Π.Θ
ΤΗΜΜΥ

ΛΥΣΕΙΣ ΑΣΚΗΣΕΩΝ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ

ΗΥ 345
2016-17-ΕΑΡ

ΑΣΚ. 1Α

$\hat{\mu}_{ML}, \hat{\sigma}_{ML} = ?$

$$p(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Γράφουμε τη συνάρτηση λογαριθμικής πιθανοφάνειας των δεδομένων:

$$\mathcal{L}(x_1, x_2, \dots, x_N; \mu, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2} \Rightarrow$$

$$\Rightarrow \begin{cases} \dfrac{\partial \mathcal{L}(\cdot)}{\partial \mu} = \displaystyle\sum_{i=1}^{N}(x_i-\mu) = 0 \Leftrightarrow \boxed{\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i} \\ \dfrac{\partial \mathcal{L}(\cdot)}{\partial \sigma^2} = -\dfrac{N}{\cancel{2}}\dfrac{1}{\sigma^2} + \dfrac{1}{\cancel{2}}\displaystyle\sum_{i=1}^{N}(x_i-\mu)^2\dfrac{1}{\sigma^4} = 0 \Leftrightarrow \boxed{\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{\mu})^2} \end{cases}$$

- Παρατηρούμε ότι $E[\hat{\mu}] = \frac{1}{N}\sum_{i=1}^{N} E[x_i] = \frac{1}{N}\sum_{i=1}^{N}\mu = \mu \Rightarrow$ UNBIASED
- Όμως, $E[\hat{\sigma}^2] = \frac{N-1}{N}\sigma^2 \Rightarrow$ BIASED, γιατί:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i-\mu)+(\mu-\hat{\mu})\right]^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 + \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}-\mu)^2 + \frac{2}{N}\sum_{i=1}^{N}(x_i-\mu)(\mu-\hat{\mu})$$

1ος ΟΡΟΣ: $E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right] = \frac{1}{N}\sum_{i=1}^{N}E(x_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma^2 = \sigma^2$

2ος ΟΡΟΣ: $E\left[\frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}-\mu)^2\right] = E\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sum_{j=1}^{N}x_j - N\mu}{N}\right)^2\right] = E\left[\frac{1}{N^3}\sum_i\sum_j(x_j-\mu)^2\right] \overset{\text{iid}}{=} \frac{1}{N}\sigma^2$

iid: $x_1, x_2 \dots x_N$

3ος ΟΡΟΣ: $E\left[+\frac{2}{N}\sum(x_i-\mu)(\mu-\hat{\mu})\right] =$

$$= E\left[-\frac{2}{N}\sum_{i=1}^{N}(x_i-\mu)\frac{\sum_{j=1}^{N}(x_j-\mu)}{N}\right] \overset{\text{iid}}{=} -\frac{2}{N^2}N\sigma^2 = -\frac{2}{N}\sigma^2$$

Αθροίζοντας: $E[\hat{\sigma}^2] = \sigma^2 - \frac{1}{N}\sigma^2 = \sigma^2\left(\frac{N-1}{N}\right)$

- Άρα το $\boxed{\tilde{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\hat{\mu})^2}$ είναι UNBIASED

ΑΣΚ. 1Β

$\hat{\theta}_{ML} = ?$

$$p(x;\theta) = \begin{cases} \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right) & ; \ x \geq 0 \\ 0 & , \ \text{αλλού} \end{cases}$$

- Γράφουμε τη συνάρτηση λογαριθμικής πιθανοφάνειας των δεδομένων:

$$\mathcal{L}(x_1, x_2, \dots, x_N; \theta) = \sum_{n=1}^{N} \ln\left(\frac{1}{\theta}\right) + \sum_{n=1}^{N} \ln\left[\exp\left(-\frac{x_i}{\theta}\right)\right] \overset{\lambda = 1/\theta}{=}$$

$$= N \ln\lambda - \lambda \sum_{i=1}^{N} x_i \Longrightarrow$$

$$\Rightarrow \frac{\partial \mathcal{L}(\cdot;\cdot)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i \Longleftrightarrow \hat{\lambda}_{ML} = \frac{N}{\sum_{i=1}^{N} x_i}$$

- Άρα $\hat{\theta}_{ML} = \frac{1}{\hat{\lambda}_{ML}} = \boxed{\frac{1}{N}\sum_{i=1}^{N} x_i}$

- Για την εκθετική κατανομή, μπορούμε να αποδείξουμε ότι:

$$E[x] = \frac{1}{\theta}\int_0^{\infty} x e^{-\frac{x}{\theta}} dx = -\int_0^{\infty} x \frac{d}{dx}\left(e^{-\frac{x}{\theta}}\right) dx = \cancel{-x e^{-\frac{x}{\theta}}\Big|_0^{\infty}} + \int_0^{\infty} e^{-\frac{x}{\theta}} dx$$

$$= -\theta e^{-\frac{x}{\theta}}\Big|_0^{\infty} = \theta$$

Άρα $E[\hat{\theta}_{ML}] = \frac{1}{N}\sum_{i=1}^{N} E[x_i] = \frac{1}{N} N\theta = \theta$

Συνεπώς η $\hat{\theta}_{ML}$ είναι $\boxed{\text{UNBIASED}}$.

