Π.Θ.
ΤΗΜΜΥ

ΛΥΣΕΙΣ ΑΣΚΗΣΕΩΝ
ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ

ΗΥ345
2017-18-ΕΑΡ

## 1.Α

$\hat{\theta}_{ML}$ για $p(x|\theta) = \begin{cases} \theta^2 x e^{-\theta x}, & x \geq 0 \\ 0, & \text{αλλιώς} \end{cases}$

με βάση $x_n, n=1 \ldots N \sim p(x|\theta)$

- Εκφράζουμε το log likelihood των N ανεξάρτητων δειγμάτων $x_1, x_2, \ldots, x_N$ που ακολουθούν την $p(x|\theta)$, ως:

$$\mathcal{L}(\theta) = \sum_{n=1}^{N} \ln p(x_n|\theta) = 2N \ln\theta + \sum_{n=1}^{N} \ln x_n - \theta \sum_{n=1}^{N} x_n$$

$$\Longrightarrow \frac{\partial \mathcal{L}(\theta)}{\partial \theta} = \frac{2N}{\theta} - \sum_{n=1}^{N} x_n$$

Θέτουμε την παράγωγο ίση με το μηδέν, οπότε $\frac{2N}{\hat{\theta}_{ML}} = \sum_{n=1}^{N} x_n \Longleftrightarrow$

$$\Longleftrightarrow \boxed{\hat{\theta}_{ML} = \frac{2}{\frac{1}{N}\sum_{n=1}^{N} x_n}}$$

## 1.Β

$\hat{\theta}_{ML}$ για $p(x|\theta) = \begin{cases} 1/\theta, & 0 \leq x \leq \theta \\ 0, & \text{αλλού} \end{cases}$

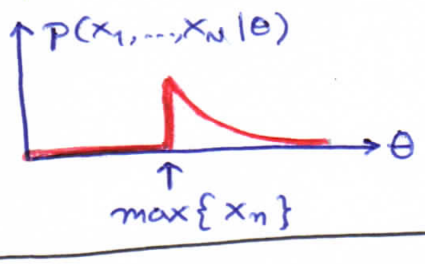
με βάση $x_n, n=1,\ldots,N \sim p(x|\theta)$

- Έστω $I(\cdot)$ η συνάρτηση δείκτη (indicator function) που ισούται με 1 αν το όρισμά της είναι αληθές, αλλιώς παίρνει την τιμή μηδέν.
- Η συνάρτηση πιθανοφάνειας (likelihood) θα είναι η:

$$p(x_1, x_2, \ldots x_n|\theta) = \prod_{n=1}^{N} p(x_n|\theta) = \prod_{n=1}^{N} \frac{1}{\theta} I(0 \leq x_n \leq \theta)$$

$$= \left(\frac{1}{\theta}\right)^N \cdot I\left(\theta \geq \max_{n=1,\ldots,N}\{x_n\}\right) \cdot I\left(\min_{n=1,\ldots,N}\{x_n\} \geq 0\right)$$

- Παρατηρούμε ότι έχει τη μορφή:




- Άρα, $\boxed{\hat{\theta}_{ML} = \max_{n=1,2,\ldots,N}\{x_n\}}$

$\boxed{2}$

$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, $M = 3$

$P(\omega_1) = P(\omega_2) = P(\omega_3) = 1/3$

$p(\underline{x}|\omega_i) \sim N(\underline{\mu}_i, \mathbb{I})$

$\underline{\mu}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\underline{\mu}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\underline{\mu}_3 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ $\Rightarrow$ Εξισώσεις επιφανειών διαχωρισμού + περιοχές αποφάσεων

- Τα discriminant functions είναι τα:

$$g_i(\underline{x}) = \ln[P(\omega_i)p(\underline{x}|\omega_i)] = -\ln(3) - \ln(2\pi) - \frac{1}{2}(\underline{x}-\underline{\mu}_i)^T(\underline{x}-\underline{\mu}_i)$$

$$= \underbrace{const}_{\text{ανεξάρτητο κλάσης}} + \underline{\mu}_i^T\underline{x} - \frac{1}{2}\underline{\mu}_i^T\underline{\mu}_i - \frac{1}{2}\underbrace{\underline{x}^T\underline{x}}_{\text{ανεξάρτητο κλάσης}}, \quad i = 1,2,3$$

- Άρα:

$$g_1(\underline{x}) = \left[c - \frac{1}{2}\underline{x}^T\underline{x}\right] + (x_1 + x_2) - 1$$

$$g_2(\underline{x}) = \left[c - \frac{1}{2}\underline{x}^T\underline{x}\right] + (x_2 - x_1) - 1$$

$$g_3(\underline{x}) = \left[c - \frac{1}{2}\underline{x}^T\underline{x}\right] - x_2 - 1/2$$

- Οι τρεις περιοχές αποφάσεων προκύπτουν ως:

– Για την $\omega_1$:

$g_1(\underline{x}) > g_2(\underline{x}) \Leftrightarrow x_1 + x_2 - 1 > x_2 - x_1 - 1 \Leftrightarrow$

& $g_1(\underline{x}) > g_3(\underline{x}) \Leftrightarrow x_1 + x_2 - 1 > -x_2 - 1/2 \Leftrightarrow$

$\boxed{x_1 > 0 \ \& \ x_1 + 2x_1 > 1/2}$

– Για την $\omega_2$:

$g_2(\underline{x}) > g_1(\underline{x}) \Leftrightarrow x_2 - x_1 - 1 > x_1 + x_2 - 1 \Leftrightarrow$

& $g_2(\underline{x}) > g_3(\underline{x}) \Leftrightarrow x_2 - x_1 - 1 > -x_2 - 1/2 \Leftrightarrow$

$\boxed{x_1 < 0 \ \& \ 2x_2 - x_1 > 1/2}$

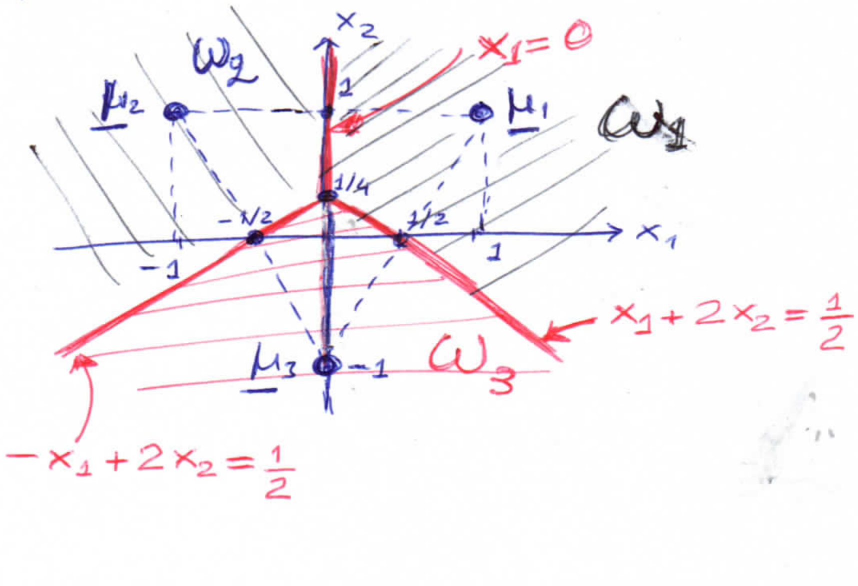
– Για την $\omega_3$:

