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 courses.pclab.ece.ntua.gr και θα πρέπει να την υποβάλετε ως ένα ενιαίο αρχείο PDF με το εξής filename format, χρησιμοποιώντας μόνο λατινικούς χαρακτήρες: pr20_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]
Διαφάνειες διαλέξεων μαθήματος

Αναλυτικές Ασκήσεις

Άσκηση 1.1: (Entropy)
Για το ακόλουθο πρόβλημα χρησιμοποιήστε το τελικό αποτέλεσμα του Προβλήματος 19 από το Κεφάλαιο 2 του βιβλίου [2], το οποίο επισυνάπτεται στο Παράρτημα Α.
(a) Υποθέστε πως γνωρίζουμε ότι μία κατανομή είναι μη μηδενική στο εύρος $x_l \le x \le x_u$. Να αποδείξετε ότι η κατανομή μέγιστης εντροπίας είναι ομοιόμορφη σε αυτό το εύρος, δηλαδή:

$$p_u(x) \sim U(x_l, x_u) = \begin{cases} 1/|x_u - x_l| & x_l \le x \le x_u \\ 0 & \text{αλλού} \end{cases}$$

(b) Υποθέστε πως γνωρίζουμε ότι μία κατανομή είναι μη μηδενική για $x \ge 0$ και ότι έχει μέση τιμή μ. Να αποδείξετε ότι η κατανομή μέγιστης εντροπίας είναι:

$$p_e(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & \text{για } x \ge 0 \\ 0 & \text{αλλού} \end{cases}$$

(c) Υποθέστε πως γνωρίζουμε ότι μία κατανομή έχει μέση τιμή μ και τυπική απόκλιση σ. Να αποδείξετε ότι η κατανομή μέγιστης εντροπίας είναι μία γκαουσιανή, δηλαδή:

$$p_g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$$

Άσκηση 1.2: (Bayes Classifiers)
Για τις παρακάτω περιπτώσεις κατανομών, να τις σχεδιάσετε στο ίδιο γράφημα (ανά δύο), να βρείτε τις περιοχές απόφασης ενός Bayesian ταξινομητή, και να υπολογίσετε την πιθανότητα λάθους $P(\text{error})$:

(α) Έστω δύο κατηγορίες που περιγράφονται από τις κανονικές κατανομές $\mathcal{N}(2,1)$ και $\mathcal{N}(3,2)$ με a-priori πιθανότητες $P(\omega_1) = 0.4$ και $P(\omega_2) = 0.6$.
(β) Έστω δύο κατηγορίες που περιγράφονται από τα χαρακτηριστικά x_1 και x_2 τα οποία κατανέμονται σύμφωνα με τις ομοιόμορφες κατανομές:

$$P(x_1, x_2|\omega_1) = \begin{cases} 0.1 & 0 \le x_1 \le 2, \quad 0 \le x_2 \le 5 \\ 0 & \text{αλλού} \end{cases}$$

$$P(x_1, x_2|\omega_2) = \begin{cases} \frac{1}{8} & 1.5 \le x_1 \le 3.5, \quad 0 \le x_2 \le 4x_1 - 6 \\ 0 & \text{αλλού} \end{cases}$$

με a-priori πιθανότητες $P(\omega_1) = \frac{1}{4}$ και $P(\omega_2) = \frac{3}{4}$.

Άσκηση 1.3: (Maximum Likelihood Estimation)
Έστω ότι η τυχαία διακριτή μεταβλητή X ακολουθεί την κατανομή Poisson:

$$X \sim \frac{\theta^x}{x!} e^{-\theta}, \quad x \in \mathbb{Z}_+ \cup \{0\}$$

(α) Να σχεδιάσετε την σ.π.π. $P(x|\theta)$ συναρτήσει του x για $\theta = 2$ και για $\theta = 10$.
(β) Εάν υποθέσετε ότι γνωρίζετε N ανεξάρτητα δείγματα $\{x_1, x_2, \ldots, x_N\}$ τα οποία προέρχονται από μία κατανομή Poisson, να βρείτε τον εκτιμητή μέγιστης πιθανοφάνειας για το θ.

