Άσκηση 1.1 (Maximum Likelihood estimation)

Η εκτίμηση μέγιστης πιθανοφάνειας για την παράμετρο θ είναι $\hat{\theta} = \underset{\theta}{\arg\max}\, p(D|\theta)$. Επειδή τα δεδομένα είναι i.i.d ισχύει ότι $p(D|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$. Επίσης, $x_i > 0$, $i = 1, 2, \dots n$ καθώς $p(x_i|\theta) = 0$ για $x_i \leq 0$. Επομένως, $u(x_i) = 1$, $i = 1, \dots n$ και $p(D|\theta) = \prod_{i=1}^{n} \theta^2 x_i \exp\{-\theta x_i\}$.

Ισχύει επίσης ότι $\hat{\theta} = \underset{\theta}{\arg\max}\, \ln p(D|\theta)$ επειδή η λογαριθμική συνάρτηση είναι γνησίως αύξουσα και δεν επηρεάζει την θέση του μεγίστου. Έχουμε ότι $\ln p(D|\theta) =$

$$= \ln \prod_{i=1}^{n} \theta^2 x_i \exp\{-\theta x_i\} = \sum_{i=1}^{n} \ln(\theta^2 x_i \exp\{-\theta x_i\}) = \sum_{i=1}^{n} (2\ln\theta + \ln x_i - \theta x_i) = 2n \ln\theta + \sum_{i=1}^{n} \ln x_i - \theta \sum_{i=1}^{n} x_i.$$

Για να υπολογίσουμε το σημείο όπου μεγιστοποιείται η log-likelihood παραγωγίζουμε ως προς θ και απαιτούμε ότι

$$\frac{\partial \ln p(D|\theta)}{\partial \theta} = 0 \Rightarrow \frac{\partial}{\partial \theta}\left(2n\ln\theta + \sum_{i=1}^{n} \ln x_i - \theta \sum_{i=1}^{n} x_i\right) = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i}.$$

Το $\hat{\theta}$ πράγματι αποτελεί σημείο μεγίστου, καθώς

$$\frac{\partial^2}{\partial \theta^2}\left(\ln p(D|\theta)\right)\Big|_{\hat{\theta}} = -\frac{2n}{\hat{\theta}^2} < 0.$$

Τελικά έχουμε ότι η εκτίμηση μέγιστης πιθανοφάνειας είναι $\hat{\theta} = \dfrac{2n}{\sum_{i=1}^{n} x_i}$ ή $\hat{\theta} = \dfrac{2}{\bar{x}}$.

Άσκηση 1.2 (Minimax Criterion)

1. Έστω R_i η περιοχή απόφασης που ο ταξινομητής προβλέπει την κλάση ω_i με $i=1,2$.

Το συνολικό ρίσκο δίνεται από $R = \int_{R_1} \lambda_{11} p(\omega_1) p(\bar{x}|\omega_1) + \lambda_{12} p(\omega_2) p(\bar{x}|\omega_2) d\bar{x} +$

$\int_{R_2} \lambda_{21} p(\omega_1) p(\bar{x}|\omega_1) + \lambda_{22} p(\omega_2) p(\bar{x}|\omega_2) d\bar{x}$. Αντικαθιστώντας $\lambda_{11} = \lambda_{22} = 0$ και $\lambda_{12} = \lambda_{21} = 1$

έχουμε ότι $R = \int_{R_1} p(\omega_2) p(\bar{x}|\omega_2) d\bar{x} + \int_{R_2} p(\omega_1) p(\bar{x}|\omega_1) d\bar{x} = (1-p(\omega_1)) \int_{R_1} p(\bar{x}|\omega_2) d\bar{x} +$

$+ p(\omega_1) \int_{R_2} p(\bar{x}|\omega_1) d\bar{x} = \int_{R_1} p(\bar{x}|\omega_2) dx - p(\omega_1) \left[\int_{R_1} p(\bar{x}|\omega_2) d\bar{x} - \int_{R_2} p(\bar{x}|\omega_1) d\bar{x} \right]$.

Θέλουμε το συνολικό ρίσκο να είναι ανεξάρτητο από τις prior κατανομές, επομένως

$$\left[\int_{R_1} p(\bar{x}|\omega_2) d\bar{x} - \int_{R_2} p(\bar{x}|\omega_1) d\bar{x} \right] = 0 \Leftrightarrow \int_{R_1} p(\bar{x}|\omega_2) d\bar{x} = \int_{R_2} p(\bar{x}|\omega_1) d\bar{x}$$

2. Η λύση δεν είναι πάντα μοναδική. Έστω για παράδειγμα $p_1(x|\omega_1) = 1,\ 0 \leq x \leq 1$,

$p_2(x|\omega_2) = 1,\ \frac{1}{2} \leq x \leq \frac{3}{2}$. Αν $R_1 = (-\infty, \frac{3}{4})$, $R_2 = (\frac{3}{4}, +\infty)$ τότε $\int_{R_1} p(x|\omega_2) = \int_{R_2} p(x|\omega_1) =$

$= \frac{3}{4}$. Αν $R_1 = (\frac{3}{4}, +\infty)$, $R_2 = (-\infty, \frac{3}{4})$, $\int_{R_1} p(x|\omega_2) = \int_{R_2} p(x|\omega_1) = \frac{1}{4}$

Άσκηση 1.3 (Bayes Error)

1. Προκύπτει άμεσα ως ειδική περίπτωση του ερωτήματος (2).

2. Το σφάλμα Bayes στις δύο διαστάσεις δίνεται από

$$P_e = \int_{R_2} p(\bar{x}|\omega_1)\, p(\omega_1)\, d\bar{x} + \int_{R_1} p(\bar{x}|\omega_2)\, p(\omega_2)\, d\bar{x} = p(\omega_1)\int_{R_2} p(\bar{x}|\omega_1)\, d\bar{x} + p(\omega_2)\int_{R_1} p(\bar{x}|\omega_2)\, d\bar{x}$$

Έστω $y = A\bar{x} \in \mathbb{R}$ η προβολή του διανύσματος $\bar{x}$ στην μία διάσταση και $\tilde{p}(y)$ η κατανομή της τυχαίας μεταβλητής y. Έχουμε ότι $\int_{\tilde{R}} \tilde{p}(y)\, dy = \int_{\tilde{R}'} p(\bar{x})\, d\bar{x}$, όπου η προβολή του $\tilde{R}'$ είναι $\tilde{R}$. Το σφάλμα Bayes στην μία διάσταση είναι $\tilde{P}_e = p(\omega_1)\int_{\tilde{R}_2} p(y|\omega_1)\, dy + p(\omega_2)\int_{\tilde{R}_1} p(y|\omega_2)\, dy$

$\Rightarrow \tilde{P}_e = p(\omega_1)\int_{\tilde{R}_2'} p(\bar{x}|\omega_1)\, d\bar{x} + p(\omega_2)\int_{\tilde{R}_1'} p(\bar{x}|\omega_2)\, d\bar{x}$. Επειδή τα $\tilde{R}_1'$, $\tilde{R}_2'$ δεν ταυτίζονται εν γένει με τα R_1, R_2 που αντιστοιχούν εξ' ορισμού στο ελάχιστο σφάλμα ταξινόμησης στις δύο διαστάσεις, προκύπτει ότι $\tilde{P}_e \geq P_e$.

