

## Επεξεργασία Δεδομένων και Αλγόριθμοι Μάθησης

### 2<sup>ο</sup> Σετ Ασκήσεων

**Πρόβλημα 1:** Θα θέλαμε να αξιολογήσουμε τις δυνατότητες της πρώτης μεθόδου kernel στο να προσεγγίζει πυκνότητες πιθανότητας. Για τον σκοπό αυτό δημιουργούμε 1000 υλοποιήσεις μιας τυχαίας μεταβλητής ομοιόμορφα κατανομημένης στο διάστημα  $[0, 1]$ . α) Προσεγγίστε την πυκνότητα πιθανότητας χρησιμοποιώντας το Gaussian kernel

$$K(x, h) = \frac{1}{\sqrt{2\pi}h} e^{-\frac{1}{2h}x^2}.$$

Ζωγραφίστε την αντίστοιχη προσέγγιση για διαφορετικές τιμές του  $h$ . Μην περιορίσετε το γράφημά σας στο  $[0, 1]$  αλλά να το επεκτείνετε πέραν των δύο αυτών ορίων. Τι παρατηρείτε σχετικά με τον τρόπο που προσεγγίζετε το διάστημα  $[0, 1]$  όπου “ζει” η τυχαία μεταβλητή καθώς τον τρόπο που προσεγγίζετε την τιμή 1 που είναι το σταθερό πλάτος της συνάρτησης πυκνότητας πιθανότητας. β) Ολοκληρώστε τα ίδια ακριβώς βήματα με το Laplacian kernel

$$K(x, h) = \frac{1}{2h} e^{-\frac{1}{h}|x|}.$$

**Πρόβλημα 2:** Το Matlab αρχείο δεδομένων hw2-2data.mat περιέχει δύο μήτρες: stars και circles όπου κάθε μια είναι μια λίστα από δι-διάστατα (2-D) διανύσματα. Κάθε 2-D διάνυσμα αντιστοιχεί σε ένα σημείο στον 2-D επίπεδο και έχει σαν ετικέτα (label) είτε “star” ή “circle”. Ενδιαφερόμαστε να δημιουργήσουμε ένα classifier ο οποίος να διακρίνει μεταξύ των δύο συνόλων. Θα χρησιμοποιήσουμε τη μέθοδο με τα kernel ώστε να βρούμε ένα διαχωριστικό σύνορο. Για το σκοπό αυτό αντιστοιχούμε την αριθμητική ετικέτα “1” στο stars και την “-1” στο circles. Καλούμε  $\phi(X)$ ,  $X = [x_1, x_2]^T$  το μετασχηματισμό που επιθυμούμε να εφαρμόσουμε στα δεδομένα ο οποίος θα υλοποιεί την εν λόγω αντιστοίχιση. Προκειμένου να προσδιορίσουμε τον  $\phi(X)$  επιλύουμε το παρακάτω πρόβλημα βελτιστοποίησης

$$\min_{\phi \in \mathcal{V}} \left\{ \sum_{X_i \in \text{stars}} (1 - \phi(X_i))^2 + \sum_{X_j \in \text{circles}} (1 + \phi(X_j))^2 + \lambda \|\phi(X)\|^2 \right\}, \quad (1)$$

όπου  $\mathcal{V}$  είναι ο διανυσματικός χώρος των συναρτήσεων που ορίζεται με τη βοήθεια του Gaussian kernel

$$K(X, Y) = e^{-\frac{1}{h}\|X-Y\|^2} = e^{-\frac{1}{h}\{(x_1-y_1)^2+(x_2-y_2)^2\}}.$$

α) Χρησιμοποιείτε το Representer Theorem για να αποδείξετε ότι στα πρώτα δύο αθροίσματα μπορούμε να αντικαταστήσουμε το  $\phi(X)$  με την ορθογώνια προβολή του  $\hat{\phi}(X)$  πάνω στον γραμμικό υπόχωρο ο οποίος δημιουργείται από τις συναρτήσεις  $K(X, X_i)$ ,  $X_i \in \text{stars}$  and  $K(X, X_j)$ ,  $X_j \in \text{circles}$  δηλαδή

$$\hat{\phi}(X) = \sum_{X_i \in \text{stars}} \alpha_i K(X, X_i) + \sum_{X_j \in \text{circles}} \beta_j K(X, X_j). \quad (2)$$

β) Για τον όρο  $\|\phi(X)\|^2$  χρησιμοποιείτε την Αρχή της Ορθογωνιότητας για να δείξετε επίσης ότι

$$\|\phi(X)\|^2 = \|\hat{\phi}(X)\|^2 + \|\phi(X) - \hat{\phi}(X)\|^2 \geq \|\hat{\phi}(X)\|^2.$$

Η παρατήρηση αυτή υπονοεί ότι μπορούμε να αντικαταστήσουμε το  $\phi(X)$  με το  $\hat{\phi}(X)$  στην αρχική συνάρτηση κόστους στο (1); Εάν ναι, εξηγήστε το ΓΙΑΤΙ; γ) Χρησιμοποιώντας τη μορφή του  $\hat{\phi}(X)$  που ορίστηκε στην (2), υπολογίστε τους βέλτιστους συντελεστές  $\alpha_i, \beta_j$ . δ) Όταν προσδιορίσετε το βέλτιστο  $\hat{\phi}(X)$  εξηγήστε πως θα το χρησιμοποιήσετε για να κατηγοριοποιήσετε ένα νέο σημείο  $X_{\text{new}}$  σε “star” ή “circle” λαμβάνοντας υπόψη ότι φυσικά η τιμή του  $\hat{\phi}(X_{\text{new}})$  δεν θα είναι ακριβώς 1 ή -1. ε) Αφού καταλήξετε στον τελικό σας κανόνα κατηγοριοποίησης στο δ) βρείτε (αριθμητικά) το διαχωριστικό σύνορο για τις δύο κλάσεις στον 2-D χώρο. Επίσης τοποθετείστε τα δεδομένα εκπαίδευσης στο 2-D επίπεδο ώστε να επιβεβαιώσετε την ποιότητα του συνόρου σας. Επαναλάβετε τη διαδικασία για διαφορετικές τιμές των παραμέτρων  $h$  και  $\lambda$ .

**Παράδοση έως Τετάρτη, 27 Μαΐου, 12:00 (μεσημέρι) ΤΟ ΑΡΓΟΤΕΡΟ!!!**