Ε.Μ.Π., Σχολή Η.Μ.& Μ.Υ.	**Αναγνώριση Προτύπων**	Ακαδ.Ετος 2017-18
Ημερ/νια: 20-12-2017	**2ο Σύνολο Αναλυτ. Ασκήσεων**	Παραδοτέο: 11-01-2018

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.

Κεφάλαια για Ανάγνωση:
Βιβλίο [1]: I, II, III, IV. Βιβλίο [2]: Κεφ. 1, 2, 3, 5, 10.
Βιβλίο [3]: Κεφ. 1, 2, 9. Βιβλίο [4]: Κεφ. 1, 2, 3, 4, 6.

Αναλυτικές Ασκήσεις

Ασκηση 2.1: (Karhunen Loeve Transform-KLT, PCA)
Σύνοψη Θεωρίας: Υποθέτομε *τυχαία διανύσματα* $\boldsymbol{x} \in \mathbb{C}^d$ (που μπορεί να παριστάνουν χαρακτηριστικά σε πρόβλημα αναγνώρισης προτύπων, γενικά με μιγαδικές τιμές). Ο αντίστοιχος χώρος Hilbert έχει εσωτερικό γινόμενο

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \mathcal{E}\{\boldsymbol{y}^H \boldsymbol{x}\}$$

Η *ενέργεια* του κάθε τυχαίου διανύσματος ισούται με

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \mathcal{E}\{\boldsymbol{x}^H \boldsymbol{x}\} = \mathcal{E}\{\|\boldsymbol{x}\|^2\}$$

Θέλομε να βρούμε ένα unitary γραμμικό μετασχηματισμό (πίνακα) $\boldsymbol{A}$ ώστε τα μετασχηματισμένα διανύσματα

$$\boldsymbol{y} = \boldsymbol{A}^H \boldsymbol{x}, \quad \boldsymbol{A}^{-1} = \boldsymbol{A}^H,$$

να έχουν δύο ιδιότητες: 1) Να έχουν ορθογώνιες συνιστώσες, και 2) αν κρατήσομε μόνο τις πρώτες $p < d$ συνιστώσες να έχομε ελάχιστο μέσο τετραγωνικό λάθος (Mean Squared Error-MSE). Η λύση και βέλτιστη επιλογή είναι ο Karhunen Loeve μετ/σμός KLT, γνωστός και ως Principal Component Analysis (PCA). Συγκεκριμένα, επιλέγομε ως στήλες του $\boldsymbol{A}$ τα ορθοκανονικά διανύσματα $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_d\}$ που είναι τα ιδιοδιανύσματα του $\boldsymbol{R}_x = \mathcal{E}\{\boldsymbol{x}\boldsymbol{x}^H\}$. Από αυτά, τα πρώτα p ιδιοδιανύσματα $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_p\}$ αντιστοιχούν στις p μεγαλύτερες ιδιοτιμές $\lambda_1 \geq \cdots \geq \lambda_p$. Η τάξησ-$p$ βέλτιστη προσέγγιση $\hat{\boldsymbol{x}}$ και το αντίστοιχο ελάχιστο MSE J είναι:

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{p} y_k \boldsymbol{e}_k, \quad y_k = \boldsymbol{e}_k^H \boldsymbol{x}$$

$$J = \mathcal{E}\{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2\}$$

Με τις ανωτέρω επιλογές, τα μετασχηματισμένα χαρακτηριστικά $\{y_i\}$ είναι ορθογώνια.

ΠΡΟΣ ΑΠΟΔΕΙΞΗ: (Inductive or Batch matrix solution of PCA)
(a) Assuming that we have solved the PCA problem for the case $p = 1$, prove the general case where p is any integer $1 < p < d$ by using the principle of induction.

(b) Show that the minimum value of the PCA error J with respect to the $\boldsymbol{e}_i$, subject to the orthonormality constraints, is obtained when the $\boldsymbol{e}_i$ are eigenvectors of the correlation (or covariance) matrix. To do this, introduce a matrix $\boldsymbol{H}$ of Lagrange multipliers, one for each constraint, so that the modified distortion measure, in matrix notation reads

$$\tilde{J} = \mathrm{Trace}\{\boldsymbol{U}^H \boldsymbol{R} \boldsymbol{U}\} + \mathrm{Trace}\{\boldsymbol{H}(\boldsymbol{I} - \boldsymbol{U}^H \boldsymbol{U})\}$$