Άσκηση 1.4 (Perceptrons)

Σχήμα 1.4(α): Γραφική παράσταση των σημείων του συνόλου δεδομένων μαζί με την διαχωριστική ευθεία που προέκυψε από την εφαρμογή του αλγορίθμου Perceptron.

Από το παραπάνω σχήμα βλέπουμε ότι οι δύο κλάσεις είναι γραμμικά διαχωρίσιμες. Επειδή η διαχωριστική ευθεία δεν μπορεί να διέρχεται από την αρχή των αξόνων, δεν επαρκεί το αρχικό διάνυσμα βαρών. Χρησιμοποιούμε επομένως το διάνυσμα βαρών $\boldsymbol{w} = [w_0\ w_1\ w_2]^\top$ και επαυξάνουμε τα διανύσματα χαρακτηριστικών από $[x_1\ x_2]^\top$ σε $[1\ x_1\ x_2]^\top$. Παρακάτω φαίνονται τα βήματα του αλγορίθμου Perceptron.

Εποχή 1

$$
\begin{aligned}
&\boldsymbol{w}(0)^\top \boldsymbol{x}_0 = [0,0,0][1,2,1.5]^\top \leq 0,\ \boldsymbol{x}_0 \in \omega_1 && \Rightarrow \boldsymbol{w}(1) = \boldsymbol{w}(0) + \boldsymbol{x}_0 = [0,0,0]^\top + [1,2,1.5]^\top = [1,2,1.5]^\top \\
&\boldsymbol{w}(1)^\top \boldsymbol{x}_1 = [1,2,1.5][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 && \Rightarrow \boldsymbol{w}(2) = \boldsymbol{w}(1) = [1,2,1.5]^\top \\
&\boldsymbol{w}(2)^\top \boldsymbol{x}_2 = [1,2,1.5][1,-3,3]^\top \leq 0,\ \boldsymbol{x}_2 \in \omega_1 && \Rightarrow \boldsymbol{w}(3) = \boldsymbol{w}(2) + \boldsymbol{x}_2 = [1,2,1.5]^\top + [1,-3,3]^\top = [2,-1,4.5]^\top \\
&\boldsymbol{w}(3)^\top \boldsymbol{x}_3 = [2,-1,4.5][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 && \Rightarrow \boldsymbol{w}(4) = \boldsymbol{w}(3) = [2,-1,4.5]^\top \\
&\boldsymbol{w}(4)^\top \boldsymbol{x}_4 = [2,-1,4.5][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 && \Rightarrow \boldsymbol{w}(5) = \boldsymbol{w}(4) = [2,-1,4.5]^\top \\
&\boldsymbol{w}(5)^\top \boldsymbol{x}_5 = [2,-1,4.5][1,-1,0]^\top \geq 0,\ \boldsymbol{x}_5 \in \omega_2 && \Rightarrow \boldsymbol{w}(6) = \boldsymbol{w}(5) - \boldsymbol{x}_5 = [2,-1,4.5]^\top - [1,-1,0]^\top = [1,0,4.5]^\top \\
&\boldsymbol{w}(6)^\top \boldsymbol{x}_6 = [1,0,4.5][1,0,0]^\top \geq 0,\ \boldsymbol{x}_6 \in \omega_2 && \Rightarrow \boldsymbol{w}(7) = \boldsymbol{w}(6) - \boldsymbol{x}_6 = [1,0,4.5]^\top - [1,0,0]^\top = [0,0,4.5]^\top \\
&\boldsymbol{w}(7)^\top \boldsymbol{x}_7 = [0,0,4.5][1,2,-0.5]^\top \leq 0,\ \boldsymbol{x}_7 \in \omega_2 && \Rightarrow \boldsymbol{w}(8) = \boldsymbol{w}(7) = [0,0,4.5]^\top \\
&\boldsymbol{w}(8)^\top \boldsymbol{x}_8 = [0,0,4.5][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 && \Rightarrow \boldsymbol{w}(9) = \boldsymbol{w}(8) = [0,0,4.5]^\top \\
&\boldsymbol{w}(9)^\top \boldsymbol{x}_9 = [0,0,4.5][1,-2,1]^\top \geq 0,\ \boldsymbol{x}_9 \in \omega_2 && \Rightarrow \boldsymbol{w}(10) = \boldsymbol{w}(9) - \boldsymbol{x}_9 = [0,0,4.5]^\top - [1,-2,1]^\top = [-1,2,3.5]^\top
\end{aligned}
$$

Εποχή 2

$$\boldsymbol{w}(10)^\top \boldsymbol{x}_0 = [-1,2,3.5][1,2,1.5]^\top \geq 0,\ \boldsymbol{x}_0 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(11) = \boldsymbol{w}(10) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(11)^\top \boldsymbol{x}_1 = [-1,2,3.5][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(12) = \boldsymbol{w}(11) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(12)^\top \boldsymbol{x}_2 = [-1,2,3.5][1,-3,3]^\top \geq 0,\ \boldsymbol{x}_2 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(13) = \boldsymbol{w}(12) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(13)^\top \boldsymbol{x}_3 = [-1,2,3.5][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(14) = \boldsymbol{w}(13) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(14)^\top \boldsymbol{x}_4 = [-1,2,3.5][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(15) = \boldsymbol{w}(14) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(15)^\top \boldsymbol{x}_5 = [-1,2,3.5][1,-1,0]^\top \leq 0,\ \boldsymbol{x}_5 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(16) = \boldsymbol{w}(15) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(16)^\top \boldsymbol{x}_6 = [-1,2,3.5][1,0,0]^\top \leq 0,\ \boldsymbol{x}_6 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(17) = \boldsymbol{w}(16) = [-1,2,3.5]^\top$$
$$\boldsymbol{w}(17)^\top \boldsymbol{x}_7 = [-1,2,3.5][1,2,-0.5]^\top \geq 0,\ \boldsymbol{x}_7 \in \omega_2 \Rightarrow \boldsymbol{w}(18) = \boldsymbol{w}(17) - \boldsymbol{x}_7 = [-1,2,3.5]^\top - [1,2,-0.5]^\top = [-2,0,4]^\top$$
$$\boldsymbol{w}(18)^\top \boldsymbol{x}_8 = [-2,0,4][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(19) = \boldsymbol{w}(18) = [-2,0,4]^\top$$
$$\boldsymbol{w}(19)^\top \boldsymbol{x}_9 = [-2,0,4][1,-2,1]^\top \geq 0,\ \boldsymbol{x}_9 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(20) = \boldsymbol{w}(19) - \boldsymbol{x}_9 = [-2,0,4]^\top - [1,-2,1]^\top = [-3,2,3]^\top$$

