Φροντιστηριακό Μάθημα στην Δεύτερη Σειρά Ασκήσεων

Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ)
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Συστημάτων και Ελέγχου
Ομάδα Επεξεργασίας Φωνής και Φυσικής Γλώσσας

Αναγνώριση Προτύπων, 2021-2022

Πίνακας Περιεχομένων

Άσκηση 2.1: Ερώτημα (α)

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:

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

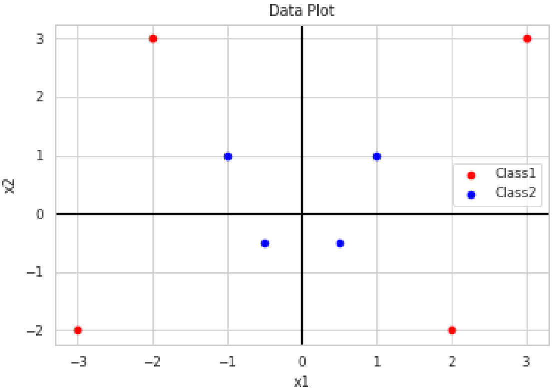
$$\omega_2 : \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.5 \\ -0.5 \end{pmatrix}, \begin{pmatrix} -0.5 \\ -0.5 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}, \qquad z_5 = z_6 = z_7 = z_8 = 1$$

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: $\phi(x_1, x_2) = [1 \quad x_1 \quad x_2 \quad \frac{x_1^2 + x_2^2 - 5}{3}]^\top$.

Άσκηση 2.1: Ερώτημα (α)

Σχεδιάζουμε γραφικά τα διανύσματα και παρατηρούμε ότι οι δύο κλάσεις **δεν** είναι γραμμικώς διαχωρίσιμες.



Άσκηση 2.1: Ερώτημα (α)

Μετασχηματίζουμε τα δεδομένα μέσω ενός μη γραμμικού μετασχηματισμού: $\phi : \mathbb{R}^2 \to \mathbb{R}^4$
Με αυτόν τον τρόπο προσπαθούμε να φέρουμε τα δεδομένα σε έναν χώρο μεγαλύτερης διαστατικότητας στον οποίο οι δύο κλάσεις θα είναι γραμμικώς διαχωρίσιμες. Η συνάρτηση ϕ είναι η εξής:

$$\phi\left(\tilde{\mathbf{x}}_n\right) = \left[1, x_1, x_2, \frac{x_1^2 + x_2^2 - 5}{3}\right]^\top$$

Έτσι, βρίσκουμε τα εξής αποτελέσματα:

$\tilde{\mathbf{x}}_n$	$\phi\left(\tilde{\mathbf{x}}_n\right)$	Κλάση
$[3, 3]^\top$	$[1, 3, 3, 13/3]^\top$	ω_1
$[2, -2]^\top$	$[1, 2, -2, 1]^\top$	ω_1
$[-3, -2]^\top$	$[1, -3, -2, 8/3]^\top$	ω_1
$[-2, 3]^\top$	$[1, -2, 3, 8/3]^\top$	ω_1
$[1, 1]^\top$	$[1, 1, 1, -1]^\top$	ω_2
$[0.5, -0.5]^\top$	$[1, 0.5, -0.5, -1.5]^\top$	ω_2
$[-0.5, -0.5]^\top$	$[1, -0.5, -0.5, -1.5]^\top$	ω_2
$[-1, 1]^\top$	$[1, -1, 1, -1]^\top$	ω_2

Άσκηση 2.1: Ερώτημα (β)

(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;

Στόχος μας είναι να ελαχιστοποιησουμε το $\|\mathbf{w}\|$ και για αυτό κατασκευάζουμε το εξής συναρτησιακό: $\mathbf{L}(\mathbf{w}, \mathbf{a}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{k=1}^{8} \alpha_k \left[z_k \mathbf{w}^t \mathbf{y}_k - 1\right]$ και θέλουμε να ελαχιστοποιήσουμε το $\mathbf{L}$ ως προς το διάνυσμα βαρών w και να το μεγιστοποιήσουμε ως προς τους συντελεστές $\alpha_k \geq 1$.
Το πρόβλημα μπορεί να αναχθεί στην μεγιστοποίηση του συναρτησιακού:
$L(\mathbf{a}) = \sum_{k=1}^{8} \alpha_i - \frac{1}{2} \sum_k^8 \sum_j^8 \alpha_k \alpha_j z_k z_j \mathbf{y}_j^t \mathbf{y}_k$
με τους εξής περιορισμούς: $\sum_{k=1}^{8} z_k, \alpha_k = 0 \quad \alpha_k \geq 0, k = 1, \ldots, 8$ για τα δεδομένα εκδπάιδευσης.
Οι δύο προηγούμενες εξισώσεις μπορούν να εκφραστούν σε μία ενιαία μορφή συναρτησιακού:

$$\mathbf{Q}(\mathbf{a}, \lambda) = \sum_{k=1}^{8} \alpha_i - \frac{1}{2} \sum_{k}^{8} \sum_{j}^{8} \alpha_k \alpha_j z_k z_j \mathbf{y}_j^t \mathbf{y}_k + \lambda \sum_{k=1}^{n} z_k \alpha_k$$

Άσκηση 2.1: Ερώτημα (β)

Με αντικατάσταση βρίσκουμε ότι:
$Q =$ -18.889$\alpha_1^2 - 5.333\alpha_1\alpha_2 + 2.445\alpha_1\alpha_3 - 15.556\alpha_1\alpha_4 + 2.667\alpha_1\alpha_5 - 5.5\alpha_1\alpha_6 - 8.5\alpha_1\alpha_7 - 3.333\alpha_1\alpha_8 + \alpha_1 - 5.0\alpha_2^2 - 1.667\alpha_2\alpha_3 + 6.333\alpha_2\alpha_4 + 1.5\alpha_2\alpha_6 - 0.5\alpha_2\alpha_7 - 4.0\alpha_2\alpha_8 + \alpha_2 - 10.556\alpha_3^2 - 8.111\alpha_3\alpha_4 - 6.667\alpha_3\alpha_5 - 3.5\alpha_3\alpha_6 - 0.5\alpha_3\alpha_7 - 0.667\alpha_3\alpha_8 + \alpha_3 - 10.556\alpha_4^2 - 0.667\alpha_4\alpha_5 - 5.5\alpha_4\alpha_6 - 3.5\alpha_4\alpha_7 + 3.333\alpha_4\alpha_8 + \alpha_4 - 2.0\alpha_5^2 - 2.5\alpha_5\alpha_6 - 1.5\alpha_5\alpha_7 - 2.0\alpha_5\alpha_8 + \alpha_5 - 1.875\alpha_6^2 - 3.25\alpha_6\alpha_7 - 1.5\alpha_6\alpha_8 + \alpha_6 - 1.875\alpha_7^2 - 2.5\alpha_7\alpha_8 + \alpha_7 - 2.0\alpha_8^2 + \alpha_8 + \lambda(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \alpha_5 - \alpha_6 - \alpha_7 - \alpha_8)$
Για την ελαχιστοποίηση του συναρτησιακού Q υπολογίζουμε το:

$$\frac{\partial Q}{\partial a_i} = 0, i = 1, \ldots, 8$$

Άσκηση 2.1: Ερώτημα (β)

Έτσι, από αυτές τις 8 διαφορετικές παραγόγους προκύπτει ένα γραμμικό σύστημα εξισώσεων που μπορεί να γραφτεί με την εξής μορφή:

$$\mathbf{AX} = \mathbf{B}$$

Όπου:

$$X = [\alpha_1, \ldots, \alpha_8, \lambda]^\top$$
$$B = [z_1, \ldots, z_8, 0]^\top$$

$$A = \begin{pmatrix} -340/9 & -16/3 & 22/9 & -140/9 & 8/3 & -11/2 & -17/2 & -10/3 & 1 \\ -16/3 & -10 & -5/3 & 19/3 & 0 & 3/2 & -1/2 & 4 & 1 \\ 22/9 & -5/3 & -190/9 & -73/9 & -20/3 & -7/2 & -1/2 & -2/3 & 1 \\ -140/9 & 19/3 & -73/9 & -190/9 & -2/3 & -11/2 & -7/2 & 10/3 & 1 \\ -8/3 & 0 & 20/3 & 2/3 & 4 & 5/2 & 3/2 & 2 & 1 \\ 11/2 & -3/2 & 7/2 & 11/2 & 5/2 & 15/4 & 13/4 & 3/2 & 1 \\ 17/2 & 1/2 & 1/2 & 7/2 & 3/1 & 13/4 & 15/4 & 5/2 & 1 \\ 10/3 & 4 & 2/3 & -10/3 & 2 & 3/2 & 5/2 & 4 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Άσκηση 2.1: Ερώτημα (β)

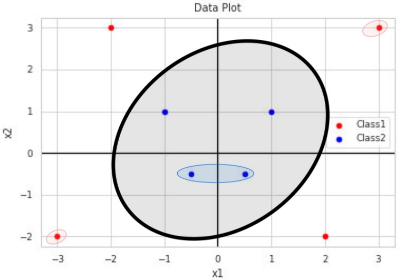
Πρώτη προσέγγιση (Σωστή λύση):

- Αν ο πίνακας A ήταν αντιστρέψιμος τότε η εξίσωση $\mathbf{AX} = \mathbf{B}$ λύνεται απλά
- Ο Πίνακας A είναι μη αντιστρέψιμος $|A| = 0$.
- Ο Πίνακας A έχει βαθμό $rank(A) = 5$, επομένως από τα 8 διανύσματα μόνο τα 4 θα χρησιμοποιηθούν στην εκπαίδευση του SVM. Αυτά τα 4 θα έχουν πολλαπλασιαστές Lagrange διαφορετικούς του μηδενός, ενώ τα υπόλοιπα 4 θα έχουν $\lambda = 0$
- Άρα, πρέπει να εντοπίσουμε τα 4 σημεία που δεν είναι SVM.

Δεύτερη προσέγγιση (Οριακά αποδεκτή λύση - Λύση σε πολλά γραπτά):

- Ο Πίνακας A είναι μη αντιστρέψιμος $|A| = 0$.
- Μπορούμε να χρησιμοποιήσουμε μια προσεγγιστική μέθοδο για την επίλυση του προβλήματος $\mathbf{AX} = \mathbf{B}$, π.χ. την λύση ελαχίστων τετραγώνων.

Άσκηση 2.1: Ερώτημα (β)



- Εντοπίζουμε γραφικά τα 4 διανύσματα που (μάλλον) θα απέχουν περισσότερο από το βέλτιστο υπερεπίπεδο.
- Μπορούμε να πειραματιστούμε με διάφορα σημεία, στόχος μας είναι να κρατήσουμε 4 διανύσματα και ο πίνακας A να είναι αντιστρέψιμος
- Έτσι, αφαιρούμε τα διανύσματα y_1, y_3, y_6, y_7 από την εκπαίδευση του SVM.