Άσκηση 1.4: (EM on GMMs)
Θεωρήστε δύο Γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας $\mathcal{N}(1.5, 0.1)$ και $\mathcal{N}(2.5, 0.2)$. Δημιουργήστε 15 δείγματα (με μία γεννήτρια τυχαίων αριθμών) σύμφωνα με τον εξής κανόνα: τα πρώτα δύο δείγματα να προέρχονται από την πρώτη Γκαουσιανή και το τρίτο δείγμα από τη δεύτερη. Ο κανόνας αυτός επαναλαμβάνεται μέχρις ότου δημιουργηθούν και τα 15 δείγματα. Η υποκείμενη συνάρτηση πυκνότητας πιθανότητας των τυχαίων δειγμάτων μπορεί να μοντελοποιηθεί ως ένα μείγμα Γκαουσιανών:

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

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 εξηγήσετε οποιεσδήποτε παραδοχές χρειαστεί να κάνετε. Τέλος, να δώσετε έναν σύντομο σχολιασμό για τα αριθμητικά αποτελέσματα που προκύπτουν.
Σημείωση: Για να λύσετε την άσκηση να πραγματοποιήσετε όλες τις πράξεις που απαιτούνται *χειροκίνητα*. Δεν ζητείται προγραμματιστική υλοποίηση.

Άσκηση 1.5: (Gradient Descent for MSE Perceptron Training)
Βασιζόμενοι στην Ενότητα 5.8.4 του βιβλίου [2], να αποδείξετε ότι το $\mathbf{a}$ που ελαχιστοποιεί το κόστος εκπαίδευσης ενός Perceptron με τη μέθοδο ελαχίστων τετραγώνων

$$J_s(\mathbf{a}) = ||\mathbf{Ya} - \mathbf{b}||^2$$

προκύπτει από την παρακάτω επαναληπτική μέθοδο gradient descent:

$$\mathbf{a}(1) \qquad \textit{arbitrary}$$

$$\mathbf{a}_{k+1} = \mathbf{a}_k - \eta_k \mathbf{Y}^\top(\mathbf{Y}\mathbf{a}_k - \mathbf{b})$$

Συγκεκριμένα, εάν χρησιμοποιήσουμε βήματα $\eta_k = 1/k$, να αποδείξετε ότι η παραπάνω επαναληπτική μέθοδος στο όριο ($\lim_{k\to\infty} \mathbf{a}_k \to \mathbf{a}$) δίνει διάνυσμα $\mathbf{a}$ το οποίο ικανοποιεί την παρακάτω εξίσωση:

$$\mathbf{Y}^\top(\mathbf{Y}\mathbf{a} - \mathbf{b}) = 0$$

Οπότε, ο ανωτέρω αλγόριθμος δίνει λύση ανεξάρτητα από το εάν ο πίνακας $\mathbf{Y}^\top\mathbf{Y}$ έχει μηδενική ορίζουσα.

Σημείωση: Γνωρίζουμε ότι αν ο πίνακας $\mathbf{Y}^\top\mathbf{Y}$ δεν έχει μηδενική ορίζουσα, τότε το διάνυσμα που ελαχιστοποιεί το κριτήριο J_s προκύπτει από τον ψευδο-αντίστροφο: $\mathbf{a} = \mathbf{Y}^\dagger\mathbf{b}$, $\mathbf{Y}^\dagger = (\mathbf{Y}^\top\mathbf{Y})^{-1}\mathbf{Y}^\top$.

Άσκηση 1.6: (Linear Classifier with Quadric Surfaces)

Θεωρήστε τη διαχωριστική επιφάνεια $g(\mathbf{x}) = \mathbf{x}^\top\mathbf{A}\mathbf{x} + \mathbf{x}^\top\mathbf{b} + c$ στην περίπτωση ενός διδιάστατου επιπέδου, όπου σας δίνεται ότι $\mathbf{A} = \begin{pmatrix} 6 & -3 \\ -3 & 2 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, και $c = -2$.