Εποχή 3

$$\boldsymbol{w}(20)^\top \boldsymbol{x}_0 = [-3,2,3][1,2,1.5]^\top \geq 0,\ \boldsymbol{x}_0 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(21) = \boldsymbol{w}(20) = [-3,2,3]^\top$$
$$\boldsymbol{w}(21)^\top \boldsymbol{x}_1 = [-3,2,3][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(22) = \boldsymbol{w}(21) = [-3,2,3]^\top$$
$$\boldsymbol{w}(22)^\top \boldsymbol{x}_2 = [-3,2,3][1,-3,3]^\top \leq 0,\ \boldsymbol{x}_2 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(23) = \boldsymbol{w}(22) + \boldsymbol{x}_2 = [-3,2,3]^\top + [1,-3,3]^\top = [-2,-1,6]^\top$$
$$\boldsymbol{w}(23)^\top \boldsymbol{x}_3 = [-2,-1,6][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(24) = \boldsymbol{w}(23) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(24)^\top \boldsymbol{x}_4 = [-2,-1,6][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(25) = \boldsymbol{w}(24) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(25)^\top \boldsymbol{x}_5 = [-2,-1,6][1,-1,0]^\top \leq 0,\ \boldsymbol{x}_5 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(26) = \boldsymbol{w}(25) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(26)^\top \boldsymbol{x}_6 = [-2,-1,6][1,0,0]^\top \leq 0,\ \boldsymbol{x}_6 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(27) = \boldsymbol{w}(26) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(27)^\top \boldsymbol{x}_7 = [-2,-1,6][1,2,-0.5]^\top \leq 0,\ \boldsymbol{x}_7 \in \omega_2 \Rightarrow \boldsymbol{w}(28) = \boldsymbol{w}(27) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(28)^\top \boldsymbol{x}_8 = [-2,-1,6][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(29) = \boldsymbol{w}(28) = [-2,-1,6]^\top$$
$$\boldsymbol{w}(29)^\top \boldsymbol{x}_9 = [-2,-1,6][1,-2,1]^\top \geq 0,\ \boldsymbol{x}_9 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(30) = \boldsymbol{w}(29) - \boldsymbol{x}_9 = [-2,-1,6]^\top - [1,-2,1]^\top = [-3,1,5]^\top$$

Εποχή 4

$$\boldsymbol{w}(30)^\top \boldsymbol{x}_0 = [-3,1,5][1,2,1.5]^\top \geq 0,\ \boldsymbol{x}_0 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(31) = \boldsymbol{w}(30) = [-3,1,5]^\top$$
$$\boldsymbol{w}(31)^\top \boldsymbol{x}_1 = [-3,1,5][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(32) = \boldsymbol{w}(31) = [-3,1,5]^\top$$
$$\boldsymbol{w}(32)^\top \boldsymbol{x}_2 = [-3,1,5][1,-3,3]^\top \geq 0,\ \boldsymbol{x}_2 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(33) = \boldsymbol{w}(32) = [-3,1,5]^\top$$
$$\boldsymbol{w}(33)^\top \boldsymbol{x}_3 = [-3,1,5][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(34) = \boldsymbol{w}(33) = [-3,1,5]^\top$$
$$\boldsymbol{w}(34)^\top \boldsymbol{x}_4 = [-3,1,5][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 \quad \Rightarrow \boldsymbol{w}(35) = \boldsymbol{w}(34) = [-3,1,5]^\top$$
$$\boldsymbol{w}(35)^\top \boldsymbol{x}_5 = [-3,1,5][1,-1,0]^\top \leq 0,\ \boldsymbol{x}_5 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(36) = \boldsymbol{w}(35) = [-3,1,5]^\top$$
$$\boldsymbol{w}(36)^\top \boldsymbol{x}_6 = [-3,1,5][1,0,0]^\top \leq 0,\ \boldsymbol{x}_6 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(37) = \boldsymbol{w}(36) = [-3,1,5]^\top$$
$$\boldsymbol{w}(37)^\top \boldsymbol{x}_7 = [-3,1,5][1,2,-0.5]^\top \leq 0,\ \boldsymbol{x}_7 \in \omega_2 \Rightarrow \boldsymbol{w}(38) = \boldsymbol{w}(37) = [-3,1,5]^\top$$
$$\boldsymbol{w}(38)^\top \boldsymbol{x}_8 = [-3,1,5][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(39) = \boldsymbol{w}(38) = [-3,1,5]^\top$$
$$\boldsymbol{w}(39)^\top \boldsymbol{x}_9 = [-3,1,5][1,-2,1]^\top \geq 0,\ \boldsymbol{x}_9 \in \omega_2 \quad \Rightarrow \boldsymbol{w}(40) = \boldsymbol{w}(39) - \boldsymbol{x}_9 = [-3,1,5]^\top - [1,-2,1]^\top = [-4,3,4]^\top$$

Εποχή 5

$$
\begin{aligned}
&\boldsymbol{w}(40)^\top\boldsymbol{x}_0 = [-4,3,4][1,2,1.5]^\top \geq 0,\ \boldsymbol{x}_0 \in \omega_1 &&\Rightarrow \boldsymbol{w}(41) = \boldsymbol{w}(40) = [-4,3,4]^\top\\
&\boldsymbol{w}(41)^\top\boldsymbol{x}_1 = [-4,3,4][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 &&\Rightarrow \boldsymbol{w}(42) = \boldsymbol{w}(41) = [-4,3,4]^\top\\
&\boldsymbol{w}(42)^\top\boldsymbol{x}_2 = [-4,3,4][1,-3,3]^\top \leq 0,\ \boldsymbol{x}_2 \in \omega_1 &&\Rightarrow \boldsymbol{w}(43) = \boldsymbol{w}(42) + \boldsymbol{x}_2 = [-4,3,4]^\top + [1,-3,3]^\top = [-3,0,7]^\top\\
&\boldsymbol{w}(43)^\top\boldsymbol{x}_3 = [-3,0,7][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 &&\Rightarrow \boldsymbol{w}(44) = \boldsymbol{w}(43) = [-3,0,7]^\top\\
&\boldsymbol{w}(44)^\top\boldsymbol{x}_4 = [-3,0,7][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 &&\Rightarrow \boldsymbol{w}(45) = \boldsymbol{w}(44) = [-3,0,7]^\top\\
&\boldsymbol{w}(45)^\top\boldsymbol{x}_5 = [-3,0,7][1,-1,0]^\top \leq 0,\ \boldsymbol{x}_5 \in \omega_2 &&\Rightarrow \boldsymbol{w}(46) = \boldsymbol{w}(45) = [-3,0,7]^\top\\
&\boldsymbol{w}(46)^\top\boldsymbol{x}_6 = [-3,0,7][1,0,0]^\top \leq 0,\ \boldsymbol{x}_6 \in \omega_2 &&\Rightarrow \boldsymbol{w}(47) = \boldsymbol{w}(46) = [-3,0,7]^\top\\
&\boldsymbol{w}(47)^\top\boldsymbol{x}_7 = [-3,0,7][1,2,-0.5]^\top \leq 0,\ \boldsymbol{x}_7 \in \omega_2 &&\Rightarrow \boldsymbol{w}(48) = \boldsymbol{w}(47) = [-3,0,7]^\top\\
&\boldsymbol{w}(48)^\top\boldsymbol{x}_8 = [-3,0,7][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 &&\Rightarrow \boldsymbol{w}(49) = \boldsymbol{w}(48) = [-3,0,7]^\top\\
&\boldsymbol{w}(49)^\top\boldsymbol{x}_9 = [-3,0,7][1,-2,1]^\top \geq 0,\ \boldsymbol{x}_9 \in \omega_2 &&\Rightarrow \boldsymbol{w}(50) = \boldsymbol{w}(49) - \boldsymbol{x}_9 = [-3,0,7]^\top - [1,-2,1]^\top = [-4,2,6]^\top
\end{aligned}
$$

