Ε.Μ.Π., Σχολή Η.Μ.& Μ.Υ. **Αναγνώριση Προτύπων** Ακαδ.Ετος 2019-20
Ημερ/νία: 27-11-2019 **1ο Σύνολο Αναλυτ. Ασκήσεων** Παραδοτέο: 11-12-2019

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:
Βιβλία: [1], [2], [3] και [4]
Διαφάνειες διαλέξεων μαθήματος

Αναλυτικές Ασκήσεις

Άσκηση 1.1: (Probabilities)
Θεωρήστε δύο Γκαουσιανές μονοδιάστατες κατανομές $\mathcal{N}(\mu_1, \sigma_1^2)$ και $\mathcal{N}(\mu_2, \sigma_2^2)$. Από τις κατανομές αυτές επιλέγουμε δύο τυχαία δείγματα x_1 και x_2, αντίστοιχα, και υπολογίζουμε το άθροισμά τους $x_3 = x_1 + x_2$. Η δειγματοληψία αυτή επαναλαμβάνεται διαρκώς.
(α) Να δείξετε ότι η προκύπτουσα κατανομή των τιμών της x_3 ακολουθεί επίσης κανονική κατανομή.
(β) Ποια είναι η μέση τιμή μ_3 αυτής της κατανομής;
(γ) Ποια είναι η διασπορά σ_3^2;
(δ) Επαναλάβετε τα παραπάνω βήματα για δύο πολυδιάστατες κατανομές $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ και $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.

Άσκηση 1.2: (Bayes Decision Theory)
Θεωρήστε ότι οι υπό συνθήκη κατανομές για ένα μονοδιάστατο πρόβλημα δύο κατηγοριών ω_1 και ω_2 δίνεται από την ακόλουθη κατανομή Cauchy:

$$p(x|\omega_i) = \frac{1}{\pi b} \cdot \frac{1}{1 + \left(\frac{x - a_i}{b}\right)^2}, \quad i = 1, 2$$

(α) Να ελέγξετε ότι η παραπάνω κατανομή είναι κανονικοποιημένη σωστά.
(β) Εάν υποθέσουμε ότι $P(\omega_1) = P(\omega_2)$, να δείξετε ότι $P(\omega_1|x) = P(\omega_2|x)$ αν $x = \frac{a_1 + a_2}{2}$. Με άλλα λόγια, ότι η διαχωριστική γραμμή απόφασης που οδηγεί στο ελάχιστο σφάλμα είναι το σημείο στη μέση των θέσεων μεγίστου των δύο κατανομών, ανεξαρτήτως του b.
(γ) Να δείξετε ότι η ελάχιστη πιθανότητα σφάλματος δίνεται από τη σχέση:

$$P(\text{error}) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left| \frac{a_1 - a_2}{2b} \right|$$

Άσκηση 1.3: (Bayes Decision Theory)
Ας υποθέσουμε ότι $p(\mathbf{x}|\omega_i) \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ για ένα πρόβλημα δύο κατηγοριών ω_1 και ω_2 και d διαστάσεων με τους ίδιους πίνακες συνδιασπορών, αλλά διαφορετικά διανύσματα για τις μέσες

τιμές και διαφορετικές εκ των προτέρων πιθανότητες. Θεωρήστε την υψωμένη στο τετράγωνο απόσταση Mahalanobis:

$$r_i^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

(α) Να δείξετε ότι το gradient της r_i^2 δίνεται από τη σχέση:

$$\nabla r_i^2 = 2\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)$$

(β) Να δείξετε ότι σε οποιοδήποτε σημείο μιας δοσμένης ευθείας που περνάει από το $\boldsymbol{\mu}_i$, το gradient ∇r_i^2 δείχνει πάντα στην ίδια διεύθυνση. Πρέπει αυτή η διεύθυνση να είναι παράλληλη με τη δοσμένη ευθεία;

(γ) Να δείξετε ότι τα ∇r_1^2 και ∇r_2^2 δείχνουν σε αντίθετες κατευθύνσεις κατά μήκος της γραμμής που συνδέει το $\boldsymbol{\mu}_1$ με το $\boldsymbol{\mu}_2$.

(δ) Να δείξετε ότι το βέλτιστο υπερεπίπεδο διαχωρισμού είναι εφαπτόμενο στα υπερελλειψοειδή σταθερής πυκνότητας πιθανότητας, στο σημείο όπου το υπερεπίπεδο διαχωρισμού τέμνει την ευθεία που συνδέει το $\boldsymbol{\mu}_1$ με το $\boldsymbol{\mu}_2$.

(ε) Σωστό ή Λάθος: Για ένα πρόβλημα δύο κατηγοριών με Γκαουσιανές κατανομές που έχουν διαφορετικές μέσες τιμές και πίνακες συνδιασπορών, και με $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, το διαχωριστικό κριτήριο απόφασης Bayes αποτελείται από το σύνολο των σημείων ίσης απόστασης Mahalanobis από τους αντίστοιχους διανυσματικούς μέσους. Να δικαιολογήσετε την απάντησή σας αναλυτικά.

Άσκηση 1.4: (Maximum Likelihood estimation)

Έστω ότι η μεταβλητή x ακολουθεί μια ομοιόμορφη κατανομή:

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

(α) Υποθέστε ότι n δείγματα $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ επιλέγονται ανεξάρτητα μεταξύ τους σύμφωνα με την κατανομή $p(x|\theta)$. Να δείξετε ότι η εκτίμηση μέγιστης πιθανοφάνειας για τη θ είναι $\max_i x_i$, δηλαδή η τιμή του μέγιστου στοιχείου του συνόλου $\mathcal{D}$.

