ΘΕΜΑΤΑ 2020-2021

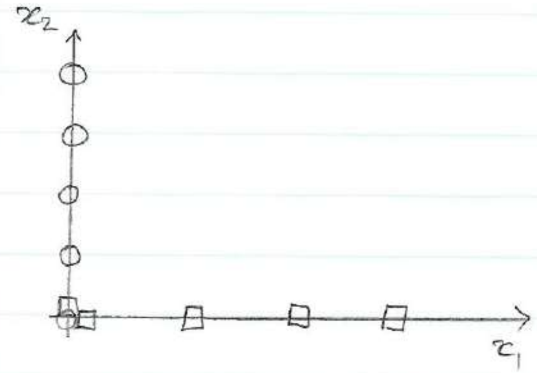
ΜΕΡΟΣ Α' - ΠΟΛΛΑΠΛΗΣ

1] $p(E,H|B) = p(H|B,E) \cdot p(E|B) \neq p(E|B) \cdot p(H|B)$

Άρα σωστή επιλογή η e.

2] $D_1 = [(0,0), (0,1), (0,2), (0,3), (0,4)] \rightarrow \circ$

$D_2 = [(1,0), (2,0), (3,0), (0,0), (0,0)] \rightarrow \square$



a. $x_1 + x_2 < 2$

Y: $(2w_1, 3w_2)$ N: $(3w_1, 2w_2)$

$i(Y) = 12/25$ $i(N) = 12/25$

$i(Q) = 12/25$

b. $x_1 > 0$

Y: $(0w_1, 3w_2)$ N: $(5w_1, 2w_2)$

$i(Y) = 0$ $i(N) = 20/49$

$i(Q) = \frac{7}{10} \cdot \frac{20}{49} = 2/7$

c. $x_2 < 1$

Y: $(1w_1, 5w_2)$ N: $(4w_1, 0w_2)$

$i(Y) = 5/18$ $i(N) = 0$

$i(Q) = \frac{6}{10} \cdot \frac{5}{18} = 1/6$

d. $x_2 < 0$

Y: $(0w_1, 0w_2)$ N: $(5w_1, 5w_2)$

$i(Y) = 0$ $i(N) = 1/2$

$i(Q) = 1/2$

e. $x_1 > 1$

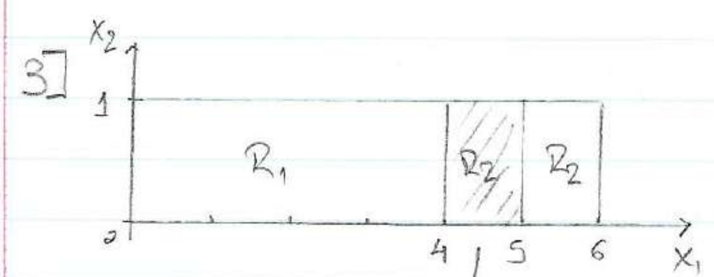
Y: $(0w_1, 2w_2)$ N: $(5w_1, 3w_2)$

$i(Y) = 0$ $i(N) = 15/32$

$i(Q) = \frac{8}{10} \cdot \frac{15}{32} = 3/8$

Τελικά, σωστή είναι η απάντηση c.

3]



εδώ υπάρχει misclassification error

$$p(\text{error}) = \int_4^5\int_0^1 d\vec{x}\, p(\vec{x}|\omega_1)p(\omega_1) = 0.5\cdot 0.2 = 10\%$$

Κάτι δεν πάει καλά είτε με τη λύση είτε με τις επιλογές...

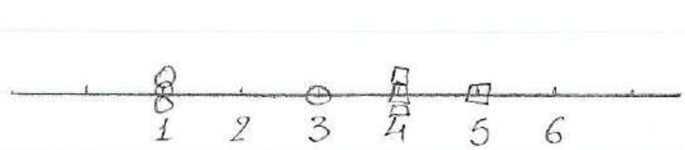
4] Το output είναι της μορφής

$$y = \Theta(w_1x_1 + w_2x_2 + b)$$ και είναι ισοδύναμο με

$$g(\vec{x}) = w_1'x_1 + w_2'x_2 + b' \rightarrow$$ ευθεία διαχωρισμού

Τα δεδομένα <u>δεν</u> ανήκουν σε γραμμικά διαχωρίσιμες κατηγορίες, συνεπώς σωστή επιλογή είναι η c.