(α) Να υπολογίσετε την αναλυτική μορφή της εξίσωσης που προκύπτει. Επίσης, να σχεδιάσετε τη διαχωριστική αυτή επιφάνεια στο επίπεδο. Τι μορφή έχει, ποια εκ των παραμέτρων $\mathbf{A}$, $\mathbf{b}$ και c καθορίζει τη μορφή της και γιατί;

(β) Να υπολογίσετε τις ιδιοτιμές, λ_1 και λ_2, και τα ιδιοδιανύσματα, $\mathbf{v}_1$ και $\mathbf{v}_2$, του πίνακα $\mathbf{A}$, αναλυτικά με εξισώσεις. Τι παρατηρείτε σχετικά με τα ιδιοδιανύσματα $\mathbf{v}_1$ και $\mathbf{v}_2$ και την καμπύλη που σχεδιάσατε;

(γ) Σε ποια κλάση ταξινομούνται τα σημεία $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ και $\mathbf{x}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$;

Άσκηση 1.7: (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.

(α) Να αποδείξετε ότι η λύση μέγιστης πιθανοφάνειας είναι ισοδύναμη με τη λύση που προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων.

(β) Να υπολογίσετε τους άγνωστους συντελεστές χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων.

(γ) 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 3 εποχές, ή μέχρι να θεωρήσετε ότι το αποτέλεσμα είναι ακριβές (το σφάλμα ανακατασκευής των δεδομένων να πέφτει κάτω από ένα κατώφλι που θα ορίσετε).

(δ) Να σχεδιάσετε σε ένα κοινό γράφημα τα δοθέντα σημεία και τις δύο παραπάνω ευθείες.

Σημείωση: Για να λύσετε την άσκηση να πραγματοποιήσετε όλες τις πράξεις που απαιτούνται *χειροκίνητα*. Δεν ζητείται προγραμματιστική υλοποίηση.

Άσκηση 1.8: (Adaptive Perceptron Learning - Adaline)
Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : \left\{ \begin{pmatrix} -1 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \end{pmatrix} \right\}$$

$$\omega_2 : \left\{ \begin{pmatrix} -4 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -3 \end{pmatrix}, \begin{pmatrix} -1 \\ -6 \end{pmatrix} \right\}$$

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 adaptive perceptron με $\rho = 1/k$ (όπου k ο αριθμός της εποχής) και αρχικό διανύσμα βαρών $\mathbf{w}(0) = [0,0,0]^\top$, για να σχεδιάσετε μία ευθεία γραμμή που να διαχωρίζει τις δύο κλάσεις. Για να γίνει αυτό, θα πρέπει να επαυξήσετε κατάλληλα και τα διανύσματα χαρακτηριστικών. Τέλος, να δοθεί με μορφή εξίσωσης η διαχωριστική καμπύλη που αντιστοιχεί στο υπολογισθέν διάνυσμα βαρών και να σχεδιαστεί στο ίδιο γράφημα με τα σημεία.
Σημείωση: Για να λύσετε την άσκηση να πραγματοποιήσετε όλες τις πράξεις που απαιτούνται *χειροκίνητα*. Δεν ζητείται προγραμματιστική υλοποίηση.

Αλγόριθμος Adaptive 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:

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

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

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

Ο ανωτέρω αλγόριθμος έχει τη 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). Αυτή η ποσότητα σταθμίζεται επιπλέον και με τον όρο $(\pm 1 - \mathbf{w}(t)^\top \mathbf{x}_t)$ ώστε να δοθεί μεγαλύτερη βαρύτητα σε σημεία που αποκλίνουν περισσότερο από τη σωστή απόφαση.

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.

Παράρτημα Α:

19. Starting from the definition of entropy (Eq. 36), derive the general equation for the maximum-entropy distribution given constraints expressed in the general form

$$\int b_k(x)p(x)\,dx = a_k, \quad k = 1, 2, ..., q$$

as follows:

(a) Use Lagrange undetermined multipliers $\lambda_1, \lambda_2, ..., \lambda_q$ and derive the synthetic function:

$$H_s = -\int p(x)\left[\ln p(x) - \sum_{k=0}^{q} \lambda_k b_k(x)\right]\,dx - \sum_{k=0}^{q} \lambda_k a_k.$$

State why we know $a_0 = 1$ and $b_0(x) = 1$ for all x.

(b) Take the derivative of H_s with respect to $p(x)$. Equate the integrand to zero, and thereby prove that the minimum-entropy distribution obeys

$$p(x) = \exp\left[\sum_{k=0}^{q} \lambda_k b_k(x) - 1\right],$$

where the $q + 1$ parameters are determined by the constraint equation.