Άσκηση 2.1: Ερώτημα (β)

Έτσι, μελετώντας μόνο τα διανύσματα y_2, y_4, y_5, y_8 γράφουμε τις εξισώσεις 4 εξισώσεις: $\frac{\partial Q}{\partial a_i} = 0, i = 2, 4, 5, 8$ και προκύπτει ένα γραμμικό σύστημα εξισώσεων που μπορεί να γραφτεί με την εξής μορφή:

$$\mathbf{AX} = \mathbf{B}$$

Όπου:

$$X = [\alpha_2, \alpha_4, \alpha_5, \alpha_8, \lambda]^\top$$

$$B = [z_2, z_4, z_5, z_8, 0]^\top$$

$$\mathbf{A} = \begin{pmatrix} -10 & 19/3 & 0 & -4 & 1 \\ 19/3 & -190/9 & -2/3 & 10/3 & 1 \\ 0 & 2/3 & 4 & 2 & 1 \\ 4 & -10/3 & 2 & 4 & 1 \\ -1 & -1 & 1 & 1 & 0 \end{pmatrix}$$

Αφού $rank(A) = 5$ τότε μπορούμε να λύσουμε ως προς X: $X = A^{-1}B$ Με

$$\mathbf{A}^{-1} = \begin{pmatrix} -157/2025 & 4/675 & 161/1350 & -193/4050 & -17/45 \\ 4/675 & -13/225 & -17/450 & 121/1360 & -1/15 \\ -161/1350 & 17/450 & 403/900 & -989/2700 & -1/30 \\ 193/4050 & -121/1350 & -989/2700 & 3307/8100 & 53/90 \\ 17/45 & 1/15 & -1/30 & 53/90 & -1 \end{pmatrix}$$

Και έτσι βρίσκουμε:

$$\alpha_2 = \frac{58}{405}, \quad \alpha_4 = \frac{14}{135}, \quad \alpha_5 = \frac{22}{135}, \quad \alpha_8 = \frac{34}{405}$$

- Αποδεκτή λύση αφού τα α_i είναι θετικά
- Άρα πράγματι τα SVM είναι τα διανύσματα: y_2, y_4, y_5, y_8

Άσκηση 2.1: Ερώτημα (γ)

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

Άσκηση 2.1: Ερώτημα (γ)

$$\mathbf{L}(\mathbf{w}, \mathbf{a}) = \frac{1}{2}\|\tilde{\mathbf{w}}\|^2 - \sum_{k=1}^{8} \alpha_k \left[z_k \mathbf{w}^t \mathbf{y}_k - 1 \right]$$

$$\frac{\partial \mathbf{L}}{\tilde{\mathbf{w}}} = 0 \Rightarrow \tilde{\mathbf{w}} = \sum_{n=1}^{8} \alpha_n z_n \tilde{\mathbf{y}}_n$$

Όπου $w = [w_0, \tilde{w}]^\top$ και $y = [1, \tilde{y}]^\top$ Έτσι, $\tilde{w} = [0, 2/9, -2/3]^\top$

- Για την εύρεση του bias w_0 χρησιμοποιούμε το γεγονός ότι για ένα από τα διανύσματα που ειναι support vectors ισχύει ότι: $z_n \mathbf{w}^\top \mathbf{y}_n = 1$
- Για το διάνυσμα $\tilde{x_2}$ μέσω τις προηγούμενης σχέσης βρίσκουμε ότι: $w_0 = 1/9$
- Άρα, $w = [1/9, 0, 2/9, -2/3]^\top$

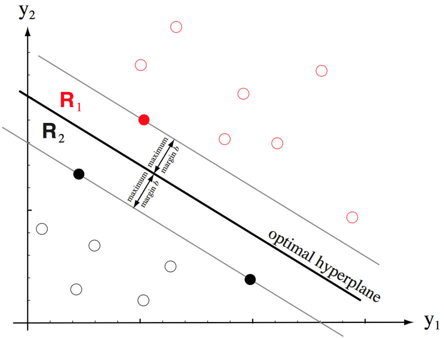
Άσκηση 2.1: Ερώτημα (γ)

- Για $y_2, y_4, y_5, y_6, y_7, y_8$ δίνουν $z_n \mathbf{w}^\top \mathbf{y}_n = 1$
- Για y_1, y_3 δίνουν $z_n \mathbf{w}^\top \mathbf{y}_n = 19/9$ (συμμετρικά)
- Θα μπορούσαμε στην ανάλυση μας ισοδύναμα, να είχαμε θεωρήσει άλλο υποσύνολο του συνόλου $y_2, y_4, y_5, y_6, y_7, y_8$ ως support vectors.
- Μπορεί να αποδειχθεί εξαρχής ότι το y_3 δεν γίνεται να είναι support vectors;

Άσκηση 2.1: Ερώτημα (δ)

(δ) Υπολογίστε το περιθώριο β του ταξινομητή.

- $\beta = \frac{1}{||\tilde{\mathbf{w}}||} = \frac{9\sqrt{10}}{20}$

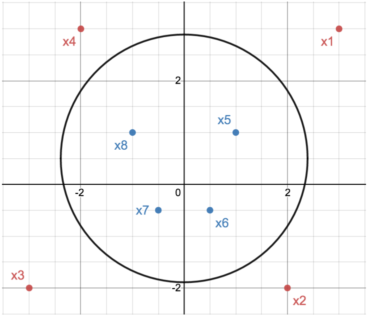


Άσκηση 2.1: Ερώτημα (ε)

(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

Άσκηση 2.1: Ερώτημα (ε)

- $g\left(x_1, x_2\right) = \mathbf{w}^\top \phi(\tilde{\mathbf{x}}) = \frac{w_3}{3} x_1^2 + \frac{w_3}{3} x_2^2 + w_1 x_1 + w_2 x_2 - \frac{5 w_3}{3} + w_0$
- $w = [1/9, 0, 2/9, -2/3]^\top$
- $g\left(x_1, x_2\right) = 0$
- Άρα επιφάνεια διαχωρισμού: $x_1^2 + \left(x_2 - \frac{1}{2}\right)^2 = \frac{23}{4}$ που είναι κύκλος με κέντρο το σημείο $(0, 0.5)$ και ακτίνα $r = \sqrt{\frac{23}{4}}$



Άσκηση 2.1: Ερώτημα (στ)

(στ) Ποια είναι τα support vectors;

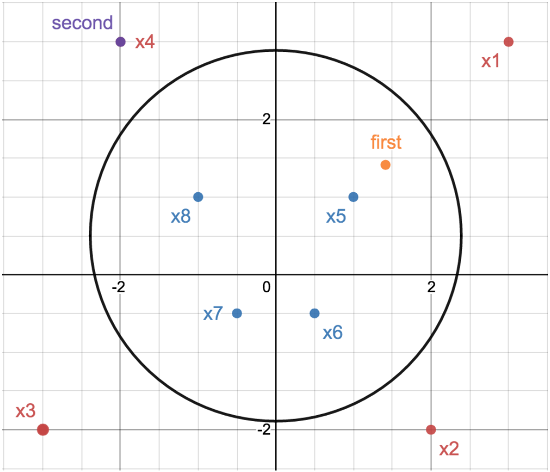
Άσκηση 2.1: Ερώτημα (στ)

- Βρήκαμε ως support vectors τα διανύσματα: y_2, y_4, y_5, y_8
- Θα μπροούσαμε στην ανάλυση μας ισοδύναμα, να είχαμε θεωρήσει άλλο υποσύνολο του συνόλου $y_2, y_4, y_5, y_6, y_7, y_8$ ως support vectors.
- Μπορεί να αποδειχθεί εξαρχής ότι το y_3 δεν γίνεται να είναι support vectors;

Άσκηση 2.1: Ερώτημα (ζ)

(ζ) Σε ποιες κατηγορίες ταξινομούνται τα σημεία $\begin{pmatrix} \sqrt{2} \\ \sqrt{2} \end{pmatrix}$ και $\begin{pmatrix} -2 \\ 3 \end{pmatrix}$;

Άσκηση 2.1: Ερώτημα (ζ)



- Το πρώτο διάνυσμα ανήκει στην κλάση ω_2
- Το δεύτερο διάνυσμααανήκει στην κλάση ω_1

Πίνακας Περιεχομένων

Άσκηση 2.2: Ερώτημα (α)

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.

Άσκηση 2.2: Ερώτημα (α)

- Σύνολο Κρυφών Καταστάσεων: $S = \{S_1, S_2, S_3\}$
- Σύνολο Παρατηρήσιμων Καταστάσεων: $\{H, T\}$
- Πίνακας Α: Ο πίνακας Α περιγράφει τις πιθανότητες το σύστημα να μεταβεί από οποιαδήποτε κατάσταση σε οποιαδήποτε άλλη. Π.χ αφού $A_{i,j} = 1/3$ αν το σύστημα βρίσκεται στην κατάσταση S_1 τότε με πιθανότητα παραμένει στην ίδια κατάσταση με πιθανότητα $1/3$ μεταβαίνει στην κατάσταση S_2 και με στην S_3.
- Πίνακας Β: Ο πίνακας Β περιγράφει τις πιθανότητες το σύστημα να εκδήλωση μια κατάσταση από το Σύνολο Παρατηρήσιμων Καταστάσεων: H,T δεδομένου της κρυφής κατάσταση που βρίσκεται το σύστημα. Π.χ Δεδομένου ότι το σύστημα βρίσκεται στην κρυφή κατάσταση S_2 είναι πιο πιθανό να παρατηρούμε την κατάσταση H.
- Οι a-priori πιθανότητες δηλώνουν την πιθανότητα να βρίσκεται το σύστημα σε μια από τις καταστάσεις, κατα την έναρξη του πειράματος. Άρα εδώ έχουμε ίση πιθανότητα.
- Στο πείραμα μας παρατηρούμε το εξής σύνολο: $O = \{H, T, H\}$

Άσκηση 2.2: Ερώτημα (α)

Αλγόριθμος Forward:

- Αρχικοποίηση: $\alpha_1(i) = \pi_i \cdot b_i(O_1)$
- Επανάληψη: $\alpha_{t+1}(j) = (\sum_{i=1}^{3} \alpha_t(i) \cdot A_{ij}) \cdot b_i(O_{t+1})$
- Τερματισμός: $P(O \mid \lambda) = \sum_{i=1}^{3} \alpha_3(i)$

Άσκηση 2.2: Ερώτημα (α)

Αρχικοποίηση:

$\alpha_1(1) = \pi_1 \cdot b_1(H) = \frac{1}{3} \cdot 0.5 = 0.1667$
$\alpha_1(2) = \pi_1 \cdot b_2(H) = \frac{1}{3} \cdot 0.75 = 0.25$
$\alpha_1(3) = \pi_1 \cdot b_3(H) = \frac{1}{3} \cdot 0.25 = 0.083$

Επανάληψη:

$t = 2\ (O_2 = T)$:
$\alpha_2(1) = \left[\sum_{i=1}^{3} \alpha_1(i) \cdot A_{i1}\right] \cdot b_1(T) = \frac{1}{3}(0.1667 + 0.25 + 0.083)0.5 = 0.0833$
$\alpha_2(2) = \left[\sum_{i=1}^{3} \alpha_1(i) \cdot A_{i2}\right] \cdot b_2(T) = \frac{1}{3}(0.1667 + 0.25 + 0.083)0.25 = 0.0416$
$\alpha_2(3) = \left[\sum_{i=1}^{3} \alpha_1(i) \cdot A_{i3}\right] \cdot b_3(T) = \frac{1}{3}(0.1667 + 0.25 + 0.083)0.75 = 0.125$