where $\boldsymbol{U}$ is a $d \times (d-p)$ matrix whose columns are given by $\boldsymbol{e}_i$. Now by minimizing $\tilde{J}$ with respect to $\boldsymbol{U}$ show that the solution satisfies $\boldsymbol{R}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{H}$. Clearly, one possible solution is that the columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{R}$, in which case $\boldsymbol{H}$ is a diagonal matrix containing the corresponding eigenvalues. To obtain the general solution, show that $\boldsymbol{H}$ can be assumed to be a symmetric matrix, and by using its eigenvector expansion show that the general solution to $\boldsymbol{R}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{H}$ gives the same value for $\tilde{J}$ as the specific solution in which the columns of $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{R}$. Because these solutions are all equivalent, it is convenient to choose the eigenvector solution.

Ασκηση 2.2: (Negentropy)
By using higher-order moments, show that the negentropy $J(y)$ of a random variable y is approximately equal to

$$J(y) \approx \frac{1}{12}(\mathcal{E}\{y^3\})^2 + \frac{1}{48}[\mathrm{kurt}(y)]^2$$

Ασκηση 2.3: (LDA)
Έστω $p_{\boldsymbol{x}}(\boldsymbol{x}|\omega_i)$ αυθαίρετες πυκνότητες πιθανότητας με μέσες τιμές (διανύσματα) $\boldsymbol{\mu}_i$ και μήτρες συμμεταβλητότητας $\boldsymbol{\Sigma}_i$ - όχι απαραίτητα κανονικές (Γκαουσιανές) - για $i = 1, 2$. Έστω $y = \boldsymbol{w}^T \boldsymbol{x}$ μια προβολή, και έστω ότι οι προκύπτουσες μονοδιάστατες πυκνότητες $p(y|\omega_i)$ έχουν μέσες τιμές μ_i και διασπορές σ_i^2. Δείξτε ότι το κριτήριο

$$J_1(\boldsymbol{w}) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

μεγιστοποιείται από το διάνυσμα βαρών

$$\boldsymbol{w} = (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

Ασκηση 2.4:
Θεωρήστε την περίπτωση όπου η κλάση ω_1 αποτελείται από τα δύο διανύσματα χαρακτηριστικών $[0,0]^T$ και $[0,1]^T$ και η κλάση ω_2 από τα $[1,0]^T$ και $[1,1]^T$. Χρησιμοποιήστε τον ακόλουθο αλγόριθμο εκπαίδευσης ενός perceptron (που είναι ισοδύναμος με τον σειριακό αλγόριθμο 4 του Κεφ.5 του [2]), με $\rho = 1$ και $\boldsymbol{w}(0) = [0,0]^T$, για να σχεδιάσετε μία ευθεία γραμμή που να διαχωρίζει τις δύο κλάσεις.

Αλγόριθμος Perceptron: Έστω $\boldsymbol{w}(t)$ η εκτίμηση του διανύσματος βάρους και $\boldsymbol{x}_{(t)}$ το αντίστοιχο διάνυσμα χαρακτηριστικών στο t-οστό βήμα επανάληψης. Ο αλγόριθμος έχει ως εξής:

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \rho \boldsymbol{x}_{(t)} \quad \text{αν} \quad \boldsymbol{x}_{(t)} \in \omega_1 \quad \text{και} \quad \boldsymbol{w}^T \boldsymbol{x}_{(t)} \le 0 \tag{1}$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \rho \boldsymbol{x}_{(t)} \quad \text{αν} \quad \boldsymbol{x}_{(t)} \in \omega_2 \quad \text{και} \quad \boldsymbol{w}^T \boldsymbol{x}_{(t)} \geq 0 \tag{2}$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) \quad \text{αλλιώς} \tag{3}$$

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 $\boldsymbol{x}_{(t)}$ (punishment = correction cost).

Ασκηση 2.5:
Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις ω_1 και ω_2:

$$\omega_1 : [0.1, -0.2]^T, [0.2, 0.1]^T, [-0.15, 0.2]^T, [1.1, 0.8]^T, [1.2, 1.1]^T$$

$$\omega_2 : [1.1, -0.1]^T, [1.25, 0.15]^T, [0.9, 0.1]^T, [0.1, 1.2]^T, [0.2, 0.9]^T$$

Ελέγξτε αν οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες, και αν όχι, σχεδιάστε ένα κατάλληλο multilayer perceptron με τους κόμβους να έχουν βηματική συνάρτηση ενεργοποίησης (step) για να ταξινομήσετε τα διανύσματα στις δύο κλάσεις.

Ασκηση 2.6:
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:

- 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.
- 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.