Αναγνώριση Προτύπων - 1η Σειρά Ασκήσεων

Ακαδ. Έτος 2021-2022

Ονοματεπώνυμο: Χρήστος Δημόπουλος

Αρ. Μητρώου: 031 17 037

e-mail adress: chrisdim1999@gmail.com

Άσκηση 1.1: (Maximum Likelihood estimation)

Έστω ότι η μεταβλητή x ακολουθεί μια κατανομή τύπου Erlang:

$$p(x|\theta) = \theta^2 x e^{-(\theta x)} u(x)$$

όπου $u(x)$ η βηματική συνάρτηση: $u(x) = \begin{cases} 1, & \text{if } x>0 \\ 0, & \text{if } x<0 \end{cases}$

Υποθέτουμε ότι η δείγματα $D = \{x_1, x_2, \ldots, x_n\}$ επιλέγονται ανεξάρτητα μεταξύ τους σύμφωνα με την κατανομή $p(x|\theta)$ (indipendent & identically distributed).

Maximum Likelihood Estimation for parameter $\theta$:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, P(D|\theta) \xrightarrow[\text{γνησίως αύξουσα}]{\ln(\cdot)} \hat{\theta} = \underset{\theta}{\operatorname{argmax}} \ln p(D|\theta)$$

$$\xRightarrow{x_i \text{ i.i.d.}} \hat{\theta} = \underset{\theta}{\operatorname{argmax}} \ln \prod_{i=1}^{n} p(x_i|\theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \ln p(x_i|\theta)$$

όπου:

$$\ln p(x_i|\theta) = \ln\left(\theta^2 x_i e^{-(\theta x_i)} u(x_i)\right) = 2\ln\theta + \ln x_i - \theta x_i + \ln u(x_i)$$

$$= 2\ln\theta - \theta x_i + \ln(x_i \cdot u(x_i))$$

Άρα: $$\ln p(D|\theta) = \sum_{i=1}^{n} \ln p(x_i|\theta) = 2n\ln\theta - \theta\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \ln(x_i \cdot u(x_i))$$

Προκειμένου να βρω το μέγιστο της άνωθι συνάρτησης, βρίσκω το σημείο μηδενισμού της παραγώγου της ως προς την παράμετρο $\theta$:

$$\frac{\partial \ln p(D|\theta)}{\partial \theta} = 0 \Rightarrow \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \quad , \theta \neq 0$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i} = \frac{2}{\bar{x}}} \quad , \text{ όπου } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ο αριθμητικός μέσος όρος των $n$ δειγμάτων.

## Άσκηση 1.2: (Minimax Criterion)

Έστω το minimax κριτήριο για τη zero-one συνάρτηση κόστους, δηλαδή $\lambda_{11} = \lambda_{22} = 0$ και $\lambda_{12} = \lambda_{21} = 1$ :

$$R(P(\omega_1)) = \lambda_{22} + (\lambda_{12} - \lambda_{22}) \int_{R_1} p(x|\omega_2) dx$$
$$+ P(\omega_1) \left[ (\lambda_{11} - \lambda_{22}) + (\lambda_{21} - \lambda_{11}) \int_{R_2} p(x|\omega_1) dx - (\lambda_{12} - \lambda_{22}) \int_{R_1} p(x|\omega_2) dx \right] \quad [1]$$

$$\overset{\text{zero-one}}{\Rightarrow} R(P(\omega_1)) = \int_{R_1} p(x|\omega_2) dx + P(\omega_1) \cdot \left[ + \int_{R_2} p(x|\omega_1) dx - \int_{R_1} p(x|\omega_2) dx \right]$$

1.) Η minimax λύση δίνεται θέτοντας την παράγωγο του άνωθι κριτηρίου ως προς την a-priori $P(\omega_1)$ ίση με μηδέν:

$$\frac{\partial R(P(\omega_1))}{\partial P(\omega_1)} = 0 \Rightarrow \int_{R_2} p(x|\omega_1) dx - \int_{R_1} p(x|\omega_2) dx = 0$$

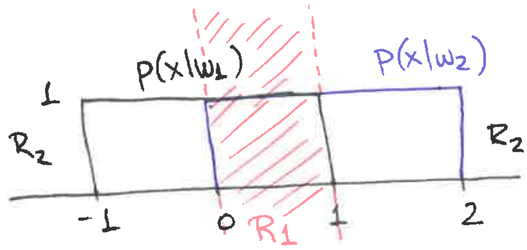
$$\Rightarrow \boxed{\int_{R_2} p(x|\omega_1) dx = \int_{R_1} p(x|\omega_2) dx}$$

2) Η παραπάνω λύση δεν είναι απαραίτητα μοναδική. Δίνεται το εξής αντιπαράδειγμα:

Έστω ίσες a-priori $P(w_1) = P(w_2) = 1/2$ και επίσης:

$$p(x|w_1) = \begin{cases} 1, & -1 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \qquad p(x|w_2) = \begin{cases} 1, & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

**$1^{st}$ Case:** $R_1 = [0, 1]$ και $R_2 = [-\infty, 0) \cup (1, +\infty)$

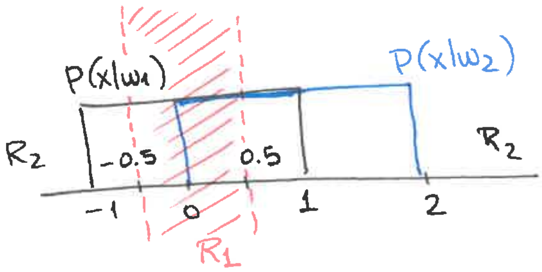



Τότε:

- $\int_{R_1} p(x|w_2)dx = \int_0^1 p(x|w_2)dx = \int_0^1 1\,dx = 1$
- $\int_{R_2} p(x|w_1)dx = \int_{-1}^{0} 1\,dx = 1$

Δηλαδή $\boxed{\int_{R_2} p(x|w_1)dx = \int_{R_1} p(x|w_2)dx}$

**$2^{nd}$ Case:** $R_1 = [-0.5, 0.5]$ και $R_2 = (-\infty, -0.5) \cup (0.5, +\infty)$




Τότε:

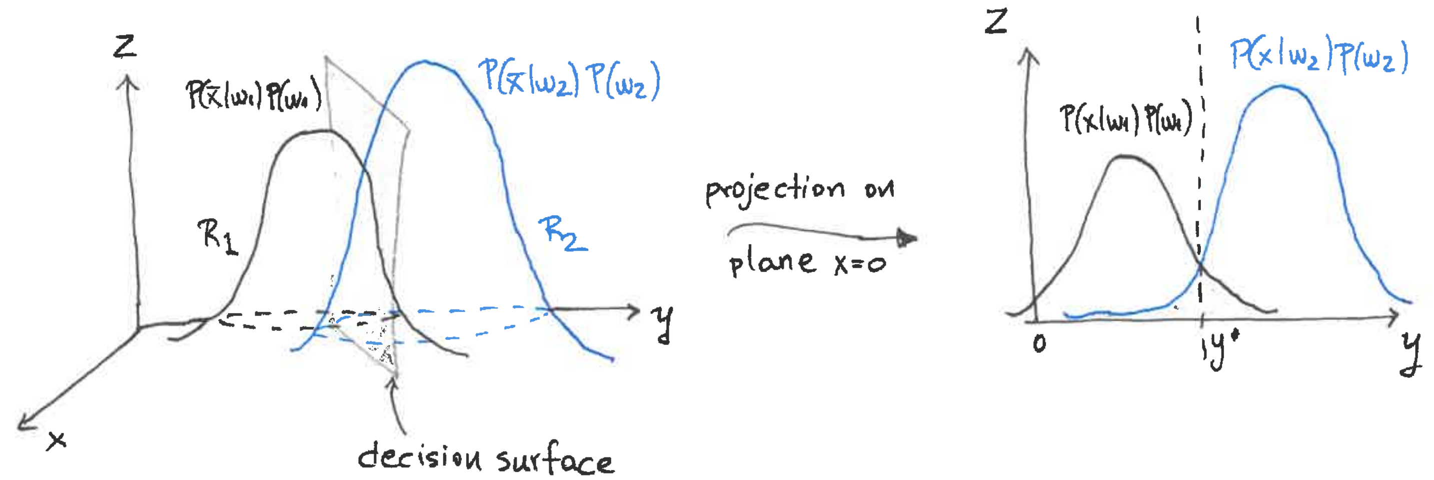
- $\int_{R_1} p(x|w_2)dx = \int_0^{0.5} 1\,dx = 0.5$
- $\int_{R_2} p(x|w_1)dx = \int_{-1}^{-0.5} 1\,dx = 0.5$

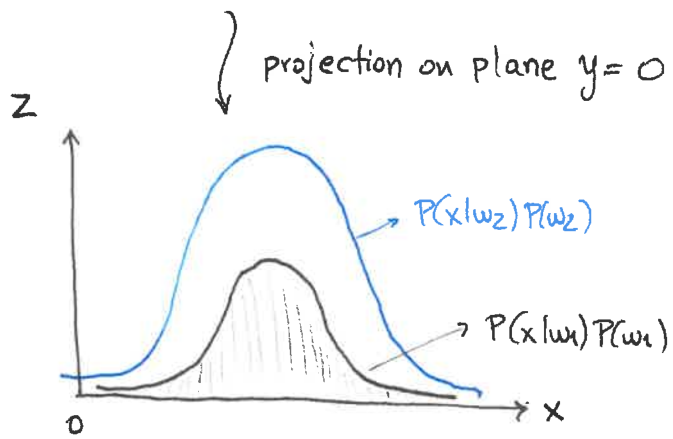
Δηλαδή $\boxed{\int_{R_2} p(x|w_1)dx = \int_{R_1} p(x|w_2)dx}$

# Άσκηση 1.3 : (Bayes Error)

Έστω δύο ανεξάρτητες κατανομές $p(\bar{x}|w_i)$ με priors $p(w_i)$, $i=1,2$. Υποθέτουμε ότι οι κατανομές αυτές ορίζονται για ένα χώρο διάστασης $d=2$.

1)




Έστω ότι οι κατανομές μας είναι κανονικές (normal distributions). Χωρίς βλάβη της γενικότητας, προβάλλουμε τις κατανομές στο επίπεδο $y=0$, δηλαδή μειώνουμε τη διαστατικότητα του προβλήματος σε $d'=d-1=1$, και υποθέτουμε ότι ισχύει $P(x|w_1)P(w_1) < P(x|w_2)\cdot P(w_2)$, $\forall x \in \mathbb{R}$. Τότε, αν ονομάσουμε $R_2$ την περιοχή στην οποία $P(x|w_2)P(w_2) > P(x|w_1)P(w_1)$, θα ισχύει:

$$P(\text{error}|1D) = \int_{R_2} P(x|w_1)P(w_1)\,dx = \int_{-\infty}^{+\infty} P(x|w_1)P(w_1)\,dx = P(w_1)\cdot \cancelto{1}{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\,dx}$$

$$= P(w_1). \quad (1)$$

Για τον χώρο διάστασης $d=2$, ορίζουμε τις περιοχές:

- $\underline{\underline{R_1}}$: όπου $P(\bar{x}|w_1)P(w_1) > P(\bar{x}|w_2)P(w_2)$
- $\underline{\underline{R_2}}$: όπου $P(\bar{x}|w_2)P(w_2) > P(\bar{x}|w_1)P(w_1)$

Τότε: $$P(error|2D) = \iint_{R_1} P(\bar{x}|w_2)P(w_2)\,d\bar{x} + \iint_{R_2} P(\bar{x}|w_1)P(w_1)\,d\bar{x} \quad (2)$$

Θα κινηθούμε με απαγωγή εις άτοπον. Έστω ότι ισχύει:

$$P(error|2D) > P(error|1D)$$

$$\overset{(1),(2)}{\Rightarrow} \iint_{R_1} P(\bar{x}|w_2)P(w_2)\,d\bar{x} + \iint_{R_2} P(\bar{x}|w_1)P(w_1)\,d\bar{x} > P(w_1)$$

$$\Rightarrow \iint_{R_1} P(\bar{x}|w_2)P(w_2)\,d\bar{x} > P(w_1)\cdot\left[1 - \iint_{R_2} P(\bar{x}|w_1)\,d\bar{x}\right]$$

$$\Rightarrow \iint_{R_1} P(\bar{x}|w_2)P(w_2)\,d\bar{x} > P(w_1)\iint_{R_1} P(\bar{x}|w_1)\,d\bar{x} \quad , \text{διότι } R_1 \cup R_2 = \mathbb{R}^2.$$

$$\Rightarrow \iint_{R_1} P(\bar{x}|w_2)P(w_2)\,d\bar{x} > \iint_{R_1} P(\bar{x}|w_1)P(w_1)\,d\bar{x}$$

Άτοπο, αφού εξ ορισμού στην περιοχή $R_1$ υποθέσαμε ότι ισχύει:

$$P(\bar{x}|w_1)P(w_1) > P(\bar{x}|w_2)P(w_2) \geq 0$$

Συνεπώς, θα ισχύει $\boxed{P(error|2D) \leq P(error|1D)}$

2) Έστω ότι οι κατανομές πλέον είναι οποιασδήποτε μορφής. Θεωρούμε μια γραμμή $\ell$ στον χώρο διάστασης $d=2$ και την προβολή της στον χώρο διάστασης $d'=1$.

