| | | |
|---|---|---|
| Ε.Μ.Π., Σχολή Η.Μ.& Μ.Υ. | **Αναγνώριση Προτύπων** | Ακαδ.Ετος 2021-22 |
| Ημερ/νία: 10-11-2021 | **1ο Σύνολο Αναλυτ. Ασκήσεων** | Παραδοτέο: 25-11-2021 |

---

Οι αναλυτικές σειρές ασκήσεων είναι ατομικές, και οι λύσεις που θα δώσετε πρέπει να αντιπροσωπεύουν μόνο την προσωπική σας εργασία. Εξηγήστε επαρκώς την εργασία σας. Αν χρησιμοποιήσετε κάποια άλλη πηγή εκτός των σημειώσεων για τη λύση σας, πρέπει να το αναφέρετε. Η παράδοση των λύσεων των αναλυτικών ασκήσεων της σειράς αυτής θα γίνει ηλεκτρονικά στο https://helios.ntua.gr και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format, χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: pr21_hwk1_AM_FirstnameLastname.pdf, όπου AM είναι ο 8-ψήφιος αριθμός μητρώου σας. Σκαναρισμένες χειρόγραφες λύσεις επιτρέπονται, αρκεί να είναι καθαρογραμμένες και ευανάγνωστες. Επίσης, στην πρώτη σελίδα των λύσεων θα αναγράφετε το ονοματεπώνυμο, AM, και email address σας.

---

**Υλικό για Ανάγνωση**:
Βιβλία: [1], [2], [3] και [4]
Υλικό διαλέξεων μαθήματος

---

## Αναλυτικές Ασκήσεις

---

**Άσκηση 1.1**: (Maximum Likelihood estimation)
Έστω ότι η μεταβλητή $x$ ακολουθεί μια κατανομή τύπου Erlang:

$$p(x|\theta) = \theta^2 x e^{(-\theta x)} u(x)$$

όπου $u(x)$ είναι η βηματική συνάρτηση

$$u(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x < 0 \end{cases}$$

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 βρείτε την εκτίμηση μέγιστης πιθανοφάνειας για την παράμετρο $\theta$.

---

**Άσκηση 1.2**: (Minimax Criterion)
Αφού μελετήσετε την ενότητα 2.3.1 από το κεφάλαιο 2 του συγγράματος [2] να λύσετε την ακόλουθη άσκηση.
Έστω το minimax κριτήριο για τη zero-one συνάρτηση κόστους, δηλαδή $\lambda_{11} = \lambda_{22} = 0$ και $\lambda_{12} = \lambda_{21} = 1$

1. Να δείξετε πως για τις περιοχές απόφασης ισχύει:

$$\int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1) \mathrm{d}\mathbf{x} = \int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2) \mathrm{d}\mathbf{x}$$

2. Είναι η λύση πάντα μοναδική; Εαν όχι δώστε ένα αντιπαράδειγμα

**Άσκηση 1.3**: (Bayes Error)
Έστω δύο ανεξάρτητες κατανομές $p(\mathbf{x}|\omega_i)$ με priors $p(\omega_i)$. Υποθέτουμε ότι οι κατανομές αυτές ορίζονται για ένα χώρο διάστασης $d = 2$.

1. Έστω οτι οι κατανομές μας είναι κανονικές (normal). Προβάλλετε τις κατανομές στη μία διάσταση και δείξτε πως το σφάλμα στη μία διάσταση είναι ίσο ή μεγαλύτερο από το σφάλμα Bayes στις δύο διαστάσεις.

2. Δείξτε το ίδιο αποτέλεσμα για οποιαδήποτε κατανομή στις δύο διαστάσεις.

---

**Άσκηση 1.4**: (Perceptrons)
Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις $\omega_1$ και $\omega_2$:

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

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

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 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 1.5**: Kullback-Leibler Divergence
Προκειμένου να συγκρίνουμε πολυδιάστατες κατανομές πιθανότητας που ορίζονται στον ίδιο χώρο έχουνε οριστεί αρκετές "μετρικές". Μια από αυτές είναι η Kullback-Leibler που ορίζεται ως εξής:

$$D_{KL}(p_1(\mathbf{x}), p_2(\mathbf{x})) = \int p_1(\mathbf{x})\ln\frac{p_1(\mathbf{x})}{p_2(\mathbf{x})}\mathrm{d}\mathbf{x}$$

Έστω τώρα ότι θέλουμε να προσεγγίσουμε μια οποιαδήποτε κατανομή $p_2(\mathbf{x})$ με μια πολυδιάστατη Γκαουσιανή $p_1(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Να δείξετε ότι οι τιμές που οδηγούν στη μικρότερη τιμής για την KL-divergence είναι οι "προφανείς" διαισθητικά τιμές:

$$\mu = \mathbb{E}_{x\sim p_2}[\mathbf{x}]$$
$$\mathbf{\Sigma} = \mathbb{E}_{x\sim p_2}[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T]$$