$t = 3\ (O_3 = H)$:
$\alpha_3(1) = \left[\sum_{i=1}^{3} \alpha_2(i) \cdot A_{i1}\right] \cdot b_1(H) = \frac{1}{3}(0.0833 + 0.0416 + 0.125)0.5 = 0.0417$
$\alpha_3(2) = \left[\sum_{i=1}^{3} \alpha_2(i) \cdot A_{i2}\right] \cdot b_2(H) = \frac{1}{3}(0.0833 + 0.0416 + 0.125)0.75 = 0.0625$
$\alpha_3(3) = \left[\sum_{i=1}^{3} \alpha_2(i) \cdot A_{i3}\right] \cdot b_3(H) = \frac{1}{3}(0.0833 + 0.0416 + 0.125)0.25 = 0.021$

Τερματισμός:

$P(O \mid \lambda) = \sum_{i=1}^{3} \alpha_3(i) = 0.125$

Άσκηση 2.2: Ερώτημα (β)

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

Άσκηση 2.2: Ερώτημα (β)

Αλγόριθμος Backward:

- Αρχικοποίηση: $\beta_3(i) = 1 \ \forall$ Κρυφή Κατάσταση
- Επανάληψη: $\beta_{t+1}(i) = \sum_{j=1}^{3} A_{ij} \cdot b_j(O_{t+1}) \cdot \beta_{t+1}(j)$
- Τερματισμός: $P(O \mid \lambda) = \sum_{i=1}^{3} \beta_1(i) \cdot \pi_i \cdot b_i(O_1)$

Άσκηση 2.2: Ερώτημα (α)

Αρχικοποίηση:
$\beta_3(1) = 1$
$\beta_3(2) = 1$
$\beta_3(3) = 1$
Επανάληψη:
$t = 2$ $(O_2 = T)$:
$\beta_2(1) = \sum_{j=1}^{3} A_{1j} \cdot b_j(H) \cdot \beta_3(j) = \frac{1}{3}(0.5 + 0.75 + 0.25) = 0.5$
3: $\beta_2(2) = \sum_{j=1}^{3} A_{2j} \cdot b_j(H) \cdot \beta_3(j) = \frac{1}{3}(0.5 + 0.75 + 0.25) = 0.5$
4: $\beta_2(3) = \sum_{j=1}^{3} A_{3j} \cdot b_j(H) \cdot \beta_3(j) = \frac{1}{3}(0.5 + 0.75 + 0.25) = 0.5$
$t = 1$ $(O_1 = T)$:
6: $\beta_1(1) = \sum_{j=1}^{3} A_{1j} \cdot b_j(T) \cdot \beta_2(j) = \frac{1}{3}(0.5 + 0.75 + 0.25)0.5 = 0.25$
7: $\beta_1(2) = \sum_{j=1}^{3} A_{2j} \cdot b_j(T) \cdot \beta_2(j) = \frac{1}{3}(0.5 + 0.75 + 0.25)0.5 = 0.25$
8: $\beta_1(3) = \sum_{j=1}^{3} A_{3j} \cdot b_j(T) \cdot \beta_2(j) = \frac{1}{3}(0.5 + 0.75 + 0.25)0.5 = 0.25$
Τερματισμός:
$P(O \mid \lambda) = \sum_{i=1}^{3} \beta_1(i) \cdot \pi_i \cdot b_i(H) = \frac{1}{2}(0.25 \cdot 0.5 + 0.25 \cdot 0.75 + 0.25 \cdot 0.25) = 0.125$
Παρατηρούμε ότι τόσο ο Forward όσο και ο Backward δίνουν την ίδια πιθανότητα $P(O \mid \lambda) = 0.125$.

Άσκηση 2.2: Ερώτημα (γ)

3. Υπολογίστε την πιο πιθανή σειρά καταστάσεων δεδομένης της ακολουθίας **O** χρησιμοποιώντας τον αλγόριθμο Viterbi

Άσκηση 2.2: Ερώτημα (γ)

Αλγόριθμος Viterbi:

- Αρχικοποίηση:
  $\delta_1(i) = \pi_i \cdot b_i(O_1)$ $\forall$ Κρυφή Κατάσταση
  $\psi_1(i) = 0$ $\forall$ Κρυφή Κατάσταση
- Επανάληψη:
  $\max_{1 \leq i \leq 3} [\delta_{t-1}(i) \cdot A_{ij}] \cdot b_j(O_t)$
  $\psi_t(j) = \arg\max_{1 \leq i \leq 3} [\delta_{t-1}(i) A_{i1}]$
- Τερματισμός:
  $P^* = \max_{1 \leq i \leq 3} [\delta_3(i)]$
  $q_3^* = \arg\max_{1 \leq i \leq 3} [\delta_3(i)]$
  Παρατηρούμε ότι ο Αλγόριθμος Viterbi αλλάζει την άθροιση σε μεγιστοποίηση.

Άσκηση 2.2: Ερώτημα (γ)

Αρχικοποίηση:

$\delta_1(1) = \pi_1 \cdot b_1(H) = \frac{1}{3} \cdot 0.5 = 0.167$

$\psi_1(1) = 0$

$\delta_1(2) = \pi_2 \cdot b_2(H) = \frac{1}{3} \cdot 0.75 = 0.25$

$\psi_1(2) = 0$

$\delta_1(3) = \pi_3 \cdot b_3(H) = \frac{1}{3} \cdot 0.25 = 0.083$

$\psi_1(3) = 0$

Επανάληψη:

$t = 2 \ (O_2 = T)$:

$\delta_2(1) = \max_{1 \le i \le 3} [\delta_1(i) A_{i1}] \cdot b_1(T) = \frac{1}{3} \cdot 0.25 \cdot 0.5 = 0.041$

$\psi_2(1) = \arg\max_{1 \le i \le 3} [\delta_1(i) A_{i1}] = 2$

$\delta_2(2) = \max_{1 \le i \le 3} [\delta_1(i) A_{i2}] \cdot b_2(T) = \frac{1}{3} \cdot 0.25 \cdot 0.25 = 0.0208$

$\psi_2(2) = \arg\max_{1 \le i \le 3} [\delta_1(i) A_{i2}] = 2$

$\delta_2(3) = \max_{1 \le i \le 3} [\delta_1(i) A_{i3}] \cdot b_3(T) = \frac{1}{3} \cdot 0.25 \cdot 0.75 = 0.0625$

$\psi_2(3) = \arg\max_{1 \le i \le 3} [\delta_1(i) A_{i3}] = 2$

Άσκηση 2.2: Ερώτημα (γ)

$t = 3$ ($O_3 = H$):
$\delta_3(1) = \max_{1 \leq i \leq 3} [\delta_2(i) A_{i1}] \cdot b_1(H) = \frac{1}{3} \cdot 0.0625 \cdot 0.5 = 0.01$
$\psi_3(1) = \arg\max_{1 \leq i \leq 3} [\delta_2(i) A_{i1}] = 3$
$\delta_3(2) = \max_{1 \leq i \leq 3} [\delta_2(i) A_{i2}] \cdot b_2(H) = \frac{1}{3} \cdot 0.0625 \cdot 0.75 = 0.0156$
$\psi_3(2) = \arg\max_{1 \leq i \leq 3} [\delta_2(i) A_{i2}] = 3$
$\delta_3(3) = \max_{1 \leq i \leq 3} [\delta_2(i) A_{i3}] \cdot b_3(H) = \frac{1}{3} \cdot 0.0625 \cdot 0.25 = 0.0052$
$\psi_3(3) = \arg\max_{1 \leq i \leq 3} [\delta_2(i) A_{i3}] = 3$
Τερματισμός:
$P^* = \max_{1 \leq i \leq 3} [\delta_3(i)] = \delta_3(2) = 0.0156$
$q_3^* = \arg\max_{1 \leq i \leq 3} [\delta_3(i)] = 2$
$q_2^* = \psi_3(2) = 3$
$q_1^* = \psi_2(3) = 2$
Άρα η πιο πιθανη σειρα καταστασεων ειναι η (S_2, S_3, S_2).

Άσκηση 2.2: Ερώτημα (δ)

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 $q_0q_1q_2$ που μόλις υπολογίσαμε και παρατηρήσεων $O_0O_1O_2$ 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).

Άσκηση 2.2: Ερώτημα (δ)

Στο Viterbi Training θέλουμε με βάση κάποια παρατηρησιακά δεδομένα να εκπαιδεύσουμε τις παραμέτρους λ του συστήματος δηλαδή τον πίνακα A τον πίνακα B και τις a-priori π.
Εδώ έχουμε **ένα** παρατηρησιακό δεδομένο: $O = HHHHHHTTHHHHHHHH$
Expectation Step (1):

- Στο E-step έχουμε απλά παρατηρήσεις τις οποίες πρέπει να αντιστοιχίσουμε στις κρυφές καταστάσεις.
- Τόσο οι πιθανότητες μετάβασης όσο και οι a-priori είναι ίσες τότε η αναμενόμενη κρυφή κατάσταση στο βήμα i είναι η $\arg\max_{1 \leq i \leq 3} b_i$
- Άρα το E-step επιστρέφει την εξής ακολουθία αναμενόμενων κρυφων καταστάσεων: $S_2, S_2, S_2, S_2, S_2, S_2, S_3, S_3, S_2, S_2, S_2, S_2, S_2, S_2, S_2, S_2$

Άσκηση 2.2: Ερώτημα (δ)

Έχουμε τα εξής:

- $O = HHHHHHTTHHHHHHHH$
- $S = S_2, S_2, S_2, S_2, S_2, S_2, S_3, S_3, S_2, S_2, S_2, S_2, S_2, S_2, S_2, S_2$

Maximization Step:
Έχουμε:

- $\alpha^*_{ij} = \frac{\#S_i \to S_j}{\#S_i \to *}$
- $\pi^*_i = \frac{P[O, q_0 = i]}{P[O]}$
  Προσοχή! Εξετάζουμε μόνο την αρχική κατάσταση κάθε παρατηρισιακού δεδομένου, εδώ έχουμε μόνο ένα παρατηρησιακό δεδομένο.
- $b^*_j(V_k) = P\left[O = V_k \mid q = i\right] = \frac{\#O_i = V_k | q = i}{\#O_i | q = i}$
  Δηλάδη δεδομένου της κατάστασης S_i πόσες παρατηρήσεις (O_i) είναι η εξεταζόμενες (V_k) προς το πλήθος των παρατηρήσεων (O_i) που παράγονται σε αυτή την κατάσταση.

Άσκηση 2.2: Ερώτημα (δ)

Έχουμε τα εξής:

- $O = HHHHHHTTHHHHHHHH$
- $S = S_2, S_2, S_2, S_2, S_2, S_2, S_3, S_3, S_2, S_2, S_2, S_2, S_2, S_2, S_2, S_2$

$A_{22}^{(1)} = \frac{12}{13}, \quad A_{23}^{(1)} = \frac{1}{13}, \quad A_{32}^{(1)} = \frac{1}{2}, \quad A_{33}^{(1)} = \frac{1}{2}$

$A_{21}^{(1)} = A_{31}^{(1)} = 0$ επειδή δεν έχουμε καμία τέτοια μετάβαση.

Το A_{1j} δεν αλλάζει αφού δεν έχουμε παρατήρηση του S_1.