Τότε:

$$P(\text{error}|2D) = \int_{\ell} \min\{P(\bar{x}|\omega_1)P(\omega_1),\ P(\bar{x}|\omega_2)P(\omega_2)\}dx$$

$$= \int_{\Gamma_2} P(\bar{x}|\omega_1)P(\omega_1)\,dx + \int_{\Gamma_1} P(\bar{x}|\omega_2)P(\omega_2)\,dx$$

όπου $\Gamma_1$: το τμήμα της γραμμής $\ell$ για το οποίο $P(\bar{x}|\omega_1)P(\omega_1) > P(\bar{x}|\omega_2)P(\omega_2)$

$\Gamma_2$: το τμήμα της γραμμής $\ell$ για το οποίο $P(\bar{x}|\omega_2)P(\omega_2) > P(\bar{x}|\omega_1)P(\omega_2)$.

Προβάλλοντας στον χώρο $d=1$, υποθέτουμε χωρίς βλάβη της γενικότητας ότι ισχύει:

$$\int_{\ell} P(x|\omega_1)P(\omega_1)\,dx < \int_{\ell} P(x|\omega_2)P(\omega_2)\,dx$$

Τότε:

$$P(\text{error}|1D) = \int_{\ell} P(x|\omega_1)P(\omega_1)\,dx = \int_{\Gamma_1\cup\Gamma_2} P(x|\omega_1)P(\omega_1)\,dx =$$

$$= \int_{\Gamma_1} P(x|\omega_1)P(\omega_1)\,dx + \int_{\Gamma_2} P(x|\omega_1)P(\omega_1)\,dx$$

Όμως στο τμήμα $\Gamma_1$ υποθέσαμε ότι $P(x|\omega_1)P(\omega_1) > P(x|\omega_2)P(\omega_2)$

$\Rightarrow P(x|\omega_1)P(\omega_1) = P(x|\omega_2)P(\omega_2) + |f(x)|$, για κάποιο $|f(x)| \geq 0$.

Άρα: $P(\text{error}|1D) = \underbrace{\int_{\Gamma_1} P(x|\omega_2)P(\omega_2)\,dx + \int_{\Gamma_2} P(x|\omega_1)P(\omega_1)\,dx}_{P(\text{error}|2D)} + \int_{\Gamma_1} |f(x)|\,dx$

$$\Rightarrow P(\text{error}|1D) = P(\text{error}|2D) + \int_{\Gamma_1} |f(x)|\,dx \Rightarrow \boxed{P(\text{error}|1D) \geq P(\text{error}|2D)}$$

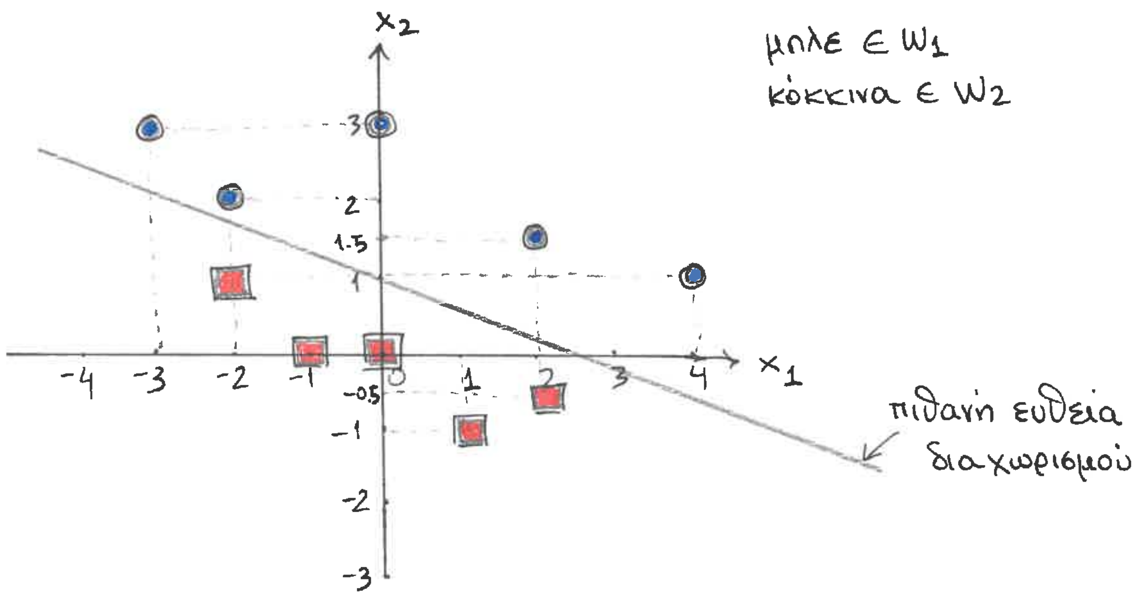
## Άσκηση 1.4: (Perceptrons)

Μας δίνονται 10 διανύσματα χαρακτηριστικών που προέρχονται από δύο κλάσεις $\omega_1$ και $\omega_2$:

$\omega_1$ : $[2, 1.5]^T$, $[4, 1]^T$, $[-3, 3]^T$, $[-2, 2]^T$, $[0, 3]^T$

$\omega_2$ : $[-1, 0]^T$, $[0, 0]^T$, $[2, -0.5]$, $[1, -1]^T$, $[-2, 1]^T$

Σε πρώτο στάδιο, σχεδιάζουμε τα παραπάνω δείγματα στο επίπεδο:




Όπως φαίνεται, τα δείγματα είναι linearly separable. Στη συνέχεια, κάνουμε χρήση του δοθέντος αλγορίθμου perceptron με $\rho = 1$ και $w(0) = [0, 0]^T$. Ωστόσο, παρατηρούμε ότι η διάσταση του συγκεκριμένου διανύσματος βαρών δεν επαρκεί για τον γραμμικό διαχωρισμό των δειγμάτων, καθώς δεν υπάρχει ευθεία που να διέρχεται από την αρχή των αξόνων (0,0) και να χωρίζει τέλεια τα δείγματα στις δύο κλάσεις $\Rightarrow$ πρέπει να προστεθεί κάποιο bias $w_0$.

Για τον λόγο αυτό, ως αρχικό διάνυσμα βαρών έχουμε το $w(0) = [0, 0, 0]^T$ και επαυξάνουμε τη διάσταση των δειγμάτων, προσθέτοντας έναν επιπλέον άσσο.

$\cdot\ x_{11} = [1, 2, 1.5]^T \in \omega_1$ και $w(0)^T x_{11} = 0 \leq 0$

άρα $w(1) = w(0) + \rho x_{11} \Rightarrow \underline{w(1) = [1, 2, 1.5]^T}$

$\cdot\ x_{12} = [1, 4, 1]^T \in \omega_1$ και $w(1)^T x_{12} = 1 + 8 + 1.5 = 10.5 > 0$

άρα $w(2) = w(1)$

$\cdot\ x_{13} = [1, -3, 3]^T \in \omega_1$ και $w(2)^T x_{13} = 1 - 6 + 4.5 \leq 0$

άρα $w(3) = w(2) + \rho x_{13} \Rightarrow \underline{w(3) = [2, -1, 4.5]^T}$

$\cdot\ x_{14} = [1, -2, 2]^T \in \omega_1$ και $w(3)^T x_{14} = 2 + 2 + 9 > 0$

άρα $w(4) = w(3)$

$\cdot\ x_{15} = [1, 0, 3]^T \in \omega_1$ και $w(4)^T x_{15} = 2 + 3 \times 4.5 > 0$

άρα $w(5) = w(4)$

$\cdot\ x_{21} = [1, -1, 0]^T \in \omega_2$ και $w(5)^T x_{21} = 2 + 2 = 4 \geq 0$

άρα $w(6) = w(5) - \rho x_{21} \Rightarrow \underline{w(6) = [1, 0, 4.5]^T}$

$\cdot\ x_{22} = [1, 0, 0]^T \in \omega_2$ και $w(6)^T x_{22} = 1 \geq 0$

άρα $w(7) = w(6) - \rho x_{22} \Rightarrow \underline{w(7) = [0, 0, 4.5]^T}$

$\cdot\ x_{23} = [1, 2, -0.5]^T \in \omega_2$ και $w(7)^T x_{23} = -0.5 \times 4.5 < 0$

άρα $w(8) = w(7)$

$\cdot\ x_{24} = [1, 1, -1]^T \in \omega_2$ και $w(8)^T x_{24} = -4.5 < 0$

άρα $w(9) = w(8)$

$\cdot\ x_{25} = [1, -2, 1]^T \in \omega_2$ και $w(9)^T x_{25} = 4.5 \geq 0$

άρα $w(10) = w(9) - \rho x_{25} \Rightarrow \underline{w(10) = [-1, 2, 3.5]^T}$

---

$\cdot\ x_{11} = [1, 2, 1.5]^T \in \omega_1$ και $w(10)^T x_{11} = -1 + 4 + 3.5 \times 1.5 > 0$

άρα $w(11) = w(10)$

· $x_{12} = [1, 4, 1]^T \in \omega_1$ και $w(11)^T x_{12} = -1+8+3.5 > 0$

άρα $w(12) = w(11)$

· $x_{13} = [1, -3, 3]^T \in \omega_1$ και $w(12)^T x_{13} = -1-6+10.5 = 3.5 > 0$

άρα $w(13) = w(12)$

· $x_{14} = [1, -2, 2] \in \omega_1$ και $w(13)^T x_{14} = -1-4+7 = +2 > 0$

άρα $w(14) = w(13)$

· $x_{15} = [1, 0, 3]^T \in \omega_1$ και $w(14)^T x_{15} = -1 + 3 \times 3.5 > 0$

άρα $w(15) = w(14)$

· $x_{21} = [1, -1, 0]^T \in \omega_2$ και $w(15)^T x_{21} = -1-2 = -3 < 0$

άρα $w(16) = w(15)$

· $x_{22} = [1, 0, 0]^T \in \omega_2$ και $w(16)^T x_{22} = -1 < 0$

άρα $w(17) = w(16)$

· $x_{23} = [1, 2, -0.5]^T \in \omega_2$ και $w(17)^T x_{23} = -1 + 4 - \frac{3.5}{2} = 1.25 \geq 0$

άρα $w(18) = w(17) - \rho x_{23} \Rightarrow \underline{w(18) = [-2, 0, 4]^T}$

· $x_{24} = [1, 1, -1] \in \omega_2$ και $w(18)^T x_{24} = -2-4 = -6 < 0$

άρα $w(19) = w(18)$

· $x_{25} = [1, -2, 1]^T \in \omega_2$ και $w(19)^T x_{25} = -2+4 = 2 \geq 0$

άρα $w(20) = w(19) - \rho x_{25} \Rightarrow \underline{w(20) = [-3, 2, 3]^T}$

· $x_{11} = [1, 2, 1.5]^T \in \omega_1$ και $w(20)^T x_{11} = -3+4+3 \times 1.5 > 0$

