία κατάλληλα επιλεγμένη σταθερά (επιλέξτε την στο τέλος…) αποδείξτε ότι η ακολουθία $||\boldsymbol{w}(t) - a\boldsymbol{w}^*||^2$ είναι φθίνουσα εργαζόμενοι ως εξής:

   (α) Αποδείξτε ότι $||\boldsymbol{w}(t+1) - a\boldsymbol{w}^*||^2 \leq ||\boldsymbol{w}(t) - a\boldsymbol{w}^*||^2 - c$, όπου c μία σταθερή ποσότητα (αντικαταστήστε στον τύπο αυτό το $\boldsymbol{w}(t+1)$ με το ίσο του από την εξίσωση ενημέρωσης του αλγορίθμου perceptron και κάνετε τις πράξεις).

   (β) Εφαρμόστε αναδρομικά τον παραπάνω τύπο και φράξτε τον αριθμό των επαναλήψεων, t.

8. Θεωρείστε ένα δισδιάστατο πρόβλημα δύο κλάσεων, ω_1 και ω_2, για το οποίο έχουμε ότι τα διανύσματα $[0,\ 0]^T$, $[0,\ 1]^T$, $[1,\ 0]^T$, $[1,\ 1]^T$ ανήκουν στην κατηγορία ω_1, ενώ τα $[2,\ 0]^T$, $[2,\ 1]^T$, $[3,\ 0]^T$, $[3,\ 1]^T$ ανήκουν στην κατηγορία ω_2. Προσδιορίστε τον βέλτιστο γραμμικό ταξινομητή ως προς το κριτήριο του αθροίσματος των τετραγώνων των σφαλμάτων (sum of error squares) και σχεδιάστε τον στο χώρο των δεδομένων. Σας θυμίζει κάτι από τον ταξινομητή Bayes η λύση που προσδιορίσατε;