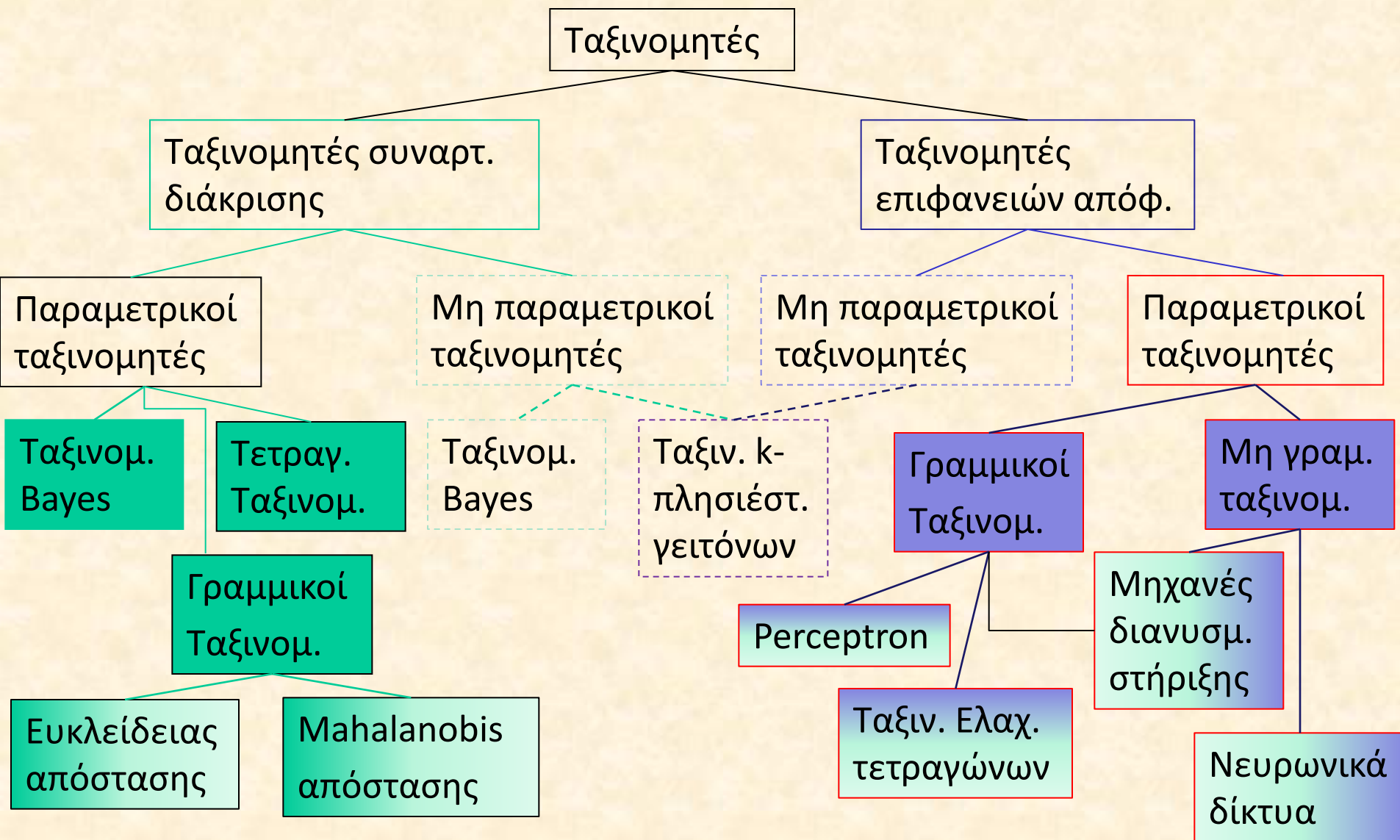


# ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

**Σέργιος Θεοδωρίδης**  
**Κωνσταντίνος Κουτρούμπας**

# “ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ



# “ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ

**Υπενθ.:**  $X$  είναι το σύνολο των δεδομένων σημείων όλων των κλάσεων

$X_j$  είναι το υποσύνολο του  $X$  που περιέχει τα διανύσματα της κλάσης  $\omega_j$ ,

$$X = X_1 \cup \dots \cup X_M$$

## Παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης

Ταξινομ. Bayes

- $g_j(x) = f(P(\omega_j) p(x|\omega_j))$
- Υπόθ.:  $p(x|\omega_j) \approx \hat{p}(x|\omega_j; \theta_j)$
- Εκτίμ.  $\theta_j$  με βάση το  $X_j$   
(ML, EM):  $X_j \rightarrow \hat{\theta}_j$
- $x \rightarrow g_j(x) = f(P(\omega_j) \hat{p}(x|\omega_j; \theta_j))$

Τετραγωνικός  
ταξινομητής

- $g_j(x) = -(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)$
- Υπόθ.:  $N(\mu_j, \Sigma_j)$
- Εκτίμ.  $\mu_j, \Sigma_j$  με βάση το  $X_j$   
(ML):  $X_j \rightarrow \hat{\mu}_j, \hat{\Sigma}_j$
- $x \rightarrow g_j(x) = (x - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1} (x - \hat{\mu}_j)$

Γραμμικός  
Ευκλείδειος  
ταξινομητής

- $g_j(x) = -(x - \mu_j)^T (x - \mu_j)$
- Υπόθ.:  $N(\mu_j, \sigma^2 I)$
- Εκτίμ.  $\mu_j$  με βάση το  $X_j$   
(ML):  $X_j \rightarrow \hat{\mu}_j$
- $x \rightarrow g_j(x) = (x - \hat{\mu}_j)^T (x - \hat{\mu}_j)$

Γραμμικός  
Mahalanobis  
ταξινομητής

- $g_j(x) = -(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)$
- Υπόθ.:  $N(\mu_j, \Sigma)$
- Εκτίμ.  $\mu_j, \Sigma$  με βάση το  $X_j$   
(ML):  $X_j \rightarrow \hat{\mu}_j, \hat{\Sigma}_j$
- Εκτίμ.  $\Sigma$  με βάση τα  $\hat{\Sigma}_j$   
(π.χ.  $\hat{\Sigma} = \frac{1}{M} \sum_{j=1}^M \hat{\Sigma}_j$ )
- $x \rightarrow g_j(x) = (x - \hat{\mu}_j)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_j)$

# “ΧΑΡΤΟΓΡΑΦΗΣΗ” ΤΟΥ ΧΩΡΟΥ ΤΩΝ ΤΑΞΙΝΟΜΗΤΩΝ

**Υπενθ.:**  $X$  είναι το σύνολο των δεδομένων σημείων όλων των κλάσεων

$X_j$  είναι το υποσύνολο του  $X$  που περιέχει τα διανύσματα της κλάσης  $\omega_j$ ,

$$X = X_1 \cup \dots \cup X_M$$

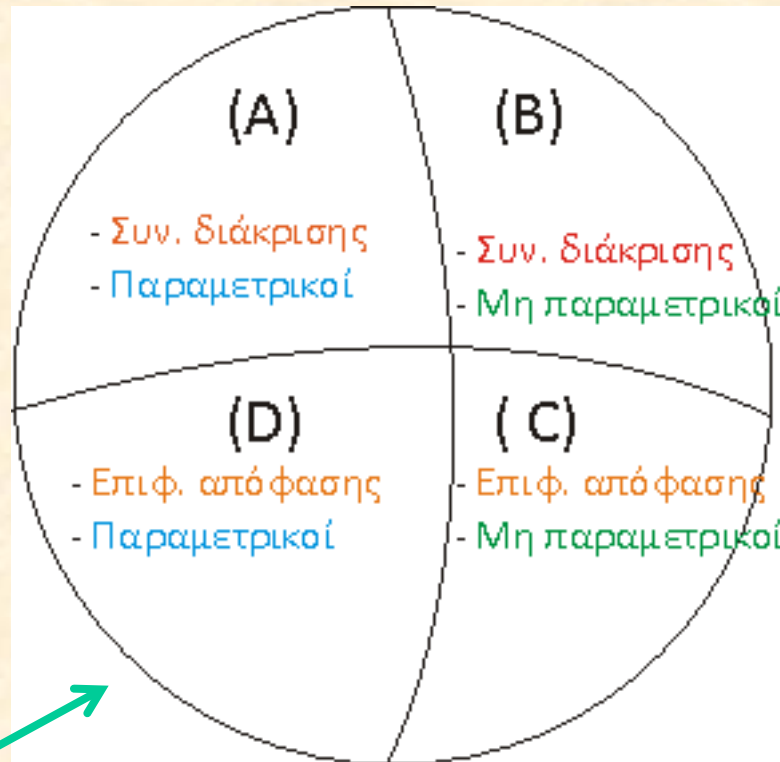
**Μη παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης**