Εποχή 6

$$
\begin{aligned}
&\boldsymbol{w}(50)^\top\boldsymbol{x}_0 = [-4,2,6][1,2,1.5]^\top \geq 0,\ \boldsymbol{x}_0 \in \omega_1 &&\Rightarrow \boldsymbol{w}(51) = \boldsymbol{w}(50) = [-4,2,6]^\top\\
&\boldsymbol{w}(51)^\top\boldsymbol{x}_1 = [-4,2,6][1,4,1]^\top \geq 0,\ \boldsymbol{x}_1 \in \omega_1 &&\Rightarrow \boldsymbol{w}(52) = \boldsymbol{w}(51) = [-4,2,6]^\top\\
&\boldsymbol{w}(52)^\top\boldsymbol{x}_2 = [-4,2,6][1,-3,3]^\top \geq 0,\ \boldsymbol{x}_2 \in \omega_1 &&\Rightarrow \boldsymbol{w}(53) = \boldsymbol{w}(52) = [-4,2,6]^\top\\
&\boldsymbol{w}(53)^\top\boldsymbol{x}_3 = [-4,2,6][1,-2,2]^\top \geq 0,\ \boldsymbol{x}_3 \in \omega_1 &&\Rightarrow \boldsymbol{w}(54) = \boldsymbol{w}(53) = [-4,2,6]^\top\\
&\boldsymbol{w}(54)^\top\boldsymbol{x}_4 = [-4,2,6][1,0,3]^\top \geq 0,\ \boldsymbol{x}_4 \in \omega_1 &&\Rightarrow \boldsymbol{w}(55) = \boldsymbol{w}(54) = [-4,2,6]^\top\\
&\boldsymbol{w}(55)^\top\boldsymbol{x}_5 = [-4,2,6][1,-1,0]^\top \leq 0,\ \boldsymbol{x}_5 \in \omega_2 &&\Rightarrow \boldsymbol{w}(56) = \boldsymbol{w}(55) = [-4,2,6]^\top\\
&\boldsymbol{w}(56)^\top\boldsymbol{x}_6 = [-4,2,6][1,0,0]^\top \leq 0,\ \boldsymbol{x}_6 \in \omega_2 &&\Rightarrow \boldsymbol{w}(57) = \boldsymbol{w}(56) = [-4,2,6]^\top\\
&\boldsymbol{w}(57)^\top\boldsymbol{x}_7 = [-4,2,6][1,2,-0.5]^\top \leq 0,\ \boldsymbol{x}_7 \in \omega_2 &&\Rightarrow \boldsymbol{w}(58) = \boldsymbol{w}(57) = [-4,2,6]^\top\\
&\boldsymbol{w}(58)^\top\boldsymbol{x}_8 = [-4,2,6][1,1,-1]^\top \leq 0,\ \boldsymbol{x}_8 \in \omega_2 &&\Rightarrow \boldsymbol{w}(59) = \boldsymbol{w}(58) = [-4,2,6]^\top\\
&\boldsymbol{w}(59)^\top\boldsymbol{x}_9 = [-4,2,6][1,-2,1]^\top \leq 0,\ \boldsymbol{x}_9 \in \omega_2 &&\Rightarrow \boldsymbol{w}(60) = \boldsymbol{w}(59) = [-4,2,6]^\top
\end{aligned}
$$

Ο αλγόριθμος ταξινομεί σωστά όλα τα σημεία και επομένως συγκλίνει. Η διαχωριστική ευθεία που προκύπτει είναι η $y = -\frac{1}{3}x + \frac{2}{3}$.

Άσκηση 1.5 (Kullback-Leibler Divergence)

Έχουμε ότι $p_2(\bar{x}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left\{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Sigma^{-1}(\bar{x}-\bar{\mu})\right\} \Rightarrow \ln p_2(\bar{x}) = \ln\left[(2\pi)^k|\Sigma|\right]^{-\frac{1}{2}} -$

$-\frac{1}{2}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu}) = -\frac{1}{2}\ln(2\pi)^k - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})$. Αντικαθιστώντας,

προκύπτει ότι $D_{KL}(p_1(\bar{x}), p_2(\bar{x})) = \int p_1(\bar{x}) \ln\frac{p_1(\bar{x})}{p_2(\bar{x})} d\bar{x} = \int p_1(\bar{x}) \ln p_1(\bar{x}) d\bar{x} - \int p_1(\bar{x}) \ln p_2(\bar{x}) d\bar{x} =$

$= \int p_1(\bar{x}) \ln p_1(\bar{x}) d\bar{x} - \mathbb{E}_{x\sim p_1}\left[\ln p_2(\bar{x})\right] d\bar{x} = I - \mathbb{E}_{x\sim p_1}\left[-\frac{1}{2}\ln(2\pi)^k - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right]$

$= I + \frac{1}{2}\mathbb{E}_{x\sim p_1}\left[\ln(2\pi)^k + \ln|\Sigma| + (\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right]$. Για ελαχιστοποίηση της $D_{KL}(p_1(\bar{x}), p_2(\bar{x}))$

απαιτούμε να μηδενίζονται οι παράγωγοι ως προς $\bar{\mu}$ και Σ αντίστοιχα. Επομένως,

$$\frac{\partial D_{KL}(p_1(\bar{x}), p_2(\bar{x}))}{\partial \bar{\mu}} = \frac{1}{2}\frac{\partial}{\partial\bar{\mu}}\mathbb{E}_{x\sim p_1}\left[\ln(2\pi)^k + \ln|\Sigma| + (\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right] =$$

$$= \frac{1}{2}\mathbb{E}_{x\sim p_1}\left[\frac{\partial}{\partial\bar{\mu}}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right] = \frac{1}{2}\mathbb{E}_{x\sim p_1}\left[-\Sigma^{-1}(\bar{x}-\bar{\mu}) - (\Sigma^{-1})^T(\bar{x}-\bar{\mu})\right] =$$

$$= \frac{1}{2}\mathbb{E}_{x\sim p_1}\left[-2\Sigma^{-1}(\bar{x}-\bar{\mu})\right] = -\Sigma^{-1}\mathbb{E}_{x\sim p_1}\left[\bar{x}-\bar{\mu}\right] = 0 \Rightarrow \mathbb{E}_{x\sim p_1}[\bar{x}-\bar{\mu}] = 0 \Rightarrow$$

$\bar{\mu} = \mathbb{E}_{x\sim p_1}[\bar{x}]$. Για τον υπολογισμό $\frac{\partial D_{KL}(p_1(\bar{x}), p_2(\bar{x}))}{\partial\Sigma}$, θα υπολογίσουμε πρώτα

τα $\frac{\partial \ln|\Sigma|}{\partial\Sigma}$, $\frac{\partial}{\partial\Sigma}(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})$. Έχουμε ότι $\frac{\partial\ln|\Sigma|}{\partial\Sigma} = \frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial\Sigma} = \frac{1}{|\Sigma|}\left[\operatorname{adj}(\Sigma)\right]^T =$

$= \frac{1}{|\Sigma|}(|\Sigma|\Sigma^{-1})^T = (\Sigma^{-1})^T = \Sigma^{-1}$. Επίσης, $\frac{\partial}{\partial\Sigma}\left[(\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right] = \frac{\partial}{\partial\Sigma}\operatorname{tr}\left((\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right)$

$= \frac{\partial}{\partial\Sigma}\operatorname{tr}\left(\Sigma^{-1}(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right) = -\Sigma^{-1}(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\Sigma^{-1}$

Έχουμε $\frac{\partial D_{KL}(p_1(\bar{x}), p_2(\bar{x}))}{\partial \Sigma} = \frac{1}{2}\frac{\partial}{\partial \Sigma}\mathbb{E}_{x\sim p_1}\left[\ln(2\pi)^k + \ln|\Sigma| + (\bar{x}-\bar{\mu})^T\Sigma^{-1}(\bar{x}-\bar{\mu})\right] =$

