
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: pr21_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:

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

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

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)

Έστω ένα κρυφό μοντέλο Markov $\lambda = (A, B, \pi)$ με τρεις καταστάσεις $1, 2, 3$ και δύο τύπους παταρατηρήσεων H και T. Δίνεται ο παρακάτω πίνακας μεταβάσεων A όπου $A_{ij} = p(q_t = j, q_{t-1} = i)$

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Δίνονται επίσης οι ακόλουθες πιθανότητες των παρατηρήσεων B

P(O|q)			
O / q	1	2	3
H	0.5	0.75	0.25
T	0.5	0.25	0.75

Οι a-priori πιθανότητες π είναι $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$

Έστω ότι η παρατηρούμενη ακολουθία είναι $\mathbf{O} = (H, T, H)$. Υπολογίστε:

1. Την πιθανότητα $P(\mathbf{O}|\lambda)$ χρησιμοποιώντας τον forward αλγόριθμο.

2. Την πιθανότητα $P(\mathbf{O}|\lambda)$ χρησιμοποιώντας τον backward αλγόριθμο.

3. Υπολογίστε την πιο πιθανή σειρά καταστάσεων δεδομένης της ακολουθίας $\mathbf{O}$ χρησιμοποιώντας τον αλγόριθμο Viterbi

4. Θέλουμε να χρησιμοποιήσουμε την ακόλουθη σειρά παρατηρήσεων

   $$HHHHHHTTHHHHHHHH$$

   για να εκπαιδεύσουμε το κρυφό μοντέλο Markov. Αντί για τον πολύπλοκο αλγόριθμο forward–backward, θα ακολουθήσουμε τον εξής ψευδο–EM αλγόριθμο που αναφέρεται συχνα ως εκπαίδευση Viterbi, (Viterbi-training).

- Expectation step: Αποκωδικοποιήση χρησιμοποιώντας τον αλγόριθμο Viterbi.
- Maximization step: Μεγιστοποιήση 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 την συνολική πιθανότητα της σειράς καταστάσεων $q_0 q_1 q_2$ που μόλις υπολογίσαμε και παρατηρήσεων $O_0 O_1 O_2$ που δίδονται. Το δεύτερο βήμα του αλγορίθμου είναι η συνήθης εκπαίδευση με μεγιστοποιήση πιθανότητας Maximum - Likelihood -training που εφαρμόζεται σε κάθε Μπεϋσιανό δίκτυο με όλες τις παραμέτρους (q, o) παρατηρήσιμες. Εκτελέστε δύο επαναλήψεις του αλγορίθμου Viterbi training χρησιμοποιώντας τη δοθείσα ακολουθία. (Hint: Μπορείτε να λύσετε την άσκηση γραφικά σε ένα Τρελλις χωρίς πολλές πράξεις).

5. Ποιές είναι οι κύριες διαφορές μεταξύ του forward–backward και Viterbi–training και ποιος αλγόριθμος αναμένεται να έχει καλύτερα αποτελέσματα.

Άσκηση 2.3: (CART)

Στην ηλεκτρονική σελίδα μιας ταξιδιωτικής υπηρεσίας έχετε εισάγει ένα αυτόματο σύστημα διαλόγου για την εξύπηρέτηση των χρηστών κατά την κράτηση εισητηρίων. Για την αξιολόγηση του συστήματος λαμβάνετε υπόψιν τις παρακάτω μετρικές από τις διαδράσεις των χρηστών με το σύστημα:

- SATISFACTION: Αν έμεινε ο χρήστης ικανοποιημένος από τη διάδραση ([Y]ES/[N]O)
- WORD_ACCURACY: Το ποσοστό των λέξεων που αναγνωρίστηκαν επιτυχώς από το σύστημα αναγνώρισης φωνής ($[0 - 100]$)
- TASK_COMPLETION: Αν ολοκληρώθηκε επιτυχώς η κράτηση ([Y]ES/[N]O)
- TASK_DURATION: Ο χρόνος που διήρκεσε η διάδραση σε λεπτά

Κάτα την αξιολόγηση ενός αυτόματου συστήματος διαλόγου λαμβάνετε τις παρακάτω απαντήσεις