Ταξινομητής  
Bayes

- $g_j(x) = f(P(\omega_j) p(x|\omega_j))$
- $x_i \rightarrow$  Εκτίμ.  $p(x|\omega_j) \approx \hat{p}(x|\omega_j; X_j)$   
(Παρ. Parzen,  $k$ -NN εκτ. πυκν. πιθ.)
- $x \rightarrow g_j(x) = f(P(\omega_j) \hat{p}(x|\omega_j; X_j))$

Ταξινομητής  $k$ -  
πλησιέστερων  
γειτόνων

$$- x \rightarrow g_j(x_i) = k^j$$



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μερικά προκαταρκτικά:

- Στην πράξη έχουμε στη διάθεσή μας ένα σύνολο δεδομένων (**training set**)

όπου 
$$X = \{(x_i, d_i), x_i \in R^l, d_i \in \{1, 2, \dots, M\}, i = 1, \dots, N\}$$

$x_i$  είναι η  $l$ -διάστατη αναπαράσταση της  $i$ -στής οντότητας ενός συνόλου  $N$  οντοτήτων (**training vector**)

$d_i$  είναι η ετικέτα της κλάσης στην οποία ανήκει το  $x_i$  (**1 για  $\omega_1$ , 2 για  $\omega_2, \dots$** ).

- Εστιάζουμε **κυρίως** στην περίπτωση **δύο κλάσεων**

(συνήθως  $\omega_1 \rightarrow +1$  ή **A** και  $\omega_2 \rightarrow -1(0)$  ή **B**).

- **Δεν υιοθετούμε κάποια υπόθεση** σχετικά με τις συναρτήσεις πυκν. πιθ. (**pdfs**) που μοντελοποιούν τις κλάσεις.

- Εστιάζουμε σε ταξινομητές που δύνανται να υλοποιήσουν **μη γραμμικούς διαχωρισμούς** μεταξύ των (δεδομένων των) κλάσεων.

-**ΣΗΜ.:** Μη γραμμικοί διαχωρισμοί μπορούν να επιτευχθούν είτε μέσω μιας **μη γραμμικής επιφανειας**, είτε μέσω **συνδυασμού μερικών γραμμικών η μη γραμμικών επιφανειών**. Σε κάθε περίπτωση ορίζεται μια **επιφάνεια απόφασης**.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Σημ.: Εκτός αν ορίζεται διαφορετικά, θεωρούμε την περίπτωση των **δύο κλάσεων**,  
i.e.,  $\omega_1 (+1)$  and  $\omega_2 (-1)$ .

**Ορισμός του προβλήματος:** Δοθέντος ενός συνόλου δεδομένων  $X$  σχεδιάσε ένα ταξινομητή που επιτυγχάνει τον **“βέλτιστο” δυνατό διαχωρισμό** των (διανυσμάτων των) δύο κλάσεων.

## Στρατηγική επίλυσης:

1. **Υιοθέτησε** ένα **συγκεκριμένο (μη γραμμικό) παραμετρικό μοντέλο** για τον ταξινομητή ( $w$  είναι το διάνυσμα που περιέχει όλες τις παραμέτρους του).
2. **Όρισε** κατάλληλη συνάρτηση κόστους (**cost function**) του  $w$ ,  $J(w)$ , η οποία εμπλέκει επίσης τα διανύσματα του  $X$ , έτσι ώστε **οι θέσεις των βέλτιστών της να αντιστοιχούν στο βέλτιστο δυνατό διαχωρισμό για το πρόβλημα.**  
**Βελτιστοποίησε** την  $J(w)$  ως προς  $w$ . Η θέση  $w$  όπου η  $J(w)$  παρουσιάζει βέλτιστο ορίζει τον καλύτερο δυνατό διαχωρισμό.

**Σημαντική παρατήρηση:** Η έννοια της φράσης **“καλύτερος δυνατός διαχωρισμός”** διαφέρει για **διαφορετικές επιλογές της  $J(w)$ .**

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Σύντομη υπενθύμιση:

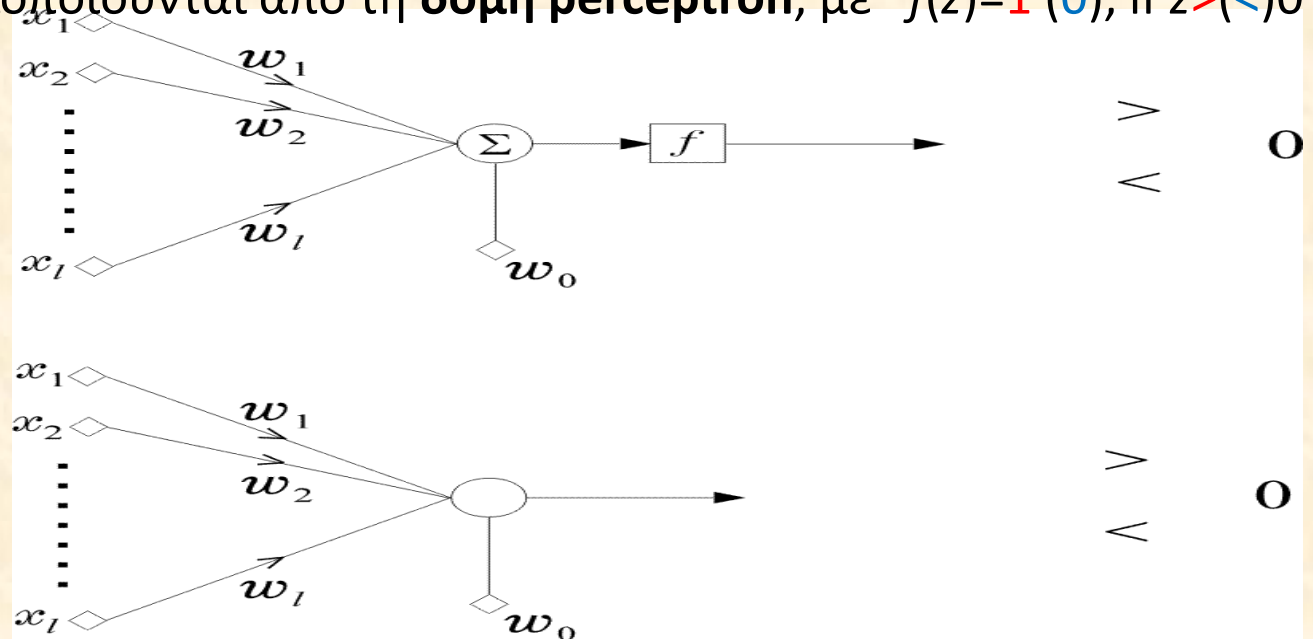
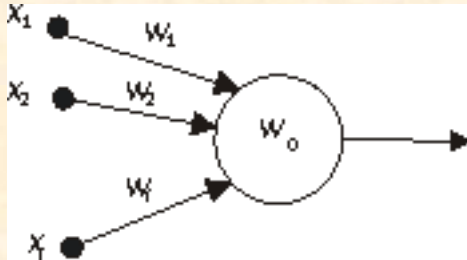
- Ένας γραμμικός ταξινομητής ορίζεται μέσω ενός υπερεπιπέδου

$$(H) : h(x, w) = w_1 x_1 + \dots + w_l x_l + w_0 = \sum_{k=1}^l w_k x_k + w_0 = w^T x + w_0 = 0$$

