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: Σχετικά κεφάλαια των βιβλίων [2], [3], [4], καθώς και οι διαφάνειες των διαλέξεων του μαθήματος.

Αναλυτικές Ασκήσεις

Άσκηση 2.1: (Expectation-Maximization)
Θεωρήστε τα δεδομένα $\mathcal{D} = \{\binom{1}{2}, \binom{4}{5}, \binom{2}{*}\}$ (το * δηλώνει μια άγνωστη τιμή χαρακτηριστικού), τα οποία δειγματοληπτούνται από μία διδιάστατη διαχωρίσιμη κατανομή $p(x_1, x_2) = p(x_1)p(x_2)$, με

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

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

α) Ξεκινώντας από μία αρχική εκτίμηση $\boldsymbol{\theta}^0 = \binom{2}{3}$ υπολογίστε αναλυτικά το $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0)$, που αποτελεί το βήμα **E** του αλγόριθμου EM.
β) Βρείτε το $\boldsymbol{\theta}$ που μεγιστοποιεί το $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^0)$, που αποτελεί το βήμα **M** του αλγόριθμου EM.
γ) Σχεδιάστε ένα γράφημα που θα δείχνει την κατανομή $p(x_1, x_2)$ πριν και μετά την εκτίμηση των παραμέτρων.

Άσκηση 2.2: (Basis vectors - Principal Component Analysis)
Έστω ότι τα διανύσματα $\mathbf{e}_i, i = 0, 1, \ldots, N-1$ αποτελούν μία ορθοκανονική βάση στον N-διάστατο χώρο. Εάν θεωρήσουμε ένα N-διάστατο τυχαίο διάνυσμα $\mathbf{x}$ με μέση τιμή $E(\mathbf{x}) = 0$ και την m-διάστατη (με $1 < m < N$) προβολή του, $\hat{\mathbf{x}}$:
α) Να δείξετε ότι το μέσο τετραγωνικό σφάλμα $E\left[||\mathbf{x} - \hat{\mathbf{x}}||^2\right]$ ελαχιστοποιείται όταν η ορθοκανονική βάση αποτελείται από τα ιδιοδιανύσματα του πίνακα συσχέτισης R_x.
β) Να δείξετε ότι για για την παραπάνω περίπτωση (α) επίσης ισχύει ότι ο m-διάστατος υποχώρος είναι εκείνος που σχηματίζεται από τα ιδιοδιανύσματα που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές.
γ) Να δείξετε ότι κατά την προβολή του διανύσματος σε αυτό τον υποχώρο μεγιστοποιείται το άθροισμα των διασπορών των στοιχείων του $\hat{\mathbf{x}}$.
Σημείωση: Προκειμένου να προκύψει μία μοναδική λύση, η ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος να γίνει υπό τη συνθήκη $\mathbf{e}_i^\top \mathbf{e}_i = 1$.
Να δικαιολογήσετε πλήρως όλα τα βήματα που πραγματοποιείτε.

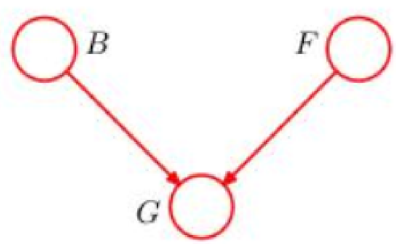
Άσκηση 2.3: (Support Vector Machines - SVM)
Μας δίνονται $n = 4$ διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : [1, 5]^\top, [-2, -4]^\top \qquad z_1 = z_2 = -1$$