5] $D_1 = (1,1,1,3) \rightarrow \circ$ $\qquad D_2 = (4,4,4,5) \rightarrow \square$



Αν θέλουμε απλώς majority vote, τότε το σημείο απόφασης είναι το $x = 2.5$, το οποίο δεν υπάρχει ως επιλογή. Εάν θέλουμε ομοφωνία, τότε τα σημεία απόφασης είναι τα $x=3.5$, $x=2.5$, με την περιοχή $[2.5, 3.5]$ να αντιστοιχεί σε «no decision». Πάλι κάτι δεν πάει καλά με τις επιλογές (ή με τη λύση).

6] Σωστή είναι η επιλογή a.

7] Τα support vectors είναι το 3 και δύο εκ των 1, 5, 7. Αφαιρώντας μόνο ένα εκ των 1, 5, 7 η λύση δεν αλλάζει. Το ίδιο ισχύει και για την αφαίρεση του 8, το οποίο ούτως ή άλλως δεν είναι support vector. Επομένως σωστή είναι η επιλογή b.

8] Το σύνολο της διασποράς αντιστοιχεί στο άθροισμα $0.5 + 1.15 + 1.3 + 1.7 + 2 = 6.65$. Η διασπορά που «ερμηνεύεται» από 3 μόνο κύριες συνιστώσες αντιστοιχεί στο άθροισμα των 3 υψηλότερων ιδιοτιμών: $1.3 + 1.7 + 2 = 5$. Έτσι, το ποσοστό είναι $5/6.65 \approx 75.19\%$, άρα σωστή είναι η επιλογή d.

9] $p(A|Q) = 0.3 \Rightarrow p(B|Q) = 0.7$
$p(B|R) = 0.3 \Rightarrow p(A|R) = 0.7$ } αφού οι πιθανότητες μετάβασης είναι ίσες,

Το decoding ισοδυναμεί με την αντιστοίχιση $B \to Q$, $A \to R$, συνεπώς $(B, A, B) \to (Q, R, Q)$, άρα σωστή είναι η επιλογή d.

10] Εφόσον το επιθυμητό επίπεδο είναι της μορφής $z = w_1 x + w_2 y$, εν γένει δε μπορεί πάντα να βρεθεί κατάλληλο $\vec{w}$ που να μηδενίζει το σφάλμα για 3 σημεία. Στην περίπτωση αυτή:

$$\left.\begin{array}{l} \bullet\ 5 = w_1 + w_2 \\ \bullet\ 11 = -w_1 + 3w_2 \end{array}\right\} \xRightarrow{(+)} 16 = 4w_2 \Rightarrow w_2 = 4, \text{ άρα } w_1 = 1, \text{ άρα } z = x_1 + 4x_2$$

Για $z = 18$ και $(x_1, x_2) = (2, 4)$ η εξ. επαληθεύεται, οπότε εδώ σωστή είναι η επιλογή d.

11] $D = (1,2,2,3,5,6,6,7)$

Αρχικοποίηση: $\mu_1 = 0$, $\mu_2 = 8$

1η iteration: $\{1,2,2,3\} \rightarrow \#1,\ \mu_1 = 2 \qquad \{5,6,6,7\} \rightarrow \#2,\ \mu_2 = 6$

2η iteration: ίδιο, οπότε επήλθε σύγκλιση και σωστή είναι η επιλογή e.

12] Η εκφώνηση θα έπρεπε να είναι «ποιο/α από τα παρακάτω ανήκει/-ουν σε marginal hyperplane», καθώς το κατά πόσο ένα σημείο αποτελεί support vector δεν καθορίζεται από αυτό (Mohri). Οπότε απαντάμε σε αυτό:

$$\text{Margin} = b = 1/\|\vec{w}\|_2$$

Για $g(\vec{x}) = x + y - 2z - 1$, ισχύουν:

$$d_a = \frac{|g(1,-1,1)|}{\|\vec{w}\|_2} = 3b, \quad d_b = 2b, \quad d_c = b \checkmark$$

$$d_e = 0$$

Άρα σωστή είναι η επιλογή c, δεδομένης της «αλλαγμένης» εκφώνησης.

