Ε.Μ.Π., Σχολή Η.Μ.& Μ.Υ.	**Αναγνώριση Προτύπων**	Ακαδ. Έτος 2020-21
Ημερ/νία: 23-12-2020	**2ο Σύνολο Αναλυτ. Ασκήσεων**	Παραδοτέο: 11-01-2021

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 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_hwk2_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: [5]

Αναλυτικές Ασκήσεις

Άσκηση 2.1: (SVM)

Μας δίνονται $N = 8$ διανύσματα χαρακτηριστικών $\tilde{\mathbf{x}}_n = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : \left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ -3 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix} \right\}, \qquad z_1 = z_2 = z_3 = z_4 = -1$$

$$\omega_2 : \left\{ \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -0.5 \\ 0.5 \end{pmatrix} \right\}, \qquad z_5 = z_6 = z_7 = z_8 = 1$$

όπου οι συντελεστές $z_n = \pm 1$ υποδεικνύουν την κατηγορία καθενός από τα δείγματα. Στην περίπτωση ενός ταξινομητή SVM, στόχος είναι η εύρεση του διανύσματος βάρους $\mathbf{w}$ με το ελάχιστο μήκος, το οποίο να υπόκειται στους περιορισμούς $z_n \mathbf{w}^\top \mathbf{y}_n \geq 1$ $(n = 1, \ldots, N)$. Τα διανύσματα $\mathbf{w}$ και $\mathbf{y}_n$ είναι επαυξημένα κατά w_0 ($\mathbf{w} = [w_0 \quad \tilde{\mathbf{w}}]^\top$) και $y_{n,0} = 1$ ($\mathbf{y}_n = [1 \quad \tilde{\mathbf{y}}_n]^\top$), αντίστοιχα.

(α) Αρχικά, ελέγξτε εάν οι δύο κλάσεις είναι γραμμικώς διαχωρίσιμες, μέσω του σχεδιασμού των παραπάνω σημείων σε ένα γράφημα. Στη συνέχεια, στην περίπτωση όπου δεν είναι γραμμικώς διαχωρίσιμες, μετατρέψτε τα διανύσματα $\tilde{\mathbf{x}}_n$ σε έναν χώρο υψηλότερων διαστάσεων, $\mathbf{y}_n = \phi(\tilde{\mathbf{x}}_n)$, χρησιμοποιώντας την εξής μορφή ϕ-functions 2ης τάξης: $\phi(x_1, x_2) = [1 \quad x_1 \quad x_2 \quad \frac{x_1^2 + x_2^2 - 5}{3}]^\top$.

(β) Να προσδιοριστούν οι συντελεστές α_n $(n = 1, \ldots, N)$ του προβλήματος ελαχιστοποίησης (Υποκεφάλαιο 5.11.1 [2], [5, SVM]). Η λύση που βρήκατε είναι δεκτή και αν ναι, γιατί;

(γ) Να υπολογιστεί το ζητούμενο διάνυσμα βαρών $\mathbf{w}$. Επαληθεύστε ότι ισχύουν όλες οι προϋποθέσεις $z_n \mathbf{w}^\top \mathbf{y}_n \geq 1$ $(n = 1, \ldots, N)$.

(δ) Υπολογίστε το περιθώριο β του ταξινομητή.

(ε) Βρείτε τη συνάρτηση διαχωρισμού $g(x_1, x_2) = \mathbf{w}^\top \phi(x_1, x_2)$ στον αρχικό χώρο x_1-x_2 και

σχεδιάστε την καμπύλη $g(x_1, x_2) = 0$ μαζί με τα 8 αρχικά σημεία σε κάποιο εργαλείο σχεδιασμού σε Η/Υ.
(στ) Ποια είναι τα support vectors;
(ζ) Σε ποιες κατηγορίες ταξινομούνται τα σημεία $\begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix}$ και $\begin{pmatrix} -2 \\ 3 \end{pmatrix}$;

Σημείωση: Επεξηγήστε αναλυτικά τη διαδικασία που ακολουθήσατε για να φτάσετε στις λύσεις σας (π.χ. τις παραδοχές, τα σύμβολα και τους τύπους που τυχόν χρησιμοποιήσατε, τη σωστή αντικατάσταση μεταβλητών και δεικτών σε αυτούς, αριθμητικά αποτελέσματα ενδιάμεσων ή βοηθητικών μεταβλητών, κ.ά.). Στην περίπτωση επαναληπτικών αλγορίθμων, τα παραπάνω ισχύουν σε κάθε επανάληψη της εκτέλεσής τους.