όπου  $w = [w_1, \dots, w_l]^T$ ,  $x = [x_1, \dots, x_l]^T$  (η γνώση των  $w$  και  $w_0$  ορίζει πλήρως το  $(H)$ )

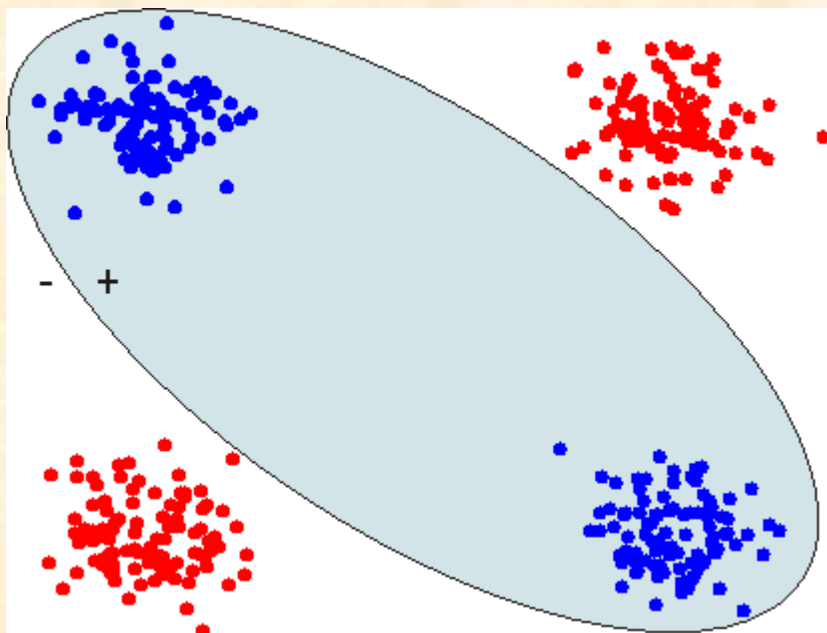
- Αν για δεδομένο  $x$  είναι  $h(x, w) = w^T x + w_0 \geq (<) 0$ , το  $x$  καταχωρείται στην κλάση  $+1 (0)$ .

- Τέτοιοι ταξινομητές υλοποιούνται από τη δομή perceptron, με  $f(z) = 1 (0)$ , if  $z > (<) 0$

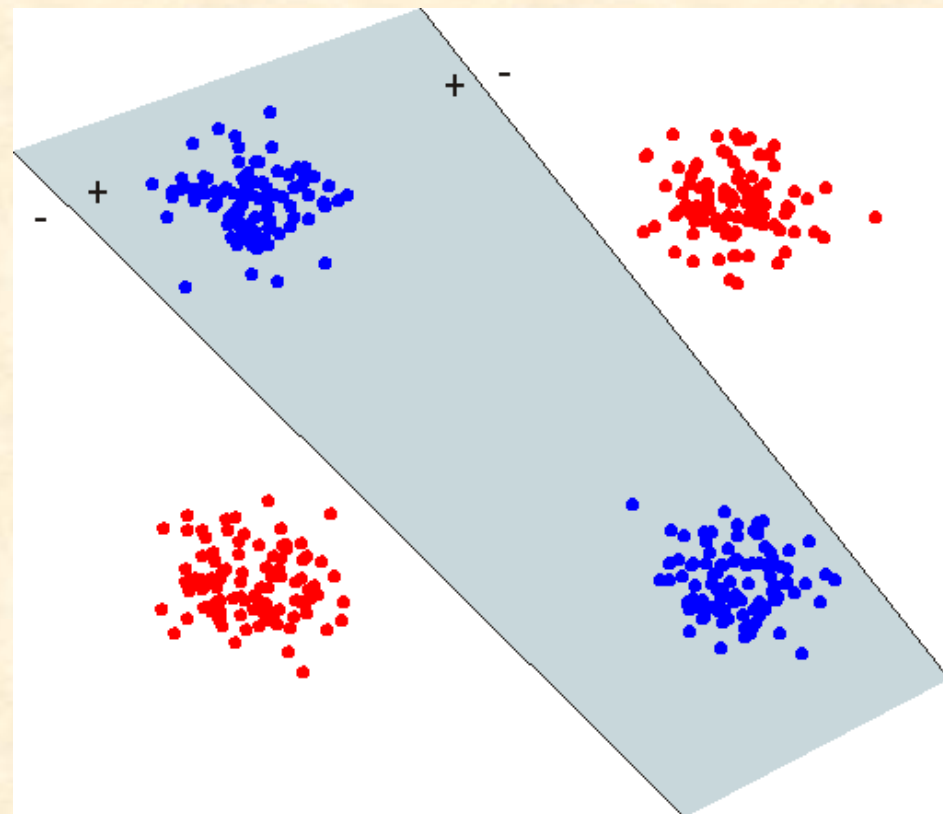




# ΜΗ ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ



Μία μη γραμμική επιφάνεια



Συνδυασμός περισσότερων της μίας επιφανειών

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

**Ερώτηση:** Με ποιο τρόπο **εισάγεται** η **μη γραμμικότητα**;

**Απάντηση:** Μέσω της **γραμμικότητας**.

**1ο σενάριο** (π.χ., **μηχανές διανυσματικής στήριξης** – SVMs)

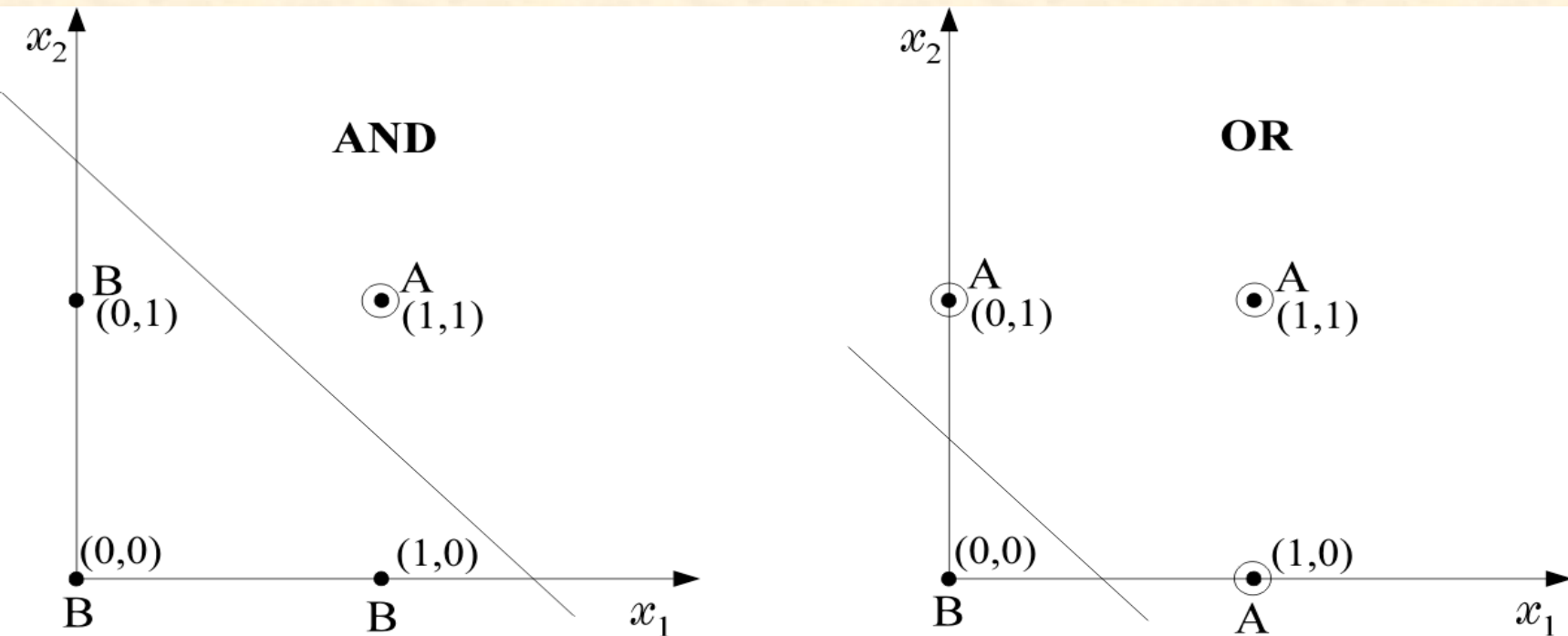
- **Απεικόνισε** τον χώρο των δεδομένων σε **ένα χώρο μεγαλύτερης διάστασης** όπου αυτά είναι πιο πιθανό να είναι (σχεδόν) γραμμικώς διαχωρίσιμα.
- **Λύσε** ένα **γραμμικό πρόβλημα** στο χώρο μεγαλύτερης διάστασης.