Άρα $\mathbf{A}^{(1)} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 12/13 & 1/13 \\ 0 & 0.5 & 0.5 \end{pmatrix}$

$\pi^{(1)} = [0, 1, 0]^\top$

$b_2^{(1)}(H) = 1 \quad b_3^{(1)}(T) = 1$

$b_2^{(1)}(T) = 0 \quad b_3^{(1)}(H) = 0$

Τα $b_1^{(1)}(H) \quad b_1^{(1)}(T)$ δεν αλλάζουν αφού δεν έχουμε παρατήρηση του S_1.

Άσκηση 2.2: Ερώτημα (δ)

Ποιοτική Προσέγγιση:

- Από την a-priori έχουμε ότι το σύστημα ξεκινάει σίγουρα από την κατάσταση S_2.
- Από τον πίνακα A βλέπουμε ότι το σύστημα δεν μπορεί να πάει από την κατάσταση S_2 στην S_1. Αν πάει στην κατάσταση S_3 με πιθανότητα $1/13$ τότε και εκεί είναι αδύνατο να πάει στην S_1. Άρα το σύστημα δεν θα πάει ποτέ στην κατάσταση S_1.
- Από τον πίνακα B βλέπουμε ότι στην S_2 παρατηρούμε μονο το H, ενώ στην S_1 παρατηρούμε μόνο το T.
- Επομένως καταλήγουμε στην ίδια ακολουθία κρυφών καταστάσεων με την πρώτη επανάληψη
- Άρα, δεν αλλάζει κάτι στο M-Step, και ο αλγόριθμος συγκλίνει.

Άσκηση 2.2: Ερώτημα (ε)

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

Άσκηση 2.2: Ερώτημα (ε)

Ο forward-backward αλγόριθμος είναι εν γένει πιο αργός λόγω των δύο περασμάτων που εκτελεί, αλλά ταυτόχρονα και πιο ακριβής στα αποτελέσματα σε σχέση με τον pseudo-EM αλγόριθμο που εφαρμόστηκε.
Άρα έχουμε trade off μεταξύ ποιότητας και ταχύτητας.
Ο αλγόριθμος Baum Welch αποτελεί μια περίπτωση του Expectation-Maximization (EM) αλγορίθμου ο οποίος μας επιτρέπει να υπολογίσουμε την πλήρη υπό συνθήκη πιθανοφάνεια μιας σειράς καταστάσεων δεδομένης μιας ακολουθίας παρατηρήσεων εκπαιδεύοντας ταυτόχρονα και τα transition και emission probabilities του HMM. Αντιθέτως το Viterbi training αποτελεί μία περίπτωση pseudo-EM αλγορίθμου πραγματοποιώντας προσεγγιστικό MLE ώστε να κερδίσει χρόνο και να μειώσει την πολυπλοκότητα. Σε αντίθεση με τον Baum Welch δεν υπολογίζει την πλήρη υπο συνθήκη πιθανοφάνεια έχοντας μικρότερη ακρίβεια. Τέλος, ο αλγόριθμος Baum Welch συγκλίνει στην χειρότερη περίπτωση σε ένα τοπικό ελάχιστο σε αντίθεση με το Viterbi training που μπορεί να συγκλίνει σε ένα οποιοδήποτε σημείο.

Πίνακας Περιεχομένων

Άσκηση 2.3: Ερώτημα (α)

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.

Άσκηση 2.3: Ερώτημα (α)

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.

Άσκηση 2.3: Ερώτημα (α)

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.

Άσκηση 2.3: Ερώτημα (α)

- Οι δύο κατηγορίες ω_1 και ω_2 που θέλω να ταξινομήσω με αυτό το δέντρο απόφασης οι ικανοποιημένοι (SATISFACTION=Y) και οι μη ικανοποιημένοι (SATISFACTION=N) χρήστες
- Οι παράμετροι είναι η *WORD_ACCURACY*, η *TASK_COMPLETION* και η *TASK_DURATION*. Οι παράμετροι *WORD_ACCURACY* και *TASK_DURATION* είναι μη κατηγορηματικοί ενώ η παράμετρος *TASK_COMPLETION* είναι κατηγορηματική.
- Οι ερωτήσεις είναι οι εξής:
  - Is *WORD_ACCURACY* $\geq a\%$?
  - Do we have *TASK_COMPLETION*?
  - Is *TASK_DURATION* $\geq a$ min?

Άσκηση 2.3: Ερώτημα (β)

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).

Άσκηση 2.3: Ερώτημα (β)

Ερωτήσεις που θα μελετήσουμε:

- Q1: Do we have *TASK_COMPLETION*?
- Q2: Is *WORD_ACCURACY* $\geq$ 97.5%?
- Q3: Is *WORD_ACCURACY* $\geq$ 92.5%?
- Q4: Is *WORD_ACCURACY* $\geq$ 87.5%?
- Q5: Is *WORD_ACCURACY* $\geq$ 82.5%?
- Q6: Is *TASK_DURATION* $\geq$ 4.5 min?
- Q7: Is *TASK_DURATION* $\geq$ 3.5 min?
- Q8: Is *TASK_DURATION* $\geq$ 2.5 min?
- Q9: Is *TASK_DURATION* $\geq$ 1.5 min?

Επιλέγουμε να θέσουμε ως α τον μέσο όρο μεταξύ δύο γειτονικών μη κατηγορηματικών τιμών που δίνονται ως απάντηση στον πίνακα με τα δεδομένα εκπαίδευσης.

Άσκηση 2.3: Ερώτημα (β)

- Θέλουμε οι ερωτήσεις να τοποθετηθούν έτσι ώστε να έχουμε το βέλτιστο δέντρο CART. Αυτό το πετυχαίνουμε βάζοντας τις ερωτήσεις με χαμηλό impurity πιο κοντά στη ρίζα.
- Impurity είναι ένα κριτήριο που δείχνει πόσο μη-καθαρός είναι ένας κόμβος, δηλαδή πόσο αναμειγμένες έχει τις κλάσεις που.
- Συνήθης μετρική για το impurity είναι το entropy: $i(Node_\kappa) = -\sum_{j=1}^{2} P(\omega_j) \log_2 P(\omega_j)$ όπου $Node_\kappa$ είναι ο κόμβος που μελετάμε και $P(\omega_j)$ είναι η πιθανότητα ενός δείγματος του κόμβου $Node_\kappa$ να ανήκει στην κλάση ω_j. Άρα $P(\omega_j) = \frac{\#\omega_j}{\#\omega_1 + \#\omega_2}$
- Το Impurity μίας ερώτησης είναι: $i(Q) = \frac{N_1 \cdot i(Node_1) + N_2 \cdot i(Node_2)}{N_1 + N_2}$, όπου N_1 είναι το πλήθος των δειγμάτων ποου κατατάσονται στον $Node_1$ με βάση την ερώτηση Q.Για το N_2 αντίστοιχα.

Άσκηση 2.3: Ερώτημα (β)

Υπολογίζουμε τα impurity των ερωτήσεων έτσι ώστε να βρούμε την ερώτηση με το χαμηλότερο impurity για να τοποθετηθεί πρώτη.

Q	Node_1	$\text{i}(\text{Node}_1)$	Node_2	$\text{i}(\text{Node}_2)$	$i(\mathrm{Q})$
Q1	$(4\omega_1, 1\omega_2)$	0.722	$(2\omega_1, 2\omega_2)$	1.0	0.846
Q2	$(2\omega_1, 0\omega_2)$	0.0	$(4\omega_1, 3\omega_2)$	0.985	**0.766**
Q3	$(3\omega_1, 1\omega_2)$	0.811	$(3\omega_1, 2\omega_2)$	0.971	0.9
Q4	$(4\omega_1, 1\omega_2)$	0.722	$(2\omega_1, 2\omega_2)$	1.0	0.846
Q5	$(5\omega_1, 2\omega_2)$	0.863	$(1\omega_1, 1\omega_2)$	1.0	0.894
Q6	$(1\omega_1, 1\omega_2)$	1.0	$(5\omega_1, 2\omega_2)$	0.863	0.894
Q7	$(3\omega_1, 1\omega_2)$	0.811	$(3\omega_1, 2\omega_2)$	0.971	0.9
Q8	$(4\omega_1, 2\omega_2)$	0.918	$(2\omega_1, 1\omega_2)$	0.918	0.918
Q9	$(5\omega_1, 3\omega_2)$	0.954	$(1\omega_1, 0\omega_2)$	0.0	0.848

- Q2: Is $\mathit{WORD\_ACCURACY} \geq 97.5\%$?
- Ο κόμβος Node_1 έχει μηδενικό impurity
- Συνεχίζουμε με την μελέτη του κόμβου Node_2 με κατανομή $(4\omega_1, 3\omega_2)$

Άσκηση 2.3: Ερώτημα (β)

Συνεχίζουμε με την μελέτη του κόμβου Node_2 με κατανομή $(4\omega_1, 3\omega_2)$.

Q	Node$_1$	i(Node$_1$)	Node$_2$	i(Node$_2$)	i(Q)
Q1	$(2\omega_1, 1\omega_2)$	0.918	$(2\omega_1, 2\omega_2)$	1.0	0.965
Q3	$(1\omega_1, 1\omega_2)$	1.0	$(3\omega_1, 2\omega_2)$	0.971	0.979
Q4	$(2\omega_1, 1\omega_2)$	0.918	$(2\omega_1, 2\omega_2)$	1.0	0.965
Q5	$(3\omega_1, 2\omega_2)$	0.971	$(1\omega_1, 1\omega_2)$	1.0	0.979
Q6	$(1\omega_1, 1\omega_2)$	1.0	$(3\omega_1, 2\omega_2)$	0.971	0.979
Q7	$(3\omega_1, 1\omega_2)$	0.811	$(1\omega_1, 2\omega_2)$	0.918	**0.857**
Q8	$(3\omega_1, 2\omega_2)$	0.971	$(1\omega_1, 1\omega_2)$	1.0	0.979
Q9	$(3\omega_1, 3\omega_2)$	1.0	$(1\omega_1, 0\omega_2)$	0.0	**0.857**

- Q7: Is *TASK_DURATION* $\geq$ 3.5 min?
- Q9: Is *TASK_DURATION* $\geq$ 1.5 min?
- Ισοβαθμία μεταξύ των δύο ερωτήσεων. Εξερευνήσουμε και τις δύο περιπτώσεις.
- Περίπτωση 1: Επιλέγουμε Q9 και συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή $(3\omega_1, 3\omega_2)$.
- Περίπτωση 2: Επιλέγουμε Q7 και συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή $(1\omega_1, 2\omega_2)$.

Άσκηση 2.3: Ερώτημα (β)

Περίπτωση 1: Επιλέγουμε Q9 και συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή $(3\omega_1, 3\omega_2)$.