άρα $w(21) = w(20)$

· $x_{12} = [1, 4, 1]^T \in \omega_1$ και $w(21)^T x_{12} = -3+8+3 = 8 > 0$

άρα $w(22) = w(21)$

· $x_{13} = [1,-3,3]^T \in \omega_1$ και $w(22)^T x_{13} = -3-6+9 = 0 \leq 0$

άρα $w(23) = w(22) + \rho x_{13} \Rightarrow \underline{w(23) = [-2,-1,6]^T}$

· $x_{14} = [1,-2,2]^T \in \omega_1$ και $w(23)^T x_{14} = -2+2+12 = 12 > 0$

άρα $w(24) = w(23)$

· $x_{15} = [1,0,3]^T \in \omega_1$ και $w(24)^T x_{15} = -2+18 = 16 > 0$

άρα $w(25) = w(24)$

· $x_{21} = [1,-1,0]^T \in \omega_2$ και $w(25)^T x_{21} = -2+1 = -1 < 0$

άρα $w(26) = w(25)$

· $x_{22} = [1,0,0]^T \in \omega_2$ και $w(26)^T x_{22} = -2 < 0$, άρα $w(27) = w(26)$

· $x_{23} = [1,2,-0.5]^T \in \omega_2$ και $w(27)^T x_{23} = -2-2-3 = -7 < 0$, άρα $w(28) = w(27$

· $x_{24} = [1,1,-1]^T \in \omega_2$ και $w(28)^T x_{24} = -2-1-6 = -9 < 0$, άρα $w(29) = w(2$

· $x_{25} = [1,-2,1]^T \in \omega_2$ και $w(29)^T x_{25} = -2+2+6 = 6 \geq 0$

άρα $w(30) = w(29) - \rho x_{25} \Rightarrow \underline{w(30) = [-3,1,5]^T}$

· $x_{11} = [1,2,1.5]^T \in \omega_1$ και $w(30)^T x_{11} = -3+2+5\times1.5 > 0$, άρα $w(31) = w(3$

· $x_{12} = [1,4,1]^T \in \omega_1$ και $w(31)^T x_{12} = -3+4+5 = 6 > 0$, άρα $w(32) = w(31)$

· $x_{13} = [1,-3,3]^T \in \omega_1$ και $w(32)^T x_{13} = -3-3+15 > 0$, άρα $w(33) = w(32)$

· $x_{14} = [1,-2,2]^T \in \omega_1$ και $w(33)^T x_{14} = -3-2+10 > 0$, άρα $w(34) = w(33)$

· $x_{15} = [1,0,3]^T \in \omega_1$ και $w(34)^T x_{15} = -3+15 > 0$, άρα $w(35) = w(34)$

· $x_{21} = [1,-1,0]^T \in \omega_2$ και $w(35)^T x_{21} = -3-3 = -6 < 0$, άρα $w(36) = w(35)$

· $x_{22} = [1,0,0]^T \in \omega_2$ και $w(36)^T x_{22} = -3 < 0$, άρα $w(37) = w(36)$

· $x_{23} = [1,2,-0.5]^T \in \omega_2$ και $w(37)^T x_{23} = -3+2-2.5 < 0$, άρα $w(38) = w(37)$

· $x_{24} = [1,1,-1]^T \in \omega_2$ και $w(38)^T x_{24} = -3+1-5 < 0$, άρα $w(39) = w(38)$

· $x_{25} = [1,-2,1] \in \omega_2$ και $w(39)^T x_{25} = -3-2+5 = 0 \geq 0$

Άρα: $w(40) = w(39) - \rho x_{25} \Rightarrow \underline{w(40) = [-4,3,4]^T}$

- $x_{11} = [1, 2, 1.5]^T \in \omega_1$ και $w(40)^T x_{11} = -4 + 6 + 4 \times 1.5 > 0$, $w(41) = w(40)$
- $x_{12} = [1, 4, 1]^T \in \omega_1$ και $w(41)^T x_{12} = -4 + 12 + 4 > 0$, $w(42) = w(41)$
- $x_{13} = [1, -3, 3]^T \in \omega_1$ και $w(42)^T x_{13} = -4 - 9 + 12 = -1 \leq 0$

  άρα $w(43) = w(42) + \rho x_{13} \Rightarrow \underline{w(43) = [-3, 0, 7]^T}$

- $x_{14} = [1, -2, 2]^T \in \omega_1$ και $w(43)^T x_{14} = -3 + 14 > 0$ άρα $w(44) = w(43)$
- $x_{15} = [1, 0, 3]^T \in \omega_1$ και $w(44)^T x_{15} = -3 + 21 > 0$, άρα $w(45) = w(44)$
- $x_{21} = [1, -1, 0]^T \in \omega_2$ και $w(45)^T x_{21} = -3 < 0$, $w(46) = w(45)$
- $x_{22} = [1, 0, 0]^T \in \omega_2$ και $w(46)^T x_{22} = -3 < 0$, $w(47) = w(46)$
- $x_{23} = [1, 2, -0.5]^T \in \omega_2$ και $w(47)^T x_{23} = -3 - 7 \cdot 0.5 < 0$, $w(48) = w(47)$
- $x_{24} = [1, 1, -1]^T \in \omega_2$ και $w(48)^T x_{24} = -3 - 7 = -10 < 0$, $w(49) = w(48)$
- $x_{25} = [1, -2, 1]^T \in \omega_2$ και $w(49)^T x_{25} = -3 + 7 = 4 \geq 0$

  άρα $w(50) = w(49) - \rho x_{25} \Rightarrow \boxed{w(50) = [-4, 2, 6]^T}$

- $x_{11} = [1, 2, 1.5]^T \in \omega_1$ και $w(50)^T x_{11} = -4 + 4 + 6 \times 1.5 > 0$, $w(51) = w(50)$
- $x_{12} = [1, 4, 1]^T \in \omega_1$ και $w(51)^T x_{12} = -4 + 8 + 6 > 0$, $w(52) = w(51)$
- $x_{13} = [1, -3, 3]^T \in \omega_1$ και $w(52)^T x_{13} = -4 - 6 + 3 \times 6 = 8 > 0$, $w(53) = w(52)$
- $x_{14} = [1, -2, 2]^T \in \omega_1$ και $w(53)^T x_{14} = -4 - 4 + 12 = 8 > 0$, $w(54) = w(53)$
- $x_{15} = [1, 0, 3]^T \in \omega_1$ και $w(54)^T x_{15} = -4 + 3 \times 6 = 14 > 0$, $w(55) = w(54)$
- $x_{21} = [1, -1, 0]^T \in \omega_2$ και $w(55)^T x_{21} = -4 - 2 = -6 < 0$, $w(56) = w(55)$
- $x_{22} = [1, 0, 0]^T \in \omega_2$ και $w(56)^T x_{22} = -4 < 0$, $w(57) = w(56)$
- $x_{23} = [1, 2, -0.5]^T \in \omega_2$ και $w(57)^T x_{23} = \cancel{-4} + \cancel{4} - 3 < 0$, $w(58) = w(57)$
- $x_{24} = [1, 1, -1]^T \in W_2$ και $w(58)^T x_{24} = -4 + 2 - 6 < 0$, $w(59) = w(58)$
- $x_{25} = [1, -2, 1]^T \in W_2$ και $w(59)^T x_{25} = -4 - 4 + 6 = -2 < 0$, $w(60) = w(59)$

Τελικώς, τα δείγματα ξεχωρίζονται πλήρως για διάνυσμα βαρών

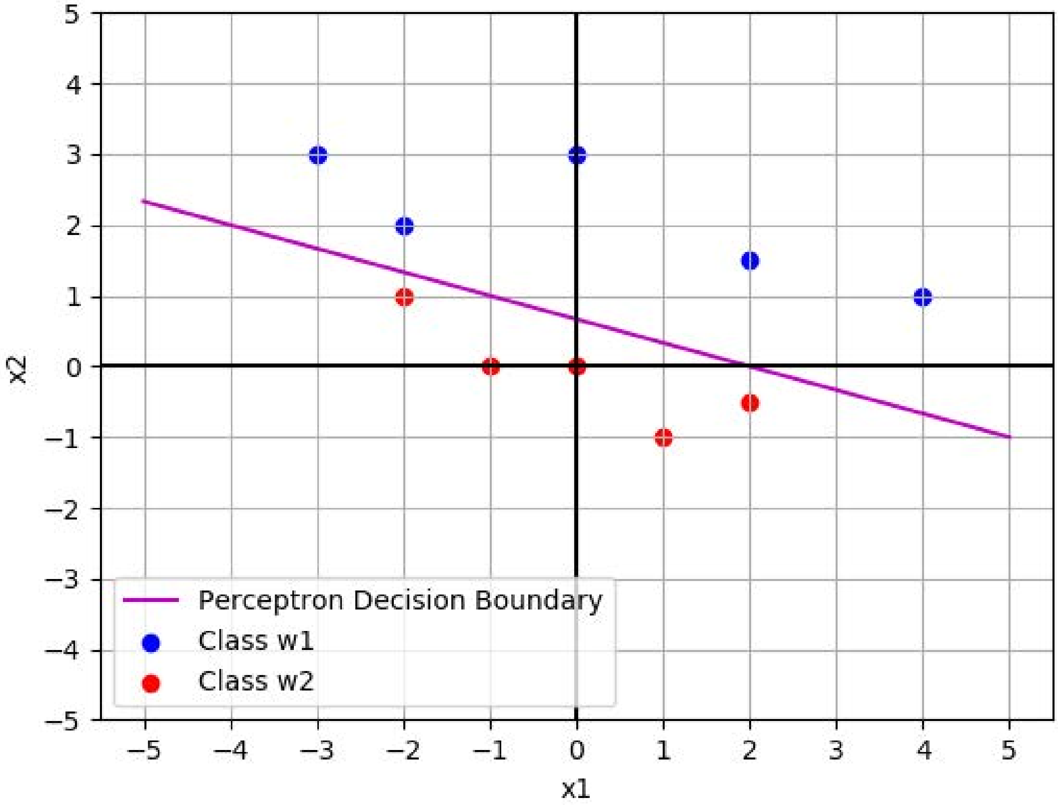
$\boxed{W = [-4, 2, 6]^T}$, μετά από 60 επαναλήψεις!

Τέλος, η διαχωριστική καμπύλη που αντιστοιχεί στο υπολογισθέν διάνυσμα βαρών είναι:

$$g(\bar{x}) = w_0 + w_1 x_1 + w_2 x_2 = 0 \Rightarrow -4 + 2x_1 + 6x_2 = 0$$

$$\Rightarrow \boxed{x_2 = -\frac{1}{3}x_1 + \frac{2}{3}}$$

Παρακάτω φαίνεται ο διαχωρισμός των δειγμάτων στις δύο κλάσεις μέσω της εν λόγω ευθείας διαχωρισμού:



# Άσκηση 1.5: Kullback-Leibler Divergence

Η μετρική Kullback-Leibler ορίζεται ως εξής:

$$D_{KL}(p_1(x), p_2(x)) = \int p_1(x) \ln \frac{p_1(x)}{p_2(x)}\, dx \qquad (I)$$

και εκφράζει την ποσότητα της πληροφορίας που χάνεται, όταν χρησιμοποιούμε την κατανομή $p_2(x)$ ως προσέγγιση της κατανομής $p_1(x)$.

Έστω ότι η $p_2(x)$ είναι κανονική κατανομή $\mathcal{N}(\mu, \Sigma)$ (multidimensional) και χρησιμοποιείται για να προσεγγίσει μια οποιαδήποτε κατανομή $p_1(x)$.