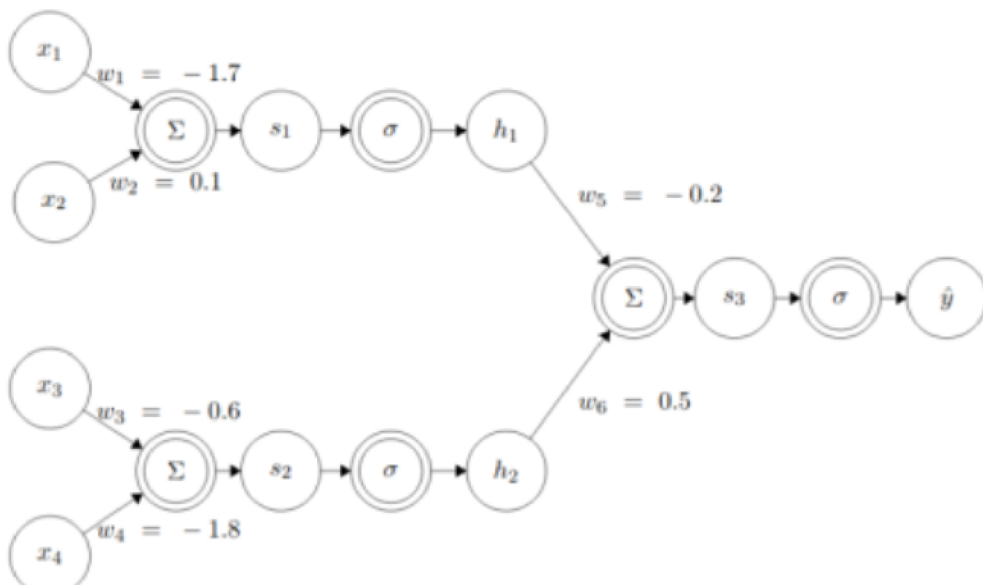
SATISFACTION	WORD_ACCURACY	TASK_COMPLETION	TASK_DURATION
Y	100	Y	3
Y	100	Y	2
Y	90	N	4
N	95	N	2
N	80	Y	5
Y	85	Y	5
Y	80	Y	1
N	85	N	3
Y	95	N	4

Με βάση αυτόν τον πίνακα αξιολόγησης θέλετε να αποφασίσετε σε ποια κατεύθυνση θα επενδύσετε για τη βελτίωση του συστήματος ώστε να μεγιστοποιήσετε την ικανοποίηση των χρηστών. Συγκεκριμένα θέλετε να χρησιμοποιήσετε τον αλγόριθμο CART για να αποφασίσετε ποια από τις παραμέτρους WORD_ACCURACY, TASK_COMPLETION και TASK_DURATION επηρεάζει περισσότερο αν ένας χρήστης έχει μείνει ικανοποιημένος ή όχι από τη διάδραση.

1. Ποιές είναι οι δύο κατηγορίες ω_1 και ω_2 που θέλω να ταξινομήσω με αυτό το δέντρο απόφασης. Ποιές είναι οι παράμετροι και οι ερωτήσεις του δέντρου απόφασης.
2. Κατασκευάστε το δυαδικό δέντρο απόφασης για τις κατηγορίες SATISFACTION=Y και SATISFACTION=N με τις παραμέτρους (ερωτήσεις) WORD_ACCURACY, TASK_COMPLETION, TASK_DURATION ώστε να ελαχιστοποιήσετε την εντροπία σε κάθε σημείο απόφασης (entropy impurity).
3. Ποιά είναι η πρώτη (ποιό σημαντική), δεύτερη, τρίτη ερώτηση του δέντρου. Πως επηρεάζει η επιτυχία της συναλλαγής, η ποιότητα του αναγνωριστή φωνής και η διάρκεια της συνομιλίας, την ικανοποίηση του συνομιλητή απο το σύστημα διαλόγου.

Άσκηση 2.4: (MLP Backpropagation)

Υποθέστε ότι έχουμε το ακόλουθο γράφο (computation graph) που περιγράφει ένα νευρωνικό δίκτυο. Οι κόμβοι που βρίσκονται σε μονό κύκλο υποδηλώνουν μεταβλητές (για παράδειγμα η x_1 είναι μια μεταβλητή εισόδου, η h_1 είναι μια ενδιάμεση μεταβλητή και $\hat{y}$ είναι μια μεταβλητή εξόδου). Οι κόμβοι που βρίσκονται μέσα σε διπλό κύκλο υποδηλώνουν συναρτήσεις (για παράδειγμα το $\sum$ υπολογίζει το άθροισμα των εισόδων του και η σ αναπαριστά τη συνάρτηση logistic $\sigma(x) = \frac{1}{1+e^{-x}}$. Οι ακμές που έχουν βάρη w_i υποδηλώνουν πολλαπλασιασμό της μεταβλητής εισόδου με το w_i.