Q	Node$_1$	i(Node$_1$)	Node$_2$	i(Node$_2$)	i(Q)
Q1	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 2\omega_2)$	1.0	1.0
Q3	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 2\omega_2)$	1.0	1.0
Q4	$(2\omega_1, 1\omega_2)$	0.918	$(1\omega_1, 2\omega_2)$	0.918	0.918
Q5	$(3\omega_1, 2\omega_2)$	0.971	$(0\omega_1, 1\omega_2)$	0.0	0.809
Q6	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 2\omega_2)$	1.0	1.0
Q7	$(3\omega_1, 1\omega_2)$	0.811	$(0\omega_1, 2\omega_2)$	0.0	**0.541**
Q8	$(3\omega_1, 2\omega_2)$	0.971	$(0\omega_1, 1\omega_2)$	0.0	0.809

- Q7: Is *TASK_DURATION* $\geq$ 3.5 min?
- Ο κόμβος Node_2 έχει μηδενικό impurity
- Συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή $(3\omega_1, 1\omega_2)$

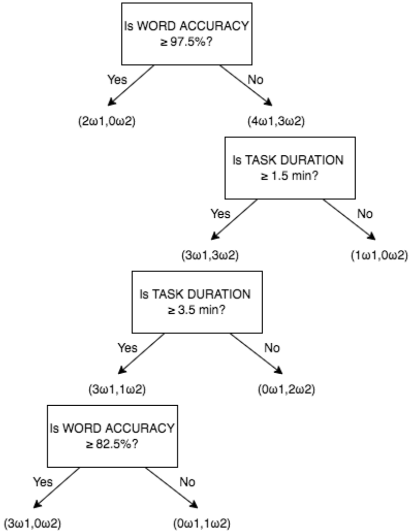
Άσκηση 2.3: Ερώτημα (β)

Μελετάμε την Περίπτωση 1 και συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή $(3\omega_1, 1\omega_2)$

Q	Node_1	$i(\mathrm{Node}_1)$	Node_2	$i(\mathrm{Node}_2)$	$i(\mathrm{Q})$
Q1	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 0\omega_2)$	0.0	0.5
Q3	$(1\omega_1, 0\omega_2)$	0.0	$(2\omega_1, 1\omega_2)$	0.918	0.689
Q4	$(2\omega_1, 0\omega_2)$	0.0	$(1\omega_1, 1\omega_2)$	1.0	0.5
Q5	$(3\omega_1, 0\omega_2)$	0.0	$(0\omega_1, 1\omega_2)$	0.0	**0.0**
Q6	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 0\omega_2)$	0.0	0.5

- Q5: Is *WORD_ACCURACY* $\geq$ 82.5%?
- Δίνει μηδενική εντροπία άρα έχει ολοκληρωθεί η ταξινόμηση

Περίπτωση 1:


Is WORD ACCURACY ≥ 97.5%?
Yes
No
(2ω1,0ω2)
(4ω1,3ω2)
Is TASK DURATION ≥ 1.5 min?
Yes
No
(3ω1,3ω2)
(1ω1,0ω2)
Is TASK DURATION ≥ 3.5 min?
Yes
No
(3ω1,1ω2)
(0ω1,2ω2)
Is WORD ACCURACY ≥ 82.5%?
Yes
No
(3ω1,0ω2)
(0ω1,1ω2)

Άσκηση 2.3: Ερώτημα (β)

Περίπτωση 2: Επιλέγουμε Q7 και συνεχίζουμε με την μελέτη του κόμβου Node_2 με κατανομή$(1\omega_1, 2\omega_2)$ και μετά τον κόμβο Node_1 με κατανομή$(3\omega_1, 1\omega_2)$.

Q	Node_1	i(Node_1)	Node_2	i(Node_2)	i(Q)
Q1	$(1\omega_1, 0\omega_2)$	0.0	$(0\omega_1, 2\omega_2)$	0.0	**0.0**
Q3	$(0\omega_1, 1\omega_2)$	0.0	$(1\omega_1, 1\omega_2)$	1.0	0.667
Q4	$(0\omega_1, 1\omega_2)$	0.0	$(1\omega_1, 1\omega_2)$	1.0	0.667
Q5	$(0\omega_1, 2\omega_2)$	0.0	$(1\omega_1, 0\omega_2)$	0.0	**0.0**
Q8	$(0\omega_1, 1\omega_2)$	0.0	$(1\omega_1, 1\omega_2)$	1.0	0.667
Q9	$(0\omega_1, 2\omega_2)$	0.0	$(1\omega_1, 0\omega_2)$	0.0	**0.0**

- Οι ερωτήσεις Q1, Q5 και Q9 είναι ισοδύναμες και έχουν μηδενική εντροπία.
- Επιλέγουμε την Q1: Do we have *TASK_COMPLETION*?
- Δίνει μηδενική εντροπία άρα έχει ολοκληρωθεί η ταξινόμηση.

Άσκηση 2.3: Ερώτημα (β)

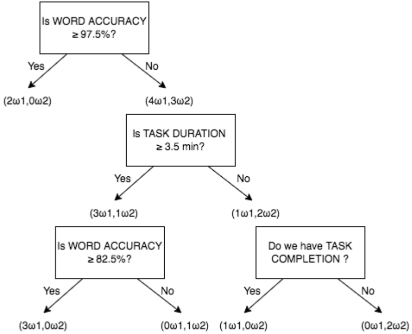
Περίπτωση 2: Επιλέγουμε Q7 και συνεχίζουμε με την μελέτη του κόμβου Node_1 με κατανομή$(3\omega_1, 1\omega_2)$.

Q	Node_1	$\mathrm{i}(\mathrm{Node}_1)$	Node_2	$\mathrm{i}(\mathrm{Node}_2)$	$i(\mathrm{Q})$
Q1	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 0\omega_2)$	0.0	0.5
Q3	$(1\omega_1, 0\omega_2)$	0.0	$(2\omega_1, 1\omega_2)$	0.918	0.689
Q4	$(2\omega_1, 0\omega_2)$	0.0	$(1\omega_1, 1\omega_2)$	1.0	0.5
Q5	$(3\omega_1, 0\omega_2)$	0.0	$(0\omega_1, 1\omega_2)$	0.0	**0.0**
Q6	$(1\omega_1, 1\omega_2)$	1.0	$(2\omega_1, 0\omega_2)$	0.0	0.5

- Q5: Is *WORD_ACCURACY* $\geq$ 82.5%?
- Δίνει μηδενική εντροπία άρα έχει ολοκληρωθεί η ταξινόμηση.

Άσκηση 2.3: Ερώτημα (β)

Περίπτωση 2:



Τα δύο δένδρα είναι ισοδύναμα καλά, με ίδιο πλήθος ερωτήσεων.

Άσκηση 2.3: Ερώτημα (γ)

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 2.3: Ερώτημα (γ)

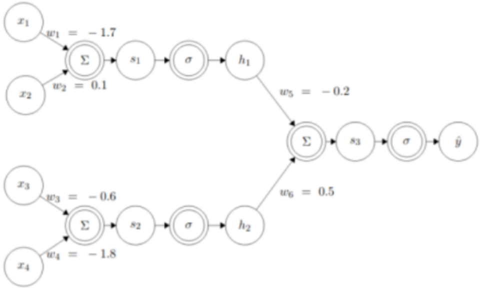
Μελετώντας τα δυο προτεινόμενα δέντρα μπορούμε να δούμε τα εξής:

- Η πρώτη και πιο σημαντική ερώτηση σχετίζεται με την παράμετρο *WORD_ACCURACY*
- Η δεύτερη πιο σημαντική ερώτηση σχετίζεται με την παράμετρο *TASK_DURATION*
- Η τρίτη και λιγότερο σημαντική ερώτηση σχετίζεται με την παράμετρο *TASK_COMPLETION* και εμφανίζεται μόνο στο δεύτερο δέντρο.
- Άρα η ικανοποιηση του συνομιλητη απο το συστημα διαλογου επιρεάζεται πρώτα από την ποιοτητα του αναγνωριστη φωνης, μετά από την διαρκεια της συνομιλιας και τέλος από την επιτυχια της συναλλαγης. Με βάση τα συγκεκριμένα δεδομένα εκπαίδευσης.

Πίνακας Περιεχομένων

Άσκηση 2.4

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 2.4

Forward Pass:

$$s_1 = w_1 \cdot x_1 + w_2 \cdot x_2 = 0.99$$

$$h_1 = \sigma(s_1) = 0.7291$$

$$s_2 = w_3 \cdot x_3 + w_4 \cdot x_4 = 4.86$$

$$h_2 = \sigma(s_2) = 0.9923$$

$$s_3 = w_5 \cdot h_1 + w_6 \cdot h_2 = 0.3503$$

$$\hat{y} = \sigma(s_3) = 0.5867$$

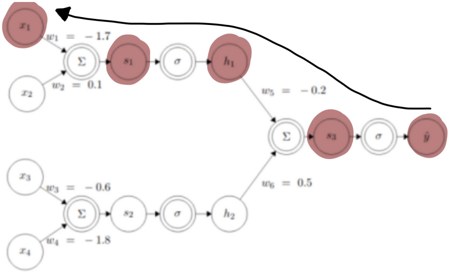
Ξέρουμε ότι:

$$\frac{\partial \sigma(x)}{\partial x} = \sigma(x) \cdot [1 - \sigma(x)]$$

$$\frac{\partial L}{\partial \hat{y}} = 2 \left\| y - \hat{y} \right\|$$

Άσκηση 2.4: Ερώτημα (β)

Backward Pass:



$$\frac{\partial L}{\partial w_1} =$$

$$\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_1} =$$

$$\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial w_1} =$$

$$\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial w_1} =$$

$$\frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1}$$

Άσκηση 2.4

Υπολογίζουμε τις μερικές παραγώγους που εμπεριέχονται στην σχέση:

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial s_3} \cdot \frac{\partial s_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial s_1} \cdot \frac{\partial s_1}{\partial w_1}$$

$$\frac{\partial s_1}{\partial w_1} = \frac{\partial (w_1 \cdot x_1 + w_2 \cdot x_2)}{\partial w_1} = x_1 = -0.5$$

$$\frac{\partial h_1}{\partial s_1} = h_1 \cdot (1 - h_1) = 0.1975$$

$$\frac{\partial s_3}{\partial h_1} = \frac{\partial (w_5 \cdot h_1 + w_6 \cdot h_2)}{\partial h_1} = w_5 = -0.2$$

$$\frac{\partial \hat{y}}{\partial s_3} = \hat{y} \cdot (1 - \hat{y}) = 0.2276$$

$$\frac{\partial L}{\partial \hat{y}} = 2 \left\| y - \hat{y} \right\| = 0.1734$$

$$\frac{\partial L}{\partial w_1} = 0.0008$$

Πίνακας Περιεχομένων

Άσκηση 2.5: Ερώτημα (α)

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:

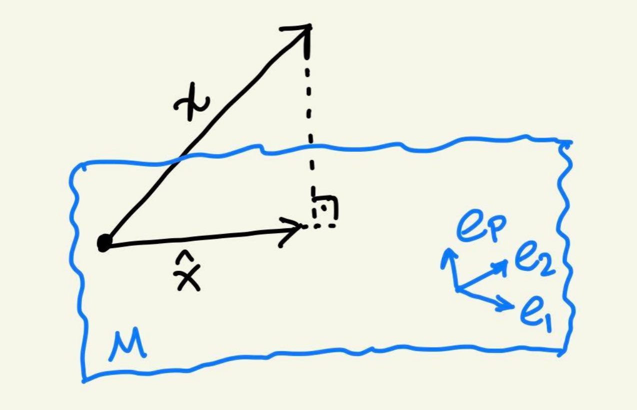
$$\hat{\mathbf{x}} = \sum_{k=1}^{p} y_k \mathbf{e}_k, \quad y_k = \mathbf{e}_k^H \mathbf{x} \tag{1}$$

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

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.

