Ε.Μ.Π., Σχολή Η.Μ.& Μ.Υ. **Αναγνώριση Προτύπων** Ακαδ.Ετος 2018-19
Ημερ/νία: 02-11-2018 **1ο Σύνολο Αναλυτ. Ασκήσεων** Παραδοτέο: 19-11-2018

Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός βιβλίου για την λύση σας, πρέπει να το αναφέρετε. Παραδίδονται γραπτώς και προσωπικώς στην Γραμματεία Εργ. Ρομποτικής (Αιθ. 2.1.12, παλαιό Κτ.Ηλεκτρ.) 9.00-14.30.

Υλικό για Ανάγνωση:
Βιβλίο [1]: I, II, III, IV. Βιβλίο [2]: Κεφ. 1, 2, 3, 5, 10.
Βιβλίο [3]: Κεφ. 1, 2, 9. Βιβλίο [4]: Κεφ. 1, 2, 3.
Διαφάνειες διαλέξεων μαθήματος

Αναλυτικές Ασκήσεις

Άσκηση 1.1: (Discriminant functions for Bayesian classifier)
Θεωρήστε την περίπτωση της ελαχιστοποίησης του σφάλματος ενός ταξινομητή Bayes και υποθέστε ότι τα δεδομένα ακολουθούν πολυδιάστατες κανονικές κατανομές με κοινό πίνακα συνδιασποράς $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ και διαφορετικές μέσες τιμές $\boldsymbol{\mu}_i$ για την κάθε κλάση ω_i. Σε αυτή την περίπτωση, η διαχωριστική συνάρτηση (discriminant function) $g_i(\mathbf{x})$ της κλάσης ω_i απλοποείται ως εξής:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) + \ln P(\omega_i)$$

Επίσης, έστω ότι η διαχωριστική επιφάνεια μεταξύ των περιοχών $\mathcal{R}_i$ και $\mathcal{R}_j$ δύο κλάσεων ω_i και ω_j δίνεται από την επίλυση της εξίσωσης $g(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0$. Να αποδείξετε ότι η διαχωριστική αυτή επιφάνεια ισοδυναμεί με το υπερεπίπεδο

$$\mathbf{w}^\top(\mathbf{x} - \mathbf{x}_0) = 0$$

όπου

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$\mathbf{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\ln [P(\omega_i)/P(\omega_j)]}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

Άσκηση 1.2: (MLE-MAP)
Θεωρήστε N ανεξάρτητα δείγματα $x_1, \ldots, x_n$ από μία μονοδιάστατη κανονική κατανομή με γνωστή διασπορά σ^2 και άγνωστη μέση τιμή μ.
α) Βρείτε τη σχέση για τον εκτιμητή μέγιστης πιθανοφάνειας (Maximum Likelihood Estimator - MLE) της μέσης τιμής μ, παρουσιάζοντας όλα τα ενδιάμεσα βήματα.
β) Βρείτε τη σχέση για τον Maximum A Posteriori (MAP) εκτιμητή της μέσης τιμής μ. Υποθέστε ότι η κατανομή της μέσης τιμής αποτελεί και η ίδια μία κανονική κατανομή με μέση τιμή υ και διασπορά β^2.
γ) Καθώς ο αριθμός των δειγμάτων N τείνει στο άπειρο, τι προκύπτει για τους MLE και MAP

εκτιμητές;

Άσκηση 1.3: (Entropies of probability distributions)
α) Να υπολογιστούν οι εντροπίες για τις ακόλουθες τρεις περιπτώσεις κατανομών: α) γκαουσιανή β) ομοιόμορφη και γ) τριγωνική. Θεωρήστε ότι οι τρεις αυτές κατανομές έχουν μηδενική μέση τιμή και την ίδια διασπορά σ^2.
β) Να υπολογιστεί η εντροπία της πολυδιάστατης γκαουσιανής κατανομής $p(\mathbf{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Άσκηση 1.4: (Bayesian parameter estimation)
Έστω ότι η κατανομή $p(\mathbf{x}|\boldsymbol{\Sigma})$ ακολουθεί την κανονική κατανομή $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπου το διάνυσμα μέσης τιμής $\boldsymbol{\mu}$ θεωρείται γνωστό και ο πίνακας συνδιασποράς $\boldsymbol{\Sigma}$ άγνωστος. Εάν έχουμε n δεδομένα $\mathbf{x}_1, \ldots, \mathbf{x}_n$, να δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας του $\boldsymbol{\Sigma}$ δίνεται από τη σχέση:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n}(\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^\top$$

Για να το αποδείξετε ακολουθήστε τα παρακάτω βήματα:
α) Αποδείξτε την ιδιότητα $\mathbf{a}^\top \mathbf{A}\mathbf{a} = \mathrm{tr}\left[\mathbf{A}\mathbf{a}\mathbf{a}^\top\right]$, όπου το ίχνος, $\mathrm{tr}\,[\mathbf{A}]$, είναι το άθροισμα των διαγώνιων στοιχείων του πίνακα $\mathbf{A}$.
(β) Δείξτε ότι η συνάρτηση πιθανοφάνειας μπορεί να γραφεί στη μορφή:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{nd/2}}|\boldsymbol{\Sigma}^{-1}|^{n/2} \exp\left[-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{k=1}^{n}(\mathbf{x}_k-\boldsymbol{\mu})(\mathbf{x}_k-\boldsymbol{\mu})^\top\right]\right]$$