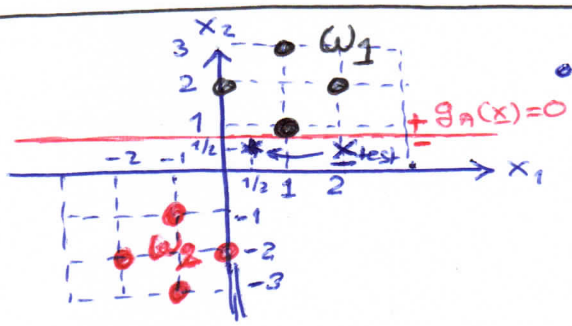
ΑΣΚ. 2(α)

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

$\omega_2 : [-1,-1]^T, [-2,-2]^T, [-1,-3]^T, [0,-2]^T$

$\underline{x}_{test} = [0.5, 0.5]^T$

ΣΧΕΔΙΟ & $g(\underline{x})$: $g(\underline{x}) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$, $g(\underline{x}_{test}) < 0$



• Η ευθεία $g_A(\underline{x}) = x_2 - \frac{3}{4} =$

$$= \underbrace{[0 \;\; 1 \;\; -\tfrac{3}{4}]}_{\underline{W}_A^T} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

ικανοποιεί τα ζητούμενα της άσκησης, δείχνοντας ότι οι κλάσεις είναι γραμμικά διαχωρίσιμες.

ΑΣΚ. 2(β)

1-NN ταξινόμηση του $\underline{x}_{test}$?

• 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 ℓ_1 (Manhattan) απόσταση

$$d_1(\underline{x}_i, \underline{x}_j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}|$$

• Εύκολα βλέπουμε πως το πιό κοντινό διάνυσμα στο $\underline{x}_{test}$ είναι το $\underline{x}_1$

με $d_1(\underline{x}_1, \underline{x}_{test}) = |1 - \frac{1}{2}| + |1 - \frac{1}{2}| = 1$

• Συνεπώς, καθώς $\underline{x}_1 \in \omega_1 \Longrightarrow \boxed{\underline{x}_{test} \in \omega_1}$

ΑΣΚ. 2(c) Percepton ?

$\underline{w}(0) = [0, 0, 0]^T$

$\rho = 1/8$

- NOTATION: Extended vectors: $\underline{x}' = [\underline{x}^T\ 1]^T$

- Προφανώς η "degenerate" ευθεία $\underline{w}(0)$ ταξινομεί και τα οκτώ σημεία λανθασμένα, καθώς $\underline{w}^T(0)\,\underline{x}'_i = 0, \quad \forall i = 1, 2, \dots, 8.$

- 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

$$\underline{w}(1)^T \underline{x}'_i = \begin{cases} 3 \text{ για } i=1 \\ 6 \text{ για } i=2 \\ 7 \text{ για } i=3 \\ 4 \text{ για } i=4 \end{cases} > 0$$

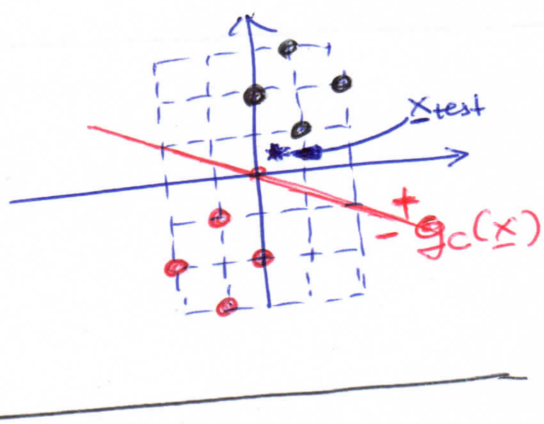
και

$$\underline{w}(1)^T \underline{x}'_i = \begin{cases} -3 \text{ για } i=5 \\ -6 \text{ για } i=6 \\ -7 \text{ για } i=3 \\ -4 \text{ για } i=4 \end{cases} < 0$$

οπότε όλα τα δείγματα εκπαίδευσης ταξινομούνται σωστά από την ευθεία με συντελεστές $\underline{w}(1)$, δηλαδή την

$$g_c(\underline{x}) = x_1 + 2x_2,$$

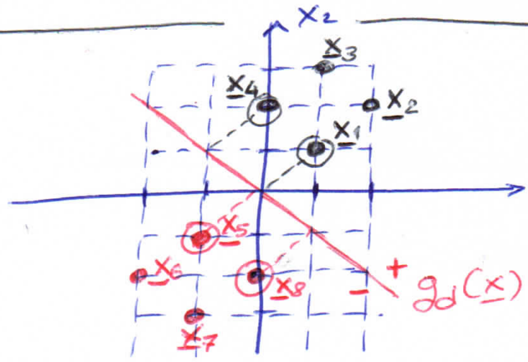
όπως φαίνεται και στο σχήμα.