Άσκηση 2.5: Ερώτημα (α)



Άσκηση 2.5: Ερώτημα (α)

Το μέσο τετραγωνικό λάθος δεν το εκφράζω ντετερμινιστικά αλλά με βάση την πιθανότητα το x να είναι ένα τυχαίο διάνυσμα και μελετώ την μέση στατιστική τιμή της διαφοράς της ευκλείδειας νόρμας του x και του $\hat{x}$.
Θέλω να προσεγγίσει το x με το $\hat{x}$ έτσι ώστε η νόρμα της διαφοράς να είναι ελάχιστη, δηλαδή το διάνυσμα που συνδέει το x με το $\hat{x}$. Με βάση την επέκταση του Πυθαγορείου θεωρήματος σε αφηρημένου χώρους Hilbert, ξέρουμε ότι η νορμα της διαφοράς θα είναι ελάχιστη όταν το $x - \hat{x}$ είναι κάθετο στο χωρο M.
Με βάση το Πυθαγόρειο Θεώρημα σε αφηρημένου χώρους Hilbert η ενέργεια της διαφοράς θα είναι η διαφορά της ενέργειας του αρχικού διανύσματος μείων την ενέργεια της προβολής του.

$$J = E\left[\|x - \hat{x}\|^2\right] = E\left[\|x\|^2\right] - E\left[\|\hat{x}\|^2\right]$$

Επειδή η βάση e_i είναι ορθοκανονική, η ενέργεια της προσέγγισης γράφεται ως:

$$E\left[\|\hat{x}\|^2\right] = \sum_{k=1}^{p} E\left[\|y_k\|^2\right]$$

Η ενέργεια του x είναι το άθροισμα της ενέργειας των προβολών του:

$$E\left[\|x\|^2\right] = \sum_{k=1}^{d} E\left[\|y_k\|^2\right]$$

Άσκηση 2.5: Ερώτημα (α)

Άρα $J = \sum_{k=p+1}^{d} E\left[\|y_k\|^2\right]$
Αφού $y_i = e_i^H x$ το τετράγωνο της προβολής του y_i που εν γένει είναι μιγαδικός αριθμός μπορεί να γραφεί ως: $\|y_k\|^2 = y_i y_i^* = e_i^H x x^H e_i$
Αφού σε αυτή τη σχέση το x καιιx^H είναι στοχαστικά, η μέση στατιστική τιμή αυτού είναι: $E\left[\|y_k\|^2\right] = e_i^H E\left[x x^H\right] e_i = e_i^H R_x e_i$ και τότε:

$$J = \sum_{i=p+1}^{d} e_i^H R_x e_i$$

Η J γίνεται ελάχιστη αν διαλέξουμε τα e_i να είναι τα ιδιοδιανύσματα του πίνακα αυτοσυσχέτισης: $R_x e_i = \lambda_i e_i$ Απόδειξη: Ελαχιστοποίηση του J ως προς το e_i με το περιορισμό: $e_i^\top e_i = 1$, με πολλαπλασιαστή Lagrange κ.λ.π.
Ο R_x είναι πίνακας συσχέτισης και επομένως είναι θετικά ημι-ορισμένος, δηλαδή είναι συμμετρικός πίνακας, και όλες οι ιδιοτιμές του είναι πραγματικές και μη αρνητικές. Άρα για τις ιδιοτιμές λ: $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_d$ (φθίνουσα σειρά)Άρα

$$J = \sum_{i=p+1}^{d} e_i^H R_x e_i = \sum_{i=p+1}^{d} e_i^H \lambda_i e_i = \sum_{i=p+1}^{d} \lambda_i$$

Το J γίνεται ελάχιστό όταν οι πρώτες p είναι οι μεγαλύτερες, λi μη αρνητικές.

Άσκηση 2.5: Ερώτημα (α)

- Βασικό βήμα: Έστω ότι ισχύει για p=1
- Επαγωγική Υπόθεση: Έστω ότι ισχύει για $p = k$
- Επαγωγικό βήμα: Θα δείξουμε ότι ισχύει για $p = k+1$

Άσκηση 2.5: Ερώτημα (α)

Λύση Άσκησης:

- Στο επαγωγικό βήμα θα έχουμε $\hat{x} = \sum_{k=1}^{p} y_k e_k + y_{p+1}\nu$
- Η μετάβαση από το p = k στο p = k+1 οδηγεί σε μείωση του μέσου στοχαστικού σφάλματος κατά μια ποσότητα $\mu = \nu^H R_x \nu$.
- Για να δώσει το ν την μέγιστη διόρθωση πρέπει να ελαχιστοποιησουμε την Λαγκρανζιανή: $L = \nu^H R_x \nu + \lambda_\nu (1 - \nu^H \nu)$ αφού $\nu^H \nu = 1$.
- Από την Λαγκρανζιανή βρίσκουμε ότι $R_x \nu = \lambda_\nu \nu$.
- Στο Επαγωγικό βήμα το στοχαστικό σφάλμα θα είναι: $J = \sum_{i=p+1}^{d} \lambda_i - \nu^H R_x \nu = \sum_{i=p+1}^{d} \lambda_i - \lambda_\nu$.
- Άρα το J γίνεται ελάχιστο όταν $\lambda_\nu = \lambda_{p+1}$ (φθίνουσα ταξινόμηση) και άρα:

$$\lambda_\nu = \lambda_{p+1}$$

$$\nu = e_{p+1}$$

$$\hat{x} = \sum_{k=1}^{p+1} y_k e_k$$

$$J = \sum_{i=p+2}^{d} \lambda_i$$

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.

Ισχύει ότι:

$$\frac{\partial}{\partial \mathbf{X}}[\mathrm{trace}(\mathbf{XA})] = \mathbf{A}^\top$$

Άσκηση 2.5: Ερώτημα (β)

$$\tilde{J} = \text{trace}\left(\mathbf{U}^H\mathbf{R}\mathbf{U}\right) + \text{trace}\left[\mathbf{H}\left(\mathbb{I} - \mathbf{U}^H\mathbf{U}\right)\right] = \text{trace}\left(\mathbf{U}^H\mathbf{R}\mathbf{U}\right) - \text{trace}\left(\mathbf{H}\mathbf{U}^H\mathbf{U}\right) + \text{trace}(\mathbf{H})$$

- Ελαχιστοποίηση J ως πρός U.

$$\frac{\partial}{\partial \mathbf{U}}\left[\text{trace}\left(\mathbf{U}^H\mathbf{R}\mathbf{U}\right)\right] - \frac{\partial}{\partial \mathbf{U}}\left[\text{trace}\left(\mathbf{H}\mathbf{U}^H\mathbf{U}\right)\right] =$$

$$\frac{\partial}{\partial \mathbf{U}}\left[\text{trace}\left(\mathbf{U}\mathbf{U}^H\mathbf{R}\right)\right] - \frac{\partial}{\partial \mathbf{U}}\left[\text{trace}\left(\mathbf{U}\mathbf{H}\mathbf{U}^H\right)\right] =$$

$$\left(\mathbf{U}^H\mathbf{R}\right)^\top - \left(\mathbf{H}\mathbf{U}^H\right)^\top = 0 \Leftrightarrow \mathbf{U}^H\mathbf{R} = \mathbf{H}\mathbf{U}^H \Leftrightarrow$$

$$\mathbf{H} = \mathbf{U}^H\mathbf{R}\mathbf{U}$$

- Ο H είναι ερμιτιανός.

$$\mathbf{R}\mathbf{U} = \mathbf{U}\mathbf{H}$$

Άσκηση 2.5: Ερώτημα (β)

$$\tilde{J} = \text{trace}\left(\mathbf{U}^H\mathbf{U}\mathbf{H}\right) - \text{trace}\left(\mathbf{H}\mathbf{U}^H\mathbf{U}\right) + \text{trace}(\mathbf{H}) = \text{trace}(\mathbf{H})$$

- Μια πιθανή λύση αποκτάται όταν οι στήλες του $\mathbf{U}$ είναι ιδιοδιανύσματα του πίνακα $\mathbf{R}_x$. Τότε ο $\mathbf{H}$ είναι ένας διαγώνιος πίνακας που περιέχει τις αντίστοιχες ιδιοτιμές.
- Στην περίπτωση που ο $\mathbf{H}$ δεν είναι διαγώνιος αλλά είναι ερμιτιανός μπορεί να παραγοντοποιηθεί ως $\mathbf{H} = \mathbf{F}^H\mathbf{\Lambda}\mathbf{F}$ όπου $\mathbf{F}$ ο πίνακας ιδιοδιανυσμάτων του $\mathbf{H}$. Ελαχιστοποιόντας το $\tilde{J}$ προκύπτει το ίδιο αποτέλεσμα

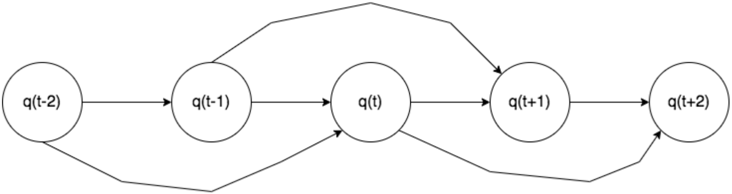
Πίνακας Περιεχομένων

Άσκηση 2.6: Ερώτημα (α)

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

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

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



Άσκηση 2.6: Ερώτημα (β)

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 ΜΜ–2 χρησιμοποιώντας την αρχή της μεγιστοποίησης της πιθανότητας των παρατηρήσεων Maximum Likelihood. Δίνεται ότι:

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

Άσκηση 2.6: Ερώτημα (β)

- (1,1,1,2,1,2,2,2,2,1,2,2,2,2,1,1,1,1,2)

$$\alpha_{ijk} = p\left(q_t = k \mid q_{t-1} = j, q_{t-2} = i\right) = \frac{\#(S_i, S_j \to S_k)}{\#(S_i, S_j \to *)}$$

$$p\left(q_t = 1 \mid q_{t-1} = 1, q_{t-2} = 1\right) = \frac{N_{111}}{N_{111} + N_{112}} = \frac{3}{5}$$

$$p\left(q_t = 1 \mid q_{t-1} = 1, q_{t-2} = 2\right) = \frac{N_{211}}{N_{211} + N_{212}} = \frac{1}{3}$$

$$p\left(q_t = 1 \mid q_{t-1} = 2, q_{t-2} = 1\right) = \frac{N_{121}}{N_{121} + N_{122}} = \frac{1}{3}$$

$$p\left(q_t = 1 \mid q_{t-1} = 2, q_{t-2} = 2\right) = \frac{N_{221}}{N_{221} + N_{222}} = \frac{2}{6} = \frac{1}{3}$$

$$p\left(q_t = 2 \mid q_{t-1} = 1, q_{t-2} = 1\right) = 1 - \frac{3}{5} = \frac{2}{5}$$

$$p\left(q_t = 2 \mid q_{t-1} = 1, q_{t-2} = 2\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$p\left(q_t = 2 \mid q_{t-1} = 2, q_{t-2} = 1\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$p\left(q_t = 2 \mid q_{t-1} = 2, q_{t-2} = 2\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

Άσκηση 2.6: Ερώτημα (β)

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

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

Άσκηση 2.6: Ερώτημα (γ)

$$p_1 = p\left(q_1 = 1, q_2 = 1, \ldots, q_{10} = 1, q_{11} = 2 \mid q_0 = 1\right)$$

$$= \frac{p\left(q_0 = 1, q_1 = 1, q_2 = 1, \ldots, q_{10} = 1, q_{11} = 2\right)}{p\left(q_0 = 1\right)} = \frac{\pi_1 \cdot \pi_{11} \cdot \overbrace{\alpha_{111} \cdot \ldots \cdot \alpha_{111}}^{9} \cdot \alpha_{112}}{\pi_1}$$