$g_3(\underline{x}) > g_1(\underline{x}) \Leftrightarrow -x_2 - 1/2 > x_1 + x_2 - 1 \Leftrightarrow$

& $g_3(\underline{x}) > g_2(\underline{x}) \Leftrightarrow -x_2 - 1/2 > x_2 - x_1 - 1 \Leftrightarrow$

$\boxed{2x_2 + x_1 < 1/2 \ \& \ 2x_2 - x_1 < 1/2}$

- Σχηματικά:



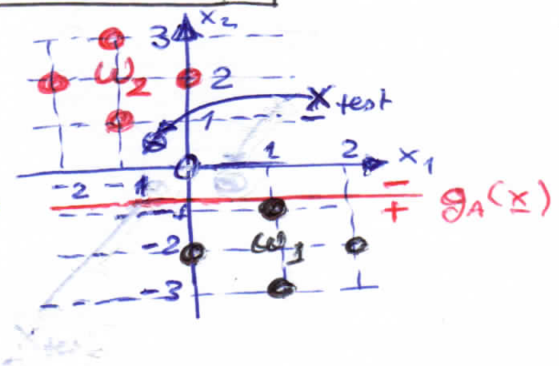
## 3

$\omega_1 : \begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}2\\-2\end{bmatrix}, \begin{bmatrix}1\\-3\end{bmatrix}, \begin{bmatrix}0\\-2\end{bmatrix}$

$\omega_2 : \begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}-2\\2\end{bmatrix}, \begin{bmatrix}-1\\3\end{bmatrix}, \begin{bmatrix}0\\2\end{bmatrix}$ , $\underline{x}_{test} = \begin{bmatrix}-1/2\\+1/2\end{bmatrix}$

### 3.A

- Σχεδιάζουμε τα δείγματα των δύο κλάσεων στον 2-D χώρο.




- Οι κλάσεις είναι προφανώς γραμμικά διαχωρίσιμες.
- Π.χ., η ευθεία $g_A(\underline{x}) = -x_2 - 3/4 =$

$$= \underbrace{\begin{bmatrix}0 & -1 & -\frac{3}{4}\end{bmatrix}}_{W_A^T} \begin{bmatrix}x_1\\x_2\\1\end{bmatrix} = 0,$$ διαχωρίζουν τις κλάσεις ως $g_A(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$

- Με βάση την $g_A(\underline{x})$, το $\boxed{\underline{x}_{test} \in \omega_2}$ καθώς $g_A(\underline{x}_{test}) = \frac{1}{2} - \frac{3}{4} = -\frac{1}{4} < 0$

### 3.B

**1-NN?**

- Στο υπόλοιπο της άσκησης, συμβολίζουμε με:

$\underline{x}_1 = [1, -1]^T$, $\underline{x}_2 = [2, -2]^T$, $\underline{x}_3 = [1, -3]^T$, $\underline{x}_4 = [0, -2]^T$

$\underline{x}_5 = [-1, 1]^T$, $\underline{x}_6 = [-2, 2]^T$, $\underline{x}_7 = [-1, 3]^T$, $\underline{x}_8 = [0, 2]^T$

- Για ευκολία στις πράξεις, θεωρούμε την $\mathcal{L}_1$ απόσταση (Manhattan distance).
- Εύκολα βλέπουμε ότι το πιο κοντινό (με βάση την $\mathcal{L}_1$) διάνυσμα στο $\underline{x}_{test}$ είναι το $\underline{x}_5$, με $d_1(\underline{x}_{test}, \underline{x}_5) = |-1 + \frac{1}{2}| + |1 - \frac{1}{2}| = 1$, άρα $\boxed{\underline{x}_{test} \in \omega_2}$ αφού $\underline{x}_5 \in \omega_2$

### 3.C

**PERCEPTRON**
$\underline{w}(0) = \underline{0}$, $\rho = 1/8$

- Χρησιμοποιούμε extended vectors, $\underline{x}' = [\underline{x}^T\ 1]^T$
- Η "degenerate" ευθεία $\underline{w}(0) = [0\ 0\ 0]^T$ ταξινομεί όλα λάθος. $(\underline{w}^T(0)\underline{x}_i' = 0)$
- Άρα, από τον αλγόριθμο του PERCEPTRON:

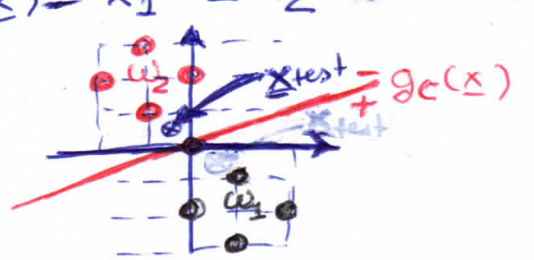
$$\underline{w}(1) = \underline{w}(0) + \rho(\underline{x}_1' + \underline{x}_2' + \underline{x}_3' + \underline{x}_4') - \rho(\underline{x}_5' + \underline{x}_6' + \underline{x}_7' + \underline{x}_8') =$$

$$= \begin{bmatrix}0\\0\\0\end{bmatrix} + \frac{1}{8}\left(\begin{bmatrix}1\\-1\\1\end{bmatrix} + \begin{bmatrix}2\\-2\\1\end{bmatrix} + \begin{bmatrix}1\\-3\\1\end{bmatrix} + \begin{bmatrix}0\\-2\\1\end{bmatrix} - \begin{bmatrix}-1\\1\\1\end{bmatrix} - \begin{bmatrix}-2\\2\\1\end{bmatrix} - \begin{bmatrix}-1\\3\\1\end{bmatrix} - \begin{bmatrix}0\\2\\1\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\0\end{bmatrix}$$

- Παρατηρούμε ότι:

$$\underline{w}^T(1)\underline{x}_i' = \begin{cases} \left.\begin{array}{l} 3\ (i=1)\\ 6\ (i=2)\\ 7\ (i=3)\\ 4\ (i=4) \end{array}\right\} > 0,\ \underline{x}_i \in \omega_1 \\ \left.\begin{array}{l} -3\ (i=5)\\ -6\ (i=6)\\ -7\ (i=7)\\ -4\ (i=8) \end{array}\right\} < 0,\ \underline{x}_i \in \omega_2 \end{cases}$$

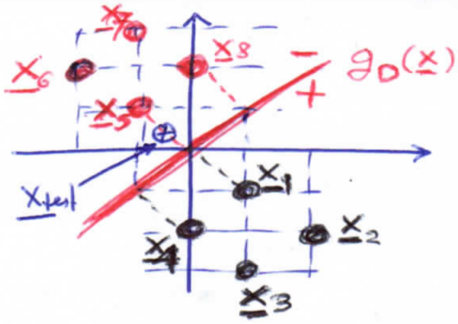
- Συνεπώς ο αλγόριθμος έχει συγκλίνει στην ευθεία $g_C(\underline{x}) = x_1 - 2x_2 = 0$ (βλέπε σχήμα).