- Το $\underline{x}_{test}$ ταξινομείται στην κλάση ω_1, καθώς

$$\underline{w}^T(1)\,\underline{x}_{test} = \frac{1}{2} + 2\frac{1}{2} = \frac{3}{2} > 0$$

ΑΣΚ. 2(d) | SVM ?



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

$$g_d(\underline{x}): \frac{1}{2}x_1 + \frac{1}{2}x_2 + 0 = 0 \iff \underline{w} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

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

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

- Θα θέλαμε να εκφράσουμε το $[w_1\ w_2]^T$ ως γραμμικό συνδυασμό των τεσσάρων αυτών σημείων:

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

$$\iff \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = \lambda_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_4 \begin{bmatrix} 0 \\ 2 \end{bmatrix} - \lambda_5 \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \lambda_8 \begin{bmatrix} 0 \\ -2 \end{bmatrix}$$

$$\iff \boxed{\begin{aligned} &\lambda_1 + \lambda_5 = 1/2 \quad (1) \\ &\lambda_1 + \lambda_5 + 2\lambda_4 + 2\lambda_8 = 1/2 \quad (2) \end{aligned}}$$

(προφανώς τα $\lambda_2, \lambda_3, \lambda_6, \lambda_7 = 0$)

όπου $\lambda_i \geq 0$

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

Από (1)+(2) $\underset{\lambda_i \geq 0}{\Rightarrow} \boxed{\lambda_4 = \lambda_8 = 0}$, οπότε η (3) δίνει $\boxed{\lambda_1 - \lambda_5 = 0}$

που μαζί με την (1) επιβάλλει $\boxed{\lambda_1 = \lambda_5 = 1/4}$

- Το w_0 επιβεβαιώνεται ίσο με 0 από τις
$$[w_1\ w_2]\,\underline{x}_1 + w_0 = 1$$
$$[w_1\ w_2]\,\underline{x}_4 + w_0 = 1$$
$$[w_1\ w_2]\,\underline{x}_5 + w_0 = -1$$
$$[w_1\ w_2]\,\underline{x}_8 + w_0 = -1$$

- Το margin είναι $\frac{2}{\|w\|} = \frac{2}{\sqrt{2}\frac{1}{2}} = \boxed{2\sqrt{2}}$

- Τέλος, $\boxed{\underline{x}_{test} \in \omega_1}$

Ασκ. 2(e) L.S.E.?

• Σχηματίζουμε τον πίνακα δεδομένων & διάνυσμα ετικετών:

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

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

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

$$\Rightarrow \underline{w} = (\underline{X}^T\underline{X}^{-1})\,\underline{X}^T\underline{y} = \frac{1}{22}\begin{bmatrix} 18 & -8 & 0 \\ -8 & 6 & 0 \\ 0 & 0 & 11 \end{bmatrix}\begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix} = \frac{1}{22}\begin{bmatrix} 4 \\ 8 \\ 0 \end{bmatrix}$$

• Ισοδύναμα, η ζητούμενη ευθεία είναι η:

$$g_c(\underline{x}) = x_1 + 2x_2$$ (όπως στο perceptron)

• Το συνολικό σφάλμα είναι (least squares error):

$$\sum_{i=1}^{8}(\underline{y}_i - \underline{x}_i^T\underline{w})^2 = 2\left[\left(1-\frac{12}{22}\right)^2 + \left(1-\frac{24}{22}\right)^2 + \left(1-\frac{28}{22}\right)^2 + \left(1-\frac{16}{22}\right)^2\right]$$

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

$$= \frac{8}{11} = 0.7273$$

• Το x_{test} ταξινομείται στην ω_1 (όπως στο σχήμα της 2(c)).

Ασκ. 2(f)

Bayesian ?
(normal class-conditional pdfs)

• Έχουμε $P(\omega_1) = P(\omega_2) = \frac{4}{8} = \frac{1}{2}$
(ισοπίθανες κλάσεις)

• Για τις κανονικές κατανομές των υπό συνθήκη πιθανοτήτων:

$$\underline{\mu}_1 = \frac{1}{4}\sum_{i=1}^{4} \underline{x}_i = \frac{1}{4}\left(\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}2\\2\end{bmatrix}+\begin{bmatrix}1\\3\end{bmatrix}+\begin{bmatrix}0\\2\end{bmatrix}\right) = \begin{bmatrix}1\\2\end{bmatrix}$$

$$[\underline{x}_1-\underline{\mu}_1, \underline{x}_2-\underline{\mu}_1, \underline{x}_3-\underline{\mu}_1, \underline{x}_4-\underline{\mu}_1] = \begin{bmatrix}0 & 1 & 0 & -1\\ -1 & 0 & 1 & 0\end{bmatrix} \Rightarrow$$

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

$$\underline{\mu}_2 = \frac{1}{4}\sum_{i=5}^{8} \underline{x}_i = \frac{1}{4}\left(\begin{bmatrix}-1\\-1\end{bmatrix}+\begin{bmatrix}-2\\-2\end{bmatrix}+\begin{bmatrix}-1\\-3\end{bmatrix}+\begin{bmatrix}0\\-2\end{bmatrix}\right) = \begin{bmatrix}-1\\-2\end{bmatrix}$$

$$[\underline{x}_5-\underline{\mu}_2, \underline{x}_6-\underline{\mu}_2, \underline{x}_7-\underline{\mu}_2, \underline{x}_8-\underline{\mu}_2] = \begin{bmatrix}0 & -1 & 0 & 1\\ 1 & 0 & -1 & 0\end{bmatrix} \Rightarrow$$

$$\Rightarrow \Sigma_2 = \frac{1}{3}\begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix}$$