Μερος Β' - Ασκησεις

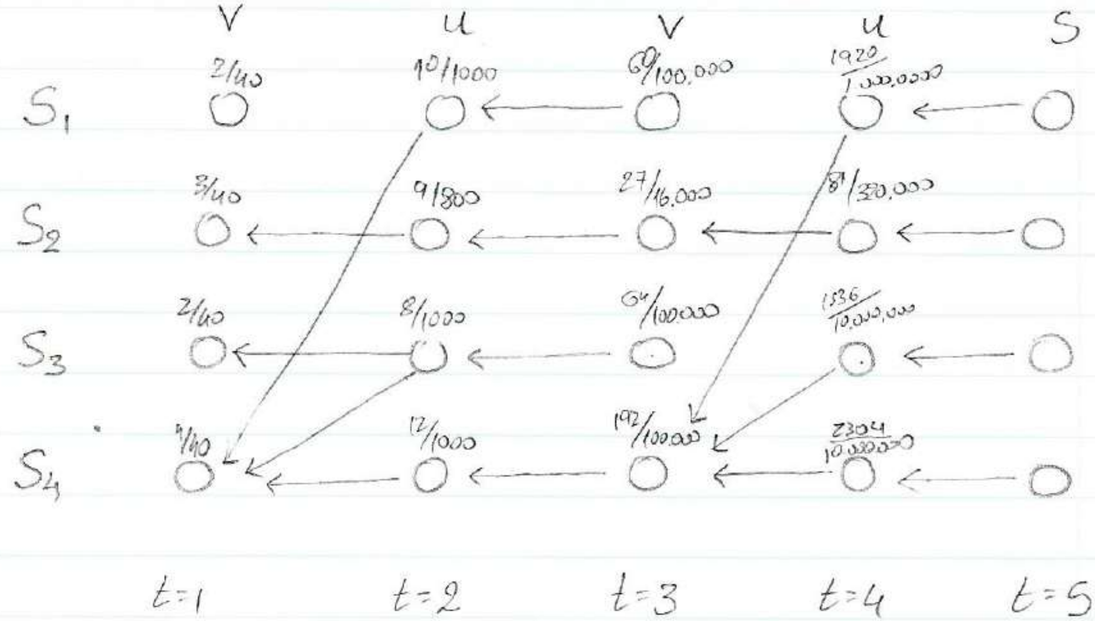
(1)
$$\underline{A} = \begin{pmatrix} 0.3 & 0.2 & 0.3 & 0.2 \\ 0.1 & 0.5 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{pmatrix} \qquad \vec{\pi} = (1/4, 1/4, 1/4, 1/4)$$

$b_{S_1}(V) = 0.2 \quad b_{S_2}(V) = 0.3 \quad b_{S_3}(V) = 0.2 \quad b_{S_4}(V) = 0.4$

$b_{S_1}(U) = 0.5 \quad b_{S_2}(U) = 0.3 \quad b_{S_3}(U) = 0.4 \quad b_{S_4}(U) = 0.3$

$b_{S_1}(S) = 0.3 \quad b_{S_2}(S) = 0.4 \quad b_{S_3}(S) = 0.4 \quad b_{S_4}(S) = 0.3$

Τα παραπάνω συνιστούν το μοντέλο λ.
Παρατηρούμενη ακολουθία: $\vec{O} = V, u, V, u, S$



Αρχικοποίηση: $\delta_1(S_1) = 1/4 \cdot 0.2 = 2/40 \qquad \delta_1(S_2) = 1/4 \cdot 0.3 = 3/40$

$\delta_1(S_3) = 1/4 \cdot 0.2 = 2/40 \qquad \delta_1(S_4) = 1/4 \cdot 0.4 = 4/40$

$t=2$: $\delta_2(S_1) = 0.5 \cdot 1/50 = \frac{1}{100}, \quad \psi_2(S_1) = S_4$

$\delta_2(S_2) = 0.3 \cdot 3/80 = 9/800, \quad \psi_2(S_2) = S_2$

$\delta_2(S_3) = 0.4 \cdot 1/50 = 1/125, \quad \psi_2(S_3) = S_3$ ή S_4

$\delta_2(S_4) = 0.3 \cdot 0.4 \cdot 0.1 = 12/1000, \quad \psi_2(S_4) = S_4$

$t=3: \ \delta_3(S_1) = 0.2 \cdot \frac{1}{100} \cdot 0.3 = \frac{6}{10000}, \quad \psi_3(S_1) = S_1$