Αντικαθιστώντας:

$$p_2(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \cdot \exp\left\{-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right\}$$

στην (I), έχουμε:

$$D_{KL}(p_1,p_2) = \int p_1(x)\ln p_1(x)\,dx + \int p_1\cdot\left[-\frac{d}{2}\cdot\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]d$$

$$= \int p_1(x)\ln p_1(x)\,dx + \int \frac{1}{2}p_1\left[-d\ln(2\pi) - \ln|\Sigma| - \cancel{\tfrac{1}{2}}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]dx$$

Αναζητάμε τις παραμέτρους $\mu$, $\Sigma$ που ελαχιστοποιούν την "απόσταση" των δύο κατανομών, όπως αυτή εκφράζεται μέσω της KL Divergence. Παραγωγίζουμε ως προς τις παραμέτρους και θέτουμε με μηδέν:

- $$\frac{\partial D_{KL}(p_1(x), p_2(x))}{\partial \mu} = -\int \Sigma^{-1}(x-\mu)\cdot p_1(x)\,dx = 0$$ (διότι $(\Sigma^{-1})^T = (\Sigma^T)^{-1}$ και $\Sigma = \Sigma^T$ λόγω συμμετρίας)

$$\Rightarrow \Sigma^{-1}\int (x-\mu)\,p_1(x)\,dx = 0$$

Υποθέτουμε ότι το μητρώο $\Sigma$ είναι non-singular ($\det\Sigma \neq 0$), οπότε:

$$\int (\underline{x}-\mu)\,p_1(\underline{x})\,d\underline{x} = 0 \Rightarrow E_{x\sim p_1}[x-\mu] = 0$$

$$\Rightarrow \boxed{\mu = E_{x\sim p_1}[x]}$$

• Για τον υπολογισμό της παραγώγου ως προς $\Sigma$ κάνουμε χρήση των εξής εξισώσεων:

$$\frac{\partial \ln|X|}{\partial X} = (X^{-1})^T = (X^T)^{-1} \quad \text{(Εξίσωση 56 του [3])}$$

$$\frac{\partial a^T X^{-1} b}{\partial X} = -X^T a b^T X^{-T} \quad \text{(Εξίσωση 61 του [3])}$$

Άρα:

$$\frac{\partial D_{KL}(p_1(x), p_2(x))}{\partial \Sigma} = -\frac{1}{2}\int p_1(x)\cdot\left[\frac{\partial \ln|\Sigma|}{\partial \Sigma} + \frac{\partial (x-\mu)^T\Sigma^{-1}(x-\mu)}{\partial \Sigma}\right]dx$$

$$= -\frac{1}{2}\int p_1(x)\cdot\left[(\Sigma^T)^{-1} - \Sigma^{-T}(x-\mu)\cdot(x-\mu)^T\Sigma^{-T}\right]dx \quad \left(\begin{matrix}\text{symmetric}\\ \Sigma^T = \Sigma\end{matrix}\right)$$

$$= -\frac{1}{2}\int p_1(x)\cdot\left[\Sigma^{-1} - \Sigma^{-1}(x-\mu)(x-\mu)^T\Sigma^{-1}\right]dx = 0$$

Πολλαπλασιάζοντας από αριστερά και από δεξιά με το μητρώο $\Sigma$:

$$\int p_1(x)\cdot\left[\Sigma - (x-\mu)(x-\mu)^T\right]dx = 0$$

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

Συνεπώς, η μετρική KL-Divergence λαμβάνει τη μικρότερη τιμή της, εφόσον η κατανομή $p_1(x)$ που θέλουμε να προσεγγίσουμε έχει την ίδια μέση τιμή και διασπορά με τη Gaussian κατανομή $p_2(x) \sim \mathcal{N}(\mu, \Sigma)$.

Άσκηση 1.6 : (Linear Regression and the LMS Algorithm)

Θεωρούμε το ακόλουθο μοντέλο γραμμικής παλινδρόμησης :

$$y = w_0 + w_1 x$$

καθώς και το παρακάτω σύνολο δεδομένων στη μορφή $(x,y)$:

$$\left\{ \begin{pmatrix} 0.38 \\ 2.05 \end{pmatrix}, \begin{pmatrix} 0.44 \\ 2.23 \end{pmatrix}, \begin{pmatrix} 0.48 \\ 2.13 \end{pmatrix}, \begin{pmatrix} 0.54 \\ 2.33 \end{pmatrix}, \begin{pmatrix} 0.58 \\ 2.67 \end{pmatrix}, \begin{pmatrix} 0.64 \\ 2.68 \end{pmatrix}, \begin{pmatrix} 0.71 \\ 2.81 \end{pmatrix}, \begin{pmatrix} 0.76 \\ 2.97 \end{pmatrix}, \begin{pmatrix} 0.82 \\ 3.12 \end{pmatrix}, \begin{pmatrix} 0.96 \\ 3.20 \end{pmatrix} \right\}$$

για το οποίο γνωρίζουμε ότι έχει προκύψει έπειτα από επίδραση Γκαουσσιανού θορύβου.

1) Το γραμμικό μοντέλο υπό την επίδραση του Gaussian Noise, μπορεί να γραφεί ως εξής:

$$y = w_0 + w_1 x + \epsilon \quad , \text{όπου} \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

άρα $y \sim \mathcal{N}(w_0 + w_1 x, \sigma^2)$

Likelihood Function : $p(y|x; w_0, w_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(\bar{y}-(w_0+w_1\bar{x}))^2}{2\sigma^2} \right\}$

Όμως, υποθέτουμε ότι τα δείγματα του συνόλου δεδομένων, έστω D, είναι i.i.d. Οπότε:

$$P(y|x; w_0, w_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{(x,y)\in D} \exp\left\{ -\frac{(y-(w_0+w_1x))^2}{2\sigma^2} \right\}$$

$$\Rightarrow \ln p(y|x; w_0, w_1, \sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2) + \sum_{(x,y)\in D} \frac{-(y-(w_0+w_1x))^2}{2\sigma^2}$$

$$\Rightarrow \ln p(y|x; w_0, w_1, \sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2) - \sum_{(x,y)\in D} \frac{(y-(w_0+w_1x))^2}{2\sigma^2}$$

Αν θεωρήσουμε $\hat{y} = w_0 + w_1 x$, τότε:

$$\ln p(y|x; w_0, w_1, \sigma^2) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{(x,y)\in D} (y-\hat{y})^2$$

Δεδομένου ότι το διάνυσμα παραμέτρων που θέλουμε να εκτιμήσουμε είναι το $\theta = (w_0, w_1)$, δηλαδή τις παραμέτρους $\theta$, η διασπορά $\sigma^2$ μπορεί να αντιμετωπίστεί ως μια απλή σταθερά στο εν λόγω πρόβλημα βελτιστοποίησης.

Επομένως:

$$\underset{\theta=(w_0,w_1)}{\operatorname{argmax}} P(y|\theta;\sigma^2) \overset{\ln(\cdot)\nearrow}{=} \underset{\theta}{\operatorname{argmax}} \ln P(y|\theta;\sigma^2)$$

$$= \underset{\theta}{\operatorname{argmax}} -\sum_{(x,y)\in D}\left(y-(w_0+w_1x)\right)^2 = \underset{\theta=(w_0,w_1)}{\operatorname{argmin}} \sum_{(x,y)\in D}\left(y-(w_0+w_1x)\right)^2 = \underset{\theta}{\operatorname{argmin}} J_{SSE}$$

Συνεπώς, το πρόβλημα maximum likelihood estimation για ~~το προ~~ την περίπτωση του linear regression, είναι ισοδύναμο με το πρόβλημα ελαχιστοποίησης του κριτηρίου SSE (Sum of Squared Errors).

2) Έστω, λοιπόν, το SSE κριτήριο για n δείγματα:

$$J_{SSE} = \sum_{i=1}^{n}\left(y_i-(w_0+w_1x_i)\right)^2$$

Για να ελαχιστοποιείται, θέτουμε τις παραγώγους ως προς $w_0$ και $w_1$ ίσες με μηδέν:

- $$\frac{\partial J_{SSE}}{\partial w_0} = 0 \Rightarrow -2\sum_{i=1}^{n}\left(y_i-\hat{w}_0-\hat{w}_1x_i\right)=0$$

$$\Rightarrow \boxed{\hat{w}_0 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{w}_1x_i\right) = \bar{y}-\hat{w}_1\bar{x}} \quad (1)$$

όπου $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

- $$\frac{\partial J_{SSE}}{w_1}=0 \Rightarrow -2\sum_{i=1}^{n}\left(y_i-\hat{w}_0-\hat{w}_1x_i\right)\cdot x_i = 0$$

$$\Rightarrow \sum_{i=1}^{n}x_i\cdot y_i - \hat{w}_0\sum_{i=1}^{n}x_i - \hat{w}_1\sum_{i=1}^{n}x_i^2 = 0$$

$$\overset{(1)}{\Rightarrow} \sum_{i=1}^{n}x_iy_i - \left(\bar{y}-\hat{w}_1\bar{x}\right)\sum_{i=1}^{n}x_i - \hat{w}_1\sum_{i=1}^{n}x_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - n \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i + \hat{w}_1 \cdot \left( n \left( \sum_{i=1}^{n} x_i \right)^2 - \sum_{i=1}^{n} x_i^2 \right) = 0$$

$$\Rightarrow \boxed{\hat{w}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - n \left( \sum_{i=1}^{n} x_i \right)^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

3) Χρησιμοποιώντας ένα από τα παραπάνω δεδομένα τη φορά, και εφαρμόζοντας τον αλγόριθμο LMS, υπολογίζουμε εκ νέου τους άγνωστους συντελεστές $\bar{w} = (w_0, w_1)^T$. Αρχικά, αυξάνουμε τη διαστατικότητα των δειγμάτων ως $\bar{x} = \begin{bmatrix} 1 \\ x_i \end{bmatrix}$, ώστε:

$$J_{SSE} = \sum_{i=1}^{n} \left( y_i - (w_0 + w_1 x_i) \right)^2 = \sum_{i=1}^{n} \left( y_i - \bar{w}^T \bar{x} \right)^2$$

<u>Learning Rate</u>: $n(k) = \frac{n}{k+1}$, όπου $\boxed{n = 1}$

Αυθαίρετη αρχικοποίηση διανύσματος βαρών $w(0) = [1 \quad 1]^T$.

<u>LMS Rule</u>: 
$$\begin{cases} w(0) = [1 \quad 1]^T \text{ arbitrarily} \\ w(k+1) = w(k) + n(k) \underbrace{\left( y(k) - \underline{w}(k)^T \cdot \underline{x}(k) \right) x(k)}_{\nabla J} \end{cases}$$