• Άρα Bayesian ταξινόμηση με βάση την: $P(\omega_1)p(\underline{x}|\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2)p(\underline{x}|\omega_2)$

δηλ:

$$\frac{1}{2}\cdot\frac{1}{2\pi}\frac{1}{\sqrt{|\Sigma_1|}}\exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_1)^T\Sigma_1^{-1}(\underline{x}-\underline{\mu}_1)\right) \underset{\omega_2}{\overset{\omega_1}{\gtrless}}$$
$$\frac{1}{2}\cdot\frac{1}{2\pi}\frac{1}{\sqrt{|\Sigma_2|}}\exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu}_2)^T\Sigma_2^{-1}(\underline{x}-\underline{\mu}_2)\right)$$

• Επειδή $P(\omega_1)=P(\omega_2)$, $\Sigma_1 = \Sigma_2 = \frac{2}{3}\cdot I$, η ευθεία διαχωρισμού δίνεται από την $\boxed{\underline{w}^T(\underline{x}-\underline{x}_0)=0}$

$$\underline{w} = \underline{\mu}_1 - \underline{\mu}_2 = \begin{bmatrix}1\\2\end{bmatrix}-\begin{bmatrix}-1\\-2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix}$$

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_1+\underline{\mu}_2) = \frac{1}{2}\left(\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}-1\\-2\end{bmatrix}\right) = \begin{bmatrix}0\\0\end{bmatrix}$$

• Άρα η ευθεία διαχωρισμού είναι η $\boxed{g_c(\underline{x}) = x_1 + 2x_2 = 0}$ (ίδια με perceptron & L.S.E.), και συνεπώς $\boxed{\underline{x}_{test} \in \omega_1}$.

ΑΣΚ. 2(g)

PCA + 1NN

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

• Συνεχίζουμε με τον πίνακα:

$$\mathbb{X} = \begin{bmatrix} 1 & 2 & 1 & 0 & -1 & -2 & -1 & 0 \\ 1 & 2 & 3 & 2 & -1 & -2 & -3 & -2 \end{bmatrix} \Rightarrow \mathbb{X}\mathbb{X}^T = \mathbb{R} = \begin{bmatrix} 12 & 16 \\ 16 & 36 \end{bmatrix} = 4\begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}$$

• Διαγωνοποιούμε τον πίνακα (η σταθερά δεν έχει σημασία όσον αφορά το ranking των ιδιοτιμών κ την κατεύθυνση των ιδιοδιανυσμάτων):

- Ιδιοτιμές: $(\lambda-3)(\lambda-9)-16=0 \iff \lambda^2 - 12\lambda + 11 = 0$

$\iff \lambda_{max} = 11, \quad \lambda_{min} = 1$ (οι ιδιοτιμές του $\mathbb{R}$ είναι οι 44 & 4 λόγω της σταθεράς 4 που πολλαπλασιάζει τον πίνακα)

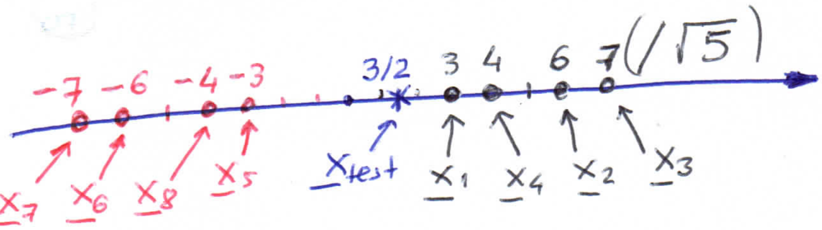
- Διαλέγοντας την μεγαλύτερη, καταλήγουμε στο ιδιοδιάνυσμα:

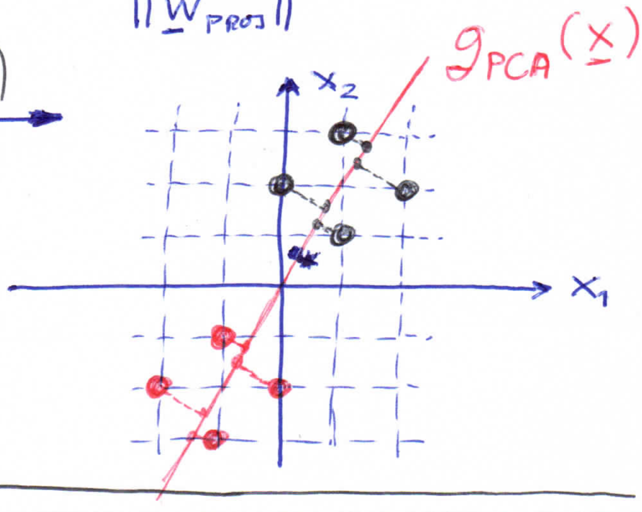
$$\begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 11\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \iff \begin{matrix} 3x_1 + 4x_2 = 11x_1 \\ 4x_1 + 9x_2 = 11x_2 \end{matrix} \iff \begin{matrix} 8x_1 = 4x_2 \\ 4x_1 = 2x_2 \end{matrix}$$