- Τέλος, $g_C(\underline{x}_{test}) = -\frac{1}{2} - \cancel{2}\frac{1}{\cancel{2}} = -\frac{3}{2} < 0 \Rightarrow \boxed{\underline{x}_{test} \in \omega_2}$

## $\boxed{3.D}$ SVM ?

• Είναι προφανές ότι η $g_D(\underline{x})$ (σχεδιασμένη δεξιά) επιτυγχάνει το μέγιστο περιθώριο διαχωρισμού:




$$g_D(\underline{x}): \ \frac{1}{2}x_1 - \frac{1}{2}x_2 + 0 = 0 \Longleftrightarrow \underline{w}_D = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix}$$

• Συνολικά, 4 σημεία εκπαίδευσης βρίσκονται στο όριο του περιθωρίου: $\underline{x}_1, \underline{x}_4, \underline{x}_5, \underline{x}_8$ :

$$\underline{w}_D^T \underline{x}_1' = \underline{w}_D^T \underline{x}_4' = 1 \ , \quad \underline{w}_D^T \underline{x}_5' = \underline{w}_D^T \underline{x}_8' = -1$$

• Θα θέλαμε να εκφράσουμε τα $[w_{D,1} \ w_{D,2}]^T$ ως γραμμικό συνδυασμό των $\underline{x}_1, \underline{x}_4, \underline{x}_5, \underline{x}_8$:

$$\begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} = \lambda_1 \cancel{y_1}^{1} \underline{x}_1 + \lambda_4 \cancel{y_4}^{1} \underline{x}_4 + \lambda_5 \cancel{y_5}^{-1} \underline{x}_5 + \lambda_8 \cancel{y_8}^{-1} \underline{x}_8$$

$$\Longleftrightarrow \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} = \lambda_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \lambda_4 \begin{bmatrix} 0 \\ -2 \end{bmatrix} - \lambda_5 \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \lambda_8 \begin{bmatrix} 0 \\ 2 \end{bmatrix} \Longleftrightarrow \boxed{\begin{array}{ll} \lambda_1 + \lambda_5 = 1/2 & (1) \\ -\lambda_1 - 2\lambda_4 - \lambda_5 - 2\lambda_8 = -\frac{1}{2} & (2) \end{array}}$$

όπου $\lambda_i \geq 0$ (προφανώς $\lambda_2 = \lambda_3 = \lambda_6 = \lambda_7 = 0$)

& $\sum_{i=1}^{8} \lambda_i y_i = 0 \Longleftrightarrow \boxed{\lambda_1 + \lambda_4 - \lambda_5 - \lambda_8 = 0}$ (3)

• Μια λύση για τις παραπάνω, είναι η $\lambda_1 = \lambda_5 = \frac{1}{4}$, $\lambda_4 = \lambda_8 = 0$ (επιβεβαιώνουν τις (1)(2)(3))

• Το $w_{D,0}$ επιβεβαιώνεται ίσο με το μηδέν από όλες τις $\underline{w}_D^T \underline{x}_1' = \underline{w}_D^T \underline{x}_4' = 1$, $\underline{w}_D^T \underline{x}_5' = \underline{w}_D^T \underline{x}_8' = -1$

• Το margin είναι: $\frac{2}{\|w_D\|} = \frac{2}{\sqrt{2}\ 1/2} = 2\sqrt{2}$

• Τέλος, $\boxed{\underline{x}_{test} \in \omega_2}$, καθώς $g_D(\underline{x}_{test}) = \frac{1}{2}(-\frac{1}{2}) - \frac{1}{2}(\frac{1}{2}) = -\frac{1}{2} < 0$

---

## $\boxed{3.E}$ L.S.E. ?

• Σχηματίζουμε τον πίνακα δεδομένων & διάνυσμα ετικετών:

$$X = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -2 & 1 \\ 1 & -3 & 1 \\ 0 & -2 & 1 \\ -1 & 1 & 1 \\ -2 & 2 & 1 \\ -1 & 3 & 1 \\ 0 & 2 & 1 \end{bmatrix}_{8\times 3}, \ \underline{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}_{8\times 1} \Rightarrow X^T X = \begin{bmatrix} 12 & -16 & 0 \\ -16 & 36 & 0 \\ 0 & 0 & 8 \end{bmatrix} = 4\begin{bmatrix} 3 & -4 & 0 \\ -4 & 9 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} 8 \\ -16 \\ 0 \end{bmatrix} = 4\begin{bmatrix} 2 \\ -4 \\ 0 \end{bmatrix}$$

$$\Longrightarrow (X^T X)^{-1} = \frac{1}{4}\ \frac{1}{22}\begin{bmatrix} 18 & 8 & 0 \\ 8 & 6 & 0 \\ 0 & 0 & 11 \end{bmatrix} \Rightarrow \underline{w}_E = (X^T X)^{-1} X^T \underline{y} = \frac{1}{22}\begin{bmatrix} 4 \\ -8 \\ 0 \end{bmatrix}$$

• Ως ευθεία, η συνάρτηση διάκρισης ταυτίζεται με αυτήν του perceptron (βλέπε σχήμα).

• Το συνολικό σφάλμα είναι: $\sum_{i=1}^{8}(y_i - x_i^T w_E)^2 = 2\left[(1-\frac{12}{22})^2 + (1-\frac{24}{22})^2 + (1-\frac{28}{22})^2 + (1-\frac{16}{22})^2\right]$

$$= \frac{2}{22^2}(100+4+36+36) = \frac{2 \cdot 176}{22 \cdot 22} = \frac{8}{11} \simeq 0.73$$

• Όπως και στην 3.C, το $\underline{x}_{test}$ ταξινομείται στην $\boxed{\omega_2}$

$$\left(g_E(\underline{x}_{test}) = \frac{1}{22}\left(4\cdot(-\tfrac{1}{2}) - 8(\tfrac{1}{2})\right) = \frac{-6}{22} < 0\right)$$

## 3.F

**Bayesian normal pdfs**

- Έχουμε $P(\omega_1) = P(\omega_2) = \frac{4}{8} = \frac{1}{2}$ (ισοπίθανες κλάσεις).
- Για τις κανονικές κατανομές των υπό συνθήκη πιθανοτήτων:
- $\mu_1 = \frac{1}{4}\sum_{i=1}^{4} x_i = \frac{1}{4}\left(\begin{bmatrix}1\\-1\end{bmatrix}+\begin{bmatrix}2\\-2\end{bmatrix}+\begin{bmatrix}1\\-3\end{bmatrix}+\begin{bmatrix}0\\-2\end{bmatrix}\right) = \begin{bmatrix}1\\-2\end{bmatrix}$
- $\mu_2 = \frac{1}{4}\sum_{i=5}^{8} x_i = \frac{1}{4}\left(\begin{bmatrix}-1\\1\end{bmatrix}+\begin{bmatrix}-2\\2\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}+\begin{bmatrix}0\\2\end{bmatrix}\right) = \begin{bmatrix}-1\\2\end{bmatrix}$