Άσκηση 2.2: (HMM)
Ζητείται να χρησιμοποιηθεί ένα HMM για να αποκωδικοποιηθεί μία απλή ακολουθία DNA μήκους T. Είναι γνωστό ότι μια ακολουθία DNA είναι μία σειρά από στοιχεία του συνόλου $\{A, C, G, T\}$. Ας υποθέσουμε ότι υπάρχει μία ακολουθία κρυμμένων καταστάσεων $\mathbf{q} = (q_1, q_2, \ldots, q_T)$, η οποία σε κάθε χρονική στιγμή t μπορεί να λαμβάνει μία εκ των δύο πιθανών καταστάσεων $\{S_1, S_2\}$, και η οποία ελέγχει τη δημιουργία της ακολουθίας DNA $\mathbf{O} = (O_1, O_2, \ldots, O_T)$. Επίσης, δίνονται οι ακόλουθες πιθανότητες μετάβασης για το HMM λ:

$$P(S_1|S_1) = 0.6 \quad P(S_2|S_1) = 0.4 \quad P(S_1|S_2) = 0.3 \quad P(S_2|S_2) = 0.7$$

οι ακόλουθες πιθανότητες των παρατηρήσεων:

$$P(A|S_1) = 0.5 \quad P(C|S_1) = 0.2 \quad P(G|S_1) = 0.2 \quad P(T|S_1) = 0.1$$

$$P(A|S_2) = 0.3 \quad P(C|S_2) = 0.1 \quad P(G|S_2) = 0.4 \quad P(T|S_2) = 0.2$$

και οι ακόλουθες a-priori πιθανότητες για τη χρονική στιγμή $t = 1$:

$$P(S_1) = 0.5 \quad P(S_2) = 0.5$$

Έστω ότι η παρατηρούμενη ακολουθία είναι $\mathbf{O} = (C, C, T, G, T, C, A, C)$. Υπολογίστε:
(α) Την πιθανότητα $P(\mathbf{O}|\lambda)$ χρησιμοποιώντας τον forward αλγόριθμο.
(β) Την πιθανότητα $P(\mathbf{O}|\lambda)$ χρησιμοποιώντας τον backward αλγόριθμο.
(γ) Τις εκ των υστέρων πιθανότητες $P(q_t = S_1|\mathbf{O}, \lambda)$ για $t = 1, \ldots, T$. Ποια είναι η πιο πιθανή κατάσταση σε κάθε χρονική στιγμή;
(δ) Το πιο πιθανό μονοπάτι κρυμμένων καταστάσεων $\mathbf{q}^* = (q_1^*, q_2^*, \ldots, q_T^*)$ χρησιμοποιώντας τον αλγόριθμο Viterbi. Οι καταστάσεις που βρήκατε στα ερωτήματα (γ) και (δ) είναι οι ίδιες; Ποια είναι η ποιοτική διαφορά τους;

Σημείωση: Επεξηγήστε αναλυτικά τη διαδικασία που ακολουθήσατε για να φτάσετε στις λύσεις σας (π.χ. τις παραδοχές, τα σύμβολα και τους τύπους που τυχόν χρησιμοποιήσατε, τη σωστή αντικατάσταση μεταβλητών και δεικτών σε αυτούς, αριθμητικά αποτελέσματα ενδιάμεσων ή βοηθητικών μεταβλητών, κ.ά.). Στην περίπτωση επαναληπτικών αλγορίθμων, τα παραπάνω ισχύουν σε κάθε επανάληψη της εκτέλεσής τους.