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

κατά μήκος δηλαδή του διανύσματος $\underline{W}_{PROJ} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

• Οι προβολές των σημείων θα είναι τα: $\dfrac{\underline{W}_{PROJ}^T \underline{x}_i}{\|\underline{W}_{PROJ}\|}$, $i = 1, 2, \ldots, 8$ & test





• Άρα, χρησιμοποιώντας 1-NN,

$\underline{x}_{test} \in \omega_1$

ΑΣΚ. 2(h)

LDA + 1NN

- Έχουμε $P(\omega_1) = P(\omega_2) = \frac{4}{8} = \frac{1}{2}$

- Από την 2(f): $\underline{\mu}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\underline{\mu}_2 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$, $\Sigma_1 = \Sigma_2 = \frac{1}{3}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

- Άρα $S_W = P(\omega_1).\Sigma_1 + P(\omega_2).\Sigma_2 = \frac{1}{3}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \frac{2}{3} I$

- Συνεπώς $\underline{W}_{PROJ}^{(LDA)} = S_W^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$

  όπου $\underline{\mu}_1 - \underline{\mu}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

  $S_W^{-1} = \frac{3}{2} I$

  $\Rightarrow \underline{W}_{PROJ}^{(LDA)} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

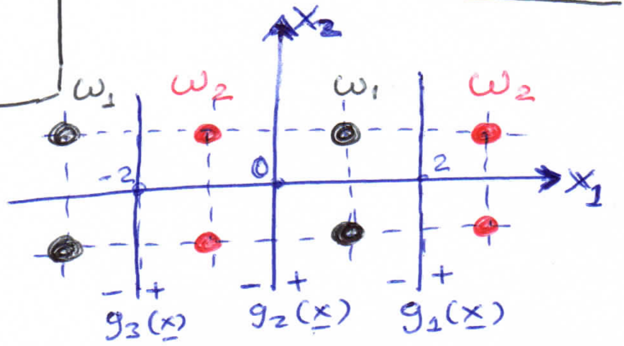
- Πρόκειται κατα συνέπεια για την ίδια ακριβώς κατεύθυνση προβολής όπως στην 2(g) (PCA), κατα συνέπεια τα διαγράμματα είναι τα ίδια όπως στην 2(g), και $\boxed{\underline{x}_{test} \in \omega_1.}$

ΑΣΚ. 3(α)

$\omega 1: \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}-3\\1\end{bmatrix}, \begin{bmatrix}-3\\-1\end{bmatrix}$

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

ΣΧΕΔΙΟ? ΓΡΑΜΜΙΚΑ ΔΙΑΧΩΡΙΣΙΜΕΣ?



- Όπως φαίνεται από το διάγραμμα, οι κλάσεις ΔΕΝ είναι γραμμικά διαχωρίσιμες.

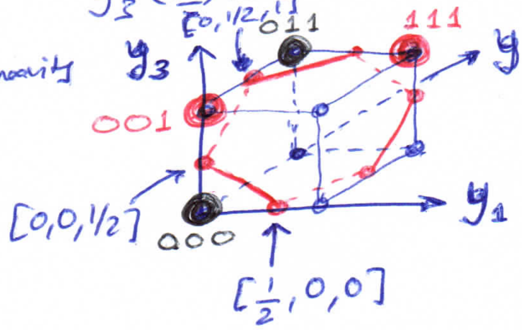
ΑΣΚ. 3(β)

PERCEPTRON ΔΥΟ ΕΠΙΠΕΔΩΝ

- Θεωρούμε τις ευθείες: $g_1(\underline{x}): x_1 - 2 = 0$
  $g_2(\underline{x}): x_1 = 0$
  $g_3(\underline{x}): x_1 + 2 = 0$

- Τα δεδομένα μετασχηματίζονται σε: όπου f: 0-1 non-linearity

$\underline{x}$	$y_1 = f(g_1(\underline{x}))$	$y_2 = f(g_2(\underline{x}))$	$y_3 = f(g_3(\underline{x}))$	ΚΛΑΣΗ
$\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\-1\end{bmatrix}$	0	1	1	ω_1
$\begin{bmatrix}-3\\1\end{bmatrix}, \begin{bmatrix}-3\\-1\end{bmatrix}$	0	0	0	ω_1
$\begin{bmatrix}3\\1\end{bmatrix}, \begin{bmatrix}3\\-1\end{bmatrix}$	1	1	1	ω_2
$\begin{bmatrix}-1\\+1\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}$	0	0	1	ω_2