<u>Για $k=0$</u>: $x(k) = [1 \quad 0.38]^T$, $y(k) = 2.05$

$$w(1) = w(0) + 1 \cdot (2.05 - 1.38) \cdot x(0)$$

$$\Rightarrow w(1) = [1 \quad 1]^T + [0.67 \quad 0.2546]^T$$

$$\Rightarrow w(1) = [1.67, \ 1.2546]^T$$

· Για $k=1$: $x(1)=[1 \quad 0.44]^T$, $y(1)=2.23$

$$w(2)=w(1)+\frac{1}{2}(2.23-(1.67+0.44\times 1.2546))\,x(1)$$

$$\Rightarrow w(2)=[1{,}67 \quad 1.2546]^T+[0.003988,\ 0.00175472]^T$$

$$\Rightarrow w(2)=[1.673988,\ 1.25635472]^T$$

· Για $k=2$: $x(2)=[1 \quad 0.48]^T$, $y(2)=2.13$

$$w(3)=w(2)+\frac{1}{3}(2.13-1.673988-0.48\times 1.25635472)\begin{bmatrix}1\\0.48\end{bmatrix}$$

$$\Rightarrow w(3)=[1.62497524,\ 1.2328286]^T$$

· Για $k=3$: $x(3)=[1 \quad 0.54]^T$, $y(3)=2.33$

$$w(4)=w(3)+\frac{1}{4}(2.33-1.62497524-1.2328286\times 0.54)\begin{bmatrix}1\\0.54\end{bmatrix}$$

$$\Rightarrow w(4)=[1.63479957,\ 1.23813373]^T$$

· Για $k=4$: $x(4)=[1 \quad 0.58]^T$, $y(4)=2.67$

$$w(5)=w(4)+\frac{1}{5}(2.67-1.63479957-1.23813373\times 0.58)\begin{bmatrix}1\\0.58\end{bmatrix}$$

$$\Rightarrow w(5)=[1.69821,\ 1.2749]^T$$

· Για $k=5$: $x(5)=[1 \quad 0.64]^T$, $y(5)=2.68$

$$w(6)=w(5)+\frac{1}{6}(2.68-1.69821-1.2749\times 0.64)\begin{bmatrix}1\\0.64\end{bmatrix}$$

$$\Rightarrow w(6)=[1.7258,\ 1.2926]^T$$

· Για $k=6$: $x(6)=[1,\ 0.71]^T$ και $y(6)=2.81$

$$w(7)=w(6)+\frac{1}{7}(2.81-1.7258-0.71\times 1.296)[1 \quad 0.71]^T$$

$$\Rightarrow w(7)=[1.74962 \quad 1.30948]^T$$

• <u>Για k=7:</u> $x(7)=[1 \quad 0.76]^T$ , $y(7)=2.97$

$$w(8) = w(7) + \frac{1}{8}(2.97 - 1.74962 - 1.30948 \times 0.76)\begin{bmatrix}1\\0.76\end{bmatrix}$$

$$\Rightarrow w(8) = [1.7777 \;,\; 1.3308]^T$$

• <u>Για k=8:</u> $x(8)=[1 \quad 0.82]^T$ , $y(8)=3.12$

$$w(9) = w(8) + \frac{1}{9}(3.12 - 1.777 - 1.3308 \times 0.82)\begin{bmatrix}1\\0.82\end{bmatrix}$$

$$\Rightarrow w(9) = [1.80565 \;,\; 1.3537]^T$$

• <u>Για k=9:</u> $x(9)=[1 \quad 0.96]^T$ και $y(9)=3.2$

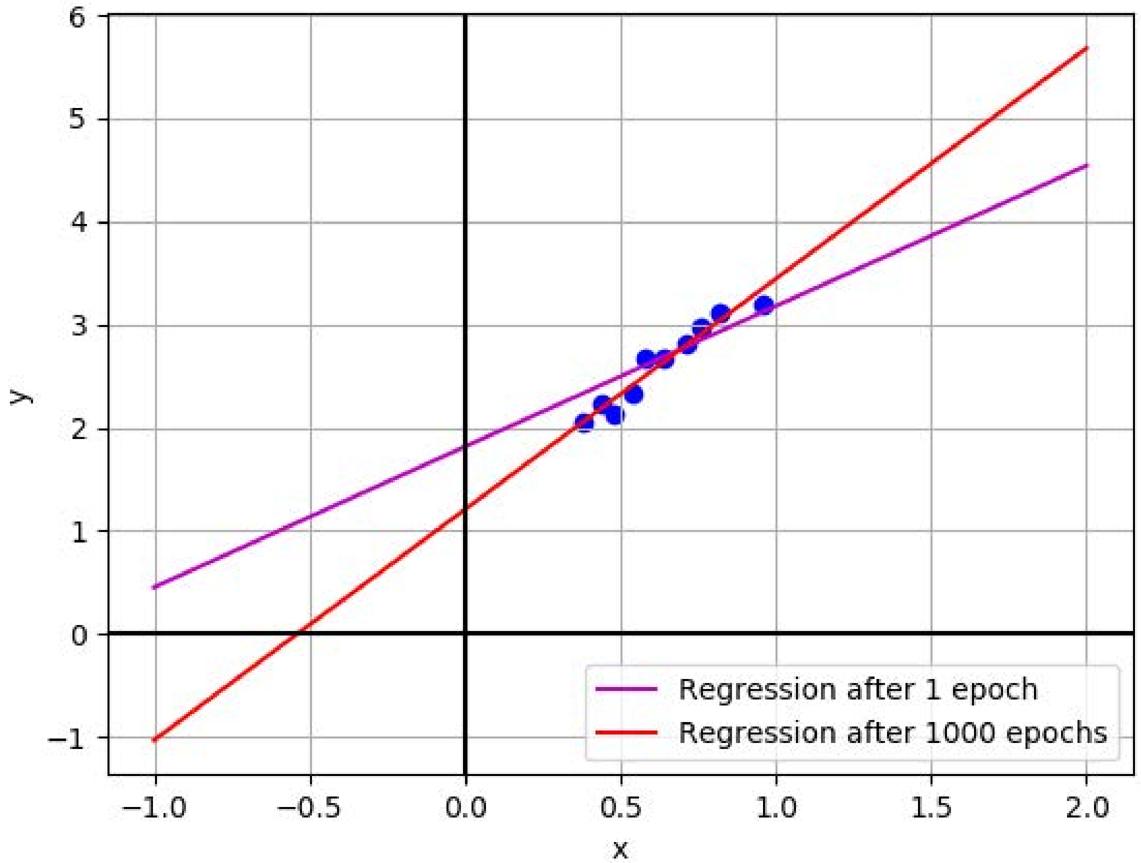
$$w(10) = w(9) + \frac{1}{10}(3.2 - 1.80565 - 1.3537 \times 0.96)\cdot\begin{bmatrix}1\\0.96\end{bmatrix}$$

$$\Rightarrow \boxed{w(10) = [1.81513 \;,\; 1.36283]^T}$$ , τέλος 1ης εποχής.

4) Γράφουμε κώδικα που να εφαρμόζει τον LMS στο συγκεκριμένο πρόβλημα για 1000 εποχές και εν τέλει βρίσκουμε:

$$\boxed{w = [1.2062337 6 \;,\; 2.23582049]^T}$$

5) Παρουσιάζουμε σε κοινό γράφημα, οι ευθείες που βρέθηκαν ως λύσεις για 1 και 1000 εποχές:



6
5
4
3
2
1
0
−1
y
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
x
Regression after 1 epoch
Regression after 1000 epochs

## Άσκηση 1.7 : (Bayes meets KNN)

Έστω δύο ταξινομητές οι οποίοι βασίζονται σε δείγματα που προέρχονται από τις δύο κατανομές :

$$P(x|w_1) = \begin{cases} 2x, & \text{for } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$p(x|w_2) = \begin{cases} 2-2x, & \text{for } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

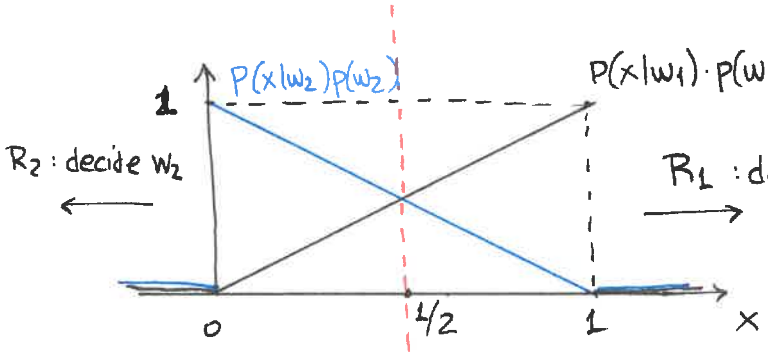
Επιπλέον οι a-priori πιθανότητες για κάθε κατηγορία είναι $p(w_1) = p(w_2) = 1/2$.

1) Κανόνας απόφασης κατά Bayes: $\hat{w} = \underset{i=1,2}{\operatorname{argmax}} P(w_i) \cdot P(x|w_i)$

Για $0 \le x \le 1$: $p(x|w_1) \cdot p(w_1) = p(x|w_2) \cdot p(w_2)$

$$\Rightarrow 2x \cdot \cancel{\tfrac{1}{2}} = (2-2x) \cdot \cancel{\tfrac{1}{2}} \Rightarrow 4x = 2 \Rightarrow \boxed{x = 1/2}$$

Decision Point




$$P(\text{error}) = \int_{-\infty}^{+\infty} P(\text{error}, x)\,dx = \int_{-\infty}^{+\infty} P(\text{error}|x) \cdot P(x)\,dx$$

$$= \int_{R_1} P(x|w_2) \cdot P(w_2)\,dx + \int_{R_2} P(x|w_1) P(w_1)\,dx$$

$$= \int_{1/2}^{1} \left(\tfrac{1}{2}\right) \cdot (2-2x)\,dx + \int_{0}^{1/2} \left(\tfrac{1}{2}\right) \cdot 2x\,dx = \int_{1/2}^{1} (1-x)\,dx + \int_{0}^{1/2} x\,dx$$

$$= \left[x - \frac{x^2}{2}\right]_{1/2}^{1} + \left[\frac{x^2}{2}\right]_{0}^{1/2} = \cancel{1} - \frac{\cancel{1}}{2} - \frac{\cancel{1}}{2} + \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = \underline{\underline{0.25}}$$

2) Χάριν ευκολίας, θα λύσουμε απευθείας το γενικότερο πρόβλημα για n το πλήθος σημεία σε καθεμία από τις κλάσεις $\omega_1$ και $\omega_2$, ώστε να βρούμε πρώτα το $P_n(e)$ και στη συνέχεια τις ειδικές περιπτώσεις $P_1(e)$ και $P_2(e)$.

Χωρίς βλάβη της γενικότητας, υποθέτουμε ότι το test σημείο που δειγματολυπτείται, έστω x, ανήκει στην κλάση $\omega_2$. Τότε η πιθανότητα λάθους για τον nearest neighbor classifier αντιστοιχεί στην πιθανότητα το πλησιέστερο στο x σημείο των train data να ανήκει στην κλάση $\omega_1$. Δηλαδή:

$$P_n(e) = \int_0^1 P(x|\omega_2) \cdot P\left[\text{κοντινότερο σημείο στον } x \in \omega_1\right] dx$$

$$= \int_0^1 P(x|\omega_2) \cdot \sum_{i=1}^{n} P\left[y_i \in \omega_1 \text{ και } y_i \text{ κοντινότερα στο } x \text{ από το } y_j, \forall j \neq i\right] dx$$

Λόγω συμμετρίας των p.d.f. στο διάστημα [0, 1], οι όροι του άνωθι αθροίσματος είναι όλοι ίσοι μεταξύ τους: (έστω με τον όρο για $i=1$)

$$P_n(e) = \int_0^1 P(x|\omega_2) \cdot n \underbrace{P\left[y_1 \in \omega_1 \text{ και } |y_1 - x| < |y_j - x|, \forall j > 1\right]}_{\text{Ανεξάρτητα Ενδεχόμενα}} dx$$

$$\Rightarrow P_n(e) = \int_0^1 P(x|\omega_2)\, n \int_0^1 P(\omega_1|y_1) \cdot P\left[|y_1 - x| < |y_j - x|, \forall j > 1\right] dy_1\, dx$$

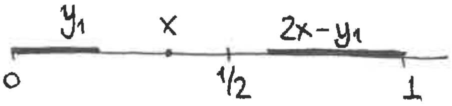
$$= \int_0^1 P(x|\omega_2) \cdot n \int_0^1 P(\omega_1|y_1) \cdot \prod_{j=2}^{n} \overbrace{P\left[|y_1 - x| < |y_2 - x|\right]}^{\text{λόγω συμμετρίας συγκρίνω απλώς } y_1 \text{ και } y_2} dy_1\, dx$$

$$= \int_0^1 P(x|\omega_2) \cdot n \int_0^1 P(\omega_1|y_1) \cdot P\left[|y_1 - x| < |y_2 - x|\right]^{n-1} dy_1\, dx$$

- Για τον όρο $P(x|\omega_2)$ στο διάστημα [0, 1] έχουμε $P(x|\omega_2) = 2 - 2x = 2(1-x)$
- Για τον όρο $P(\omega_1|y_1)$ στο διάστημα [0, 1]: $P(\omega_1|y_1) = \dfrac{P(y_1|\omega_1) \cdot P(\omega_1)}{P(y_1)} = \dfrac{2y_1 \cdot \frac{1}{2}}{1} = y_1.$