$$\omega_2 : [1, 3]^\top, [-1, 5]^\top \qquad z_3 = z_4 = 1$$

όπου οι συντελεστές $z_k = \pm 1$ υποδεικνύουν την κατηγορία καθενός από τα δείγματα. Στην περίπτωση ενός ταξινομητή SVM, στόχος είναι η εύρεση του διανύσματος βάρους $\mathbf{a}$ με το ελάχιστο μήκος, το οποίο να υπόκειται στους περιορισμούς $z_k \mathbf{a}^\top \mathbf{y}_k \geq 1$ $(k = 1, \dots, n)$. Τα διανύσματα $\mathbf{a}$ και $\mathbf{y}_k$ είναι επαυξημένα κατά a_0 ($\mathbf{a} = [a_0 \quad \mathbf{a}_r]^\top$) και $y_{k,0} = 1$ ($\mathbf{y} = [1 \quad \mathbf{y}_{k,r}]^\top$), αντίστοιχα.
α) Αρχικά, ελέγξτε εάν οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες, μέσω του σχεδιασμού των παραπάνω σημείων σε ένα γράφημα. Στη συνέχεια, μετατρέψτε τα διανύσματα αυτά σε ένα χώρο υψηλότερων διαστάσεων, χρησιμοποιώντας την απλούστερη μορφή ϕ-functions 2ης τάξης: $1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2$, όπου το $\sqrt{2}$ έχει χρησιμοποιηθεί για κανονικοποίηση.
β) Να προσδιοριστούν οι συντελεστές α_k $(k = 1, \dots, n)$ του προβλήματος ελαχιστοποίησης (Υποκεφάλαιο 5.11.1 [2]). Η λύση που βρήκατε είναι δεκτή και αν ναι, γιατί;
γ) Να υπολογιστεί το ζητούμενο διάνυσμα βαρών $\mathbf{a}$. Επαληθεύστε ότι ισχύουν όλες οι προϋποθέσεις $z_k \mathbf{a}^\top \mathbf{y}_k \geq 1$ $(k = 1, \dots, n)$.
δ) Υπολογίστε το περιθώριο β του ταξινομητή.
ε) Βρείτε τη συνάρτηση διαχωρισμού $g(x_1, x_2)$ στον αρχικό χώρο x_1-x_2 και σχεδιάστε την καμπύλη $g(x_1, x_2) = 0$ μαζί με τα 4 αρχικά σημεία, είτε προσεγγιστικά στο χέρι, είτε σε κάποιο εργαλείο σχεδιασμού σε Η/Υ.
στ) Ποια είναι τα support vectors;

Άσκηση 2.4: (Conditional independence)



Σχήμα 1: Μπαταρία (battery B), ντεπόζιτο (fuel tank F) και δείκτης (gauge G) ενός αυτοκινήτου.

Θεωρήστε το σύστημα καυσίμων ενός αυτοκινήτου, όπως φαίνεται στο Σχήμα 1, με γνωστές τις παρακάτω πιθανότητες:

$$p(B = 1) = 0.95$$

$$p(F = 1) = 0.8$$

$$p(G = 1|B = 1, F = 1) = 0.95$$

$$p(G = 1|B = 1, F = 0) = 0.3$$

$$p(G = 1|B = 0, F = 1) = 0.25$$

$$p(G = 1|B = 0, F = 0) = 0.2$$

Υποθέστε ότι αντί να παρατηρείτε την κατάσταση του δείκτη καυσίμων G απ'ευθείας, ο δείκτης διαβάζεται από τον οδηγό D και μας μεταφέρει την τιμή. Οι δύο πιθανές καταστάσεις είναι είτε ότι διαβάστηκε ο δείκτης ως γεμάτος ($D = 1$) είτε ως άδειος ($D = 0$). Ο οδηγός είναι σχετικά αναξιόπιστος, όπως εκφράζεται από τις ακόλουθες πιθανότητες:

$$p(D = 1|G = 1) = 0.8$$

$$p(D = 0|G = 0) = 0.8$$

Σύμφωνα με τον οδηγό, ο μετρητής καυσίμων φαίνεται άδειος, δηλαδή παρατηρούμε ότι $D = 0$.
α) Υπολογίστε την πιθανότητα το ντεπόζιτο F να είναι άδειο, με δεδομένη μόνο αυτή την παρατήρηση.
β) Υπολογίστε την πιθανότητα το ντεπόζιτο F να είναι άδειο, εάν επιπλέον η μπαταρία B είναι άδεια. Η δεύτερη πιθανότητα είναι μεγαλύτερη ή μικρότερη από την πρώτη; Σχολιάστε το αποτέλεσμα και εξηγήστε γιατί.