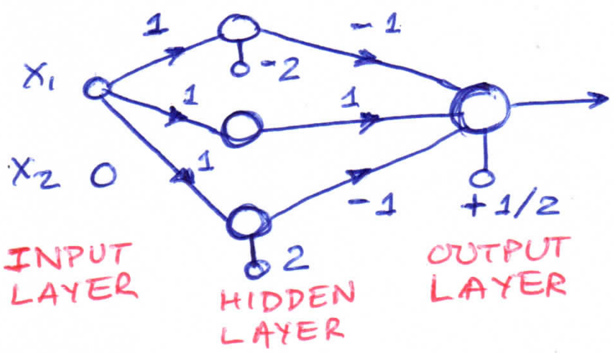


- Το ζητούμενο επίπεδο διαχωρισμού (ένα από τα άπειρα...) έχει σχεδιαστεί πάνω δεξιά. Περνάει από τα εξής σημεία:
  $[\frac{1}{2}, 0, 0], [0, 0, 1/2], [0, 1/2, 1]$
- Λύνοντας ως προς τη γενική εξίσωση $\alpha y_1 + \beta y_2 + \gamma y_3 + \delta = 0$ (όταν δέσουμε τα 3 σημεία σε αυτήν), παίρνουμε: $\alpha = -2\delta,\ \gamma = -2\delta,\ \beta = 2\delta$

οπότε με $\delta = +1/2$ η εξίσωση γίνεται:

$$g(\underline{y}): \boxed{-y_1 + y_2 - y_3 + 1/2 = 0} \Rightarrow g(\underline{y}) \begin{cases} > 0 & \forall \underline{y} \in \omega_1 \\ < 0 & \forall \underline{y} \in \omega_2 \end{cases}$$

- Άρα το perceptron με 1 hidden layer είναι το:



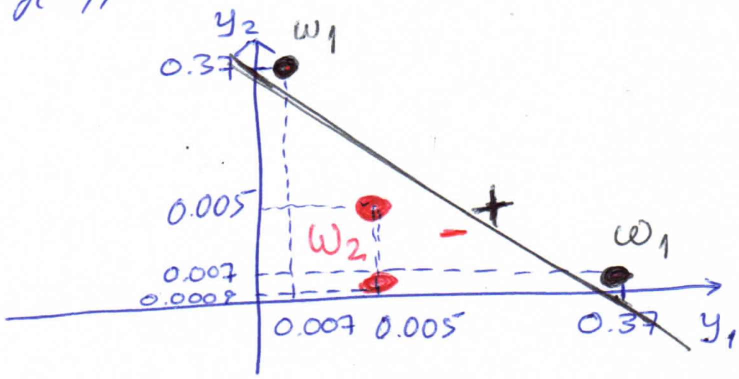
ΑΣΚ. 3(c)

RBF

- Τοποθετούμε δύο κέντρα στις θέσεις $c_1=\begin{bmatrix}1\\0\end{bmatrix}$, $c_2=\begin{bmatrix}-3\\0\end{bmatrix}$, κοντά δηλαδή στα σημεία της κλάσης ω_1, εκμεταλλευόμενοι και την συμμετρία ως προς τον άξονα των x_1.
- Για απλοποίηση στις πράξεις, θέτουμε $\sigma_1^2=\sigma_2^2=1/2$.
- Άρα $y_1 = \exp(-\|x-c_1\|^2)$, $y_2 = \exp(-\|x-c_2\|^2)$
- Τα 8 σημεία εκπαίδευσης μετασχηματίζονται σε:

$\underline{x}$	y_1	y_2	ΚΛΑΣΗ
$[1,\pm1]^T$	$e^{-1}=0.3679$	$e^{-5}=0.0067$	ω_1
$[-3,\pm1]^T$	$e^{-5}=0.0067$	$e^{-1}=0.3679$	ω_1
$[3,\pm1]^T$	$e^{-3}=0.0498$	$e^{-7}=0.0009$	ω_2
$[-1,\pm1]^T$	$e^{-3}=0.0498$	$e^{-3}=0.0498$	ω_2

- Σχεδιάζουμε τα σημεία στον καινούργιο χώρο των $[y_1, y_2]$, και βλέπουμε ότι πλέον είναι γραμμικά διαχωρίσιμα.



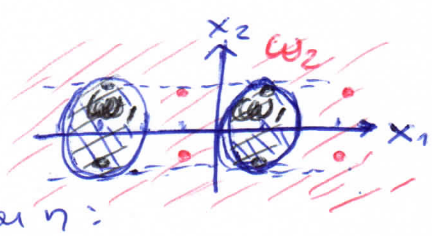
- Μια από τις άπειρες ευθείες που το πετυχαίνουν αυτό είναι η $g(\underline{y})$:

$$\boxed{y_1 + y_2 - e^{-1} = 0}$$

καθώς $(e^{-1}+e^{-5})-e^{-1} = e^{-5} > 0$

ενώ $e^{-3}+e^{-7}-e^{-1} = -0.3172 < 0$

και $e^{-3}+e^{-3}-e^{-1} = -0.2683 < 0$



- Η επιφάνεια διαχωρισμού στον αρχικό χώρο είναι η:

$$e^{-[(x_1-1)^2+x_2^2]} + e^{-[(x_1+3)^2+x_2^2]} - e^{-1} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} 0$$