$\delta_3(S_2) = 0.3 \cdot \frac{9}{800} \cdot 0.5 = \frac{27}{16000}, \quad \psi_3(S_2) = S_2$

$\delta_3(S_3) = 0.2 \cdot \frac{8}{1000} \cdot 0.4 = \frac{64}{100000}, \quad \psi_3(S_3) = S_3$

$\delta_3(S_4) = 0.4 \cdot \frac{12}{1000} \cdot 0.4 = \frac{192}{100.000}, \quad \psi_3(S_4) = S_4$

$t=4: \ \delta_4(S_1) = 0.5 \cdot \frac{192}{100.000} \cdot 0.2 = \frac{192}{1.000.000}, \quad \psi_4(S_1) = S_4$

$\delta_4(S_2) = 0.3 \cdot \frac{27}{16.000} \cdot 0.5 = \frac{81}{320.000}, \quad \psi_4(S_2) = S_2$

$\delta_4(S_3) = 0.4 \cdot \frac{192}{100.000} \cdot 0.2 = \frac{1536}{10.000.000}, \quad \psi_4(S_3) = S_4$

$\delta_4(S_4) = 0.3 \cdot \frac{192}{100.000} \cdot 0.4 = \frac{2304}{10.000.000}, \quad \psi_4(S_4) = S_4$

$t=5: \ \delta_5(S_1) = 0.3 \cdot \frac{1920}{10000000} \cdot 0.3 = 0.00001728 \quad \psi_5(S_1) = S_1$

$\delta_5(S_2) = 0.4 \cdot \frac{81}{320.000} \cdot 0.5 = 0.000050625, \quad \psi_5(S_2) = S_2$

$\delta_5(S_3) = 0.4 \cdot \frac{1536}{1000000} \cdot 0.4 = 0.000024576, \quad \psi_5(S_3) = S_3$

$\delta_5(S_4) = 0.3 \cdot \frac{2304}{10.000.00} \cdot 0.4 = 0.000027648, \quad \psi_5(S_4) = S_4$

Αφού $max = \delta_5(S_2)$, το backtracing μας οδηγεί στην

$$Q^* = S_2 \to S_2 \to S_2 \to S_2 \to S_2$$

Τέλος, ισχύει: $p^* = p(v,u,v,u,s, S_2, S_2, S_2, S_2, S_2 \mid \lambda) =$

$= \pi_{S_2} A_{S_2,S_2} \cdot A_{S_2,S_2} A_{S_2,S_2} A_{S_2,S_2} \cdot b_{S_2}(v)\, b_{S_2}(u)\, b_{S_2}(v)\, b_{S_2}(u)\, b_{S_2}(s)$

$= \frac{1}{4} \cdot \left(\frac{1}{2}\right)^4 \cdot \left(\frac{3}{10}\right)^4 \cdot \frac{4}{10} \Rightarrow p^* = \frac{81}{16} \cdot 10^{-9}$

2

$$\left.\begin{matrix} 7 & 5 & 4 \\ 7 & 4 & 6 \\ 6 & 4 & 7 \end{matrix}\right\}$$

Καθώς το max-pooling παίρνει το μέγιστο στοιχείο ενός 3x3 υπο-πίνακα.

③ ω_1: $\vec{\mu}_1 = (2,1)$, $\Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, άρα $|\Sigma_1| = 4$

$$p(\vec{x}|\omega_1) = \frac{1}{2\pi\sqrt{4}} \cdot \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} (\vec{x}-\vec{\mu}_1)\right) =$$

$$= \frac{1}{4\pi} \cdot \exp\left(-\frac{1}{4}\|\vec{x}-\vec{\mu}_1\|^2\right) = \frac{1}{4\pi} \cdot \exp\left[-\frac{(x_1-2)^2+(x_2-1)^2}{4}\right]$$