- $[x_1-\mu_1,\ x_2-\mu_1,\ x_3-\mu_1,\ x_4-\mu_1] = \begin{bmatrix}0 & 1 & 0 & -1\\ 1 & 0 & -1 & 0\end{bmatrix} \Rightarrow$

$$\Rightarrow \Sigma_1 = \frac{1}{3}\begin{bmatrix}0 & 1 & 0 & -1\\ 1 & 0 & -1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\\ 0 & -1\\ -1 & 0\end{bmatrix} = \frac{1}{3}\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix} = \frac{2}{3}I$$

- Παρόμοια, $\Sigma_1 = \Sigma_2$
- Επειδή έχουμε δύο ισοπίθανες κλάσεις, κ $\Sigma_1 = \Sigma_2 = \begin{bmatrix}\sigma^2 & 0\\ 0 & \sigma^2\end{bmatrix}$ ($\sigma^2 = 2/3$), η ευθεία διαχωρισμού των δύο κλάσεων θα είναι η $w^T(x - x_0)$ με $w = \mu_1 - \mu_2 = \begin{bmatrix}1\\-2\end{bmatrix} - \begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}2\\-4\end{bmatrix}$ κ $x_0 = \frac{1}{2}(\mu_1 + \mu_2) = \begin{bmatrix}0\\0\end{bmatrix}$
- Οπότε $g_F(x)$: $2x_1 - 4x_2 = 0$ ή απλοποιώντας:

  $g_F(x)$: $x_1 - 2x_2 = 0$
- Πρόκειται πάλι για την ίδια ευθεία όπως στην 3.C (perceptron), και $\boxed{x_{test} \in \omega_2}$

---

## 3.H

**LDA + 1NN**

- Από την 3.F έχουμε:

  $P(\omega_1) = P(\omega_2) = 1/2$

  $\mu_1 = \begin{bmatrix}1\\-2\end{bmatrix}$, $\mu_2 = \begin{bmatrix}-1\\2\end{bmatrix}$, $\Sigma_1 = \Sigma_2 = \frac{1}{3}\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix} = \frac{2}{3}I$

- Άρα: $S_W = P(\omega_1)\Sigma_1 + P(\omega_2)\Sigma_2 = \frac{2}{3}I$
- Συνεπώς: $W^{(LDA)}_{PROJ} = S_W^{-1}\cdot(\mu_1 - \mu_2) = \frac{3}{2}I \cdot \begin{bmatrix}2\\-4\end{bmatrix} = \begin{bmatrix}3\\-6\end{bmatrix}$
- Πρόκειται για την ίδια κατεύθυνση προβολής όπως για τον PCA (βλέπε επόμενη σελίδα, 3.G). Συνεπώς και ίδια ταξινόμηση του $x_{test}$, στην $\boxed{\omega_2}$.

---

$\boxed{3.G}$

$\boxed{PCA + 1NN}$

• Παρατηρούμε ότι:

$$\frac{1}{8}\sum_{i=1}^{8} \underline{x}_i = \frac{1}{8}\left(\begin{bmatrix}1\\-1\end{bmatrix}+\begin{bmatrix}2\\-2\end{bmatrix}+\begin{bmatrix}1\\-3\end{bmatrix}+\begin{bmatrix}0\\-2\end{bmatrix}+\begin{bmatrix}-1\\1\end{bmatrix}+\begin{bmatrix}-2\\2\end{bmatrix}+\begin{bmatrix}-1\\3\end{bmatrix}+\begin{bmatrix}0\\2\end{bmatrix}\right)=\begin{bmatrix}0\\0\end{bmatrix}$$

• Συνεχίζουμε με τον πίνακα:

$$\mathbb{X}=\begin{bmatrix}1 & 2 & 1 & 0 & -1 & -2 & -1 & 0\\ -1 & -2 & -3 & -2 & 1 & 2 & 3 & 2\end{bmatrix} \Rightarrow \mathbb{R}=\frac{1}{8}\mathbb{X}\mathbb{X}^T=\frac{1}{2}\begin{bmatrix}3 & -4\\ -4 & 9\end{bmatrix}$$

• Διαγωνιοποιούμε τον πίνακα (η σταθερά 1/2 δεν παίζει ρόλο στο ranking των ιδιοτιμών και στις κατευθύνσεις των ιδιοδιανυσμάτων), οπότε:

– ΙΔΙΟΤΙΜΕΣ: $(\lambda-3)(\lambda-9)-16=0 \Leftrightarrow \lambda^2-12\lambda+11=0 \Rightarrow \lambda_{max}=11$, $\lambda_{min}=1$

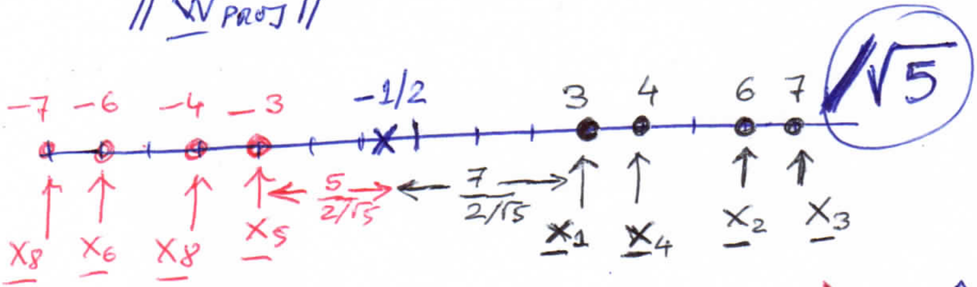
– Διαλέγοντας την μεγαλύτερη, καταλήγουμε στο ιδιοδιάνυσμα:

$$\begin{bmatrix}3 & -4\\ -4 & 9\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=11\begin{bmatrix}x_1\\x_2\end{bmatrix} \Longleftrightarrow \begin{matrix}3x_1-4x_2=11x_1\\ -4x_1+9x_2=11x_2\end{matrix} \Leftrightarrow \begin{matrix}8x_1=-4x_2\\ -4x_1=2x_2\end{matrix} \Leftrightarrow$$

$\Longleftrightarrow 2x_1+x_2=0$, δηλαδή $\underline{W}_{PCA}=\begin{bmatrix}2\\1\\0\end{bmatrix}$ η ευθεία της προβολής.