(β) Υποθέστε ότι επιλέγονται $n = 5$ δείγματα από την κατανομή και ότι η μέγιστη τιμή αυτών ισούται με $\max_i x_i = 0.6$. Να σχεδιάσετε την πιθανοφάνεια $p(\mathcal{D}|\theta)$ στο εύρος τιμών $0 \le \theta \le 1$. Να εξηγήσετε γιατί δεν απαιτείται να γνωρίζουμε τις τιμές των υπόλοιπων τεσσάρων σημείων.

Άσκηση 1.5: (k-Nearest Neighbors)

Θεωρήστε ότι το $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ είναι ένα σύνολο από n ανεξάρτητα και επισημειωμένα δείγματα και ότι το $\mathcal{D}_k(\mathbf{x}) = \{\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_k\}$ περιλαμβάνει τους k κοντινότερους γείτονες του $\mathbf{x}$. Ο κανόνας για την ταξινόμηση του $\mathbf{x}$ σύμφωνα με τους k κοντινότερους γείτονες είναι να ταξινομηθεί το $\mathbf{x}$ στην κατηγορία που "εκπροσωπείται" περισσότερο στο $\mathcal{D}_k(\mathbf{x})$. Θεωρήστε επίσης ότι μελετάμε ένα πρόβλημα δύο κατηγοριών με $P(\omega_1) = P(\omega_2) = \frac{1}{2}$ και ότι οι υπό συνθήκη πυκνότητες πιθανότητας $p(\mathbf{x}|\omega_i)$ είναι ομοιόμορφες εντός μοναδιαίων υπερσφαιρών με απόσταση δέκα μονάδων μεταξύ τους.

(α) Να δείξετε ότι αν το k είναι περιττός αριθμός, τότε η μέση πιθανότητα λάθους δίνεται από τη σχέση:

$$P_n(e) = \frac{1}{2^n} \sum_{j=0}^{\frac{k-1}{2}} \binom{n}{j}$$

(β) Να δείξετε ότι για αυτή την περίπτωση ο κανόνας του ενός ($k = 1$) κοντινότερου γείτονα παρουσιάζει μικρότερο ρυθμό σφαλμάτων σε σχέση με τον κανόνα των k κοντινότερων γειτόνων με $k > 1$.
(γ) Εάν το k επιτρέπεται να αυξάνει καθώς αυξάνεται και το n, αλλά περιορίζεται από τη σχέση $k < a\sqrt{n}$, να δείξετε ότι $P_n(e) \to 0$ καθώς το $n \to \infty$.

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

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

$$\omega_2 : [-4, 2]^\top, [-2, 1]^\top, [-2, -1]^\top, [-1, -3]^\top, [-1, -6]^\top$$

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 με $\rho = 1$ και $\mathbf{w}(0) = [0, 0]^\top$, για να σχεδιάσετε μία ευθεία γραμμή που να διαχωρίζει τις δύο κλάσεις.
Εάν το συγκεκριμένο διάνυσμα βαρών $\mathbf{w}$ δεν επαρκεί (να σχολιαστεί με μία πρόταση το γιατί), τότε να χρησιμοποιηθεί ως αρχικό διάνυσμα βαρών το $\mathbf{w}(0) = [0, 0, 0]^\top$, κάνοντας την κατάλληλη επαύξηση ταυτόχρονα και στα διανύσματα χαρακτηριστικών.
Τέλος, να δοθεί με μορφή εξίσωσης η διαχωριστική καμπύλη που αντιστοιχεί στο υπολογισθέν διάνυσμα βαρών.

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).

Άσκηση 1.7: (EM on GMMs)
Θεωρήστε τρεις Γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας $\mathcal{N}(1.0, 0.1)$, $\mathcal{N}(3.0, 0.1)$ και $\mathcal{N}(2.0, 0.2)$. Δημιουργήστε 500 δείγματα σύμφωνα με τον εξής κανόνα: τα πρώτα δύο δείγματα να προέρχονται από τη δεύτερη Γκαουσιανή, το τρίτο δείγμα από την πρώτη, και το τέταρτο δείγμα από την τελευταία Γκαουσιανή. Ο κανόνας αυτός επαναλαμβάνεται μέχρις ότου δημιουργηθούν και τα 500 δείγματα. Η υποκείμενη συνάρτηση πυκνότητας πιθανότητας των τυχαίων δειγμάτων μπορεί να μοντελοποιηθεί ως ένα μείγμα Γκαουσιανών:

$$\sum_{i=1}^{3} \mathcal{N}(\mu_i, \sigma_i^2) P_i$$

Να χρησιμοποιήσετε τον αλγόριθμο Expectation-Maximization (EM) και τα παραχθέντα δείγματα προκειμένου να εκτιμήσετε τις άγνωστες παραμέτρους μ_i, σ_i^2, P_i. Να δώσετε ένα σύντομο σχολιασμό για τα αριθμητικά αποτελέσματα που προκύπτουν.
Για να λύσετε την άσκηση μπορείτε να αναπτύξετε ρουτίνες σε όποια γλώσσα προγραμματισμού επιθυμείτε, αρκεί να συνοδεύσετε τον κώδικά σας με κάποια σχόλια για τη λειτουργία του. Εναλλακτικά, μπορείτε να χρησιμοποιήσετε ένα μικρότερο σύνολο δειγμάτων, και να επιχειρήσετε να δείξετε τη λειτουργία του αλγόριθμου EM χειροκίνητα, εξηγώντας οποιεσδήποτε παραδοχές χρειαστεί να κάνετε σε αυτή την περίπτωση.

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.