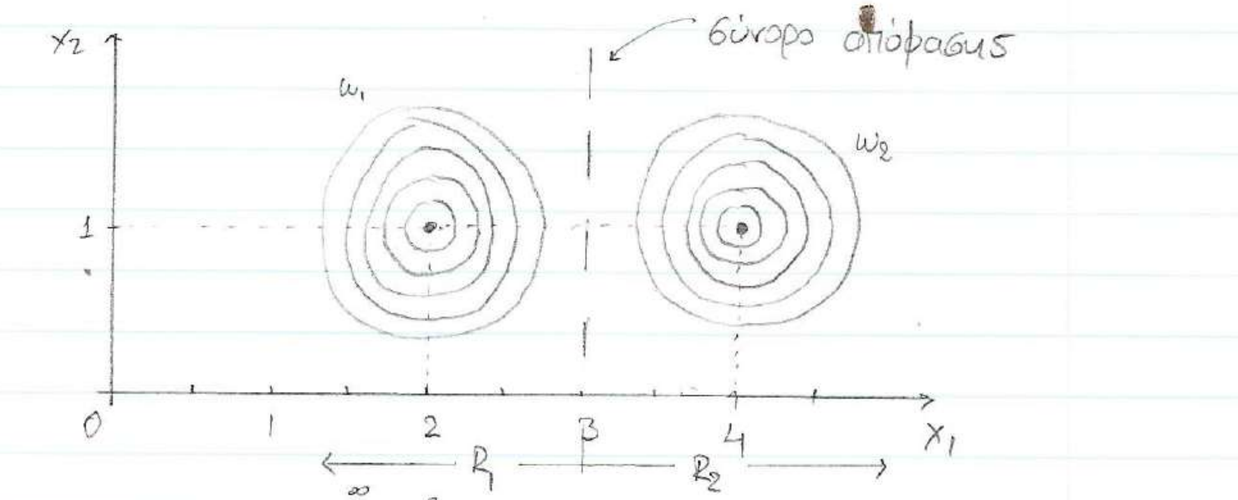
με $p(\omega_1) = 1/2$.

ω_2: $\vec{\mu}_2 = (4,1)$ και αντιστοίχως προκύπτει

$$p(\vec{x}|\omega_2) = \frac{1}{4\pi} \cdot \exp\left[-\frac{(x_1-4)^2+(x_2-1)^2}{4}\right], \text{ με } p(\omega_2) = 1/2$$

Decision Boundary: $p(\omega_1)\, p(\vec{x}|\omega_1) = p(\omega_2)\, p(\vec{x}|\omega_2) \Leftrightarrow$

$$(x_1-2)^2+(x_2-1)^2 = (x_1-4)^2+(x_2-1)^2 \Leftrightarrow |x_1-2| = |x_1-4| \Leftrightarrow x_1 = 3$$



Έστω $f(c) = \frac{1}{\sqrt{2\pi}} \int_c^{\infty} e^{-\frac{x^2}{2}} dx$. Για το σφάλμα κατά Bayes θα ισχύει:

$$p(\text{error}) = \int_{R_1} p(\omega_2)\, p(\vec{x}|\omega_2)\, d\vec{x} + \int_{R_2} p(\omega_1)\, p(\vec{x}|\omega_1)\, d\vec{x} =$$

$$= \frac{1}{8\pi} \int_{-\infty}^{+\infty} dx_2 \exp\left[-\left(\frac{x_2-1}{2}\right)^2\right] \left\{ \int_{-\infty}^{3} dx_1 \exp\left[-\left(\frac{x_1-4}{2}\right)^2\right] + \int_{3}^{+\infty} dx_1 \exp\left[-\left(\frac{x_1-2}{2}\right)^2\right] \right\}$$

Όμως $\int_{-\infty}^{+\infty} dx_2 \cdot \exp\left(-\frac{1}{2}\frac{(x_2-1)^2}{\sqrt{2}^2}\right) = \sqrt{2\pi}\cdot\sqrt{2}\cdot \int_{-\infty}^{+\infty} dx_2 \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}} e^{-\frac{1}{2}\frac{(x_2-1)^2}{\sqrt{2}^2}}$

$= 2\sqrt{\pi}\cdot \int_{-\infty}^{+\infty} N(1, \sqrt{2}^2)\, dx = 2\sqrt{\pi}$