Θεωρήστε ότι η συνάρτηση για το L2 Loss δίνεται από τη σχέση $L(y, \hat{y}) = \|y - \hat{y}\|_2^2$. Επίσης, υποθέστε ότι μας δίνονται τα δεδομένα ενός δείγματος $(x_1, x_2, x_3, x_4) = (-0.5, 1.4, 0.9, -3)$ με τιμή για το πραγματικό label ίση με 0.5. Χρησιμοποιήστε τον αλγόριθμο backpropagation για να υπολογίσετε τη μερική παράγωγο $\frac{\partial L}{\partial w_1}$.
Σημείωση: το gradient για τη συνάρτηση L2 loss είναι ίσο με $2\|y - \hat{y}\|$

Άσκηση 2.5: (KLT - PCA)

Υποθέτουμε τυχαία διανύσματα $\mathbf{x} \in \mathbb{C}^d$ (που μπορεί να παριστάνουν χαρακτηριστικά σε πρόβλημα αναγνώρισης προτύπων, γενικά με μιγαδικές τιμές). Ο αντίστοιχος χώρος Hilbert έχει εσωτερικό γινόμενο

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

Η ενέργεια του κάθε τυχαίου διανύσματος ισούται με

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

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

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

να έχουν δύο ιδιότητες:

1. Να έχουν ορθογώνιες συνιστώσες
2. αν κρατήσουμε μόνο τις πρώτες $p < d$ συνιστώσες να έχουμε ελάχιστο μέσο τετραγωνικό λάθος (Mean Squared Error – MSE)

Η λύση και βέλτιστη επιλογή είναι ο Karhunen Loeve μετασχηματισμός KLT, γνωστός και ως Principal Component Analysis (PCA). Συγκεκριμένα, επιλέγομε ως στήλες του $\mathbf{A}$ τα ορθοκανονικά διανύσματα $\{\mathbf{e_1}, ..., \mathbf{e_d}\}$ που είναι τα ιδιοδιανύσματα του $R_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$. Από αυτά, τα πρώτα p ιδιοδιανύσματα $\{\mathbf{e_1}, ..., \mathbf{e_p}\}$ αντιστοιχούν στις p μεγαλύτερες ιδιοτιμές $\lambda_1 \geq ... \geq \lambda_p$. Η τάξης-$p$ βέλτιστη προσέγγιση $\hat{x}$ και το αντίστοιχο ελάχιστο MSE J είναι:

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

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

Με τις ανωτέρω επιλογές, τα μετασχηματισμένα χαρακτηριστικά $\{y_i\}$ είναι ορθογώνια.
ΠΡΟΣ ΑΠΟΔΕΙΞΗ: Inductive or Batch matrix solution of PCA

1. Υποθέτοντας ότι έχουμε λύσει το πρόβλημα PCA για την περίπτωση του $p = 1$, αποδείξτε τη γενική περίπτωση όπου το p είναι οποιοσδήποτε αριθμός $1 < p < d$ χρησιμοποιώντας επαγωγή.
2. Δείξτε ότι η ελάχιστη τιμή για το σφάλμα J της PCA ως πρός τα $\mathbf{e}_i$, και δεδομένου του περιορισμού ορθοκανονικότητας, αποκτάται όταν τα $\mathbf{e}_i$ είναι ιδιοδιανύσματα του πίνακα συσχέτισης (συνδιασποράς). Για να πετύχουμε αυτό πρέπει να εισάγουμε τον πίνακα πολλαπλασιαστών Lagrange $\mathbf{H}$, ένα για κάθε περιορισμό, έτσι ώστε η τροποποιημένη μετρική παραμόρφωσης να δίνεται από τον τύπο

$$\tilde{J} = \mathtt{trace}\{\mathbf{U}^H \mathbf{R} \mathbf{U}\} + \mathtt{trace}\{\mathbf{H}(\mathbf{I} - \mathbf{U}^H \mathbf{U})\}$$