**2ο σενάριο** (π.χ., **νευρωνικά δίκτυα**)

- Μέσω της κατάλληλης εισαγωγής υπερεπιπέδων **απεικόνισε διαδοχικά τον αρχικό χώρο** σε **άλλους χώρους** όπου το πρόβλημα θα γίνεται όλο και «περισσότερο γραμμικώς διαχωρίσιμο».
- **Λύσε** το **γραμμικό πρόβλημα** στον τελευταίο από τους παραπάνω χώρους.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ

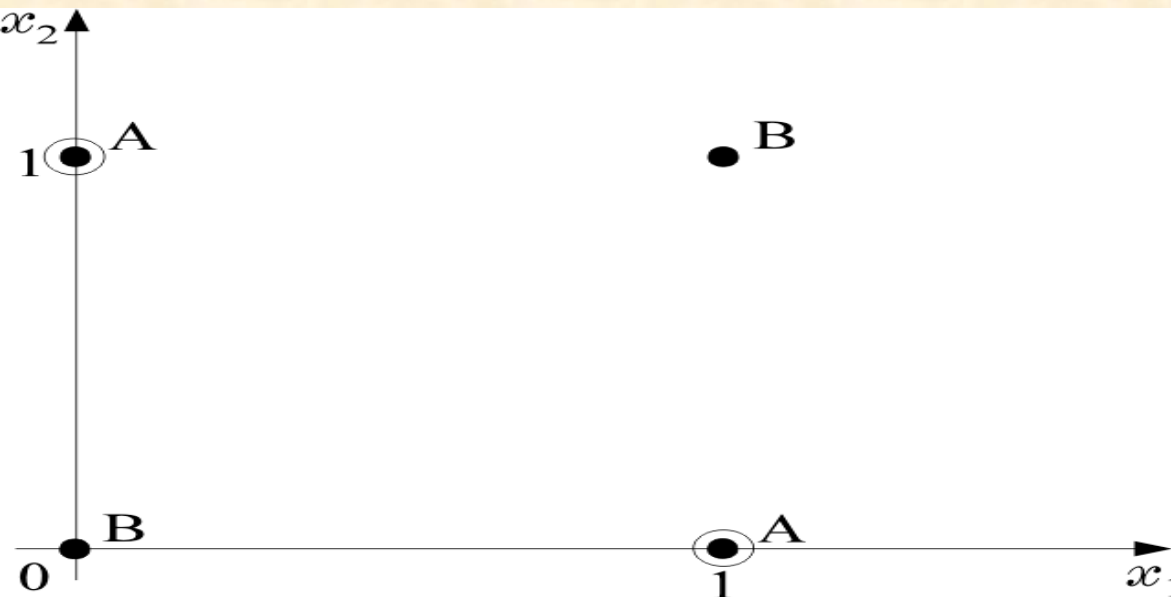
Τα προβλήματα **AND** και **OR** είναι γραμμικώς διαχωρίσιμα.



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ

Το πρόβλημα exclusive OR (**XOR**)

$x_1$	$x_2$	XOR	Class
0	0	0	B
0	1	1	A
1	0	1	A
1	1	0	B



Δεν υπάρχει καμία γραμμή (υπερεπίπεδο) που **διαχωρίζει** την κλάση **A** από την κλάση **B**.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

**Διατύπωση προβλήματος:** Έστω ένα μη γραμμικώς διαχωρίσιμο πρόβλημα ταξινόμησης στο χώρο  $R^l$ . Υπάρχουν  $k$  συναρτήσεις  $f_i(\cdot)$  μέσω των οποίων ο αρχικός χώρος μπορεί να απεικονιστεί σε ένα νέο  $k$ -διάστατο χώρο, όπου το πρόβλημα να είναι γραμμικώς διαχωρίσιμο;

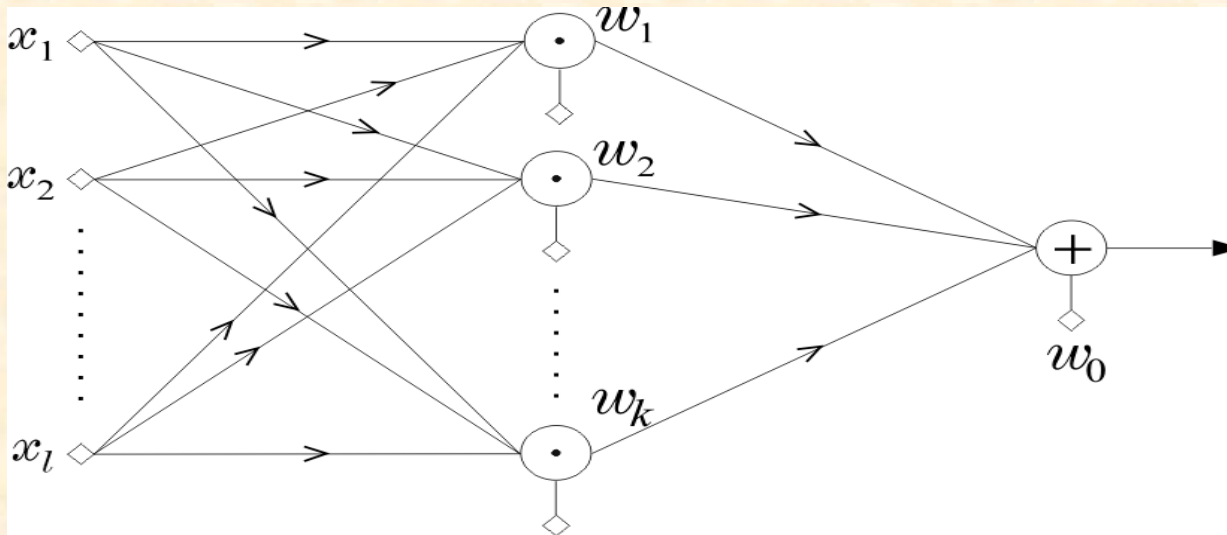
$$\underline{x} \rightarrow \underline{y} = \begin{bmatrix} f_1(\underline{x}) \\ \dots \\ f_k(\underline{x}) \end{bmatrix}$$

Αν συμβαίνει αυτό, η μη γραμμική επιφάνεια-σύνορο στον αρχικό χώρο ( $\mathbf{x}$ ) μπορεί να γραφεί ως γραμμική στο μετασχηματισμένο χώρο ( $\mathbf{y}$ ).