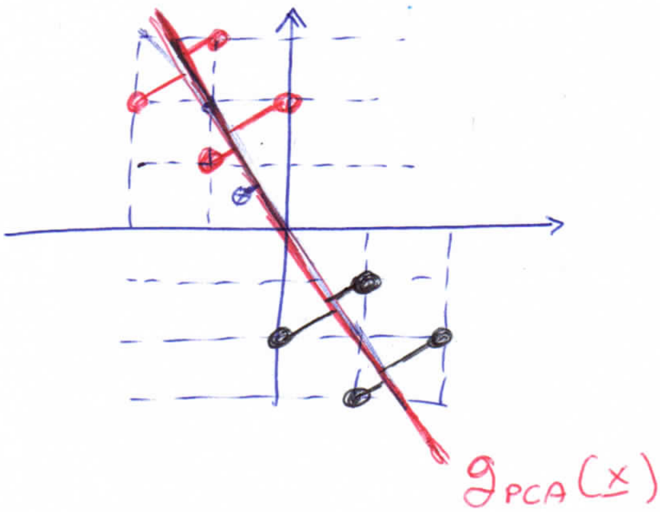
• Άρα $W_{PROJ}=\begin{bmatrix}1\\-2\end{bmatrix}$, και κατά συνέπεια οι προβολές των σημείων θα είναι τα

$$\frac{\underline{W}_{PROJ}^T\,\underline{x}_i}{\|\underline{W}_{PROJ}\|}, \quad i=1,2,\dots,8 \ \& \ test.$$




• Άρα χρησιμοποιώντας 1-NN:

$\boxed{\underline{x}_{test} \in \omega_2}$



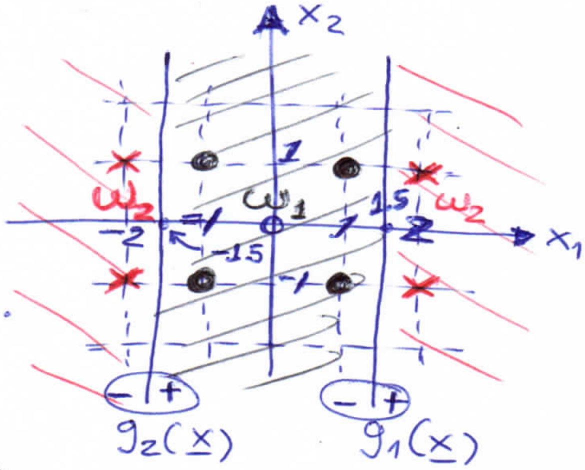
# 4

$\omega_1: \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}-1\\1\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}$

$\omega_2: \begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}2\\-1\end{bmatrix}, \begin{bmatrix}-2\\1\end{bmatrix}, \begin{bmatrix}-2\\-1\end{bmatrix}$

## 4.A. ΣΧΕΔΙΟ

Προφανώς οι δύο κλάσεις δεν είναι γραμμικά διαχωρίσιμες.




## 4.B. PERCEPTRON WITH 2 LAYERS

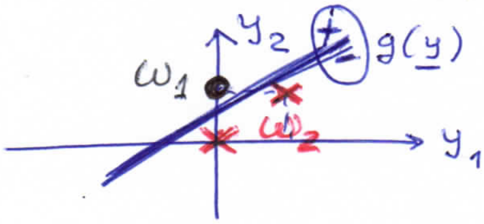
• Έστω οι δύο ευθείες:

$$g_1(\underline{x}): 1 \cdot x_1 + 0 \cdot x_2 - \frac{3}{2} = 0$$

$$g_2(\underline{x}): 1 \cdot x_1 + 0 \cdot x_2 + \frac{3}{2} = 0$$

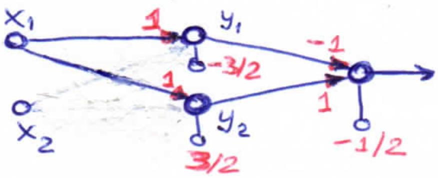
• Έστω $y_1, y_2$ οι ~~1-0~~ έξοδοι των παραπάνω ταξινομητών, με αποτέλεσμα το mapping:

| $(x_1$ | $x_2)$ | $\mapsto$ | $(y_1$ | $y_2)$ | |
|---|---|---|---|---|---|
| 1 | $\pm 1$ | $\to$ | 0 | 1 | $\omega_1$ |
| -1 | $\pm 1$ | $\to$ | 0 | 1 | $\omega_1$ |
| 2 | $\pm 1$ | $\to$ | 1 | 1 | $\omega_2$ |
| -2 | $\pm 1$ | $\to$ | 0 | 0 | $\omega_2$ |




• Τα τρία σημεία είναι γραμμικά διαχωρίσιμα από την $g(\underline{y}): -y_1 + y_2 - \frac{1}{2} = 0$

• Διάγραμμα του 2-layer perceptron.




• Περιοχές απόφασης στον αρχικό χώρο: Βλέπε σχήμα 4.A.

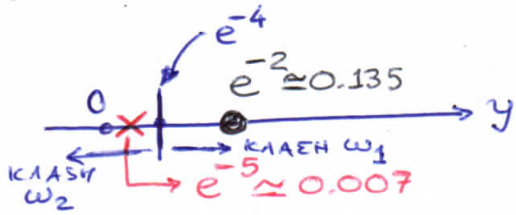
## 4.C. RBF

• Χρησιμοποιούμε μία συνάρτηση RBF με κέντρο $\underline{c} = \begin{bmatrix}0\\0\end{bmatrix}$ και $\sigma^2 = 1/2$.

• Άρα $y = \exp(-\|x\|^2) = \exp(-x_1^2 - x_2^2)$

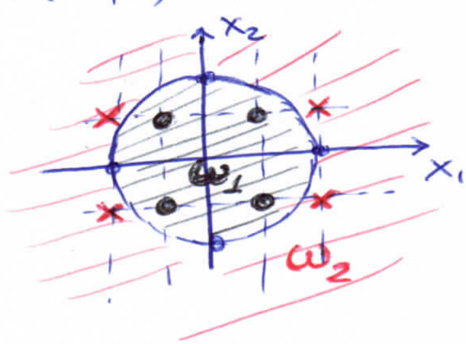
• Έχουμε λοιπόν το mapping:

| $x_1$ | $x_2$ | $\longrightarrow$ | $y$ | ΚΛΑΣΗ |
|---|---|---|---|---|
| $\pm 1$ | $\pm 1$ | $\longrightarrow$ | $e^{-2}$ | $\omega_1$ |
| $\pm 2$ | $\pm 1$ | $\longrightarrow$ | $e^{-5}$ | $\omega_2$ |




• Βλέπουμε ότι οι δύο κλάσεις είναι πλέον διαχωρίσιμες, π.χ. από το κατώφλι (threshold) $e^{-4} \simeq 0.0183$

• Άρα $y \underset{\omega_2}{\overset{\omega_1}{\gtrless}} e^{-4} \Leftrightarrow x_1^2 + x_2^2 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} 4$

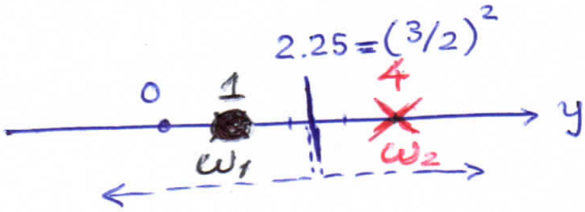


## 4.D. POLYNOMIAL

- Θα μπορούσε κανείς να λύσει το πρόβλημα με διάφορους τρόπους, π.χ. ένα 2-D mapping στον χώρο των $y_1 = x_1^2$, $y_2 = x_2^2$, οπότε θα προκύψει κάποια επιφάνεια διαχωρισμού παρόμοια με (καμπύλη) την περίπτωση του RBF (κύκλος, όπως στην 4.C.)