$= \frac{1}{2}\mathbb{E}_{x\sim p_1}\left[\Sigma^{-1} - \Sigma^{-1}(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\Sigma^{-1}\right] = 0 \Rightarrow \mathbb{E}_{x\sim p_1}\left[\mathbb{I} - (\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\Sigma^{-1}\right] = 0$

$\Rightarrow \mathbb{E}_{x\sim p_1}\left[\Sigma - (\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right] = 0 \Rightarrow \Sigma = \mathbb{E}_{x\sim p_1}\left[(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right]$

Τελικά, προκύπτει ότι $\mu = \mathbb{E}_{x\sim p_1}[x]$ και $\Sigma = \mathbb{E}_{x\sim p_1}\left[(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right]$

Άσκηση 1.6 (Linear Regression and the LMS algorithm)

1. Έχουμε ότι $y_i = w_0 + w_1 x_i + \varepsilon_i$, όπου $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ ο Γκαουσιανός θόρυβος, για $i = 1, 2, \ldots n$. Επομένως, $y_i \sim \mathcal{N}(w_0 + w_1 x_i, \sigma^2)$. Η εκτίμηση μέγιστης πιθανοφάνειας δίνεται από $\hat{w} = \underset{w}{\arg\max}\, p(D \mid w_0, w_1) = \underset{w}{\arg\max}\, \ln p(D \mid w_0, w_1) =$

$$= \underset{w}{\arg\max}\, \ln \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left(y_i - w_0 - w_1 x_i \right)^2 \right\} = \underset{w}{\arg\max} \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \right.$$

$$\left. - \frac{1}{2\sigma^2} \left(y_i - w_0 - w_1 x_i \right)^2 \right] = \underset{w}{\arg\max} \left(\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i \right)^2 \right) =$$

$$= \underset{w}{\arg\max}\, -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i \right)^2 = \underset{w}{\arg\min} \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i \right)^2,$$

δηλαδή ισούται με την εκτίμηση ελαχίστων τετραγώνων.

2. Για να υπολογίσουμε τους άγνωστους συντελεστές, απαιτούμε οι αντίστοιχες μερικές παράγωγοι ως προς αυτούς να είναι ίσες με μηδέν. Επομένως, $\frac{\partial \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2}{\partial w_0} = 0 \Rightarrow$

$$\Rightarrow \sum_{i=1}^{n} 2(y_i - w_0 - w_1 x_i)(-1) = 0 \Rightarrow n w_0 = \sum_{i=1}^{n} y_i - w_1 \sum_{i=1}^{n} x_i \Rightarrow w_0 = \bar{y} - w_1 \bar{x}.$$

Επίσης, έχουμε ότι

$$\frac{\partial \sum_{i=1}^{n} (y_i - w_0 - w_1 x_i)^2}{\partial w_1} = \sum_{i=1}^{n} 2(y_i - w_0 - w_1 x_i)(-x_i) = \sum_{i=1}^{n} (-2x_i y_i + 2x_i w_0 + 2 w_1 x_i^2) = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} (-x_i y_i + x_i (\bar{y} - w_1 \bar{x}) + w_1 x_i^2) = \sum_{i=1}^{n} (-x_i y_i + x_i \bar{y} - w_1 x_i \bar{x} + w_1 x_i^2) =$$

$$= \sum_{i=1}^{n} (-x_i y_i) + n\bar{y}\bar{x} - n w_1 \bar{x}^2 + w_1 \sum_{i=1}^{n} x_i^2 = 0 \Rightarrow w_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}.$$

Αντικαθιστώντας τις τιμές για τα (x_i, y_i) έχουμε ότι $\bar{x} = 0.631$, $\bar{y} = 2.619$, $\sum_{i=1}^{10} x_i y_i = 17.1873$, $\sum_{i=1}^{10} x_i^2 = 4.2817$.

Τελικά, $w_1 = 2.204$ και $w_0 = 1.228$, οπότε $\hat{y} = 1.228 + 2.204x$

3. Η ανανέωση των βαρών στον αλγόριθμο LMS δίνεται από την σχέση

$w^{(k+1)} = w^{(k)} + p^{(k)} \left(y^{(k)} - w^{(k)} \cdot x^{(k)} \right) x^{(k)}$. Θεωρούμε ότι $w^{(0)} = [0 \;\; 0]^T$ και $p^{(k)} = \frac{1}{k}$, $k \geq 1$.

Επομένως, έχουμε:

$$w^{(1)} = [0 \;\; 0]^T + 1 \cdot \left(2.05 - [0 \;\; 0][1 \;\; 0.38]^T\right)[1 \;\; 0.38]^T = [2.05 \;\; 0.779]^T$$

$$w^{(2)} = [2.05 \;\; 0.779]^T + \frac{1}{2}\left(2.23 - [2.05 \;\; 0.779][1 \;\; 0.44]^T\right)[1 \;\; 0.44]^T = [1.969 \;\; 0.743]^T$$

$$w^{(3)} = [1.969 \;\; 0.743]^T + \frac{1}{3}\left(2.13 - [1.969 \;\; 0.743][1 \;\; 0.48]^T\right)[1 \;\; 0.48]^T = [1.904 \;\; 0.712]^T$$

$$w^{(4)} = [1.904 \;\; 0.712]^T + \frac{1}{4}\left(2.33 - [1.904 \;\; 0.712][1 \;\; 0.54]^T\right)[1 \;\; 0.54]^T = [1.914 \;\; 0.718]^T$$

$$w^{(5)} = [1.914 \;\; 0.718]^T + \frac{1}{5}\left(2.67 - [1.914 \;\; 0.718][1 \;\; 0.58]^T\right)[1 \;\; 0.58]^T = [1.982 \;\; 0.757]^T$$

$$w^{(6)} = [1.982 \;\; 0.757]^T + \frac{1}{6}\left(2.68 - [1.982 \;\; 0.757][1 \;\; 0.64]^T\right)[1 \;\; 0.64]^T = [2.018 \;\; 0.78]^T$$

$$w^{(7)} = [2.018 \;\; 0.78]^T + \frac{1}{7}\left(2.81 - [2.018 \;\; 0.78][1 \;\; 0.71]^T\right)[1 \;\; 0.71]^T = [2.052 \;\; 0.804]^T$$

$$w^{(8)} = [2.052 \;\; 0.804]^T + \frac{1}{8}\left(2.97 - [2.052 \;\; 0.804][1 \;\; 0.76]^T\right)[1 \;\; 0.76]^T = [2.09 \;\; 0.833]^T$$

$$w^{(9)} = [2.09 \;\; 0.833]^T + \frac{1}{9}\left(3.12 - [2.09 \;\; 0.833][1 \;\; 0.82]^T\right)[1 \;\; 0.82]^T = [2.129 \;\; 0.865]^T$$

$$w^{(10)} = [2.129 \;\; 0.865]^T + \frac{1}{10}\left(3.2 - [2.129 \;\; 0.865][1 \;\; 0.96]^T\right)[1 \;\; 0.96]^T = [2.153 \;\; 0.888]^T$$

Το διάνυσμα βαρών μετά την πρώτη εποχή είναι $[2.153 \;\; 0.888]^T$.