Άσκηση 2.5: (Hidden Markov Models)
Έστω ένα κρυφό μοντέλο Markov με τρεις καταστάσεις 1, 2, 3 και δύο τύπους παρατηρήσεων H και T. Δίνεται ο παρακάτω πίνακας μετάβασης, $a_{ij} = p(q_t = j|q_{t-1} = i)$:

$$\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.2 & 0.1 & 0.7 \end{pmatrix}$$

Δίνονται επίσης οι ακόλουθες πιθανότητες των παρατηρήσεων

$P(O\|q)$			
O q	1	2	3
H	0.5	0.7	0.25
T	0.5	0.3	0.75

Οι πιθανότητες a-priori είναι ίσες με $\pi_1 = p(q_0 = 1) = \pi_2 = p(q_0 = 2) = \pi_3 = p(q_0 = 3) = \frac{1}{3}$.
α) Υπολογίστε την πιθανότητα $P(O_0$=H, O_1=H, O_2=T$)$.
β) Υπολογίστε την πιο πιθανή σειρά καταστάσεων δεδομένης της σειράς παρατηρήσεων (O_0=H, O_1=H, O_2=T), χρησιμοποιώντας τον αλγόριθμο Viterbi.
γ) Θέλουμε να χρησιμοποιήσουμε την ακόλουθη σειρά παρατηρήσεων HHTTTHHHTTTTH-HH για να εκπαιδεύσουμε το κρυφό μοντέλο Markov. Αντί για τον πολύπλοκο αλγόριθμο forward-backward, θα ακολουθήσουμε τον εξής ψευδο-EM αλγόριθμο που αναφέρεται συχνά ως εκπαίδευση Viterbi (Viterbi-training).
Expectation step: Αποκωδικοποίηση χρησιμοποιώντας τον αλγόριθμο Viterbi.
Maximization step: Μεγιστοποίηση συνολικής πιθανότητας καταστάσεων και παρατηρήσεων ML.
Δηλαδή, ο αλγόριθμος εκπαίδευσης Viterbi βρίσκει την πιο πιθανή σειρά καταστάσεων δεδομένης της σειράς παρατηρήσεων, και στη συνέχεια μεγιστοποιεί τη συνολική πιθανότητα της σειράς καταστάσεων $q_0q_1q_2$ που μόλις υπολογίσαμε και παρατηρήσεων $O_0O_1O_2$ που δίνονται.
Το δεύτερο βήμα του αλγορίθμου είναι η συνήθης εκπαίδευση με μεγιστοποίηση πιθανότητας Maximum-Likelihood training που εφαρμόζεται σε κάθε Μπεϋσιανό δίκτυο με όλες τις παραμέτρους (q, O) παρατηρήσιμες.
δ) Ποιες είναι οι κύριες διαφορές μεταξύ του forward-backward και Viterbi-training και ποιος

αλγόριθμος αναμένεται να έχει καλύτερα αποτελέσματα;

Άσκηση 2.6: (Cross-entropy)
Για την ακόλουθη cross-entropy συνάρτηση κόστους

$$J = -\sum_{i=1}^{N} \sum_{k=1}^{k_L} y_k(i) \ln \frac{\hat{y}_k(i)}{y_k(i)}$$

όπου k_L είναι ο αριθμός των κόμβων εξόδου, N ο αριθμός των δειγμάτων, $y_k(i)$ οι επιθυμητές πιθανότητες στην έξοδο του multilayer perceptron, και $\hat{y}_k(i)$ οι πραγματικές πιθανότητες στην έξοδο:

α) Δείξτε ότι η ελάχιστη τιμή της για δυαδικές επιθυμητές τιμές εξόδου (δηλ. τιμές 0,1) είναι μηδενική και προκύπτει όταν οι πραγματικές έξοδοι είναι ίσες με τις επιθυμητές.

β) Δείξτε ότι η cross-entropy συνάρτηση κόστους εξαρτάται από τα σχετικά σφάλματα εξόδου.

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.