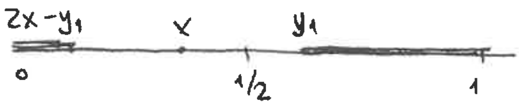
· Για τον όρο $P[|y_1-x|<|y_2-x|]$ μελετάμε τις εξής περιπτώσεις διάταξης:

Περιπτωση 1: $x\in[0,1/2]$, $0<y_1<x$



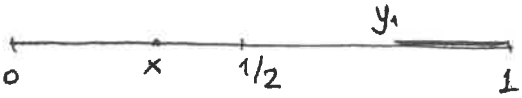

$$P[|y_1-x|<|y_2-x|]=1+2y_1-2x$$

Περιπτωση 2: $x\in[0,1/2]$, $x<y_1<2x$



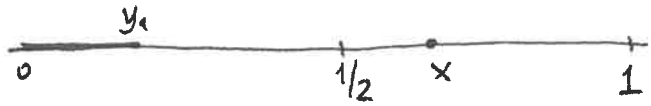

$$P[|y_1-x|<|y_2-x|]=1+2x-2y_1$$

Περιπτωση 3: $x\in[0,1/2]$, $2x<y_1<1$



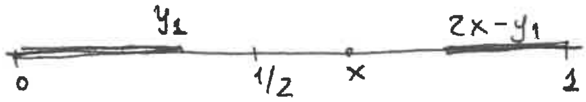

$$P[|y_1-x|<|y_2-x|]=1-y_1$$

Περιπτωση 4: $x\in[1/2,1]$, $0<y_1<2x-1$



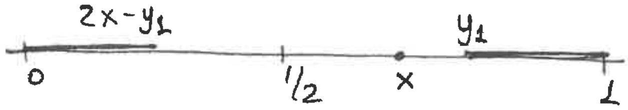

$$P[|y_1-x|<|y_2-x|]=y_1$$

Περιπτωση 5: $x\in[1/2,1]$, $2x-1<y_1<x$




$$P[|y_1-x|<|y_2-x|]=1+2y_1-2x$$

Περιπτωση 6: $x\in[1/2,1]$, $x<y_1<1$




$$Pr[|y_1-x|<|y_2-x|]=1+2x-2y_1$$

Επομένως, κάνουμε την εξής διάσπαση:

$$P_n(e)=\int_0^{1/2}P(x|\omega_2)\cdot n\cdot\Big\{\int_0^{x}\underbrace{P(\omega_1|y_1)\cdot(1+2y_1-2x)^{n-1}}_{\text{Case 1}}dy_1+\int_x^{2x}\underbrace{P(\omega_1|y_1)\cdot(1+2x-2y_1)^{n-1}}_{\text{Case 2}}dy_1$$

$$+\int_{2x}^{1}\underbrace{P(\omega_1|y_1)\cdot(1-y_1)^{n-1}}_{\text{Case 3}}dy_1\Big\}dx+\int_{1/2}^{1}P(x|\omega_2)\cdot n\Big\{\int_0^{2x-1}\underbrace{P(\omega_1|y_1)\cdot y_1^{n-1}}_{\text{Case 4}}dy_1+$$

$$+\int_{2x-1}^{x}\underbrace{P(\omega_1|y_1)\cdot(1+2y_1-2x)^{n-1}}_{\text{Case 5}}dy_1+\int_x^{1}\underbrace{P(\omega|y_1)\cdot(1+2x-2y_1)^{n-1}}_{\text{Case 6}}dy_1\Big\}dx$$

△ Για τις περιπτώσεις 1 και 5 πρέπει να υπολογίσουμε το ολοκλήρωμα:

$$I_1(a,b) = \int_a^b P(w_1|y_1)\cdot(1+2y_1-2x)^{n-1}dy_1 = \int_a^b y_1\cdot(1+2y_1-2x)^{n-1}dy_1$$

Θέτω $u = 1+2y_1-2x \overset{=u(y_1)}{}$ (αλλαγή μεταβλητής). Τότε: $du = 2dy_1 \Rightarrow dy_1 = \frac{du}{2}$

και $y_1 = \frac{u-1+2x}{2}$. Οπότε:

$$I_1(a,b) = \int_{u(a)}^{u(b)} \left(\frac{u-1+2x}{2}\right)\cdot(\cancel{1}+u-\cancel{1}+\cancel{2x}-\cancel{2x})^{n-1}\frac{du}{2} = \int_{u(a)}^{u(b)} \frac{1}{4}(u+2x-1)\cdot u^{n-1}du$$

$$= \int_{u(a)}^{u(b)} \frac{1}{4}\cdot\left[u^n+(2x-1)\cdot u^{n-1}\right]du = \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1}+(2x-1)\cdot\frac{u^n}{n}\right)\right]_{u(a)}^{u(b)}$$

Άρα:

- Στην περίπτωση 1: $I_1(0,x) = \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1}+(2x-1)\frac{u^n}{n}\right)\right]_{u(0)=1-2x}^{u(x)=1}$

$$\Rightarrow I_1(0,x) = \frac{1}{4}\cdot\left(\frac{1}{n+1}+(2x-1)\cdot\frac{1}{n} - \frac{(1-2x)^{n+1}}{n+1} - (2x-1)\cdot\frac{(1-2x)^n}{n}\right)$$

$$= \frac{1}{4}\cdot\left[\frac{1-(1-2x)^{n+1}}{n+1} + \frac{(2x-1)}{n}\cdot\left(1-(1-2x)^n\right)\right]$$

- Στην περίπτωση 5: $I_1(2x-1,x) = \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1}+(2x-1)\frac{u^n}{n}\right)\right]_{u(2x-1)=2x-1}^{u(x)=1}$

$$= \frac{1}{4}\left[\frac{1}{n+1}+(2x-1)\cdot\frac{1}{n} - \frac{(2x-1)^{n+1}}{n+1} - \frac{(2x-1)^{n+1}}{n}\right]$$

▷ Για τις περιπτώσεις 2 και 6 υπολογίζουμε το ολοκλήρωμα:

$$I_2(a,b) = \int_a^b P(w_1|y_1)(1+2x-2y_1)^{n-1}\,dy_1 = \int_a^b y_1\cdot(1+2x-2y_1)^{n-1}\,dy_1$$

Αλλαγή Μεταβλητής $u(y_1) = 1+2x-2y_1 \Rightarrow y_1 = \frac{1+2x-u}{2}$ και $dy_1 = -\frac{du}{2}$

Οπότε:

$$I_2(a,b) = \int_{u(a)}^{u(b)} \frac{1+2x-u}{2}\cdot u^{n-1}\left(-\frac{du}{2}\right) = \int_{u(a)}^{u(b)} \frac{1}{4}\cdot\left[u^n-(1+2x)\cdot u^{n-1}\right]du$$

$$= \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1} + (1+2x)\cdot\frac{u^n}{n}\right)\right]_{u(a)}^{u(b)}$$

• Στην περίπτωση 2: $I_2(x,2x) = \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1} - (1+2x)\frac{u^n}{n}\right)\right]_{u(x)=1}^{u(2x)=1-2x}$

$$= \frac{1}{4}\cdot\left(\frac{(1-2x)^{n+1}}{n+1} - (1+2x)\cdot\frac{(1-2x)^n}{n} - \frac{1}{n+1} + \frac{1}{n}\cdot(1+2x)\right)$$

$$= \frac{1}{4}\cdot\left(\frac{(1-2x)^{n+1}-1}{n+1} + \frac{(1+2x)\cdot\left(1-(1-2x)^n\right)}{n}\right)$$

• Στην περίπτωση 6: $I_2(x,1) = \left[\frac{1}{4}\left(\frac{u^{n+1}}{n+1} - (1+2x)\frac{u^n}{n}\right)\right]_{u(x)=1}^{u(1)=2x-1}$

$$= \frac{1}{4}\cdot\left[\frac{(2x-1)^{n+1}}{n+1} - (1+2x)\cdot\frac{(2x-1)^n}{n} - \frac{1}{n+1} + \frac{(1+2x)}{n}\right]$$

Έπειτα, απομένουν δύο ακόμα περιπτώσεις: <u>Περίπτωση 3 και 4</u>

$$\cdot I_3 = \int_{2x}^{1} P(w_1|y_1)(1-y_1)^{n-1}\,dy_1 = \int_{2x}^{1} y_1\cdot(1-y_1)^{n-1}\,dy_1$$

Αλλαγή μεταβλητής $u(y_1)=1-y_1 \Rightarrow y_1 = 1-u$, $dy_1=-du$ και $u(2x)=1-2x$, $u(1)=$

Οπότε:

$$I_3 = \int_{1-2x}^{0} (1-u)\cdot u^{n-1}(-du) = \int_{1-2x}^{0} \left(u^{n}-u^{n-1}\right)du = \left[\frac{u^{n+1}}{n+1}-\frac{u^{n}}{n}\right]_{1-2x}^{0}$$

$$= \frac{(1-2x)^{n}}{n} - \frac{(1-2x)^{n+1}}{n+1}$$

$$\cdot I_4 = \int_{0}^{2x-1} P(w_1|y_1)\cdot y_1^{n-1}\,dy_1 = \int_{0}^{2x-1} y_1\cdot y_1^{n-1}\,dy_1 = \int_{0}^{2x-1} y_1^{n}\,dy_1 = \left[\frac{y_1^{n+1}}{n+1}\right]_{0}^{2x-1}$$

$$= \frac{(2x-1)^{n+1}}{n+1}$$

Οπότε, μοιράζοντας τα ολοκληρώματα στις δύο μεγαλύτερες περιπτώσεις $x\in[0,1/2]$ και $x\in[1/2,1]$:

$$P_n(e) = \int_{0}^{1/2} 2n(1-x)\cdot\left[\frac{1}{4}\left(\cancel{\frac{1-(1-2x)^{n+1}}{n+1}} + \frac{(2x-1)}{n}\cdot\left(1-(1-2x)\right)^{n}\right) + \frac{(1-2x)^{n}}{n} - \frac{(1-2x)^{n+1}}{n+1}\right.$$

$$\left. + \frac{1}{4}\left(\cancel{\frac{(1-2x)^{n+1}-1}{n+1}} + \frac{(1+2x)\left(1-(1-2x)^{n}\right)}{n}\right)\right]dx$$

$$+ \int_{1/2}^{1} 2n(1-x)\cdot\left[\frac{1}{4}\left(\cancel{\frac{1-(1-2x)^{n+1}}{n+1}} + \frac{(2x-1)}{n}\left(1-(1-2x)\right)^{n} + \cancel{\frac{(1-2x)^{n+1}-1}{n+1}} + \frac{(1+2x)\left(1-(1-2x)^{n}\right)}{n}\right)\right.$$

$$\left. + \frac{(2x-1)^{n+1}}{n+1}\right]dx$$

$$\Rightarrow P_n(e) = \int_0^{1/2} \underbrace{2n(1-x)\cdot\left[\frac{x}{n} + \frac{1}{2n}(1-2x)^n + \left(\frac{1}{2n}-\frac{1}{n+1}\right)(1-2x)^{n+1}\right]}_{A} dx$$

$$+ \int_{1/2}^{1} \underbrace{2n(1-x)\cdot\left[\frac{x}{n} - \frac{1}{2n}(2x-1)^n - \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot(2x-1)^{n+1}\right]}_{B} dx$$

· Αλλαγή μεταβλητής στο A:

Θέτω $u(x) = 1-2x$, οπότε $x = \frac{1-u}{2}$ και $dx = -\frac{du}{2}$, ενώ $u(0) = 1$, $u(\frac{1}{2}) = 0$.