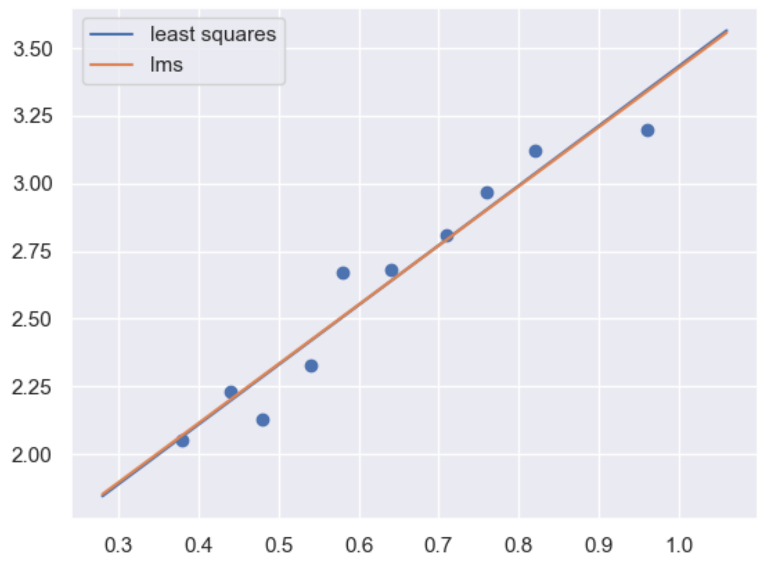
4. Ο κώδικας για τον αλγόριθμο LMS φαίνεται παρακάτω:

```
def lms(X, Y, epochs=2000):
    """
    Least Mean Squares algorithm

    Parameters:
    X (numpy.ndarray): 2-dimensional numpy array
    Y (numpy.ndarray): 1-dimensional numpy array

    Returns:
    numpy.ndarray: 1-dimensional array of weights
    """
    weights = np.zeros_like(X[0])
    eta = 1
    k = 0
    for i in range(epochs):
        for x, y in zip(X, Y):
            k += 1
            # weights += (eta/k) * (y - weights @ x.T ) * x
            # Constant learning rate has better convergence
            weights += 0.01 * (y - weights @ x.T ) * x
    return weights
```

5. Η ευθεία που προκύπτει με την μέθοδο των ελαχίστων τετραγώνων είναι η $y_{lsq} = 1.228 + 2.204x$ και η ευθεία που προκύπτει με την παραπάνω υλοποίηση του αλγορίθμου LMS είναι η $y_{lms} = 1.240 + 2.186x$.



Σχήμα 1.6: Γραφική παράσταση των σημείων του συνόλου δεδομένων, καθώς και των ευθειών που προκύπτουν με την μέθοδο των ελαχίστων τετραγώνων (μπλε) και του αλγορίθμου LMS (κόκκινη).

Άσκηση 1.7 (Bayes meets k-NN)

1. Ο κανόνας απόφασης κατά Bayes δίνεται από $P(\omega_1)p(x|\omega_1) \gtrless P(\omega_2)p(x|\omega_2) \Rightarrow$

$\Rightarrow \frac{1}{2}2x \gtrless \frac{1}{2}(2-2x) \Rightarrow x \gtrless 0.5$. Οι περιοχές απόφασης είναι $R_1 = [0.5, 1]$ και

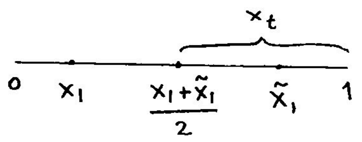
$R_2 = [0, 0.5]$. Το σφάλμα ταξινόμησης είναι $P_e = p(\omega_1)\int_{R_2} p(x|\omega_1)\,dx +$

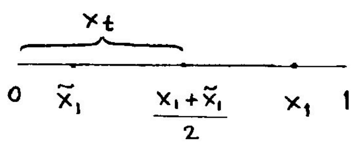
$$+ p(\omega_2)\int_{R_1} p(x|\omega_2)\,dx = \frac{1}{2}\int_0^{0.5} 2x\,dx + \frac{1}{2}\int_{0.5}^{1}(2-2x)\,dx = \frac{1}{2}\left[x^2\right]_0^{0.5} + \frac{1}{2}\left[2x-x^2\right]_{0.5}^{1} =$$

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \Rightarrow P_e = 0.25$$

2. Έστω x_1 το τυχαίο σημείο από την κατηγορία ω_1, $\tilde{x}_1$ από την ω_2 και x_t το σημείο αποτίμησης από την κατηγορία ω_1. Η ταξινόμηση είναι εσφαλμένη στις εξής περιπτώσεις: αν $0 \le x_1 \le \tilde{x}_1 \le 1$ και $x_t > \frac{x_1+\tilde{x}_1}{2}$ ή αν $1 \ge x_1 > \tilde{x}_1 \ge 0$ και $x_t < \frac{x_1+\tilde{x}_1}{2}$

Σχηματικά,





Επομένως, $$p_1(e) = \int_{\tilde{x}_1=0}^{1}\int_{x_1=0}^{\tilde{x}_1}\int_{x_t=\frac{x_1+\tilde{x}_1}{2}}^{1} p(\tilde{x}_1|\omega_2)\; p(x_1|\omega_1)\; p(x_t|\omega_1)\, dx_t\, dx_1\, d\tilde{x}_1 +$$

$$\int_{x_1=0}^{1}\int_{\tilde{x}_1=0}^{x_1}\int_{x_t=0}^{\frac{x_1+\tilde{x}_1}{2}} p(x|\omega_1)\, p(\tilde{x}_1|\omega_2)\, p(x_t|\omega_1)\; dx_t\, d\tilde{x}_1\, dx_1$$

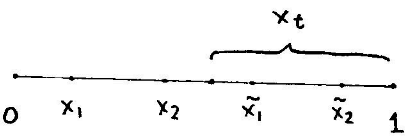
$$= \int_{\tilde{x}_1=0}^{1}\int_{x_1=0}^{\tilde{x}_1}\int_{x_t=\frac{x_1+\tilde{x}_1}{2}}^{1} (2-2\tilde{x}_1)(2x_1)(2x_t)\; dx_t\, dx_1\, d\tilde{x}_1 +$$

$$+ \int_{x_1=0}^{1}\int_{\tilde{x}_1=0}^{1}\int_{x_t=0}^{\frac{x_1+\tilde{x}_1}{2}} (2x_1)(2-2\tilde{x}_1)(2x_t)\; dx_t\, d\tilde{x}_1\, dx_1 =$$

$$= \frac{43}{360} + \frac{83}{360} = \frac{7}{20} \Rightarrow P_1(e) = 0.35$$

3. Έστω x_1, x_2 τα τυχαία σημεία της κατηγορίας ω_1, $\tilde{x}_1, \tilde{x}_2$ της κατηγορίας ω_2 και x_t το σημείο αποτίμησης από την ω_1. Θα υποθέσουμε ότι $x_1 \leq x_2$ και $\tilde{x}_1 \leq \tilde{x}_2$, καθώς οι περιπτώσεις $x_1 \leq x_2 \wedge \tilde{x}_2 \leq \tilde{x}_1$, $x_2 \leq x_1 \wedge \tilde{x}_1 \leq \tilde{x}_2$, $x_2 \leq x_1 \wedge \tilde{x}_2 \leq \tilde{x}_1$ είναι ισοδύναμες λόγω συμμετρίας. Διακρίνουμε τις ακόλουθες υποπεριπτώσεις:

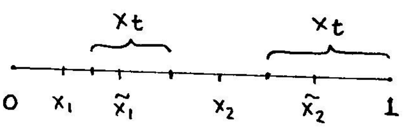
i.