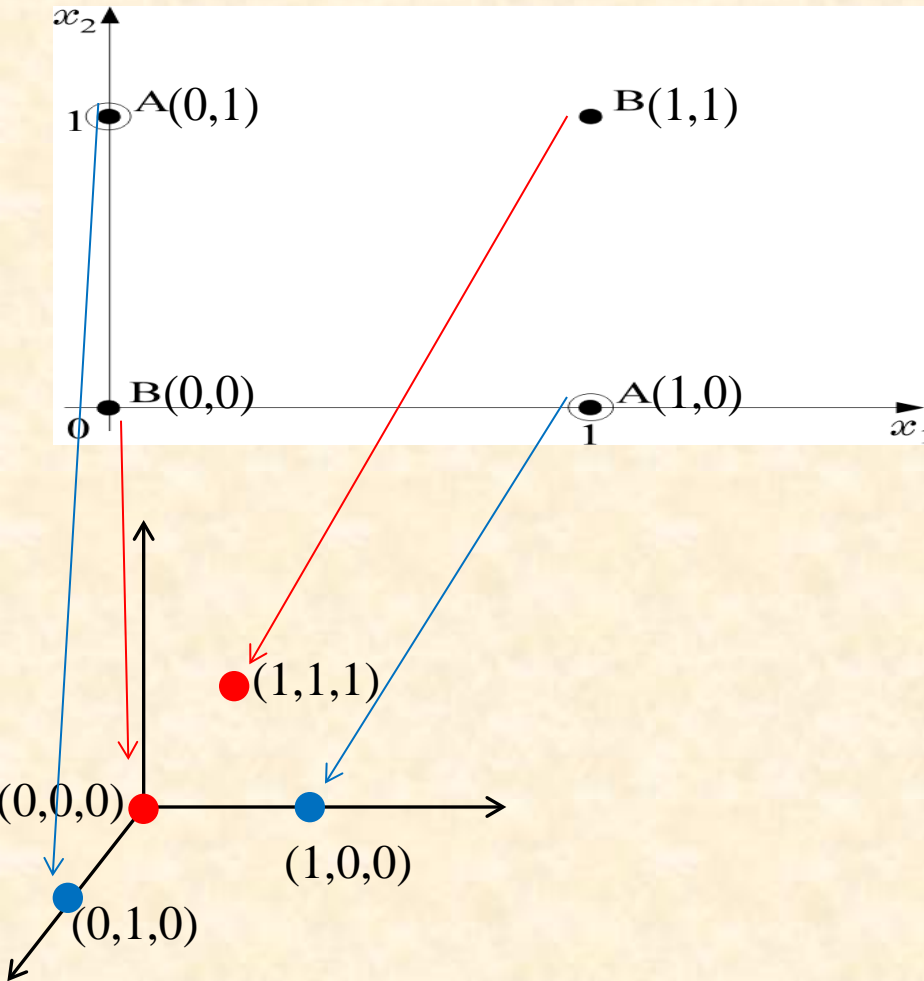
$$g(\underline{x}) \cong w_0 + \sum_{i=1}^k w_i f_i(\underline{x}) \quad (><) \quad 0$$

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

**ΣΗΜ.:** Σχηματικά η παραπάνω απεικόνιση  $\mathbf{x} \rightarrow \mathbf{y}$  φαίνεται στο παρακάτω **διάγραμμα**, που περιλαμβάνει δύο επίπεδα επεξεργασίας. Στο πρώτο επιτυγχάνεται η απεικόνιση του αρχικού χώρου σε ένα νέο χώρο (κάθε κόμβος αντιστοιχεί σε μια συνάρτηση  $f_i(\cdot)$ ). Το δεύτερο αντιστοιχεί σε ένα γραμμικό ταξινομητή.



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ



Έστω ο μετασχηματισμός

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix}$$

Στο νέο χώρο η διαδικασία ταξινόμησης γίνεται γραμμική

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

**Παράδειγμα 1:** Έστω

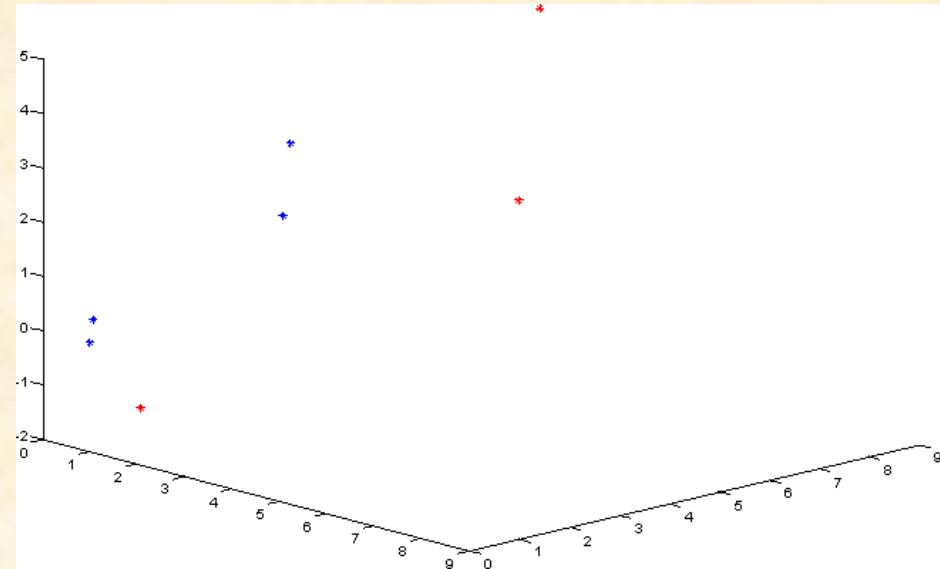
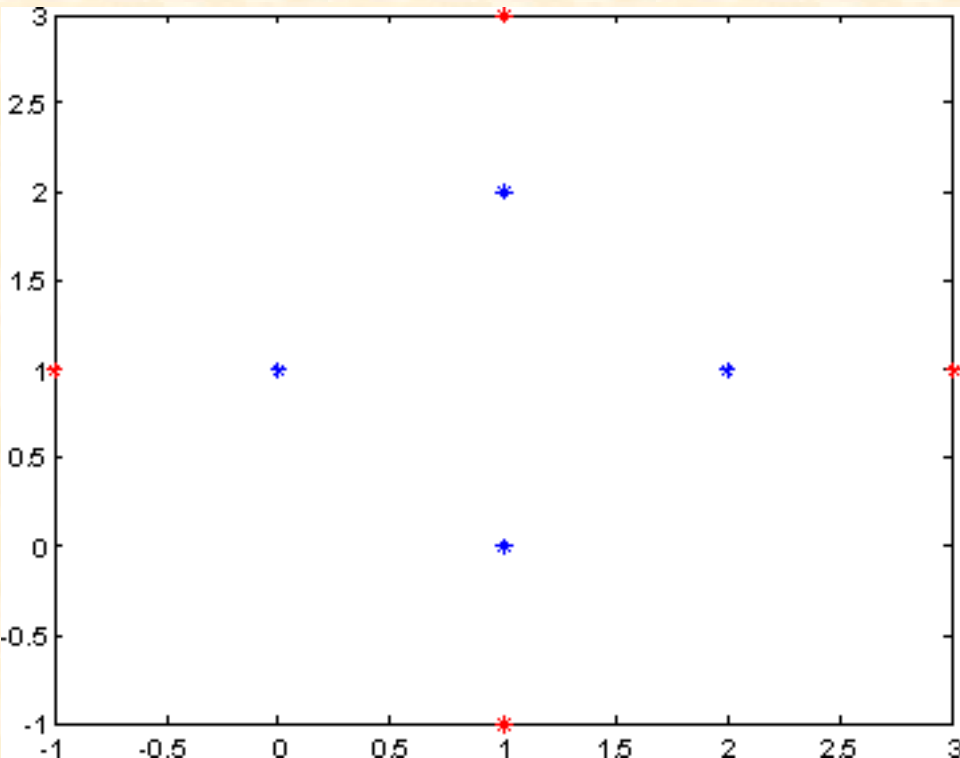
$\mathbf{x}_1=(2,1)$ ,  $\mathbf{x}_2=(1,2)$ ,  $\mathbf{x}_3=(0,1)$ ,  $\mathbf{x}_4=(1,0)$ ,  $\mathbf{x}_5=(3,1)$ ,  $\mathbf{x}_6=(1,3)$ ,  $\mathbf{x}_7=(-1,1)$ ,  $\mathbf{x}_8=(1,-1)$ .

Τα πρώτα 4 διανύσματα ανήκουν στην κλάση **+1** ενώ τα υπόλοιπα στην κλάση **-1**.