---

**Άσκηση 1.6**: (Linear Regression and the LMS Algorithm)
Θεωρήστε το ακόλουθο μοντέλο γραμμικής παλινδρόμησης:

$$y = w_0 + w_1 x$$

καθώς και το παρακάτω σύνολο δεδομένων στη μορφή $(x, y)$:

$$\left\{\begin{pmatrix}0.38\\2.05\end{pmatrix},\begin{pmatrix}0.44\\2.23\end{pmatrix},\begin{pmatrix}0.48\\2.13\end{pmatrix},\begin{pmatrix}0.54\\2.33\end{pmatrix},\begin{pmatrix}0.58\\2.67\end{pmatrix},\begin{pmatrix}0.64\\2.68\end{pmatrix},\begin{pmatrix}0.71\\2.81\end{pmatrix},\begin{pmatrix}0.76\\2.97\end{pmatrix},\begin{pmatrix}0.82\\3.12\end{pmatrix},\begin{pmatrix}0.96\\3.20\end{pmatrix}\right\}$$

για το οποίο γνωρίζουμε ότι έχει προκύψει έπειτα από την επίδραση γκαουσιανού θορύβου.

1. Να αποδείξετε ότι η λύση μέγιστης πιθανοφάνειας είναι ισοδύναμη με τη λύση που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων.
2. Να υπολογίσετε τους άγνωστους συντελεστές χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων.
3. Χρησιμοποιώντας ένα από τα παραπάνω δεδομένα τη φορά, και εφαρμόζοντας τον αλγόριθμο Least Mean Squares (LMS) (Υποκεφάλαιο 5.8 του βιβλίου [2]), να υπολογίσετε εκ νέου τους άγνωστους συντελεστές. Να επαναληφθεί η διαδικασία για τουλάχιστον μία εποχή χειροκίνητα.
4. Γράψτε πρόγραμμα που να εφαρμόζει τον LMS στο συγκεριμένο πρόβλημα, ώστε να βρείτε μια καλύτερη προσέγγιση.
5. Να σχεδιάσετε σε ένα κοινό γράφημα τις ευθείες που βρήκατε σαν λύσεις.

---

**Άσκηση 1.7**: (Bayes meets k-NN)
Έστω δύο ταξιμονητές οι οποίοι βασίζονται σε δείγματα που προέρχονται από δύο κατανομές

$$p(x|\omega_1) = \begin{cases} 2x & \text{for} \quad 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 2-2x & \text{for} \quad 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Επιπλέον οι a-priori πιθανότητες για κάθε κατηγορία είναι $p[\omega_1] = p[\omega_2] = 1/2$.

1. Ποιος είναι ο κανόνας απόφασης (decision rule) κατά Bayes και ποιο το σφάλμα ταξινόμησης Bayes;

2. Έστω ότι δειγματοληπτούμε ένα τυχαίο σημείο από την κατηγορία $\omega_1$ και ένα από την $\omega_2$. Με βάση τα σημεία αυτά κατασκευάζουμε έναν ταξινομητή κοντινότερου γείτονα (nearest-neighbor classifier). Υποθέτουμε τώρα πως δειγματοληπτούμε ένα επιπλέον test σημείο για μια από τις δύο κατηγορίες, π.χ $\omega_1$. Βρείτε το αναμενόμενο σφάλμα (expected error) $P_1(e)$ (ο δείκτης υποδηλώνει το πλήθος σημείων) μέσω ολοκλήρωσης.

3. Επαναλάβετε τη διαδικασία για δύο σημεία εκπαίδευσης (training) από κάθε κατηγορία και ένα μόνο σημείο αποτίμησης (test point) και βρείτε το $P_2(e)$.

4. Γενικεύστε για $n$ το πλήθος σημεία, ώστε να βρείτε το $P_n(e)$.

5. Μελετήστε τι συμβαίνει στο όριο $\lim_{n\to\infty} P_n(e)$ σε σχέση με το Bayes error.

---

**Άσκηση 1.8**: (EM)

Θεωρήστε τα δεδομένα

$$D = \{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}4\\5\end{pmatrix}, \begin{pmatrix}1\\\star\end{pmatrix}\}$$

όπου το $\star$ υποδηλώνει μια άγνωστη τιμή χαρακτηριστικού. Υποθέτουμε πως τα σημεία αυτά έχουνε προέρθει από μια διδιάστατη διαχωρίσιμη κατανομή $p(x_1, x_2) = p(x_1)p(x_2)$, με

$$p(x_1) \sim \begin{cases} \frac{1}{\theta_1} e^{(-\theta_1 x_1)} & \text{εαν} \quad x_1 \geq 0 \\ 0 & \text{αλλιώς} \end{cases}$$

$$p(x_2) \sim U(0, \theta_2) = \begin{cases} 1/\theta_2 & \text{εαν} \quad 0 \leq x_2 \leq \theta_2 \\ 0 & \text{αλλιώς} \end{cases}$$

1. Ξεκινώντας από μια αρχική εκιμηση $\theta^0 = \binom{2}{3}$ υπολογίστε αναλυτικά το $Q(\theta, \theta_0)$ που αποτελεί το βήμα Ε του αλγορίθμου ΕΜ.

2. Βρείτε το $\theta$ που μεγιστοποιεί το $Q(\theta, \theta^0)$ που αποτελεί το βήμα Μ του αλγορίθμου ΕΜ.

3. Σχεδιάστε ένα γράφημα που να δείχνει την κατανομή $p(x_1, x_2)$ πριν και μετά την εκτίμηση των παραμέτρων.

---

**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.