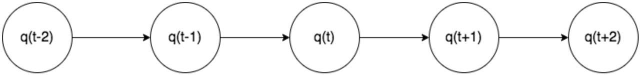
$$= \pi_{11} \cdot {\alpha_{111}}^9 \cdot \alpha_{112} = \underbrace{\pi_{11}}_{0.25} \cdot \left(\frac{3}{5}\right)^9 \cdot \frac{2}{5} \simeq 0.001$$

Άσκηση 2.6: Ερώτημα (δ)

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 2.6: Ερώτημα (δ)

- (1,1,1,2,1,2,2,2,2,1,2,2,2,2,1,1,1,1,2)

$$\alpha_{ij} = p\left(q_t = j \mid q_{t-1} = i\right) = \frac{\#(S_i \to S_j)}{\#(S_i \to *)}$$

$$\alpha_{11} = p\left(q_t = 1 \mid q_{t-1} = 1\right) = \frac{N_{11}}{N_{11} + N_{12}} = \frac{5}{9}$$

$$\alpha_{22} = p\left(q_t = 2 \mid q_{t-1} = 2\right) = \frac{N_{22}}{N_{22} + N_{21}} = \frac{2}{3}$$

$$\alpha_{12} = p\left(q_t = 2 \mid q_{t-1} = 1\right) = 1 - \frac{5}{9} = \frac{4}{9}$$

$$\alpha_{21} = p\left(q_t = 1 \mid q_{t-1} = 2\right) = 1 - \frac{2}{3} = \frac{1}{3}$$

Άσκηση 2.6: Ερώτημα (δ)

$$p_2 = p\left(q_1 = 1, q_2 = 1, \ldots, q_{10} = 1, q_{11} = 2 \mid q_0 = 1\right)$$
$$= \frac{p\left(q_0 = 1, q_1 = 1, q_2 = 1, \ldots, q_{10} = 1, q_{11} = 2\right)}{p\left(q_0 = 1\right)} = \frac{\pi_1 \cdot \overbrace{\alpha_{11} \cdot \ldots \cdot \alpha_{11}}^{10} \cdot \alpha_{12}}{\pi_1}$$
$$= {\alpha_{11}}^{10} \cdot \alpha_{12} = \left(\frac{5}{9}\right)^{10} \cdot \frac{4}{9} \simeq 0.00124$$

- Παρατηρούμε οτι η νέα τιμή της πιθανότητας της ακολουθίας είναι λίγο μεγαλύτερη σε σχέση με το MM-2 μοντέλο, αλλά ανήκει στην ίδια τάξη μεγέθους.
- Το MM-2 βγάζει πιό ρεαλιστικά αποτελέσματα από το απλοϊκό μοντέλο MM-1

Πίνακας Περιεχομένων

Άσκηση 2.7: Ερώτημα (α)

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 $S_B = P(\omega_1)P(\omega_2)(\vec{\mu_2}-\vec{\mu_1})(\vec{\mu_2}-\vec{\mu_1})^T$.

Άσκηση 2.7: Ερώτημα (α)

Στο πρόβλημα με τις 2 κλάσεις ισχύει $\mid \text{Classes} \mid = 2$, $P(\omega_1) + P(\omega_2) = 1$ και το S_B δίνεται από τη σχέση:

$$S_B = P(\omega_1)(\mu_1 - \mu)(\mu_1 - \mu)^T + P(\omega_2)(\mu_2 - \mu)(\mu_2 - \mu)^T$$

Αφού $\mu = \sum_{i=1}^{\mid \text{Classes} \mid} P(\omega_i)\mu_i = P(\omega_1)\mu_1 + P(\omega_2)\mu_2$ έχουμε ότι :

$$\mu_1 - \mu = \mu_1 - P(\omega_1)\mu_1 - P(\omega_2)\mu_2 = (1 - P(\omega_1))\mu_1 - P(\omega_2)\mu_2 = P(\omega_2)(\mu_1 - \mu_2)$$
$$\mu_2 - \mu = -P(\omega_1)\mu_1 + (1 - P(\omega_2))\mu_2 = P(\omega_1)(\mu_2 - \mu_1)$$

Χρησιμοποιόντας τις προηγούμενες εξισώσεις το S_B εκφράζεται ως:

$$\begin{aligned} S_B &= P(\omega_1)P^2(\omega_2)(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T + P(\omega_2)P^2(\omega_1)(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \\ &= P(\omega_1)P(\omega_2)(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T = P(\omega_1)P(\omega_2)(\mu_2 - \mu_1)(\mu_2 - \mu_1)^T \end{aligned}$$

Άσκηση 2.7: Ερώτημα (β)

(β) Βασιζόμενοι στο υποερώτημα **(α)**, να βρείτε το ιδιοδιάνυσμα του πίνακα $S_W^{-1}S_B$ και την ιδιοτιμή του.

Άσκηση 2.7: Ερώτημα (β)

Ψάχνουμε το ιδιοδιανυσμα και την ιδιοτιμή του $S_w^{-1} S_B$.
Έστω u το ιδιοδιάνυσμα και λ η ιδιοτιμή $S_w^{-1} S_B u = \lambda u$.

$$S_w^{-1} S_B u = P(\omega_1) P(\omega_2) S_w^{-1} (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T u$$

$$= \left[P(\omega_1) P(\omega_2) (\mu_2 - \mu_1)^T u \right] S_w^{-1} (\mu_2 - \mu_1) \Leftrightarrow$$

$$S_w^{-1} S_B u = \left[P(\omega_1) P(\omega_2) (\mu_2 - \mu_1)^T u \right] S_w^{-1} (\mu_2 - \mu_1)$$

Άρα:

$$u = S_w^{-1} (\mu_2 - \mu_1)$$

$$\lambda = P(\omega_1) P(\omega_2) (\mu_2 - \mu_1)^T S_w^{-1} (\mu_2 - \mu_1)$$

Πίνακας Περιεχομένων

Άσκηση 2.8: Ερώτημα (α)

(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:

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

Υπενθυμίζουμε τις ιδιότητες της μέσης τιμής:

- $\mathbb{E}[c_1 X + c_2 X] = c_1 \mathbb{E}[X] + c_2 \mathbb{E}[Y]$
- Αν X, Y στοχαστικώς ανεξάρτητες τ.μ. τότε: $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Άσκηση 2.8: Ερώτημα (α)

$$\mathrm{kurt}(x+y) = \mathbb{E}\left[(x+y)^4\right] - 3\left(\mathbb{E}\left[(x+y)^2\right]\right)^2$$

$$\begin{aligned}
\mathbb{E}\left[(x+y)^4\right] &= \mathbb{E}\left[x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4\right] \\
&= \mathbb{E}\left[x^4\right] + 4\mathbb{E}\left[x^3\right]\underbrace{\mathbb{E}[y]}_{=0} + 6\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right] + 4\underbrace{\mathbb{E}[x]}_{=0}\mathbb{E}\left[y^3\right] + \mathbb{E}\left[y^4\right] \\
&= \mathbb{E}\left[x^4\right] + \mathbb{E}\left[y^4\right] + 6\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right]
\end{aligned}$$

$$\begin{aligned}
-3\left(\mathbb{E}\left[(x+y)^2\right]\right)^2 &= -3\left(\mathbb{E}\left[x^2 + 2xy + y^2\right]\right)^2 = -3(\mathbb{E}\left[x^2\right] + 2\underbrace{\mathbb{E}[x]\mathbb{E}[y]}_{=0} + \mathbb{E}\left[y^2\right])^2 \\
&= -3\left(\mathbb{E}\left[x^2\right]\right)^2 - 6\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right] - 3\left(\mathbb{E}\left[y^2\right]\right)^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{kurt}(x+y) &= \mathbb{E}\left[(x+y)^4\right] - 3\left(\mathbb{E}\left[(x+y)^2\right]\right)^2 \\
&= \mathbb{E}\left[x^4\right] + \mathbb{E}\left[y^4\right] + 6\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right] - 3\left(\mathbb{E}\left[x^2\right]\right)^2 - 6\mathbb{E}\left[x^2\right]\mathbb{E}\left[y^2\right] - 3\left(\mathbb{E}\left[y^2\right]\right)^2 \\
&= \mathbb{E}\left[x^4\right] - 3\left(\mathbb{E}\left[x^2\right]\right)^2 + \mathbb{E}\left[y^4\right] - 3\left(\mathbb{E}\left[y^2\right]\right)^2 \\
&= \mathrm{kurt}(x) + \mathrm{kurt}(y)
\end{aligned}$$

Άσκηση 2.8: Ερώτημα (β1)

(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.

- $\mathbb{E}\left[s_1\right] = 0$
- $\mathbb{E}\left[s_1^2\right] = 1$

Άσκηση 2.7: Ερώτημα (β1)

Έχουμε ότι:

$$\mathbb{E}\left[x\right] = \sum_{i=1}^{N} w_i \mathbb{E}\left[s_i\right] = w_i \sum_{i=1}^{N} \mathbb{E}\left[s_i\right] = 0$$

$$\mathbb{E}\left[s_i s_j\right] = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$$

Με βάση αυτά εκφράζουμε την διασπορά ως:

$$\mathrm{Var}[x] = \mathbb{E}\left[x^2\right] - \underbrace{(\mathbb{E}[x])^2}_{=0} = \mathbb{E}\left[\sum_{i=1}^{N} w_i s_i \sum_{j=1}^{N} w_j s_j\right] = \mathbb{E}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j s_i s_j\right]$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \mathbb{E}\left[s_i s_j\right] = \sum_{i=1}^{N} {w_i}^2$$

Άρα:

$$\sum_{i=1}^{N} w_i^2 = 1$$

Άσκηση 2.8: Ερώτημα (β2)

(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 2.8: Ερώτημα (β2)

$$\mathrm{Var}[x] = 1 \Rightarrow \sum_{i=1}^{N} w_i^2 = 1 \Rightarrow \sum_{i=1}^{N} w^2 = 1 \Rightarrow w = \frac{1}{\sqrt{N}}$$

$$-\alpha \leq \mathrm{kurt}\,(s_i) \leq \alpha$$

$$|\,\mathrm{kurt}(x)| = \left|\mathrm{kurt}\left(\sum_{i=1}^{N} w_i s_i\right)\right| = \left|\mathrm{kurt}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} s_i\right)\right| = \frac{1}{N^2}\left|\sum_{i=1}^{N} \mathrm{kurt}\,(s_i)\right|$$

$$\leq \frac{1}{N^2}\sum_{i=1}^{N} \alpha = \frac{\alpha}{N} \Leftrightarrow -\frac{\alpha}{N} \leq \mathrm{kurt}(x) \leq \frac{\alpha}{N}$$

Έχουμε ότι: $\lim_{N\to\infty}\left(-\frac{a}{N}\right) = \lim_{N\to\infty}\left(\frac{a}{N}\right) = 0$

Άρα από κριτήριο παρεμβολής: $\lim_{N\to\infty}[\mathrm{kurt}(x)] = 0$

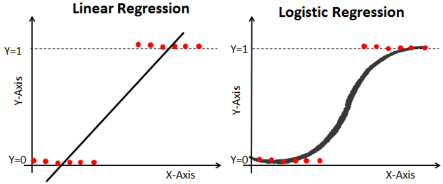
Πίνακας Περιεχομένων

Άσκηση 2.9: Ερώτημα (α)

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 $n = \{1, 2, \dots, N\}$. 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 $\mathbf{w}$.