ΑΣΚ. 3 (D)

DECISION TREE

- Οι πιθανές ερωτήσεις στους κόμβους του δέντρου είναι οι:

  Q_1: $x_1 > 2$

  Q_2: $x_1 > 0$

  Q_3: $x_1 > -2$

  Q_4: $x_2 > 0$

- Στο root note, $I_0 = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = 1$

- Αν ερωτηθεί η Q_4, προφανώς $I_{Q_4}(t_Y) = I_{Q_4}(t_N) = 1$, δηλαδή $\Delta I = 0$

- Τα Q_1 & Q_3 είναι "συμμετρικές" με:

  $I_{Q_1}(t_Y) = 0, \quad I_{Q_1}(t_N) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6}$

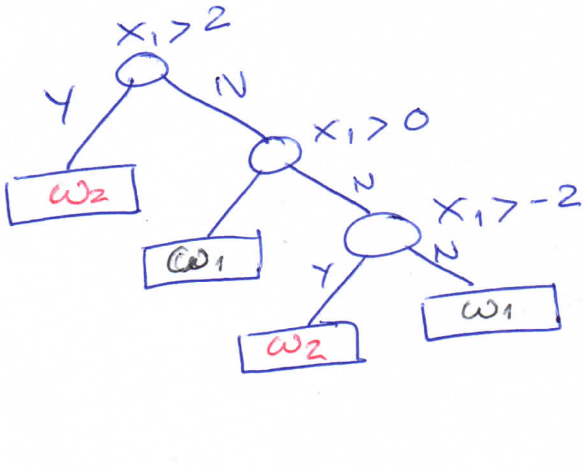
  Άρα $\Delta I = 1 - \cancel{\frac{1}{3}\cdot 0} - \frac{2}{3}\left[-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right] =$

  $= 1 + \frac{2}{3}\log\frac{1}{3} + \frac{4}{9} = 0.3878$

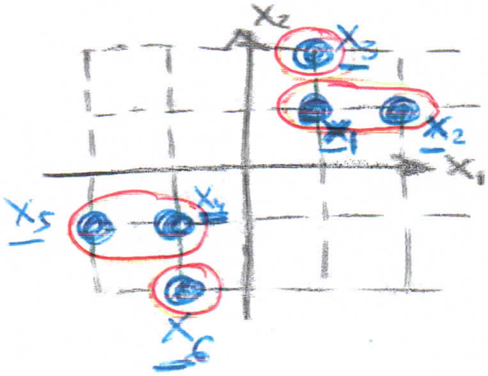
- Το Q_2 δίνει $I_{Q_1}(t_Y) = I_{Q_1}(t_N) = 1$, άρα $\Delta I = 0$

- Συνεπώς διαλέγουμε την Q_1 (ή ισοδύναμα, την Q_3):

- Οι επόμενη ερώτηση (Q_2 ή Q_3) έχει συμμετρία, οπότε:



ΑΣΚ 4Α

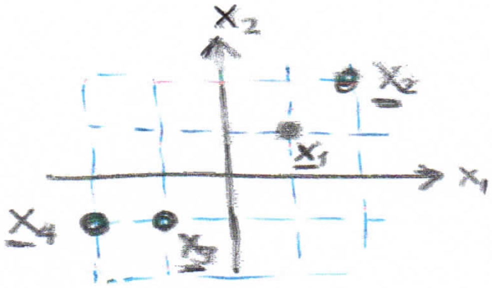


$\Rightarrow$ BSAS CLUSTERING (L_1 distance, $\theta = 1.25$)

ΒΗΜΑ	ΟΜΑΔΕΣ / ΣΥΛΛΟΓΙΣΤΙΚΗ
1:	$\{\underline{x}_1\}$
2:	$\{\underline{x}_1, \underline{x}_2\}$, καθώς $d(\underline{x}_1, \underline{x}_2) = 1 < \theta$ ΜΕΣΟ $\underline{c}_1 = [1.5, 1.0]^T$
3:	$\{\underline{x}_1, \underline{x}_2\}, \{\underline{x}_3\}$ καθώς $d(\underline{x}_3, \underline{c}_1) = 1.5 > \theta$
4:	$\{\underline{x}_1, \underline{x}_2\}, \{\underline{x}_3\}, \{\underline{x}_4\}$ καθώς $d(\underline{x}_4, \underline{x}_3) = 5 > \theta$ & $d(\underline{x}_4, \underline{c}_1) = 4.5 > \theta$
5:	$\{\underline{x}_1, \underline{x}_2\}, \{\underline{x}_3\}, \{\underline{x}_4, \underline{x}_5\}$ καθώς $d(\underline{x}_4, \underline{x}_5) = 1 < \theta$ & $d(\underline{x}_5, \underline{x}_3) > \theta$, $d(\underline{x}_5, \underline{c}_1) > \theta$ ΜΕΣΟ $\underline{c}_2 = \frac{\underline{x}_4 + \underline{x}_5}{2} = [-1.5, -1.0]^T$
6:	$\boxed{\{\underline{x}_1, \underline{x}_2\}, \{\underline{x}_3\}, \{\underline{x}_4, \underline{x}_5\}, \{\underline{x}_6\}}$ καθώς $d(\underline{x}_6, \underline{c}_1), d(\underline{x}_6, \underline{c}_2), d(\underline{x}_6, \underline{x}_3) > \theta$