- Πιο εύκολα ωστόσο, παρατηρούμε ότι και με κατάλληλο 1-D mapping μπορεί να επιλυθεί το πρόβλημα, $y = x_1^2$, οπότε:

| $x_1$ | $x_2$ | $\longrightarrow$ | $y$ | ΚΛΑΣΗ |
|---|---|---|---|---|
| $\pm 1$ | $\pm 1$ | $\longrightarrow$ | 1 | $\omega_1$ |
| $\pm 2$ | $\pm 1$ | $\longrightarrow$ | 4 | $\omega_2$ |




- Βλέπουμε ότι οι δύο κλάσεις είναι πλέον διαχωρίσιμες, π.χ. από το κατώφλι $\left(\frac{3}{2}\right)^2 = 2.25$.

- Άρα $y \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \left(\frac{3}{2}\right)^2 \Longleftrightarrow x_1^2 \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \left(\frac{3}{2}\right)^2 \Longleftrightarrow$ $-3/2 < x_1 < 3/2 \rightarrow \omega_1$

  $x_1 < -3/2$ ή $x_2 > 3/2 \rightarrow \omega_2$

- Οι περιοχές απόφασης στον αρχικό χώρο είναι οι ίδιες με το PERCEPTRON (Σχήμα 4.A).

## 4.E DECISION TREE

- Το σύνολο των πιθανών ερωτήσεων είναι

  $Q_1: x_1 > -3/2$

  $Q_2: x_1 > 0$

  $Q_3: x_1 > 3/2$

  $Q_4: x_2 > 0$

- Διαισθητικά, λόγω συμμετρίας, η $Q_4$ δεν θα ερωτηθεί, καθώς δεν μεταβάλλει το impurity
- Το ίδιο ισχύει και για την $Q_2$ (στο αρχικό στάδιο)
- Οπότε, ως αρχική ερώτηση, η $Q_1$ ή $Q_3$ αποτελούν τις καλύτερες — έστω ρωτάται η $Q_1$ (και οι δύο θα μεταβάλλουν το impurity με το ίδιο ποσό).
- Στη συνέχεια μπορεί να ερωτηθεί η $Q_2$ ή $Q_3$, ωστόσο η $Q_3$ είναι η optimal.

- Πραγματικά, για το root node ($t_0$):

$I_{Q_4}(t_0Y) = 1$

$I_{Q_4}(t_0N) = 1$

$I_0 = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = -\log_2 2^{-1} = 1$

άρα $\Delta I_{Q_4} = 0$. Ομοίως, $\Delta I_{Q_2} = 0$

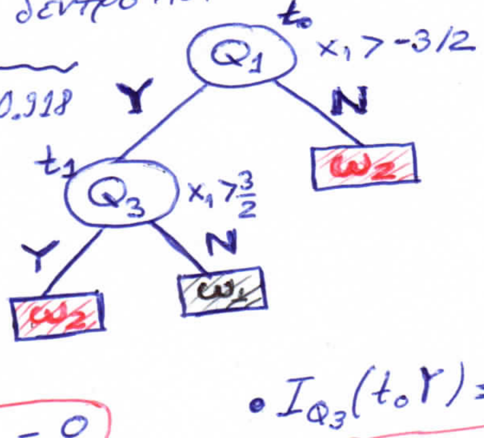
$I_{Q_1}(t_0Y) = -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6} = \log_2 3 - \frac{2}{3} = 0.918$

$I_{Q_1}(t_0N) = 0$

Άρα $\Delta I_{Q_1} = 1 - \frac{2}{8}\cdot 0 - \frac{6}{8}\, 0.918 = 0.311$

Ομοίως και $\Delta I_{Q_3} = 0.311$

- Το δέντρο λοιπόν θα είναι το:




- Άρα η $Q_1$ ή $Q_3$ είναι optimal — έστω διαλέγουμε την $Q_1$.

- Για τον κόμβο $t_1$: $\Delta I_{Q_4} = 0$, $\Delta I_{Q_1} = 0$

$\left.\begin{array}{l} I(t_1) = 0.918 \\ I_{Q_2}(t_0Y) = 1 \\ I_{Q_2}(t_0N) = 0 \end{array}\right\} \Rightarrow \Delta I_{Q_2} = 0.918 - \frac{2}{3} - \frac{1}{3}\,0 = 0.2513$

- $I_{Q_3}(t_0Y) = I_{Q_3}(t_0N) = 0$

  $\Rightarrow \Delta I_{Q_3} = 0.918$

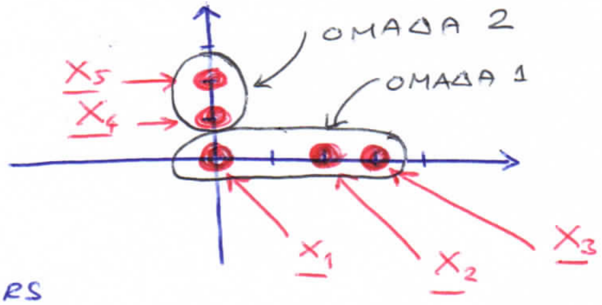
- Άρα διαλέγουμε την $Q_3$ όπως φαίνεται στο σχήμα
- Όλοι οι κόμβοι τώρα έχουν $I(t) = 0$.

# 5

$\underline{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \underline{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, x_3 \begin{bmatrix} 3 \\ 0 \end{bmatrix}, x_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, x_5 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$