(γ) Έστω ότι $\mathbf{A} = \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}$ και $\lambda_1, \ldots, \lambda_n$ οι ιδιοτιμές του $\mathbf{A}$. Δείξτε ότι το παραπάνω αποτέλεσμα οδηγεί στη σχέση

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{nd/2}|\hat{\boldsymbol{\Sigma}}|^{n/2}}(\lambda_1 \cdots \lambda_d)^{n/2} \exp\left[-\frac{n}{2}(\lambda_1 + \cdots + \lambda_d)\right]$$

(δ) Ολοκληρώστε την απόδειξη δείχνοντας ότι η πιθανοφάνεια μεγιστοποιείται με την επιλογή $\lambda_1 = \cdots = \lambda_d = 1$. Εξηγήστε αναλυτικά.

Άσκηση 1.5: (Probabilistic discriminative models: Logistic Regression)
Η συνάρτηση logistic sigmoid ορίζεται ως

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

με παράγωγο που αποδεικνύεται εύκολα ότι είναι

$$\frac{d\sigma}{da} = \sigma(1-\sigma)$$

Για ένα σύνολο δεδομένων $\{\phi_n, t_n\}$, όπου $t_n \in \{0,1\}$ και $\phi_n = \phi(\mathbf{x}_n)$, με $n = 1, \ldots, N$, η συνάρτηση πιθανότητας μπορεί να γραφεί ως

$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} y_n^{t_n} \{1 - y_n\}^{1-t_n}$$

όπου $\mathbf{t} = (t_1, \ldots, t_N)^\top$ και $y_n = p(\mathcal{C}_1|\phi_n)$. Τότε, η συνάρτηση λάθους cross-entropy θα είναι της μορφής

$$E(\mathbf{w}) = -\ln p(\mathbf{t}|\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

όπου $y_n = \sigma(a_n)$ και $a_n = \mathbf{w}^\top \phi_n$. Να αποδείξετε ότι η παράγωγος της συνάρτησης λάθους για το μοντέλο logistic regression ισούται με

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (y_n - t_n)\phi_n$$

Άσκηση 1.6: (Perceptrons)
Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : [2, 1.5]^\top, [4, 1]^\top, [3, 3]^\top, [-2, 2]^\top, [0, 3]^\top$$

$$\omega_2 : [-1, 0]^\top, [0, 0]^\top, [2, -0.5]^\top, [1, -1]^\top, [-2, 1]^\top$$

Αρχικά, ελέγξτε εάν οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες, μέσω του σχεδιασμού των παραπάνω σημείων σε ένα γράφημα. Στη συνέχεια, χρησιμοποιήστε τον ακόλουθο αλγόριθμο εκπαίδευσης ενός perceptron (που είναι ισοδύναμος με τον σειριακό αλγόριθμο 4 του Κεφ. 5 του [2]), με $\rho = 1$ και $\mathbf{w}(0) = [0, 0]^\top$, για να σχεδιάσετε μία ευθεία γραμμή που να διαχωρίζει τις δύο κλάσεις.
Αλγόριθμος Perceptron: Έστω $\mathbf{w}(t)$ η εκτίμηση του διανύσματος βάρους και $\mathbf{x}_t$ το αντίστοιχο διάνυσμα χαρακτηριστικών στο t-οστό βήμα επανάληψης. Ο αλγόριθμος έχει ως εξής:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \rho \mathbf{x}_t \quad \text{αν} \quad \mathbf{x}_t \in \omega_1 \quad \text{και} \quad \mathbf{w}(t)^\top \mathbf{x}_t \leq 0$$

$$\mathbf{w}(t+1) = \mathbf{w}(t) - \rho \mathbf{x}_t \quad \text{αν} \quad \mathbf{x}_t \in \omega_2 \quad \text{και} \quad \mathbf{w}(t)^\top \mathbf{x}_t \geq 0$$

$$\mathbf{w}(t+1) = \mathbf{w}(t) \quad \text{αλλιώς}$$

Ο ανωτέρω αλγόριθμος έχει την μορφή αλγορίθμων τύπου reward and punishment. Δηλαδή, αν το τωρινό δείγμα εκπαίδευσης ταξινομηθεί σωστά, τότε δεν γίνεται τίποτα (reward = no action). Αλλιώς, αν το δείγμα δεν ταξινομηθεί σωστά, η τιμή του διανύσματος βάρους μεταβάλλεται προσθέτοντας (αφαιρώντας) μία τιμή ανάλογη του $\mathbf{x}_t$ (punishment = correction cost).

ΒΙΒΛΙΟΓΡΑΦΙΑ:
[1] Γ. Καραγιάννης και Γ. Σταϊνχάουερ, *Αναγνώριση Προτύπων και Μάθηση Μηχανών*, ΕΜΠ, 2001.
[2] R. O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, Wiley, 2001.
[3] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[4] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th Edition Academic Pres, Elsevier, 2009. *Ελληνική μετάφραση*: απόδοση-επιμέλεια-πρόλογος ελληνικής έκδοσης Α. Πικράκης, Κ. Κουτρουμπάς, Θ. Γιαννακόπουλος, Επιστημονικές Εκδόσεις Π.Χ. Πασχαλίδησ-Broken Hill Publishers LTD, 2012.