$$I_{(i)} = \int_{\tilde{x}_2=0}^{1} \int_{\tilde{x}_1=0}^{\tilde{x}_2} \int_{x_2=0}^{\tilde{x}_1} \int_{x_1=0}^{x_2} \int_{x_t=\frac{x_2+\tilde{x}_1}{2}}^{1} \overbrace{p(\tilde{x}_2|\omega_2)\, p(\tilde{x}_1|\omega_2)\, p(x_2|\omega_1)\, p(x_1|\omega_1)\, p(x_t|\omega_1)}^{A}\, dx_t\, dx_1\, dx_2\, d\tilde{x}_1\, d\tilde{x}_2 = \int_{\tilde{x}_2=0}^{1} \int_{\tilde{x}_1=0}^{\tilde{x}_2} \int_{x_2=0}^{\tilde{x}_1} \int_{x_1=0}^{x_2} \int_{x_t=\frac{x_2+\tilde{x}_1}{2}}^{1} (2-2\tilde{x}_2)(2-2\tilde{x}_1)(2x_2)(2x_1)(2x_t)$$

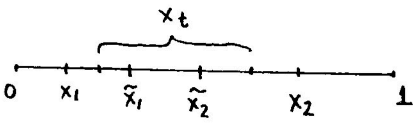
$$dx_t\, dx_1\, dx_2\, d\tilde{x}_1\, d\tilde{x}_2 = \frac{131}{50400}$$

ii.

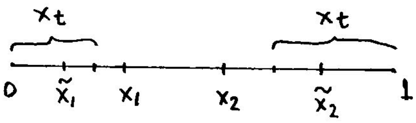


$$I_{(ii)} = \int_{\tilde{x}_2=0}^{1} \int_{x_2=0}^{\tilde{x}_2} \int_{\tilde{x}_1=0}^{x_2} \int_{x_1=0}^{\tilde{x}_1} \int_{x_t=\frac{x_1+\tilde{x}_1}{2}}^{\frac{\tilde{x}_1+x_2}{2}} A + \int_{\tilde{x}_2=0}^{1} \int_{x_2=0}^{\tilde{x}_2} \int_{\tilde{x}_1=0}^{x_2} \int_{x_1=0}^{\tilde{x}_1} \int_{x_t=\frac{x_2+\tilde{x}_2}{2}}^{1} A$$

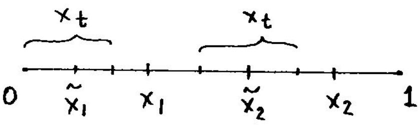
$$= \frac{2081}{453600}$$

iii. 

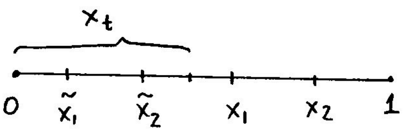
$$I_{(iii)} = \frac{31}{4536}$$

iv. 

$$I_{(iv)} = \frac{71}{5670}$$

v. 

$$I_{(v)} = \frac{1187}{64800}$$

vi. 

$$I_{(vi)} = \frac{673}{16800}$$

Οι συμμετρικές περιπτώσεις που προκύπτουν με εναλλαγή των δεικτών 1,2 στα x_1, x_2 και στα $\tilde{x}_1, \tilde{x}_2$ δίνουν αντίστοιχα αποτελέσματα. Επομένως, τελικά

$$P_2(e) = 4\left(I_{(i)} + I_{(ii)} + I_{(iii)} + I_{(iv)} + I_{(v)} + I_{(vi)} \right) = 4 \cdot \frac{107}{1260} = 0.33968 \Rightarrow$$

$$P_2(e) = 0.33968.$$

4. Έστω $X_i,\ i=1,\dots n$ οι τυχαίες μεταβλητές της ω_1 με p.d.f $f_X(x)$ και c.d.f $F_X(x)$, και $\tilde{X}_i,\ i=1,\dots n$ οι τυχαίες μεταβλητές της ω_2 με p.d.f $f_{\tilde{X}}(x)$ και c.d.f $F_{\tilde{X}}(x)$. Για το σημείο αποτίμησης έχουμε ότι $X_t \sim f_X(x)$. Παρατηρούμε ότι για δεδομένο σημείο αποτίμησης x_t, η πιθανότητα εσφαλμένης ταξινόμησης δίνεται από $\mathbb{P}\left[\min_i |\tilde{X}_i - x_t| \le \min_i |X_i - x_t| \;\middle|\; X_t = x_t\right]$. Επομένως συνολικά η πιθανότητα εσφαλμένης ταξινόμησης είναι

$$\int_0^1 \mathbb{P}\left[\min_i |\tilde{X}_i - x_t| \le \min_i |X_i - x_t| \;\middle|\; X_t = x_t\right] \mathbb{P}[X_t = x_t]\, dx_t =$$

$$\int_0^1 \mathbb{P}\left[\min_i |\tilde{X}_i - x_t| \le z \;\middle|\; X_t = x_t, \min_i |X_i - x_t| = z\right] \mathbb{P}\left[\min_i |X_i - x_t| = z\right] \cdot$$

$\cdot\, \mathbb{P}[X_t = x_t]\, dz\, dx_t$. Έχουμε ότι $\mathbb{P}\left[\min_i |\tilde{X}_i - x_t| \le z\right] = 1 - \mathbb{P}\left[\min_i |\tilde{X}_j - x_t| \ge z\right] =$

$$= 1 - \prod_i \mathbb{P}\left[|\tilde{X}_i - x_t| \ge z\right] = 1 - \prod_i \left(\mathbb{P}[\tilde{X}_i - x_t \ge z] + \mathbb{P}[\tilde{X}_i - x_t \le -z]\right) =$$

$$= 1 - \prod_i \left(\mathbb{P}[\tilde{X}_i \ge x_t + z] + \mathbb{P}[\tilde{X}_i \le x_t - z]\right) =$$

$$= 1 - \prod_i \left(1 - F_{\tilde{X}}(x_t + z) + F_{\tilde{X}}(x_t - z)\right) = 1 - \left(1 - F_{\tilde{X}}(x_t + z) + F_{\tilde{X}}(x_t - z)\right)^n$$

Παραγωγίζοντας, λαμβάνουμε ότι $\mathbb{P}\left[\min_i |X_i - x_t| = z\right] = -n\Big(1 - F_X(x_t + z) +$ $F_X(x_t - z)\Big)^{n-1}\left(-f_X(x_t + z) - f_X(x_t - z)\right) = n\left(1 - F_X(x_t + z) + F_X(x_t - z)\right)^{n-1}$ $\cdot \left(f_X(x_t + z) + f_X(x_t - z)\right)$.

Επομένως, έχουμε ότι

$$P_n(e) = \int_0^1 \int_0^1 n\left[1 - \left(1 - F_{\tilde{X}}(x_t+z) + F_{\tilde{X}}(x_t-z)\right)^n\right]\left[1 - F_X(x_t+z) + F_X(x_t-z)\right]^{n-1} \cdot$$

$$\cdot \left(f_X(x_t+z) + f_X(x_t-z)\right) f_X(x_t)\, dz\, dx_t$$

5. Όταν $n \to \infty$, για δεδομένο σημείο αποτίμησης $X_t = x_t$, η πιθανότητα ορθής ταξινόμησης δίνεται από $\dfrac{f_X(x_t)}{f_X(x_t) + f_{\tilde{X}}(x_t)} = \dfrac{2x_t}{2} = x_t$

Επομένως η πιθανότητα εσφαλμένης ταξινόμισης θα είναι

$$\lim_{n\to\infty} P_n(e) = 1 - \int_0^1 x_t\, \mathbb{P}[X_t = x_t]\, dx_t = 1 - \int_0^1 x_t (2x_t)\, dx_t = 1 - \frac{2}{3} = \frac{1}{3}$$

$\Rightarrow \lim_{n\to\infty} P_n(e) = \frac{1}{3}$, η οποία είναι μεγαλύτερη από το σφάλμα κατά Bayes P^*.