## 5.A

BSAS
$\mathcal{L}_1$, $\theta = 2.1$
centroids
$k_{max} = inf.$

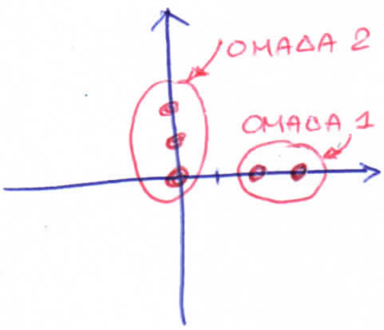



| ITERATION/STEP | CLUSTERS |
|---|---|
| [1] | $\{\{\underline{x}_1\}\}$ |
| [2] | $\{\{\underline{x}_1, \underline{x}_2\}\}$ γιατί $d(\underline{x}_1, \underline{x}_2) = 2 < \theta$ |
| | ⟶ ΚΕΝΤΡΟ $\underline{C}_1 = [1, 0]^T = \dfrac{\underline{x}_1 + \underline{x}_2}{2}$ |
| [3] | $\{\{\underline{x}_1, \underline{x}_2, \underline{x}_3\}\}$ γιατί $d(\underline{x}_3, \underline{C}_1) = 2 < \theta$ |
| | ⟶ ΚΕΝΤΡΟ $C_1' = \dfrac{\underline{x}_1 + \underline{x}_2 + \underline{x}_3}{3} = \left[\frac{5}{3}, 0\right]^T \simeq \begin{bmatrix} 1.66 \\ 0 \end{bmatrix}$ |
| [4] | $\{\{\underline{x}_1, \underline{x}_2, \underline{x}_3\}, \{\underline{x}_4\}\}$ γιατί $d(\underline{x}_4, \underline{c}_1') = 2.66 > \theta$ |
| [5] | $\{\{\underline{x}_1, \underline{x}_2, \underline{x}_3\}, \{\underline{x}_4, \underline{x}_5\}\}$ γιατί $d(\underline{x}_5, c_1') = 3.66$, $d(\underline{x}_5, \underline{x}_4) = 1 < 3.66$ & $< \theta$. |

Η ομαδοποίηση έχει σχεδιαστεί στο σχήμα.

## 5.B

K-means
INIT by 5.A.
$\mathcal{L}_2$ distance




**INITIALIZATION:**

$\{\{\underline{x}_1, \underline{x}_2, \underline{x}_3\}, \{\underline{x}_4, \underline{x}_5\}\}$ (από 5.A.)

**CENTROIDS:**

$\underline{C}_1^{(1)} = \dfrac{\underline{x}_1 + \underline{x}_2 + \underline{x}_3}{3} = \begin{bmatrix} 5/3 \\ 0 \end{bmatrix}, \quad \underline{C}_2^{(1)} = \dfrac{\underline{x}_4 + \underline{x}_5}{2} = \begin{bmatrix} 0 \\ 3/2 \end{bmatrix}$

**RE-CLUSTERING / ASSIGNMENT:**

$d(\underline{x}_1, \underline{c}_1^{(1)}) = 5/3, \; d(\underline{x}_1, \underline{c}_2^{(1)}) = 3/2 \Rightarrow \underline{x}_1 \rightarrow \underline{c}_2^{(1)}$

$d(\underline{x}_2, \underline{c}_1^{(1)}) = 1/3, \; d(\underline{x}_2, \underline{c}_2^{(1)}) = 5/2 \Rightarrow \underline{x}_2 \rightarrow \underline{c}_1^{(1)}$

$d(\underline{x}_3, \underline{c}_1^{(1)}) = 4/3, \; d(\underline{x}_3, \underline{c}_2^{(1)}) = \frac{3}{2}\sqrt{5} \Rightarrow \underline{x}_3 \rightarrow \underline{c}_1^{(1)}$

$d(\underline{x}_4, \underline{c}_1^{(1)}) = \sqrt{34}/3, \; d(\underline{x}_4, \underline{c}_2^{(1)}) = 1/2 \Rightarrow \underline{x}_4 \rightarrow \underline{c}_2^{(1)}$

$d(\underline{x}_5, \underline{c}_1^{(1)}) = \sqrt{61}/3, \; d(\underline{x}_5, \underline{c}_2^{(1)}) = 1/2 \Rightarrow \underline{x}_5 \rightarrow \underline{c}_2^{(1)}$

ΑΡΑ: $\{\{\underline{x}_2, \underline{x}_3\}, \{\underline{x}_1, \underline{x}_4, \underline{x}_5\}\}$

**CENTROIDS:**

$\underline{C}_1^{(2)} = \dfrac{\underline{x}_2 + \underline{x}_3}{2} = \begin{bmatrix} 5/2 \\ 0 \end{bmatrix}, \quad \underline{C}_2^{(2)} = \dfrac{\underline{x}_1 + \underline{x}_4 + \underline{x}_5}{2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Εύκολα βλέπουμε ότι στο επόμενο assignment δεν αλλάζουν οι ομάδες. —

**5.C** Agglomerative clustering $L_1$ single vs. complete link

• Υπολογίζουμε πρώτα τον πίνακα ανομοιότητας των δεδομένων με βάση την απόσταση $L_1$.

$$\begin{array}{c|ccccc} & x_1 & x_2 & x_3 & x_4 & x_5 \\ \hline x_1 & 0 & 2 & 3 & 1 & 2 \\ x_2 & 2 & 0 & 1 & 3 & 4 \\ x_3 & 3 & 1 & 0 & 4 & 5 \\ x_4 & 1 & 3 & 4 & 0 & 1 \\ x_5 & 2 & 4 & 5 & 1 & 0 \end{array}$$

• ΑΛΓΟΡΙΘΜΟΣ SINGE LINK

- ΟΜΑΔΟΠΟΙΗΣΗ $x_4$ με $x_5$ (απόσταση 1)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & \{x_4, x_5\} \\ \hline x_1 & 0 & 2 & 3 & 1 \\ x_2 & 2 & 0 & 1 & 3 \\ x_3 & 3 & 1 & 0 & 4 \\ \{x_4, x_5\} & 1 & 3 & 4 & 0 \end{array}$$

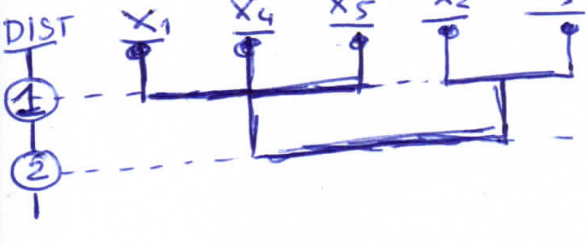
- ΟΜΑΔΟΠΟΙΗΣΗ $x_2$ με $x_3$ (απόσταση 1)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|ccc} & x_1 & \{x_2, x_3\} & \{x_4, x_5\} \\ \hline x_1 & 0 & 2 & 1 \\ \{x_2, x_3\} & 2 & 0 & 3 \\ \{x_4, x_5\} & 1 & 3 & 0 \end{array}$$

- ΟΜΑΔΟΠΟΙΗΣΗ $x_1$ με $\{x_4, x_5\}$ (απόσταση 1)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|cc} & \{x_2 x_3\} & \{x_1 x_4 x_5\} \\ \hline \{x_2 x_3\} & 0 & 2 \\ \{x_1 x_4 x_5\} & 2 & 0 \end{array}$$

- ΠΛΗΡΗΣ ΟΜΑΔΑ $\{x_1\ x_2\ x_3\ x_4\ x_5\}$ (απόσταση 2)
- ΔΕΝΔΡΟΓΡΑΜΜΑ




• ΑΛΓΟΡΙΘΜΟΣ COMPLETE LINK

- ΟΜΑΔΟΠΟΙΗΣΗ $x_4$ με $x_5$ (απόσταση 1)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & \{x_4, x_5\} \\ \hline x_1 & 0 & 2 & 3 & 2 \\ x_2 & 2 & 0 & 1 & 4 \\ x_3 & 3 & 1 & 0 & 5 \\ \{x_4, x_5\} & 2 & 4 & 5 & 0 \end{array}$$