Οπότε:

$$A = \int_1^0 \cancel{2}n\left(\frac{1+u}{2}\right)\cdot\left[\frac{1-u}{2n} + \frac{1}{2n}\cdot u^n + \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}\right]\left(-\frac{du}{\cancel{2}}\right)$$

$$= \int_0^1 \left(\frac{1+u}{2}\right)\cdot n\cdot\left[\frac{1-u}{2n} + \frac{1}{2n}\cdot u^n + \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}\right] du$$

· Αλλαγή μεταβλητής στο B:

Θέτω $u(x) = 2x-1 \Rightarrow x = \frac{u+1}{2}$ και $dx = \frac{du}{2}$, ενώ $u(1/2) = 0$, $u(1) = 1$

Οπότε:

$$B = \int_0^1 \cancel{2}n\cdot\left(\frac{1-u}{2}\right)\cdot\left[\frac{u+1}{2n} - \frac{1}{2n}\cdot u^n - \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}\right]\cdot\frac{du}{\cancel{2}}$$

$$= \int_0^1 \left(\frac{1-u}{2}\right)\cdot n\cdot\left[\frac{u+1}{2n} - \frac{1}{2n}\cdot u^n - \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}\right] du$$

Επομένως:

$$P_n(e) = A + B = \int_0^1 \frac{n}{2}\cdot\left[\cancel{\frac{1-u}{2n}} + \cancel{\frac{1}{2n}\cdot u^n} + \cancel{\left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}} + \cancel{\frac{u+1}{2n}} - \cancel{\frac{1}{2n}\cdot u^n} - \cancel{\left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}}\right] + \frac{n\cdot u}{2}\cdot\left[\frac{1-u}{2n} + \frac{1}{2n}\cdot u^n + \left(\frac{1}{2n}-\frac{1}{n+1}\right)u^{n+1}\right.$$

$$\left. - \frac{u+1}{2n} + \frac{1}{2n}\cdot u^n + \left(\frac{1}{2n}-\frac{1}{n+1}\right)\cdot u^{n+1}\right] du$$

$$\Rightarrow P_n(e) = \int_0^1 \frac{1-u^2}{2} + \frac{u^{n+1}}{2} + \left(\frac{1}{2n} - \frac{1}{n+1}\right)\cdot n \cdot u^{n+2}\, du$$

$$\Rightarrow P_n(e) = \left[\frac{u}{2} - \frac{u^3}{6} + \frac{u^{n+2}}{2(n+2)} + \left(\frac{1}{2n} - \frac{1}{n+1}\right)\cdot n \cdot \frac{u^{n+3}}{n+3}\right]_0^1$$

$$\Rightarrow P_n(e) = \frac{1}{2} - \frac{1}{6} + \frac{1}{2(n+2)} + \frac{(n+1-2n)}{2\cancel{n}\cdot(n+1)} \cdot \frac{\cancel{n}}{n+3}$$

$$\Rightarrow P_n(e) = \frac{1}{3} + \frac{1}{2(n+2)} + \frac{1-n}{(n+1)\cdot(n+3)\cdot 2}$$

$$\Rightarrow \boxed{P_n(e) = \frac{1}{3} + \frac{(n+1)(n+3)+(1-n)(n+2)}{2(n+1)(n+2)(n+3)}}$$

Συνεπώς για $n=1$ το αναμενόμενο σφάλμα είναι:

$$P_1(e) = \frac{1}{3} + \frac{\cancel{2\cdot 4}+0}{\cancel{2}\cdot 2\cdot 3\cdot \cancel{4}} = \frac{1}{3} + \frac{1}{6} = \frac{3}{6}$$

$$\Rightarrow \boxed{P_1(e) = \frac{1}{2} \text{ ή } 50\%}$$

2) Για $n=2$ έχουμε:

$$P_2(e) = \frac{1}{3} + \frac{3\cdot 5 - 4}{2\cdot 3\cdot 4\cdot 5} = \frac{1}{3} + \frac{11}{120} = \frac{40+11}{120}$$

$$\Rightarrow \boxed{P_2(e) = \frac{51}{120} \approx 0.425}$$

3) Στη γενικότερη περίπτωση των $n$ σημείων δείξαμε ότι:

$$P_n(e) = \frac{1}{3} + \frac{(n+1)(n+3)+(1-n)(n+2)}{2(n+1)(n+2)(n+3)}$$

4) $P_{NN} = \lim_{n \to +\infty} P_n(e) = \lim_{n \to +\infty} \cancelto{0}{\frac{(n+1)(n+2)+(1-n)(n+2)}{2(n+1)(n+2)(n+3)}} + \frac{1}{3} = \frac{1}{3}$

Ενώ στο Ερώτημα 1 βρήκαμε για τον Bayes Classifier :

$$P_{bayes} = \frac{1}{4} = 0.25$$

Παρατηρούμε ότι ισχύει $P_{NN} > P_{bayes}$, για $n \to +\infty$, ενώ μάλιστα ικανοποιείται και η Εξίσωση 40 στο Κεφ 4.5 του [2] για $C=2$ κλάσεις :

$$P_{bayes} \leq P_{NN} \leq P_{bayes}\left(2 - \frac{C}{C-1} \cdot P_{bayes}\right)$$

$$\Longrightarrow \boxed{\frac{1}{4} \leq \frac{1}{3} \leq \frac{3}{8}}$$

## Άσκηση 1.8: (EM)

Θεωρούμε τα δεδομένα:

$$D = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ * \end{pmatrix} \right\}$$

όπου το $*$ υποδηλώνει μια άγνωστη τιμή χαρακτηριστικού. Υποθέτουμε πως τα σημεία αυτά έχουν προέλθει από μια διδιάστατη διαχωρίσιμη κατανομή $p(x_1,x_2) = p(x_1)p(x_2)$ με:

$$p(x_1) \sim \begin{cases} \frac{1}{\theta_1} e^{-\theta_1 x_1} &, \text{εαν } x_1 \geq 0 \\ 0 &, \text{αλλιώς} \end{cases}$$

$$p_2(x_2) \sim U(0,\theta_2) = \begin{cases} 1/\theta_2 &, \text{εάν } 0 \leq x_2 \leq \theta_2 \\ 0 &, \text{αλλιώς} \end{cases}$$

1) Expectation Step:

Θεωρούμε αρχική εκτίμηση παραμέτρων $\theta^{(0)} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$. Τότε:

$$Q(\theta;\theta^{(0)}) = E_{x_{32}}\left[\ln p(D_g, D_b; \theta) \mid \theta^{(0)}, D_g\right]$$

όπου $D_g = \{\underline{x_1}, \underline{x_2}, x_{31}\}$ και $D_b = \{x_{32}\}$.

$$Q(\theta;\theta^{(0)}) = \int_{-\infty}^{+\infty} \left[\ln p(\underline{x_1}|\theta) + \ln p(\underline{x_2}|\theta) + \ln p(\underline{x_3}|\theta)\right] \cdot p(x_{32}|\theta^{(0)}, x_{31}=1)\, dx_{32}$$

Εν γένει:

$$\int_{-\infty}^{+\infty} p(x_{32}|\theta^{(0)}, x_{31}=1)\, dx_{32} = \int_0^{\theta_2^{(0)}} \frac{1}{\theta_1^{(0)}} e^{-\theta_1^{(0)} \cdot x_{31}} \cdot \frac{1}{\theta_2^{(0)}}\, dx_{32}$$

$$= \int_0^3 \frac{1}{2} e^{-2} \cdot \frac{1}{3}\, dx_{32} = \frac{e^{-2}}{2}$$

Οπότε :

$$Q(\theta;\theta^{(0)}) = \frac{1}{2e^2}\cdot\left[\ln p(\underline{x}_1|\theta) + \ln p(\underline{x}_2|\theta)\right] + \underbrace{\int_{-\infty}^{+\infty} \ln p(\underline{x}_3|\theta)\cdot p(x_{32}|\theta^{(0)}, x_{31}=1)\, dx_{32}}_{I}$$

$$= \frac{1}{2e^2}\cdot\left[\ln\left(\frac{1}{\theta_1}e^{-\theta_1}\cdot\frac{1}{\theta_2}\right) + \ln\left(\frac{1}{\theta_1}\cdot e^{-4\theta_1}\cdot\frac{1}{\theta_2}\right)\right] + I$$

$$= \frac{1}{2e^2}\cdot\left[-\theta_1 - \ln\theta_1 - \ln\theta_2 - 4\theta_1 - \ln\theta_1 - \ln\theta_2\right] + I$$

$$= \frac{1}{2e^2}\cdot\left[-5\theta_1 - 2\ln\theta_1 - 2\ln\theta_2\right] + I$$

Για τον υπολογισμό του ολοκληρώματος $I$ διακρίνουμε τις εξής υποπεριπτώσεις :

(i) $\theta_2 \leq 3$ και ~~[illegible]~~ προφανώς $\theta_2 \geq 0$

$$I = \int_0^{\theta_2} \ln\left(\frac{1}{\theta_1}\cdot e^{-\theta_1}\cdot\frac{1}{\theta_2}\right)\cdot\frac{1}{2}\cdot e^{-2}\cdot\frac{1}{3}\, dx_{32}$$

$$= \frac{e^{-2}}{2}\cdot\left[\frac{1}{3}\theta_2\cdot\ln\left(\frac{1}{\theta_1}\cdot e^{-\theta_1}\cdot\frac{1}{\theta_2}\right)\right]$$

(ii) $\theta_2 \geq 3$ :

$$I = \int_0^{3} \ln\left(\frac{1}{\theta_1}\cdot e^{-\theta_1}\frac{1}{\theta_2}\right)\cdot\frac{1}{2}e^{-2}\cdot\frac{1}{3}\, dx_{32}$$

$$= \frac{e^{-2}}{2}\cdot\ln\left(\frac{1}{\theta_1}\cdot e^{-\theta_1}\cdot\frac{1}{\theta_2}\right)$$

(iii) Σε κάθε άλλη περίπτωση $I = 0$.

Επομένως :

(i) <u>Για $0 \leq \theta_2 \leq 3$</u> :

$$Q(\theta;\theta^{(0)}) = \frac{1}{2e^2}\cdot\left[-5\theta_1 - 2\ln\theta_1 - 2\ln\theta_2 - \frac{1}{3}\theta_2\ln\theta_1 - \frac{1}{3}\theta_1\theta_2 - \frac{1}{3}\theta_2\ln\theta_2\right]$$

(ii) <u>Για $\theta_2 \geq 3$</u> :

$$Q(\theta;\theta^{(0)}) = \frac{1}{2e^2}\cdot\left[-5\theta_1 - 2\ln\theta_1 - 2\ln\theta_2 - \ln\theta_1 - \theta_1 - \ln\theta_2\right]$$

$$\Rightarrow Q(\theta;\theta^{(0)}) = \frac{1}{2e^2}\cdot\left[-6\theta_1 - 3\ln\theta_1 - 3\ln\theta_2\right]$$

2) Αρχικά, επιθυμούμε να ικανοποιείται η συνθήκη κανονικοποίησης πιθανότητας για την $p(x_1)$, δηλαδή:

$$\int_{-\infty}^{+\infty} p(x_1)\,dx_1 = 1 \Rightarrow \int_0^{+\infty} \frac{1}{\theta_1}\cdot e^{-\theta_1 x_1}\,dx_1 = 1$$

$$\Rightarrow \left[-\frac{1}{\theta_1^2}\cdot e^{-\theta_1 x_1}\right]_0^{+\infty} = 1 \Rightarrow 0 + \frac{1}{\theta_1^2}\cdot\cancelto{1}{e^{-\theta_1\cdot 0}} = 1$$

$$\Rightarrow \theta_1^2 = 1 \overset{\theta_1 \geq 0}{\Longrightarrow} \boxed{\theta_1 = 1}$$

Οπότε :

(i) <u>Για $0 \leq \theta_2 \leq 3$</u> :

$$Q(\theta;\theta^{(0)}) = \frac{1}{2e^2}\cdot\left[-5 - 2\cdot 0 - 2\ln\theta_2 - \frac{1}{3}\theta_2\cdot\left(0 + 1\cdot\theta_2 + \ln\theta_2\right)\right]$$

$$= \frac{1}{2e^2}\cdot\left[-5 - 2\ln\theta_2 - \frac{1}{3}\theta_2 - \frac{1}{3}\theta_2\ln\theta_2\right]$$

<u>Maximisation Step</u>: $\frac{\partial Q(\theta;\theta^{(0)})}{\partial\theta_2} = 0 \Rightarrow \frac{-2}{\theta_2} - \frac{1}{3} - \frac{1}{3}\ln\theta_2 - \frac{1}{3} = 0$

<u>Άτοπο</u> για $\theta_2 \in (0,3]$

Παρατηρούμε, ωστόσο, ότι για $\theta_2 \in [0, 3]$ ισχύει $\frac{\partial Q}{\partial \theta_2} < 0$, οπότε η συνάρτηση $Q(\theta; \theta^{(0)})$ είναι γνησίως φθίνουσα και λαμβάνει την υψηλότερη τιμή της για $\theta_2 = 0$, οπότε $Q = +\infty$.