Άσκηση 2.3: (CART)
Θεωρήστε τα παρακάτω $N = 16$ δείγματα, τα οποία προέρχονται από δύο κατηγορίες ω_1 και ω_2. Τα δείγματα αυτά αποτελούνται από μία δυαδική τιμή που αφορά το χρώμα του δείγματος (Άσπρο/Κόκκινο), καθώς και δύο δεκαδικές τιμές που σχετίζονται με δύο χαρακτηριστικά του:

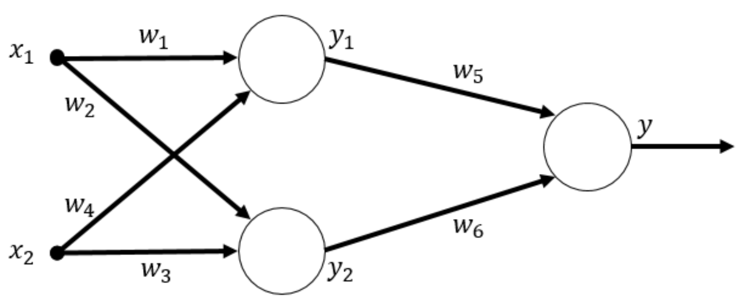
Χρώμα	Χαρ. 1	Χαρ. 2
A	1.2	5.7
A	3.2	5.1
A	1.5	3.4
A	2.1	4.0
K	3.5	3.8
K	1.9	5.7
K	5.2	2.6
K	4.3	4.5

Χρώμα	Χαρ. 1	Χαρ. 2
A	1.7	5.5
A	0.5	4.4
A	3.9	3.1
A	4.0	3.9
K	5.6	2.4
K	4.8	5.3
K	5.2	3.2
K	4.3	0.5

Να κατασκευάσετε ένα δυαδικό δέντρο CART ($B = 2$), το οποίο θα στοχεύει στον διαχωρισμό των δύο κλάσεων, χρησιμοποιώντας την entropy impurity.

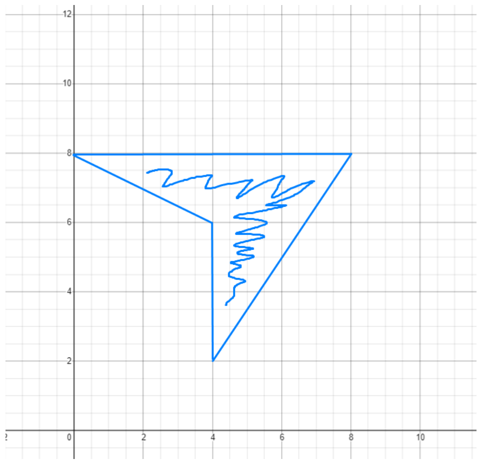
Σημείωση: Επεξηγήστε αναλυτικά τη διαδικασία που ακολουθήσατε για να φτάσετε στις λύσεις σας (π.χ. τις παραδοχές, τα σύμβολα και τους τύπους που τυχόν χρησιμοποιήσατε, τη σωστή αντικατάσταση μεταβλητών και δεικτών σε αυτούς, αριθμητικά αποτελέσματα ενδιάμεσων ή βοηθητικών μεταβλητών, κ.ά.). Στην περίπτωση επαναληπτικών αλγορίθμων, τα παραπάνω ισχύουν σε κάθε επανάληψη της εκτέλεσής τους.

Άσκηση 2.4: (MLP Backpropagation and Geometrical)
Υποθέστε ότι έχουμε το ακόλουθο νευρωνικό δίκτυο με τιμές βαρών $w_1 = 0.5, w_2 = 0.3, w_3 = -0.2, w_4 = -1, w_5 = 0.5, w_6 = -0.4$. Οι τιμές bias σε όλους τους νευρώνες είναι $b = 0.5$. Οι έξοδοι του κάθε νευρώνα είναι της μορφής $y_i = \sigma(\sum_{\text{inputs}} + b)$, όπου $\sigma(x) = \frac{1}{1+e^{-x}}$ είναι η συνάρτηση logistic.