Έστω ο μετασχηματισμός

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix}$$

Στο νέο χώρο η διαδικασία ταξινόμησης γίνεται γραμμική





# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

**Παράδειγμα 2:** Έστω

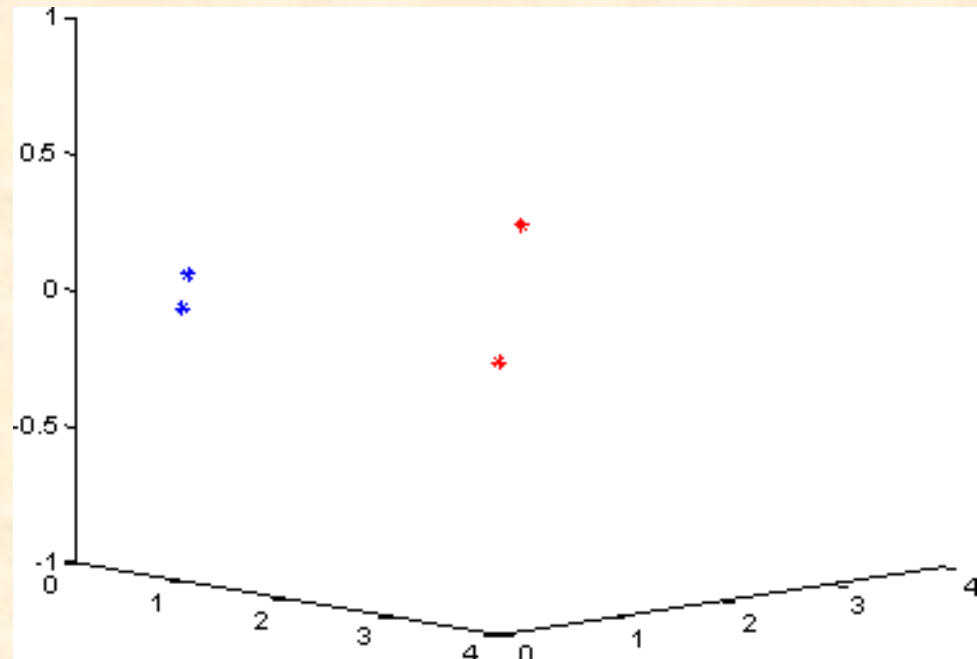
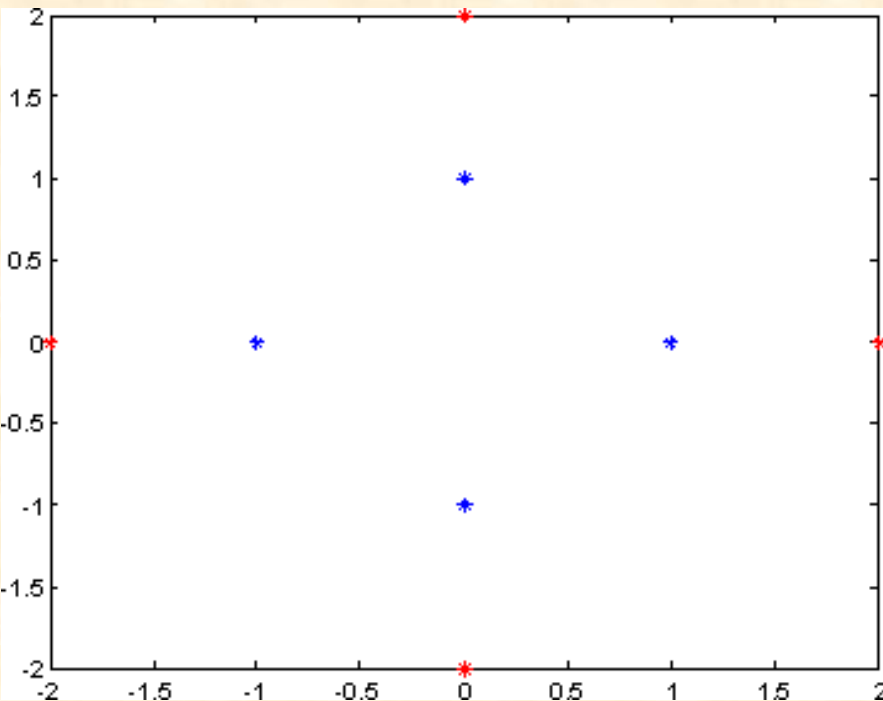
$\mathbf{x}_1=(1,0)$ ,  $\mathbf{x}_2=(0,1)$ ,  $\mathbf{x}_3=(-1,0)$ ,  $\mathbf{x}_4=(0,-1)$ ,  $\mathbf{x}_5=(2,0)$ ,  $\mathbf{x}_6=(0,2)$ ,  $\mathbf{x}_7=(-2,0)$ ,  $\mathbf{x}_8=(0,-2)$ .

Τα πρώτα 4 διανύσματα ανήκουν στην κλάση **+1** ενώ τα υπόλοιπα στην κλάση **-1**.

Έστω ο μετασχηματισμός

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix}$$

Στο νέο χώρο η διαδικασία ταξινόμησης γίνεται γραμμική

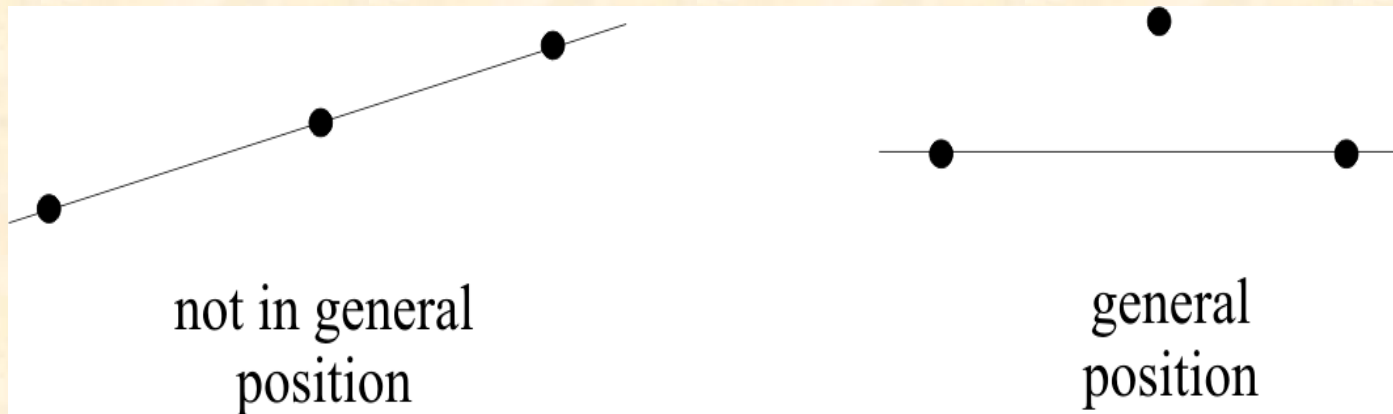


## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Χωρητικότητα του  $l$ -διάστατου χώρου σε γραμμικές διχοτομήσεις

Έστω  $N$  σημεία στον  $R^l$  χώρο που υποθέτουμε ότι βρίσκονται σε γενική θέση, δηλαδή:

Καμία ομάδα  $l + 1$  τέτοιων σημείων δεν βρίσκεται σε έναν  $l - 1$  διάστατο χώρο.



## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

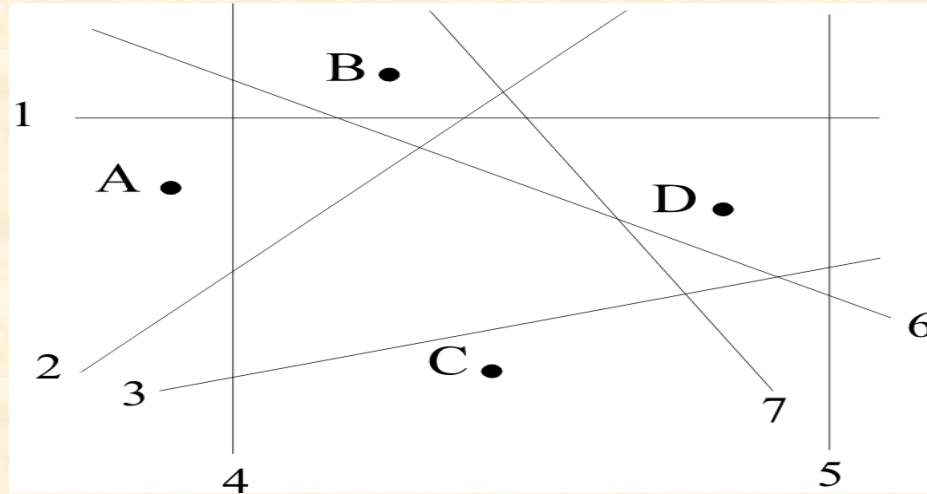
Χωρητικότητα του  $l$ -διάστατου χώρου σε γραμμικές διχοτομήσεις