Επίσης, ικανοποιείται η ανισότητα $P^* \leq \lim_{n\to\infty} P_n(e) \leq 2P^*(1-P^*)$ καθώς

$$\frac{1}{4} \leq \frac{1}{3} \leq \frac{3}{8}.$$

Άσκηση 1.8 (EM)

Έστω ότι $D = D_{good} \cup D_{bad}$, όπου $D_{good} = \{\bar{x}_1, \bar{x}_2, x_{31}\}$ και $D_{bad} = \{x_{32}\}$

1. Για το Βήμα E του αλγορίθμου EM έχουμε ότι $Q(\theta,\theta^0) = \mathbb{E}_{D_{bad}}\left[\ln p(D|\theta) \,|\, D_{good}, \theta^0\right]$

$$\Rightarrow Q(\theta,\theta^0) = \mathbb{E}_{x_{32}}\left[\ln p(\bar{x}_1|\theta) + \ln p(\bar{x}_2|\theta) + \ln p(\bar{x}_3|\theta) \,\Big|\, D_{good}, \theta^0\right] \Rightarrow$$

$$\Rightarrow Q(\theta,\theta^0) = \ln p(\bar{x}_1|\theta) + \ln p(\bar{x}_2|\theta) + \mathbb{E}_{x_{32}}\left[\ln p(\bar{x}_3|\theta) \,|\, D_{good}, \theta^0\right] \quad (1)$$

Έχουμε για $i=1,2$ ότι $\ln p(\bar{x}_i|\theta) = \ln\left(\theta_1 e^{-\theta_1 x_{i1}} \frac{1}{\theta_2}\right)$, $0 \le x_{i2} \le \theta_2$. Επομένως,

$\ln p(\bar{x}_i|\theta) = \ln\theta_1 - \ln\theta_2 - \theta_1 x_{i1}$, $0 \le x_{i2} \le \theta_2$. Έχουμε λοιπόν $\ln p(\bar{x}_1|\theta) + \ln p(\bar{x}_2|\theta) =$

$$= (\ln\theta_1 - \ln\theta_2 - \theta_1) + (\ln\theta_1 - \ln\theta_2 - 4\theta_1) = 2\ln\theta_1 - 2\ln\theta_2 - 5\theta_1, \quad \theta_2 \ge \max(x_{12}, x_{22}) = 5$$

$\Rightarrow \ln p(\bar{x}_1|\theta) + \ln p(\bar{x}_2|\theta) = 2\ln\theta_1 - 2\ln\theta_2 - 5\theta_1$, $\theta_2 \ge 5$. Επίσης, έχουμε ότι

$$\mathbb{E}_{x_{32}}\left[\ln p(\bar{x}_3|\theta) \,|\, D_{good}, \theta^0\right] = \int_{-\infty}^{\infty} \ln p(\bar{x}_3|\theta)\, p(x_{32}|\theta^0)\, dx_{32} = \int_0^{\theta_2^0} (\ln\theta_1 - \ln\theta_2 - \theta_1 x_{31}) \frac{1}{\theta_2^0} dx_{32}$$

$$= \int_0^3 (\ln\theta_1 - \ln\theta_2 - \theta_1 x_{31}) \frac{1}{3} dx_{32} = \ln\theta_1 - \ln\theta_2 - \theta_1 x_{31} = \ln\theta_1 - \ln\theta_2 - \theta_1$$ Επομένως,

$$Q(\theta,\theta^0) = (2\ln\theta_1 - 2\ln\theta_2 - 5\theta_1) + (\ln\theta_1 - \ln\theta_2 - \theta_1) = 3\ln\theta_1 - 3\ln\theta_2 - 6\theta_1, \quad \theta_2 \ge 5.$$

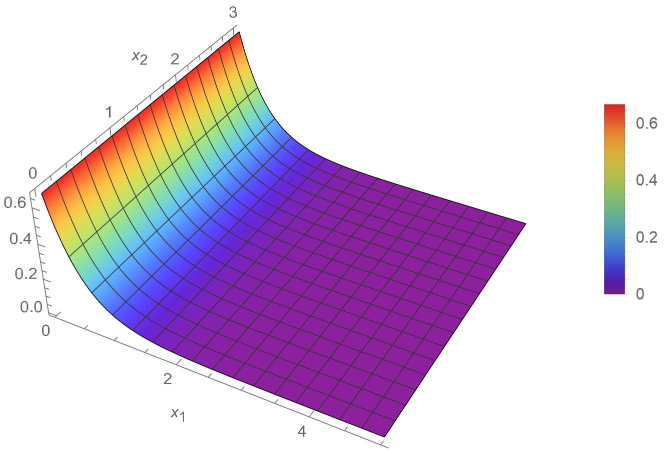
2. Για την μεγιστοποίηση της $Q(\theta,\theta_0)$ απαιτούμε $\frac{\partial Q(\theta,\theta_0)}{\partial\theta_1} = 0 \Rightarrow \frac{3}{\theta_1} - 6 = 0 \Rightarrow \theta_1 = \frac{1}{2}$

Επειδή $\left.\frac{\partial^2 Q(\theta,\theta_0)}{\partial^2\theta_1}\right|_{\theta_1=\frac{1}{2}} < 0$ το $\theta_1 = \frac{1}{2}$ είναι πράγματι σημείο τοπικού μεγίστου. Επίσης,

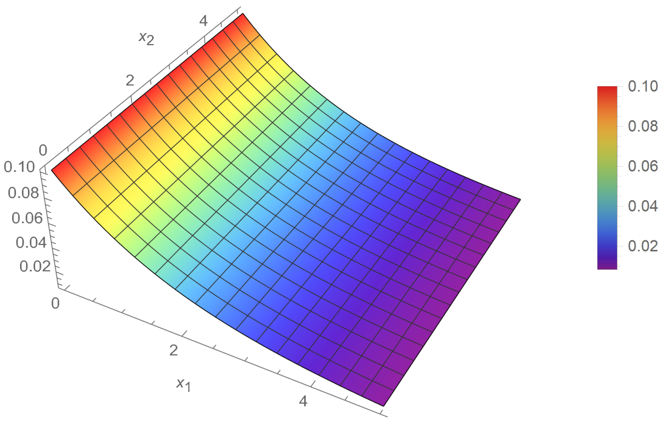
παρατηρούμε ότι η $Q(\theta,\theta_0)$ είναι γν. φθίνουσα ως προς θ_2 και μεγιστοποιείται όταν

$\theta_2 = 5$. Τελικά $\theta = \binom{1/2}{5}$.

3.



Σχήμα 1.8(a): Γραφική παράσταση της κατανομής $p(x_1, x_2)$ πριν από την εκτίμηση παραμέτρων.



Σχήμα 1.8(b): Γραφική παράσταση της κατανομής $p(x_1, x_2)$ μετά από την εκτίμηση παραμέτρων.