Για τα δεδομένα ενός δείγματος $(x_1, x_2) = (-0.7, 1.2)$ με τιμή για το πραγματικό του label ίση με $\hat{y} = 0.5$:
(α1) Υπολογίστε την έξοδο y χρησιμοποιώντας τον feedforward αλγόριθμο.
(α2) Χρησιμοποιήστε τον αλγόριθμο backpropagation για να υπολογίσετε τη μερική παράγωγο $\frac{\partial L}{\partial w_1}$. Θεωρήστε ότι η συνάρτηση για το L2 loss δίνεται από τη σχέση $L(y, \hat{y}) = ||y - \hat{y}||_2^2$.
(α3) Να εφαρμόσετε μία επανάληψη του αλγόριθμου gradient descent $\Delta w_i \leftarrow -\eta \frac{\partial L}{\partial w_i}$ με $\eta = 0.1$ για την εύρεση των νέων τιμών των βαρών w_1, w_4 και w_6, καθώς και των νέων τιμών bias b για τους 3 νευρώνες του MLP.

(β) Να κατασκευάσετε το δικό σας multilayer perceptron που να ταξινομεί σύμφωνα με την παρακάτω περιοχή απόφασης. Στην απάντησή σας θα πρέπει να εξηγείτε και γιατί το MLP που προτείνετε όντως αντιστοιχεί στη ζητούμενη περιοχή απόφασης.



Σημείωση: Επεξηγήστε αναλυτικά τη διαδικασία που ακολουθήσατε για να φτάσετε στις λύσεις σας (π.χ. τις παραδοχές, τα σύμβολα και τους τύπους που τυχόν χρησιμοποιήσατε, τη σωστή αντικατάσταση μεταβλητών και δεικτών σε αυτούς, αριθμητικά αποτελέσματα ενδιάμεσων ή βοηθητικών μεταβλητών, κ.ά.). Στην περίπτωση επαναληπτικών αλγορίθμων, τα παραπάνω ισχύουν σε κάθε επανάληψη της εκτέλεσής τους.

Άσκηση 2.5: (Probabilistic PCA)
Θεωρήστε την prior κατανομή μίας κρυφής μεταβλητής $\mathbf{z} \in \mathbb{R}^M$ που ακολουθεί Γκαουσιανή κατανομή, η οποία, χωρίς βλάβη της γενικότητας, μπορεί να θεωρηθεί ότι έχει μηδενική μέση τιμή και μοναδιαίο πίνακα συνδιασποράς:

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

Παρομοίως, η υπό συνθήκη κατανομή της παρατηρούμενης μεταβλητής $\mathbf{x} \in \mathbb{R}^D$, δεδομένης της κρυφής μεταβλητής $\mathbf{z}$, είναι επίσης Γκαουσιανή, της μορφής:

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

όπου $\mathbf{W}$ είναι ένας $D \times M$ πίνακας και $\boldsymbol{\mu}$ ένα D-διάστατο διάνυσμα. Τέλος, η περιθώρια συνάρτηση πιθανότητας είναι επίσης Γκαουσιανή και δίνεται από τη σχέση:

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

όπου ο $D \times D$ πίνακας συνδιασποράς $\mathbf{C}$ ορίζεται ως $\mathbf{C} = \mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I}$.
Θεωρήστε ότι έχουμε ένα σύνολο δεδομένων $\mathbf{X} = \{\mathbf{x}_n\}$, το οποίο αποτελείται από N ανεξάρτητες παρατηρήσεις, καθεμία από τις οποίες σχετίζεται με μία κρυφή μεταβλητή $\mathbf{z}_n$.
(α) Να αποδείξετε ότι η μεγιστοποίηση της log likelihood $\ln p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$ για το probabilistic PCA, που περιγράφθηκε μέσω των παραπάνω εξισώσεων, ως προς την παράμετρο $\boldsymbol{\mu}$, οδηγεί στο αποτέλεσμα $\boldsymbol{\mu}_{ML} = \bar{\mathbf{x}}$, όπου $\bar{\mathbf{x}}$ είναι ο αριθμητικός μέσος του συνόλου δεδομένων.
(β) Υπολογίζοντας τις δεύτερες παραγώγους της log likelihood $\ln p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$, ως προς την παράμετρο $\boldsymbol{\mu}$, να αποδείξετε ότι το stationary σημείο $\boldsymbol{\mu}_{ML} = \bar{\mathbf{x}}$ αποτελεί το μοναδικό μέγιστο.