**Θεώρημα Cover:** Ο αριθμός των ομαδοποιήσεων που μπορούν να υλοποιηθούν από  $(l-1)$ -διάστατα **υπερεπίπεδα** προκειμένου να διαχωριστούν  $N$  σημεία σε δύο κλάσεις είναι

$$O(N, l) = 2 \sum_{i=0}^l \binom{N-1}{i}, \quad \binom{N-1}{i} = \frac{(N-1)!}{(N-1-i)!i!}$$

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Παράδειγμα: Για  $N=4$ ,  $l=2$ , είναι  $O(4,2)=14$



Σημείωση: Ο συνολικός αριθμός δυνατών ομαδοποιήσεων είναι  $2^4=16$

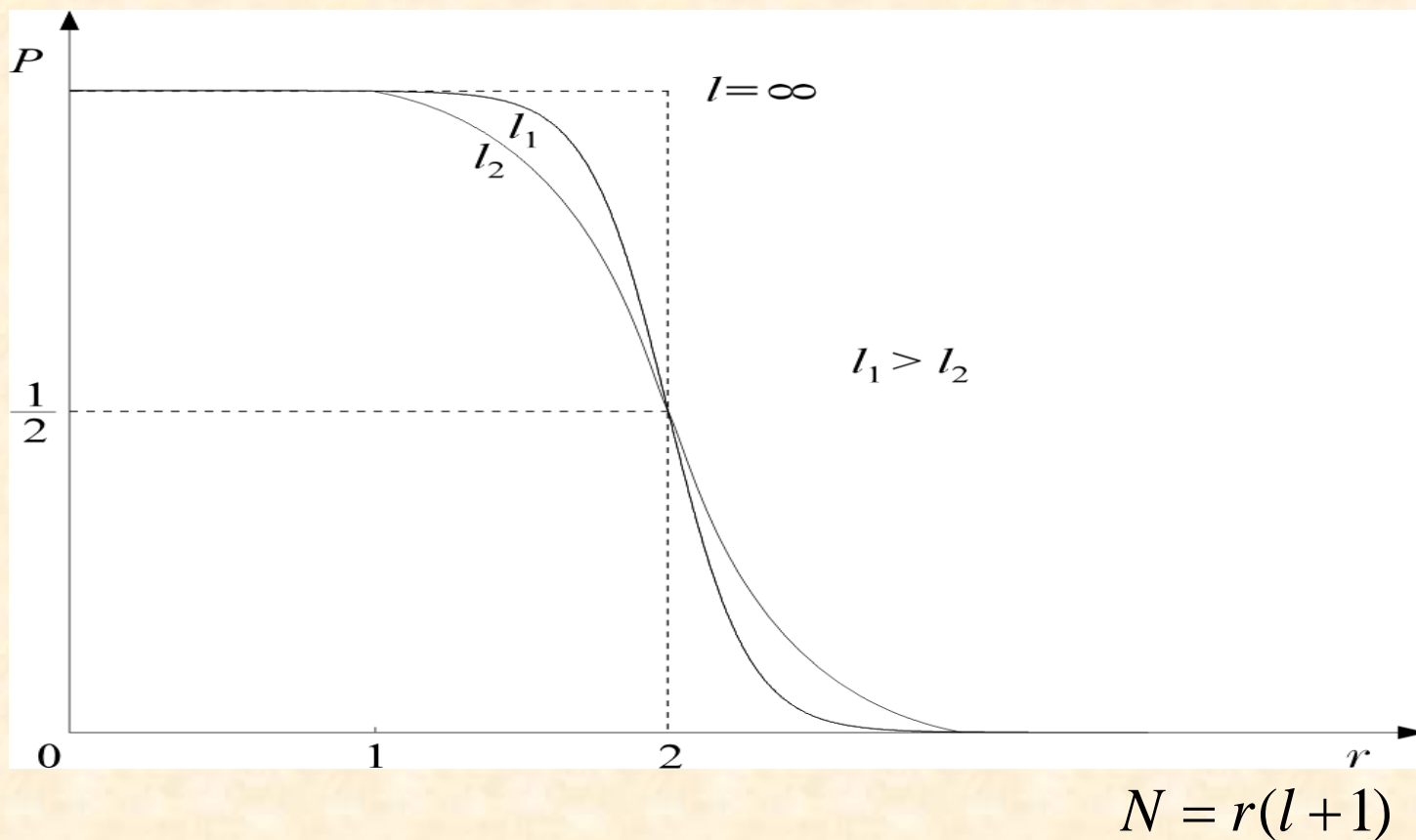
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
B	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
C	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
D	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Χωρητικότητα του  $l$ -διάστατου χώρου σε γραμμικές διχοτομήσεις

Η πιθανότητα ομαδοποίησης  $N$  σημείων σε δύο γραμμικώς διαχωρίσιμες κλάσεις είναι

$$\frac{O(N, l)}{2^N} = P_N^l$$



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Χωρητικότητα του  $l$ -διάστατου χώρου σε γραμμικές διχοτομήσεις

Έτσι, η πιθανότητα να έχουμε  $N$  σημεία σε γραμμικώς διαχωρίσιμες κλάσεις τείνει στο 1, για μεγάλο  $l$ , υπό την προϋπόθεση ότι  $N < 2(l + 1)$

Συνεπώς, απεικονίζοντας σε χώρο υψηλότερης διάστασης, αυξάνουμε την πιθανότητα γραμμικού διαχωρισμού, υπό την προϋπόθεση ότι ο χώρος δεν παρουσιάζει μεγάλη πυκνότητα σε σημεία δεδομένων.

## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Ερώτηση: Πώς τα παραπάνω μπορούν να χρησιμοποιηθούν στην πράξη;

Απάντηση:

Δοθέντος ενός συνόλου  $Y = \{(\mathbf{x}_n, d_n), \mathbf{x}_n \in R^l, d_n \in \{-1, 1\} n=1, \dots, N\}$ :

Απεικόνισε τα  $\mathbf{x}_n$  ως εξής:

$R^l \ni \mathbf{x}_n \rightarrow \varphi(\mathbf{x}_n) \in R^K$ , με  $K \gg l$

Όσο *μεγαλύτερο* είναι το  $K$ , τόσο *μεγαλύτερη* είναι η *πιθανότητα* να καταστεί το πρόβλημα *γραμμικώς διαχωρίσιμο*.

Όρισε  $\varphi(\mathbf{x}) = [\varphi_1(\mathbf{x}), \dots, \varphi_K(\mathbf{x})]^T$

Προσδιόρισε το υπερεπίπεδο

(H) :  $\mathbf{w}^T \varphi(\mathbf{x}) + w_0 = 0$

που *διαχωρίζει με βέλτιστο τρόπο* τα δεδομένα στο *μετασχηματισμένο χώρο*.

**Γενικό σχόλιο:**

Στην περίπτωση *προβλημάτων ταξινόμησης*, η *απεικόνιση* σε *χώρους υψηλότερης διάστασης* καθιστά το *πρόβλημα* περισσότερο “*γραμμικώς διαχωρίσιμο*”.

## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

$$Y = \{(\mathbf{x}_n, d_n), \mathbf{x}_n \in \mathbb{R}^l, d_n \in \{-1, 1\} \ n=1, \dots, N\}$$

**Παράδειγμα:**

Τα δεδομένα στο **επίπεδο** στο διπλανό σχήμα **δεν** είναι **γραμμικώς διαχωρίσιμα**.