Επίσης, $\int_{-\infty}^{3} dx_1 \cdot \exp\left(-\frac{1}{2}\left(\frac{x_1-4}{\sqrt{2}}\right)^2\right) \overset{u=\frac{x_1-4}{\sqrt{2}}}{=\!=\!=} \sqrt{2}\int_{-\infty}^{-1/\sqrt{2}} du \cdot e^{-\frac{u^2}{2}}$

$\overset{u\to -u}{=\!=} \sqrt{2}\int_{1/\sqrt{2}}^{+\infty} du \cdot e^{-\frac{u^2}{2}} = \sqrt{2}\, f(1/\sqrt{2})\cdot\sqrt{2\pi}$

Τέλος, $\int_{3}^{+\infty} dx_1 \cdot \exp\left(-\frac{1}{2}\left(\frac{x_1-2}{\sqrt{2}}\right)^2\right) \overset{u=\frac{x_1-2}{\sqrt{2}}}{=\!=\!=} \sqrt{2}\int_{1/\sqrt{2}}^{+\infty} du \cdot e^{-\frac{u^2}{2}} = \sqrt{2}\, f(1/\sqrt{2})\sqrt{2\pi}$

Βάσει αυτών, προκύπτει $p(error) = \frac{1}{8\pi}\cdot(2\sqrt{\pi})^2 \cdot 2\sqrt{2}\, f\left(\frac{1}{\sqrt{2}}\right) \Rightarrow$

$p(error) = \sqrt{2}\, f\left(\frac{1}{\sqrt{2}}\right)$

Προχωρούμε στις προβολές. Για την προβολή στον x, το άνυσμα πάνω στο οποίο προβάλλουμε είναι το $(1, 0)^T$, οπότε ισχύει $\tilde{\mu} = \mu_x$ και $\tilde{\sigma} = \sigma_i^2$. Άρα

$p(\tilde{x}|\omega_1) = \frac{1}{\sqrt{4\pi}} \cdot \exp\left[-\frac{1}{2}\frac{(x-2)^2}{2}\right]$ και

$p(\tilde{x}|\omega_2) = \frac{1}{\sqrt{4\pi}} \cdot \exp\left[-\frac{1}{2}\frac{(x-4)^2}{2}\right]$

Ξανά, το σημείο απόφασης είναι το $x = 3$. Σε ό,τι αφορά το σφάλμα, παρατηρούμε ότι τα σχετικά ολοκληρώματα είναι ίδια με τα άνω. Επίσης,

$\frac{1}{2}\cdot\frac{1}{\sqrt{4\pi}} = 2\sqrt{\pi} \circ \frac{1}{8\pi}$ (ελέγχω τους συντελεστές),

πράγμα που σημαίνει πως και εδώ $p(error) = \sqrt{2}\, f(\frac{1}{\sqrt{2}})$
Γιατί? Διότι η προβολή στον x διατηρεί όλη την πληροφορία, λόγω της συμμετρίας των $p(x|w_i)$. Αυτό σημαίνει πως η προβολή στον y θα πρέπει να δίνει μηδενική πληροφορία και άρα μέγιστο σφάλμα. Πράγματι, εκεί οι προκύπτουσες καμπύλες έχουν 100% ολική λύψη, οπότε η ταξινόμηση γίνεται τυχαία. Μαθηματικά:

$$\tilde{\mu} = \mu_y = 1, \quad \tilde{\sigma}^2 = \sigma_i^2 = 2, \text{ οπότε}$$

$$p(\tilde{y}|w_1) = \frac{1}{\sqrt{4\pi}} \exp\left[-\frac{1}{2}\frac{(y-1)^2}{2}\right] \text{ και}$$

$$p(\tilde{y}|w_2) = \frac{1}{\sqrt{4\pi}} \exp\left[-\frac{1}{2}\frac{(y-1)^2}{2}\right], \text{ δηλαδή } p(\tilde{y}|w_1) = p(\tilde{y}|w_2).$$

Αποτέλεσμα της προβολής αυτής είναι, λοιπόν, οι κλάσεις να μοιάζουν σαν να αντιστοιχούν στην ίδια κατανομή.