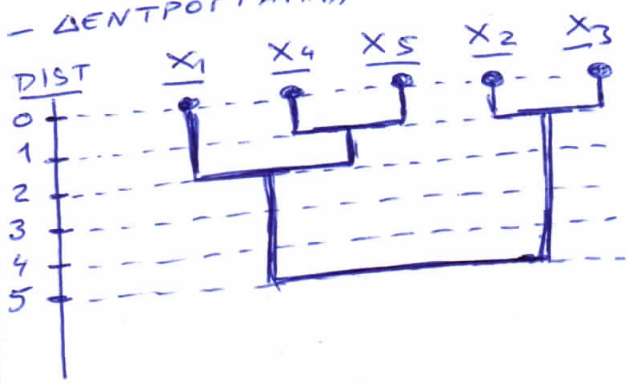
- ΟΜΑΔΟΠΟΙΗΣΗ $x_2$ με $x_3$ (απόσταση 1)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|ccc} & x_1 & \{x_2, x_3\} & \{x_4, x_5\} \\ \hline x_1 & 0 & 3 & 2 \\ \{x_2, x_3\} & 3 & 0 & 5 \\ \{x_4, x_5\} & 2 & 5 & 0 \end{array}$$

- ΟΜΑΔΟΠΟΙΗΣΗ $x_1$ με $\{x_4, x_5\}$ (απόσταση 2)
- ΝΕΟΣ ΠΙΝΑΚΑΣ

$$\begin{array}{c|cc} & \{x_2, x_3\} & \{x_1, x_4, x_5\} \\ \hline \{x_2, x_3\} & 0 & 5 \\ \{x_1, x_4, x_5\} & 5 & 0 \end{array}$$

- ΤΕΛΙΚΗ ΟΜΑΔΟΠΟΙΗΣΗ (απόσταση 5)
- ΔΕΝΤΡΟΓΡΑΜΜΑ:



## 6

**ΕΚΠΑΙΔΕΥΣΗ:**

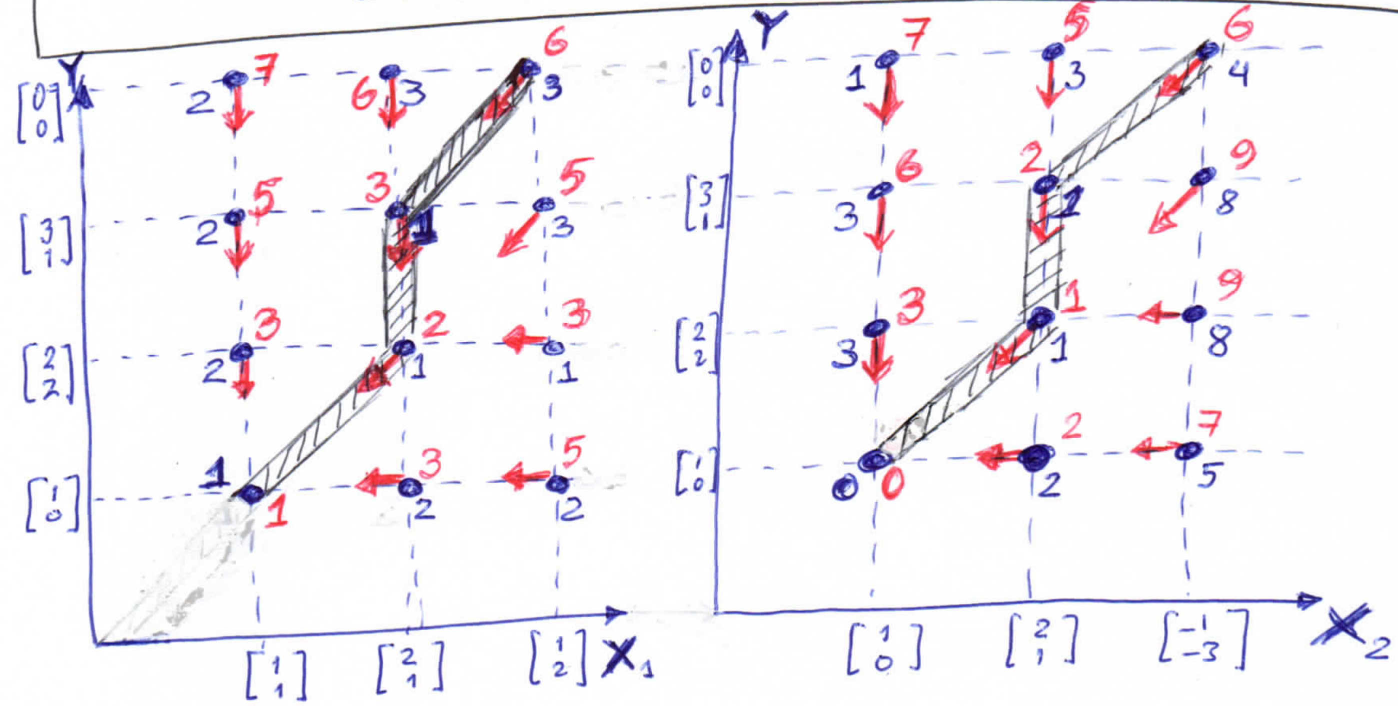
$\omega_1: X_1 = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

$\omega_2: X_2 = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -3 \end{bmatrix}$

**ΔΟΚΙΜΗ:**

$Y = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & 0 \end{bmatrix} \Rightarrow Y \in \omega_1$ ή $\omega_2$?

με βάση: DTW, $L_1$, SAKOE-CHIBA

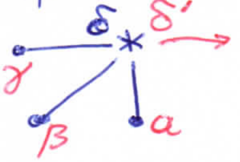



- Για να βρούμε τις αποστάσεις DTW:

– Βρίσκουμε pair-wise τις $L_1$ αποστάσεις διανυσμάτων (πλέγμα 3×4 μπλε τιμές)

– Στήλη-στήλη (από αριστερά προς τα δεξιά) και κάτω προς τα πάνω

  * Υπολογίζουμε το ελάχιστο κόστος με βάση SAKOE-CHIBA περιορισμούς




$\delta' \rightarrow \min\{\alpha, \beta, \gamma\} + \delta = \delta'$

  * Σημειώνουμε το καινούργιο κόστος (με κόκκινο)
  * Σημειώνουμε με βέλος από "πού έχει έρθει" αυτό (δηλ. από το $\min\{\alpha, \beta, \gamma\}$)

– Το ολικό κόστος είναι στο πάνω-δεξιά σημείο

– Αντιστοιχεί στο μονοπάτι με backtracking ("μαύρο" μονοπάτι).

- Στο συγκεκριμένο πρόβλημα:

$$d_{DTW}(X_1, Y) = d_{DTW}(X_2, Y) = 6$$

άρα το Y τυχαία ταξινομείται σε οποιαδήποτε από τις 2 κλάσεις.