Άσκηση 2.6: (NMF)
Θεωρήστε την ακόλουθη (μη συμμετρική) συνάρτηση απόστασης (divergence) για D-διάστατα διανύσματα που αποτελούνται αποκλειστικά από μη αρνητικά στοιχεία:

$$\mathcal{D}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{D} \left[u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i \right] \tag{1}$$

(α) Έστω ότι $\mathbf{X}$ είναι ένας πίνακας διαστάσεων $D \times N$, ο οποίος απαρτίζεται από ένα σύνολο $\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$ N στηλών ή εισόδων και $\mathbf{W}$ ένας πίνακας διαστάσεων $D \times K$, ο οποίος απαρτίζεται από ένα σύνολο $\{\boldsymbol{\mu}_k \in \mathbb{R}^D\}_{k=1}^K$ K στηλών ή κέντρων ή προτύπων. Επίσης, θεωρήστε ότι οι δείκτες $y_{kn} \in \{0, 1\}$ απαρτίζουν έναν δυαδικό πίνακα συντελεστών ενεργοποίησης $\mathbf{H}$, διαστάσεων $K \times N$, του οποίου οι στήλες αθροίζουν στη μονάδα. Έστω η εξής συνάρτηση κόστους συσταδοποίησης:

$$\mathcal{E}(\mathbf{H}, \boldsymbol{\mu}) = \sum_{k=1}^{K} \sum_{n=1}^{N} y_{kn} \mathcal{D}(\boldsymbol{\mu}_k, \mathbf{x}_n)$$

Να δείξετε ότι το πρότυπο $\boldsymbol{\mu}_k$ που ελαχιστοποιεί αυτή τη συνάρτηση κόστους, για γνωστά y_{kn}, μπορεί να ερμηνευθεί ως ο γεωμετρικός μέσος (στοιχείο προς στοιχείο) των εισόδων που έχουν αντιστοιχηθεί με το k-στό πρότυπο.
(β) Κατασκευάστε έναν επαναληπτικό αλγόριθμο, ανάλογο με εκείνον του K-means, ο οποίος θα ελαχιστοποιεί αυτή τη συνάρτηση κόστους.

Άσκηση 2.7: (ICA)
(α) Για μία τυχαία μεταβλητή x, με μηδενική μέση τιμή, η κύρτωση ορίζεται ως:

$$\mathrm{kurt}(x) := E[x^4] - 3(E[x^2])^2$$

Να δείξετε ότι για δύο στατιστικώς ανεξάρτητες τυχαίες μεταβλητές x και y, με μηδενικές μέσες τιμές, ισχύει η παρακάτω σχέση:

$$\mathrm{kurt}(x + y) = \mathrm{kurt}(x) + \mathrm{kurt}(y)$$

(β1) Θεωρήστε ότι σας δίνεται ένα σύνολο από N στατιστικώς ανεξάρτητες τυχαίες μεταβλητές s_i, με μηδενικές μέσες τιμές, μοναδιαίες διασπορές, και τιμές a_i για τις κυρτώσεις τους που κυμαίνονται από $-a$ έως a, για κάποια άγνωστη αλλά σταθερή τιμή a. Οι τυχαίες μεταβλητές s_i αναμειγνύονται μέσω σταθερών βαρών w_i ως εξής:

$$x := \sum_{i}^{N} w_i s_i$$

Να προσδιορίσετε τους περιορισμούς που πρέπει να ικανοποιούνται για τα βάρη w_i, ώστε και το μείγμα των τυχαίων μεταβλητών, x, να έχει μοναδιαία διασπορά.
(β2) Να αποδείξετε ότι η κύρτωση ενός ομοιόμορφα σταθμισμένου μείγματος ($w_i = w_j, \ \forall i, j$) N τυχαίων μεταβλητών συγκλίνει στο μηδέν, καθώς το N απειρίζεται. Θεωρήστε ότι το μείγμα, x, έχει μοναδιαία διασπορά.

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.