Άσκηση 2.9: Ερώτημα (α)

Έστω το σύνολο δεδομένων $\{\varphi_n, t_n\}$ με $t_n \epsilon \{0, 1\}$ μεγέθους N που ακολουθεί σ.π.π $f(n, w) = y_n^{t_n}(1 - y_n)^{t_n}$.
Η πιθανοφάνεια αυτου είναι:

$$\ell(\mathbf{w}) = \prod_{n=1}^{N} y_n^{t_n} (1 - y_n)^{1-t_n}$$

Λογαριθμίζοντας την πιθανοφάνεια έχουμε:

$$\begin{aligned}
\ln \ell(\mathbf{w}) &= \ln \prod_{n=1}^{N} y_n^{t_n} (1 - y_n)^{1-t_n} = \sum_{n=1}^{N} \ln \left[y_n^{t_n} (1 - y_n)^{1-t_n} \right] \\
&= \sum_{n=1}^{N} \left[\ln y_n^{t_n} + \ln (1 - y_n)^{1-t_n} \right] \\
&= \sum_{n=1}^{N} \left[t_n \ln y_n + (1 - t_n) \ln (1 - y_n) \right] = -E(\mathbf{w})
\end{aligned}$$

Άρα η μεγιστοποίηση της πιθανοφανειας ισοδυναμεί με την ελαχιστοποίηση της συνάρτησης cross-entropy.

Άσκηση 2.9: Ερώτημα (α)

$$-\nabla E(\mathbf{w}) = \sum_{n=1}^{N} \nabla \left[t_n \ln y_n + (1 - t_n) \ln (1 - y_n) \right]$$
$$= \sum_{n=1}^{N} \left(\frac{t_n}{y_n} - \frac{1 - t_n}{1 - y_n} \right) \nabla y_n = \sum_{n=1}^{N} \left(\frac{t_n}{y_n} - \frac{1 - t_n}{1 - y_n} \right) y_n (1 - y_n) \phi_n$$
$$= \sum_{n=1}^{N} \frac{t_n - y_n}{y_n (1 - y_n)} y_n (1 - y_n) \phi_n = \sum_{n=1}^{N} (t_n - y_n) \phi_n$$

Ετσι, η μεγιστοποίηση της πιθανοφάνειας βρίσκεται από την εξίσωση:

$$-\nabla E(\mathbf{w}) = 0 \Rightarrow y_n = t_n \Rightarrow \sigma(w^\top \varphi_n) = \begin{cases} 1, t_n = 1 \\ 0, t_n = 0 \end{cases} \quad \Rightarrow w^\top \varphi_n = \begin{cases} +\infty, t_n = 1 \\ -\infty, t_n = 0 \end{cases}$$

Άρα το μέτρο του διανύσματος w απειρίζεται.

Άσκηση 2.9: Ερώτημα (β)

(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.

Άσκηση 2.9: Ερώτημα (β)

- $R = R_s R_s$ με$R_s = \sqrt{y_n(1 - y_n)}$
- $R = R_s^\top R_s$ αφού R_s διαγώνιος
- $H = \Phi^\top H \Phi = \Phi^\top R_s^\top R_s \Phi = (\Phi R_s)^\top (\Phi R_s)$

Για μη μηδενικό διάνυσμα v έχουμε:

$$\mathrm{x}^\top \cdot H \cdot x = x^\top \left(R_s \Phi\right)^\top \cdot R_s \cdot \Phi \cdot x = \|R_s \cdot \Phi \cdot x\|^2 \geq 0$$

- Επειδή ο αριθμός των δειγμάτων είναι πολύ μεγαλύτερος από τον αριθμό των διαστάσεων οι στήλες του πίνακα Φ είναι γραμμικά ανεξάρτητες $\Phi \cdot x \neq 0$

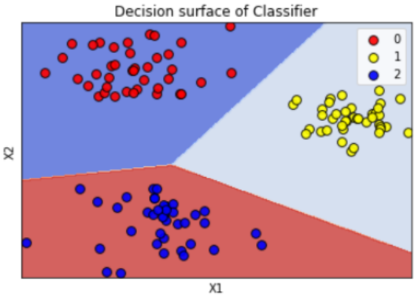
$\|R_s \cdot \Phi \cdot x\|^2 > 0$
Συνεπώς ο πίνακας H ειναι θετικά ορισμένος. Αν ο Hessian πίνακας μίας συνάρτησης είναι θετικά ορισμένος τότε η συνάρτηση είναι (αυστηρά) κυρτή από το οποίο μπορούμε να συνάγουμε ότι το ελάχιστο της θα είναι μοναδικό.

Άσκηση 2.9: Ερώτημα (γ)

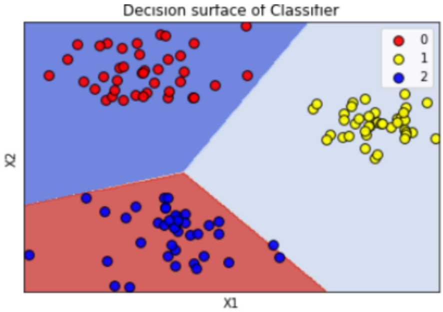
(γ) Να γράψετε κώδικα που θα υλοποιεί τον iterative reweighted least squares (IWLS) αλγόριθμο για logistic regression. Χρησιμοποιώντας τον αλγόριθμο αυτό, να υπολογίσετε και να σχεδιάσετε τα διαχωριστικά επίπεδα απόφασης που αντιστοιχούν στο σύνολο δεδομένων του προβλήματος τριών κλάσεων που έχει ανεβεί 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 2.9: Ερώτημα (γ)

- Logistic Regression (καλύτερα αποτελέσματα)

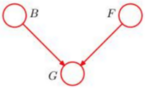


- Ταξινόμηση με βάση τα Ελάχιστα Τετράγωνα



Πίνακας Περιεχομένων

Άσκηση 2.10: Ερώτημα (α)



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

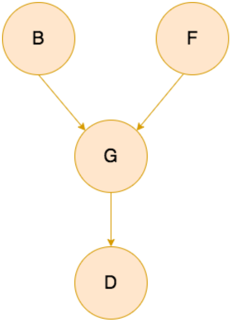
Θεωρείστε το σύστημα καυσίμων ενός αυτοκινήτου, όπως φαίνεται στο Σχήμα 1, με γνωστές τις παρακάτω πιθανότητες:

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

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:

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

Άσκηση 2.10: Ερώτημα (α)



B	0	1
p(B)	0.05	0.95

F	0	1
p(F)	0.2	0.8

B, F \ G	0	1
(0,0)	0.8	0.2
(0,1)	0.75	0.25
(1,0)	0.7	0.3
(1,1)	0.05	0.95

G \ D	0	1
0	0.8	0.2
1	0.2	0.8

- B και F ειναι ανεξάρτητα
- B και F δεν είναι ανεξαρτητα δεδομένου του G

Άσκηση 2.10: Ερώτημα (α)

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.

Άσκηση 2.10: Ερώτημα (α)

$$
\begin{aligned}
p(F=0 \mid D=0) &= \frac{p(D=0 \mid F=0)\cdot p(F=0)}{p(D=0)} = \frac{p(F=0)}{p(D=0)} \sum_G p(D=0, G \mid F=0) \\
&= \frac{p(F=0)}{p(D=0)} \sum_G p(D=0 \mid G, F=0)\cdot p(G \mid F=0) \\
&= \frac{p(F=0)}{p(D=0)} \sum_G p(D=0 \mid G)\cdot p(G \mid F=0)
\end{aligned}
$$

- Κανόνας του Bayes
- Κανόνας δεσμευμένης πιθανότητας
- Μαρκοβιανή ιδιότητα $p(D \mid G, F) = p(D \mid G)$

Άσκηση 2.10: Ερώτημα (α)

$$p(G=0)=\sum_{B,F} p(G=0,B,F)=\sum_{B,F} p(G=0\mid B,F)p(B\mid F)p(F)$$
$$=\sum_{B,F} p(G=0\mid B,F)p(B)p(F)=0.008+0.133+0.03+0.038=0.209$$

$$p(G=0)=0.6746$$

- B,F ανεξάρτητα

$$p(G=0\mid F=0)=\sum_{B} p(G=0,B\mid F=0)=\sum_{B} p(G=0\mid B,F=0)p(B\mid F=0)$$
$$=\sum_{B} p(G=0\mid B,F=0)p(B)=0.04+0.665=0.705$$

$$p(G=0\mid F=0)=0.295$$

- Άρα τελικά:

$$p(F=0\mid D=0)=0.3829$$

Άσκηση 2.10: Ερώτημα (α)

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.

Άσκηση 2.10: Ερώτημα (β)

$$
\begin{aligned}
p(F=0 \mid D=0, B=0) &= \frac{p(D=0, B=0, F=0)}{p(D=0, B=0)} \\
&= \frac{1}{p(D=0, B=0)} \sum_G p(D=0 \mid G, B=0, F=0) p(G, B=0, F=0) \\
&= \frac{1}{p(D=0, B=0)} \sum_G p(D=0 \mid G) p(G, B=0, F=0) \\
&= \frac{p(B=0) p(F=0)}{p(D=0, B=0)} \sum_G p(D=0 \mid G) p(G \mid B=0, F=0)
\end{aligned}
$$

- Κανόνας του Bayes
- Κανόνας δεσμευμένης πιθανότητας
- Μαρκοβιανή ιδιότητα $p(D \mid G, B, F) = p(D \mid G)$

Άσκηση 2.10: Ερώτημα (β)

$$p(G=0 \mid B=0) = \sum_F p(G=0, F \mid B=0) = \sum_F p(G=0 \mid B=0, F) p(F \mid B=0)$$
$$= \sum_F p(G=0 \mid B=0, F) p(F) = 0.16 + 0.6 = 0.76$$

$$p(G=1 \mid B=0) = 0.24$$

$$p(D=0, B=0) = p(D=0 \mid B=0) p(B=0) = p(B=0) \sum_G p(D=0, G \mid B=0)$$
$$= p(B=0) \sum_G p(D=0 \mid G, B=0) p(G \mid B=0)$$
$$= p(B=0) \sum_G p(D=0 \mid G) p(G \mid B=0) = 0.05(0.608 + 0.048)$$
$$= 0.0328$$

- Άρα τελικά:

$$p(F=0 \mid D=0, B=0) = 0.2073$$

Άσκηση 2.10: Ερώτημα (β)

Παρατηρούμε:

- $p(F = 0 \mid D = 0) > p(F = 0 \mid D = 0, B = 0)$
- Αυτά συμβαίνει γιατί στην πρώτη περίπτωση, η ένδειξη D = 0 καθιστά αρκετά πιθανή την αιτία F = 0, ενώ εάν γνωρίζαμε οτι B = 0 δεν είναι πιθανό να ισχύει ταυτόχρονα το άλλο ανεξάρτητο και σπάνιο γεγονός F = 0, έχουμε ήδη βρεί δηλαδή μια αιτία για την ένδειξη D = 0 (explaining away).