Άρα έχουμε $\boxed{\text{4 ΟΜΑΔΕΣ}}$

ΑΣΚ. $\boxed{4B}$



$\Rightarrow$ AGGLOMERATIVE CLUSTERING

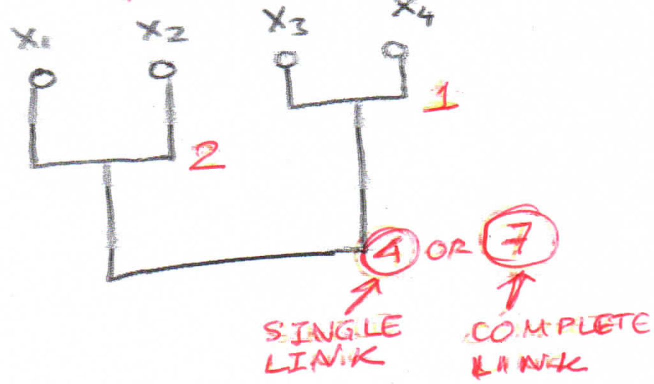
ℓ_1 DISTANCE

SINGLE/COMPLETE-LINK

ΠΙΝΑΚΑΣ ΑΝΟΜΟΙΟΤΗΤΑΣ

	x_1	x_2	x_3	x_4
x_1	0	2	4	5
x_2	2	0	6	7
x_3	4	6	0	1
x_4	5	7	1	0

DENDROGRAMS



$\{x_3, x_4\}$

SINGLE LINK

	x_1	x_2	$x_3 x_4$
x_1	0	2	4
x_2	2	0	6
$x_3 x_4$	4	6	0

COMPLETE LINK

	x_1	x_2	$x_3 x_4$
x_1	0	2	5
x_2	2	0	7
$x_3 x_4$	5	7	0

$\{x_1, x_2\}$

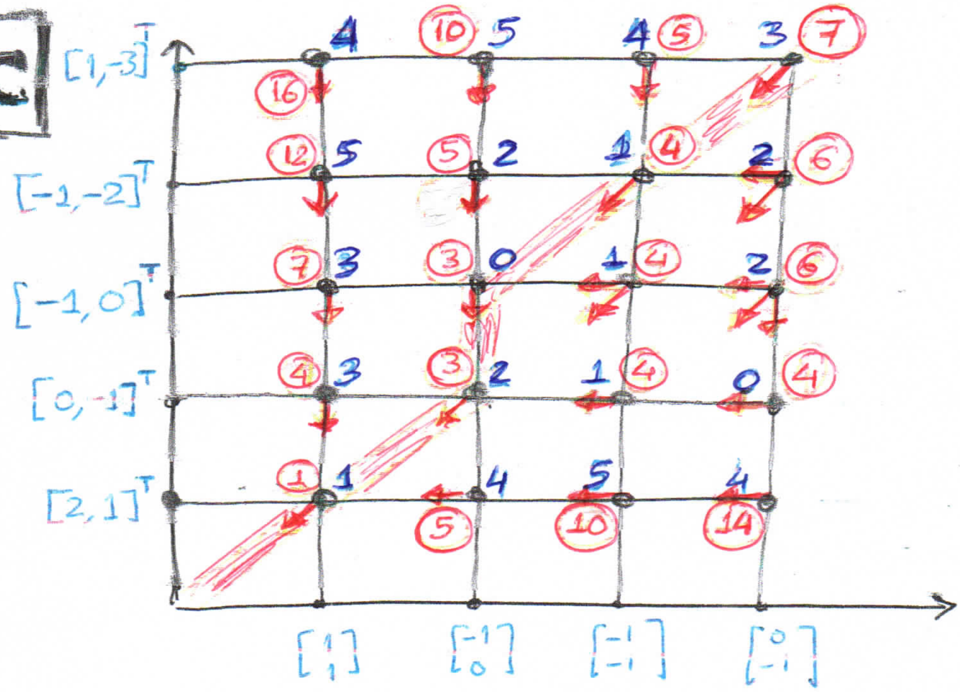
	$x_1 x_2$	$x_3 x_4$
$x_1 x_2$	0	4
$x_3 x_4$	4	0

	$x_1 x_2$	$x_3 x_4$
$x_1 x_2$	0	7
$x_3 x_4$	7	0

$\{x_1\ x_2\ x_3\ x_4\}$

ΑΣΚ

4c



- Συνολικό κόστος είναι 7 με το επισημειωμένο μονοπάτι με κόκκινο
- Έχουμε υπολογίσει όλες τις αποστάσεις $d(\underline{x}, \underline{y})$ στο πλέγμα (με μπλε).
- Επίσης, για κάθε σημείο του πλέγματος, το ελάχιστο κόστος μετάβασης στον κόμβο (με κόκκινο) και την κατεύθυνση της μετάβασης αυτής (από πού "έρχεται" το ελάχιστου κόστους μονοπάτι).