όπου το $\mathbf{U}$ είναι ένας πίνακας $d \times (d-p)$, του οποίου οι στήλες είναι τα διανύσματα $\mathbf{e}_i$. Τώρα, ελαχιστοποιώντας το $\tilde{J}$ ως προς το $\mathbf{U}$, δείξτε ότι η λύση ικανοποιεί την εξίσωση $\mathbf{RU} = \mathbf{UH}$. Προφανώς, μια πιθανή λύση αποκτάται όταν οι στήλες του $\mathbf{U}$ είναι ιδιοδιανύσματα του πίνακα $\mathbf{R}$, στην οποία περίπτωση ο $\mathbf{H}$ είναι ένας διαγώνιος πίνακας που περιέχει τις αντίστοιχες ιδιοτιμές. Για να αποκτήσετε μια γενική λύση, δείξτε ότι ο $\mathbf{H}$ μπορεί να θεωρηθεί ένας συμμετρικός πίνακας, και χρησιμοποιώντας την τεχνική επέκτασης ιδιοδιανυσμάτων δείξτε ότι η γενική λύση της $\mathbf{RU} = \mathbf{UH}$ οδηγεί στην ίδια τιμή του $\tilde{J}$ με την ειδική λύση όπου οι στήλες του $\mathbf{U}$ είναι τα ιδιοδιανύσματα του $\mathbf{R}$. Δείχνοντας ότι όλες οι λύσεις είναι ισοδύναμες μπορούμε να επιλέξουμε τη βολική λύση των ιδιοδιανυσμάτων.

Άσκηση 2.6: (Graphical Models)
Δίδεται ένα μοντέλο Markov με μνήμη 2 (MM–2) με δύο καταστάσεις 1 και 2. Γνωρίζουμε ότι για μία σειρά απο καταστάσεις του μοντέλου MM–2 ισχύει η σχέση ανεξαρτησίας

$$p(q_t|q_{t-1}, q_{t-2}, q_{t-3}, ...) = p(q_t|q_{t-1}, q_{t-2})$$

1. Σχεδιάστε το Μπεϋζιανό δίκτυο που αντιστοιχεί στο μοντέλο Markov με μνήμη 2 (MM–2)

2. Παρατηρούμε την ακόλουθη σειρά από καταστάσεις

$$(1, 1, 1, 2, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 1, 1, 1, 2)$$

   Υπολογίστε τις πιθανότητες μετάβασης του MM–2 χρησιμοποιώντας την αρχή της μεγιστοποίησης της πιθανότητας των παρατηρήσεων Maximum Likelihood. Δίνεται ότι:

$$p(q_0 = 1) = p(q_0 = 2) = 0.5$$

3. Υπολογίστε την πιθανότητα να μείνουμε στην κατάσταση 1, 10 φορές δεδομένου ότι είμαστε ήδη στην κατάσταση 1, δηλαδή

$$p(q_1 = 1, q_2 = 1, q_3 = 1, ..., q_{10} = 1, q_{11} = 2|q_0 = 1)$$

   4. θεωρείστε ότι τα δεδομένα που δίνονται στο ερώτημα 2 είναι για ένα μοντέλο Markov με μνήμη 1 (MM-1).

   (α΄) Υπολογίστε τον πίνακα μετάβασης του νέου μοντέλου.

   (β΄) Υπολογίστε πάλι την πιθανότητα του ερωτήματος 3 και συγκρίνετε τα αποτελέσματα.

Άσκηση 2.7: (Linear Discriminant Analysis)
Στο μάθημα είδαμε ότι η Linear Discriminant Analysis (LDA) βασίζεται στην ανάστροφη σχέση των μητρών (πινάκων) S_W και S_B:

$$S_W = \sum_{i=1}^{|Classes|} \mathbb{E}_{x|x\in\omega_i}\left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T\right]$$

$$S_B = \sum_{i=1}^{|Classes|} P(\omega_i)(\vec{\mu_i} - \vec{\mu})(\vec{\mu_i} - \vec{\mu})^T$$

όπου το ω_i αναπαριστά μια κλάση με μέση τιμή $\vec{\mu_i}$, $|Classes|$ είναι το πλήθος των κλάσεων και $\vec{\mu}$ είναι η μέση τιμή όλων των δειγμάτων.
(α) Δείξτε ότι στην περίπτωση διαχωρισμού δύο κλάσεων ω_1 και ω_2, ο πίνακας S_B μπορεί να γραφτεί στη μορφή $S_B = P(\omega_1)P(\omega_2)(\vec{\mu_2} - \vec{\mu_1})(\vec{\mu_2} - \vec{\mu_1})^T$.
(β) Βασιζόμενοι στο υποερώτημα **(α)**, να βρείτε το ιδιοδιάνυσμα του πίνακα $S_W^{-1}S_B$ και την ιδιοτιμή του.

Άσκηση 2.8: (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.