(ii) Για $\theta_2 \geq 3$:

$$Q(\theta; \theta^{(0)}) = -\frac{1}{2e^2} \cdot \left[6 + 3\ln\theta_2\right]$$

$$\frac{\partial Q}{\partial \theta_2} = -\frac{3}{2e^2\theta_2} < 0, \ \forall \theta_2 \geq 3 \rightarrow \text{επίσης γνησίως φθίνουσα.}$$

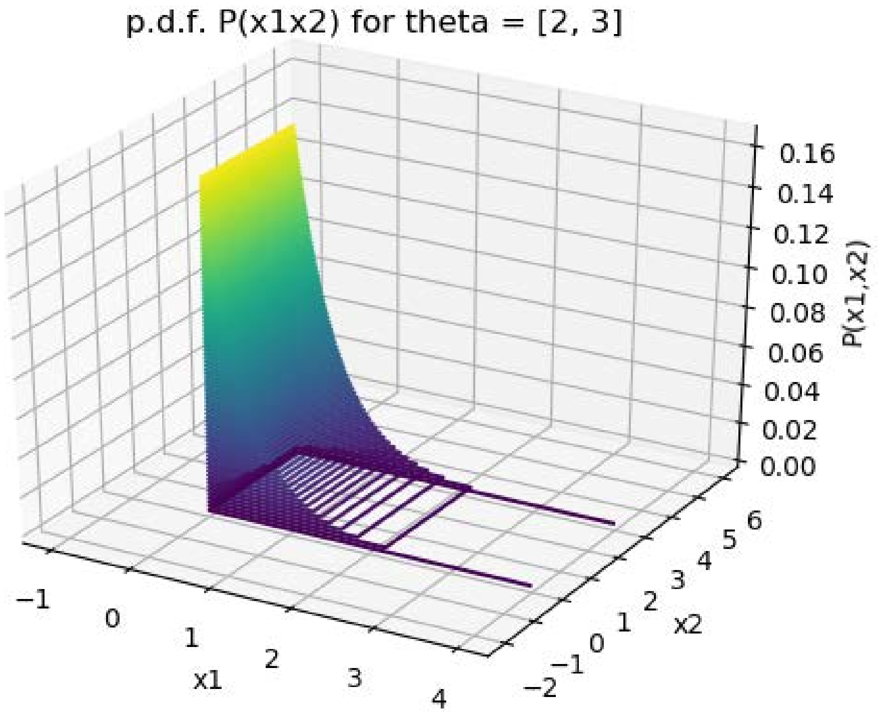
Άρα $\underset{\theta_2}{\operatorname{argmax}} Q = 3$, όπου $Q(\theta_2 = 3) \approx -\frac{1}{2e^2}(9.2$

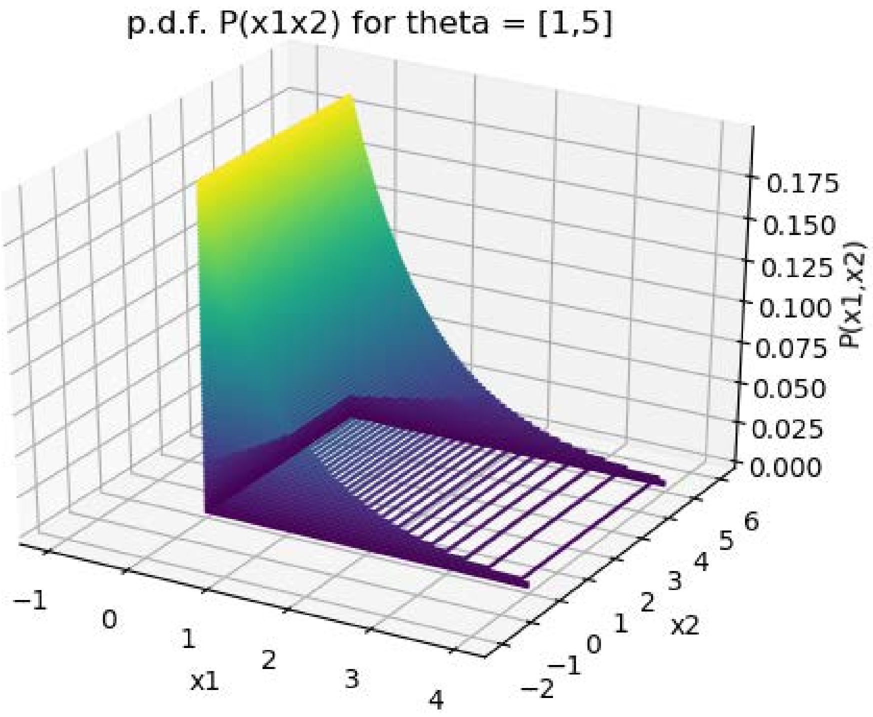
Σε κάθε περίπτωση, όμως, μπορούμε να συμπεράνουμε πως η συνάρτηση $Q(\theta; \theta^{(0)})$ δεν παρουσιάζει κάποιο τοπικό μέγιστο ως προς $\theta_2$, αφού $\lim_{\theta_2 \to 0} Q(\theta) = +\infty$.

Συνεπώς, επιλέγουμε ως $\theta_2^{(1)} = \max_i x_2^{(i)} = 5$.

Άρα $\boxed{\theta^{(1)} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}}$ το νέο διάνυσμα παραμέτρων.

3)






## References

[1] R. O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, Wiley, 2001, Chapter 2: Bayesian decision theory.

[2] R. O. Duda, P.E. Hart and D.G. Stork, Pattern Classification, Wiley, 2001, Chapter 4: Nonparametric techniques.

[3] The Matrix Cookbook, K.B. Petersen, M.S Pedersen, November 15, 2012

# 1 Supplementary Code

```
"""
Pattern Recognition - Flow S
Exercise 1.4: Perceptron
Christos Dimopoulos - 03117037
"""

import numpy as np
import matplotlib.pyplot as plt
plt.close('all')

rho = 1 # learning rate
weights = np.array([0,0,0]) # initial weight vector
# Input Data
xdata = np.array([[1, 2, 1.5],
                  [1, 4, 1],
                  [1, -3, 3],
                  [1, -2, 2],
                  [1, 0, 3],
                  [1, -1, 0],
                  [1, 0, 0],
                  [1, 2,-0.5],
                  [1, 1,-1],
                  [1, -2, 1]
                  ]) #always starts with 1

# 0 --> w1 class, 1 --> w2 class
y_labels = [0,0,0,0,0, 1,1,1,1,1]

counter = 0
j = 0
while counter!=10: # that is all 10 samples are classified correctly
    for i in range(xdata.shape[0]):
        j+=1
        inner_product = np.dot(weights, xdata[i])

        if (y_labels[i]==0 and inner_product <=0):
            weights = weights +rho*xdata[i]
            counter = 0
        elif (y_labels[i]==1 and inner_product>=0):
            weights = weights - rho*xdata[i]
            counter = 0
        else:
            counter=counter+1
            if counter==10:
                break

print('After '+str(j)+' iterations the Weight Vector is [w0, w1, w2] = ',weights)

# Plot Result
blacks = [(2,1.5),(4,1),(-3,3),(-2,2),(0,3)] # w1 class
whites = [(-1,0),(0,0),(2,-0.5),(1,-1),(-2,1)] # w2 class

def coords(data):
```

```python
    xs, ys = [], []
    for point in data:
        xs.append(point[0])
        ys.append(point[1])
    return xs, ys

black_x, black_y = coords(blacks)
plt.figure()
plt.grid()
plt.scatter(black_x, black_y, color='b', label = 'Class w1')

white_x, white_y = coords(whites)
plt.scatter(white_x, white_y, color='r', label = 'Class w2')

# draw perceptron
N = 5
x1 = np.linspace(-N,N,100)
x2= (-weights[0]-weights[1]*x1)/weights[2]

plt.plot(x1, x2, color = 'm', label ='Perceptron Decision Boundary')
plt.xlabel('x1')
plt.ylabel('x2')
plt.xticks(np.arange(-N, N+1, 1))
plt.yticks(np.arange(-N, N+1, 1))
plt.axvline(0, color='k')
plt.axhline(0, color = 'k')
plt.legend()
plt.show()
```

```python
"""
Pattern Recognition - Flow S
Exercise 1.6: Linear Regression
Christos Dimopoulos - 03117037
"""

import numpy as np
import matplotlib.pyplot as plt
plt.close('all')

data = [(0.38, 2.05), (0.44, 2.23), (0.48, 2.13), (0.54, 2.33), (0.58, 2.67), (0.64, 2.68),
        (0.71, 2.81), (0.76, 2.97), (0.82, 3.12), (0.96, 3.2)]

x = np.array([[i[0]] for i in data])
y = np.array([[i[1]] for i in data])

xnew = np.concatenate((np.ones((10,1)),x),axis=1)

h = 1 # learning rate

weights = np.array([1, 1]) # random initialization

# LSE Algorithm
for epochs in range(1000):
    for i in range(len(xnew)): # one sample at a time
        # Gradient of MSE
        gradient = (y[i]-np.dot(weights, xnew[i]))*xnew[i]
```

```
        weights = weights + (h/(i+1))*gradient
    if epochs==0:
        weights_first = weights #collect weights after 1st epoch for comparison

print('After 1 epoch: [w0,w1] = ',weights_first)
print('After 1000 epochs: [w0,w1] = ',weights)

# Plot Result
plt.figure()
plt.grid()
# Draw samples
plt.scatter(x, y, color='b')

# Draw Line after 1st Epoch
x1 = np.linspace(-1,2,100)
x2= weights_first[0]+weights_first[1]*x1
plt.plot(x1, x2, color = 'm', label = 'Regression after 1 epoch')

# Draw line after 1000 epochs
N = 3
x1 = np.linspace(-1,2,100)
x2= weights[0]+weights[1]*x1
plt.plot(x1, x2, color = 'r', label = 'Regression after 1000 epochs')

plt.xlabel('x')
plt.ylabel('y')
plt.axvline(0, color='k')
plt.axhline(0, color = 'k')
plt.legend()
plt.show()
```

```
"""
Pattern Recognition - Flow S
Exercise 1.8: EM
Christos Dimopoulos - 03117037
"""
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
plt.close('all')

x1 = np.linspace(-1,4,1000)
x2 = np.linspace(-2,6,1000)

theta0 = [2, 3]
theta1 = [1, 5]

# Exponential Distribution
def exponential(x1, theta):
    return (1/theta[0])*np.exp(-theta[0]*x1)*(x1>=0)

def uniformal(x2, theta):
    return 1/theta[1]*(x2>=0)*(x2<=theta[1])

# Plot p.d.f. before EM
X1, X2 = np.meshgrid(x1, x2)
```

```
p1 = exponential(X1, theta0)
p2 = uniformal(X2, theta0)
p_beforeEM = p1*p2

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.contour3D(X1, X2, p_beforeEM, 100)
ax.set_xlabel('x1')
ax.set_ylabel('x2')
ax.set_zlabel('P(x1,x2)')
ax.set_title('p.d.f. P(x1x2) for theta = [2, 3]')

# Plot p.d.f. after EM
p1 = exponential(X1, theta1)
p2 = uniformal(X2, theta1)
p_afterEM = p1*p2

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.contour3D(X1, X2, p_afterEM, 100)
ax.set_xlabel('x1')
ax.set_ylabel('x2')
ax.set_zlabel('P(x1,x2)')
ax.set_title('p.d.f. P(x1x2) for theta = [1,5]')
```