Άσκηση 2.9: (Logistic Regression)
Θεωρήστε το πρόβλημα logistic regression για ένα σύνολο δεδομένων $\{\boldsymbol{\phi}_n, t_n\}$, όπου $t_n \in \{0, 1\}$ και $\boldsymbol{\phi}_n = \boldsymbol{\phi}(\mathbf{x}_n)$ είναι οι κατηγορίες και οι συναρτήσεις βάσης, αντίστοιχα, για δείγματα $n = \{1, 2, \ldots, N\}$. Η συνάρτηση σφάλματος $E(\mathbf{w})$, η οποία αναφέρεται συνήθως και ως cross-entropy, ορίζεται ως:

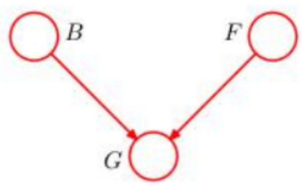
$$E(\mathbf{w}) = -\sum_{n=1}^{N}\{t_n \ln y_n + (1 - t_n)\ln(1 - y_n)\}$$

όπου $\mathbf{w}$ είναι το διάνυσμα βαρών, $y_n = \sigma(\mathbf{w}^\top \boldsymbol{\phi}_n)$ η έξοδος του μοντέλου logistic regression στο διάνυσμα εισόδου $\mathbf{x}_n$, και $\sigma(a) = \frac{1}{1+\exp(-a)}$ η logistic sigmoid συνάρτηση.
(α) Να δείξετε ότι για ένα γραμμικώς διαχωρίσιμο σύνολο δεδομένων, η λύση μέγιστης πιθανοφάνειας για το μοντέλο logistic regression αντιστοιχεί στην εύρεση ενός διανύσματος $\mathbf{w}$, για το οποίο η επιφάνεια απόφασης $\mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}) = 0$ διαχωρίζει τις κλάσεις, απειρίζοντας ταυτόχρονα το μέτρο του διανύσματος $\mathbf{w}$.
(β) Η Hessian μήτρα για το logistic regression δίνεται από τη σχέση:

$$\mathbf{H} = [\mathbf{\Phi}^\top \mathbf{R} \mathbf{\Phi}]$$

όπου $\mathbf{\Phi}$ είναι ο πίνακας των χαρακτηριστικών και $\mathbf{R}$ είναι ένας διαγώνιος πίνακας με στοιχεία $y_n(1 - y_n)$.
Να δείξετε ότι η Hessian μήτρα $\mathbf{H}$ είναι θετικώς ορισμένη. Ως εκ τούτου, δείξτε ότι η συνάρτηση σφάλματος είναι κυρτή συνάρτηση του $\mathbf{w}$ και ότι έχει μοναδικό ελάχιστο.
(γ) Να γράψετε κώδικα που θα υλοποιεί τον iterative reweighted least squares (IWLS) αλγόριθμο για logistic regression. Χρησιμοποιώντας τον αλγόριθμο αυτό, να υπολογίσετε και να σχεδιάσετε τα διαχωριστικά επίπεδα απόφασης που αντιστοιχούν στο σύνολο δεδομένων του προβλήματος τριών κλάσεων που έχει ανεβεί στο mycourses (αρχείο `MLR.data`). Οι δύο πρώτες στήλες περιλαμβάνουν τα διανύσματα χαρακτηριστικών, ενώ η τρίτη την κλάση. Συγκρίνετε τα αποτελέσματα με εκείνα που θα προέκυπταν εάν εφαρμοζόταν ταξινόμηση με βάση τα ελάχιστα τετράγωνα, σχεδιάζοντας τα αντίστοιχα διαχωριστικά επίπεδα απόφασης.
Σημείωση: Λεπτομέρειες από θεωρία του logistic regression και IWLS μπορούν να βρεθούν στις σχετικές διαφάνειες του μαθήματος και τις ενότητες 4.3.2, 4.3.3, 4.3.4 από [3].

Άσκηση 2.10: (Conditional Independence)



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

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

$$\begin{aligned}
p(B = 1) &= 0.95 \\
p(F = 1) &= 0.8 \\
p(G = 1 | B = 1, F = 1) &= 0.95 \\
p(G = 1 | B = 1, F = 0) &= 0.3 \\
p(G = 1 | B = 0, F = 1) &= 0.25 \\
p(G = 1 | B = 0, F = 0) &= 0.2
\end{aligned}$$

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:

$$\begin{aligned}
p(D = 1 | G = 1) &= 0.8 \\
p(D = 0 | G = 0) &= 0.8
\end{aligned}$$

Σύμφωνα με τον οδηγό, ο μετρητής καυσίμων φαίνεται άδειος, δηλαδή παρατηρούμε ότι $D = 0$.

1. Υπολογίστε την πιθανότητα το ντεπόζιτο F να είναι άδειο, με δεδομένη μόνο αυτή την παρατήρηση.
2. 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.

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.