Απεικονίζοντας το επίπεδο στον 3-διάστατο χώρο μέσω του μετασχηματισμού

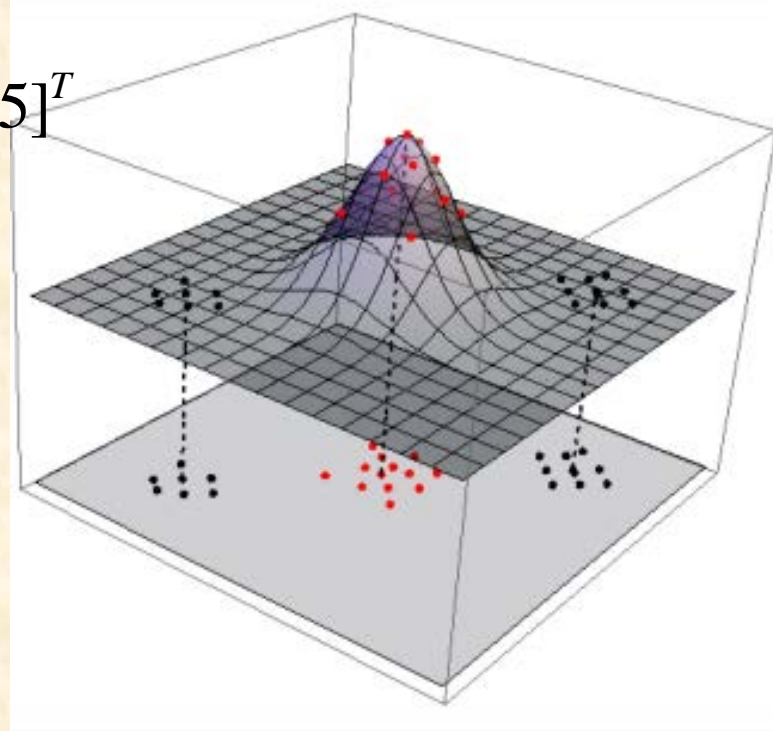
$$[x_1, x_2]^T \xrightarrow{\varphi} [x_1, x_2, 4 \exp(-(x_1^2 + x_2^2)/3) + 5]^T$$

Τα σημεία των δύο κλάσεων στον **3-διάστατο χώρο** είναι **γραμμικώς διαχωρίσιμα**.

**ΣΗΜ.:** Τα σημεία στον 3-d χώρο βρίσκονται σε παραβολοειδή επιφάνεια (περιγράφεται από δύο ανεξάρτητες παραμέτρους).

Στην πραγματικότητα,

**Ενσωματώνουμε** μια  **$l$ -διάστατη μη γραμμική επιφάνεια** (όπου ανήκουν τα δεδομένα) σε έναν  **$K$ -διάστατο χώρο**, ώστε **τα δεδομένα των δύο κλάσεων να καταστούν γραμμικώς διαχωρίσιμα**.





## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μια δημοφιλής επιλογή για τις  $\varphi_k(\cdot)$ : **πολυωνυμικές συναρτήσεις**, π.χ., για πολυώνυμα τάξεως  $r=2$  έχουμε

$$w_0 + \sum_{i=1}^l w_i x_i + \sum_{i=1}^l \sum_{j=1}^l w_{ij} x_i x_j = 0$$

**Σχόλια:** Ο αριθμός των **ελεύθερων παραμέτρων**  $K$  αυξάνει με ταχύ ρυθμό καθώς αυξάνονται τα  $l$  και  $r \Rightarrow$

**Αυξημένες υπολογιστικές απαιτήσεις** για τον **υπολογισμό** του  $\theta^T \varphi(x)$ .

- ❖ Μια **καλή προσέγγιση** μπορεί να απαιτεί **αυξημένη τάξη πολυωνύμου**  $r$  ( $\Rightarrow$  **αυξημένο**  $K$ ).
- ❖ Άλλες δημοφιλείς **επιλογές** για τις  $\varphi(\cdot)$ , είναι οι **τριγωνομετρικές συναρτήσεις**, τα **splines** και οι **συναρτήσεις ακτινικής βάσης**.
- ❖ Τα **γενικευμένα γραμμικά μοντέλα** είναι **γραμμικά** ως προς τις **ελεύθερες παραμέτρους**.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

Ας θυμηθούμε ότι η πιθανότητα να έχουμε γραμμικώς διαχωρίσιμες κλάσεις αυξάνει καθώς αυξάνει και η διάσταση των χώρου χαρακτηριστικών. Έστω η απεικόνιση:

$$x \in R^l \rightarrow y \in R^k, \quad k > l$$

Τότε μπορούμε να χρησιμοποιήσουμε SVM στο χώρο  $R^k$ .

Υπενθυμίζουμε ότι το **δυϊκό πρόβλημα που θα λύσουμε** θα είναι το

$$\underset{\lambda \geq 0}{\text{maximize}} \left( \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j d_i d_j y_i^T y_j \right)$$

ΟΠΟΥ  $y_i \in R^k$

Ο ταξινομητής στο μετασχηματισμένο χώρο ( $\underline{y}$ ) εκφράζεται ως

$$g(\underline{y}) = w^T \underline{y} + w_0 = \sum_{i=1}^{N_s} \lambda_i d_i y_i^T \underline{y}$$

ΟΠΟΥ  $x \rightarrow y \in R^k$

Έτσι, εμπλέκονται εσωτερικά γινόμενα σε ένα χώρο υψηλής διάστασης → **υψηλή υπολογιστική πολυπλοκότητα.**

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

- Κάτι έξυπνο: Υπολόγισε τα εσωτερικά γινόμενα στον χώρο **υψηλής** διάστασης σαν συνάρτηση των εσωτερικών γινομένων στο χώρο **χαμηλής** διάστασης!!!
- Είναι αυτό ΔΥΝΑΤΟΝ;; Ναι. Να ένα παράδειγμα

$$\text{Εστω } x = [x_1, x_2]^T \in R^2$$

$$\text{Εστω } x \rightarrow y = \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix} \in R^3$$

Τότε είναι εύκολο να δείξουμε ότι

$$y_i^T y_j = (x_i^T x_j)^2$$

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

➤ **Θεώρημα του Mercer:** Έστω  $\mathbf{x} \in R^l$  και μια απεικόνιση  $f$

$$\mathbf{x} \rightarrow f(\mathbf{x}) \in H,$$

όπου  $H$  είναι ένας **χώρος Hilbert**<sup>(1)</sup>. Για την πράξη του εσωτερικού γινομένου στο χώρο  $H$ , που συμβολίζεται με  $\langle \cdot, \cdot \rangle$ , ισχύει

$$\langle f(\mathbf{x}), f(\mathbf{z}) \rangle = K(\mathbf{x}, \mathbf{z}),$$

όπου  $K(\mathbf{x}, \mathbf{z})$ , μία συνεχής **συμμετρική** συνάρτηση (**πυρήνας** – **kernel**) που ικανοποιεί τη συνθήκη

$$\int_C \int_C K(\mathbf{x}, \mathbf{z}) g(\mathbf{x}) g(\mathbf{z}) d\mathbf{x} d\mathbf{z} \geq 0$$

για κάθε  $g(\mathbf{x})$ ,  $\mathbf{x} \in C \subset R^l$  με  $\int g^2(\underline{x}) d\underline{x} < +\infty$ ,

➤ **Αντίστροφα:** Κάθε συνεχής, συμμετρική συνάρτηση  $K(\mathbf{x}, \mathbf{z})$  που ικανοποιεί τις παραπάνω συνθήκες, αντιστοιχεί στο εσωτερικό γινόμενο **ΚΑΠΟΙΟΥ** χώρου.

<sup>(1)</sup> **Χώρος Hilbert:** Πλήρης γραμμικός (**πεπερασμένης** ή **άπειρης διάστασης**) χώρος εξοπλισμένος με μια πράξη εσωτερικού γινομένου. Όταν η διάσταση είναι πεπερασμένη έχουμε έναν Ευκλείδειο χώρο.

## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

Παραδείγματα συναρτήσεων πυρήνων

Συναρτήσεις ακτινωτής βάσης (Radial basis functions) (για κατάλληλες τιμές του  $\sigma$ )

$$K(\underline{x}, \underline{z}) = \exp\left(-\frac{\|\underline{x} - \underline{z}\|^2}{\sigma^2}\right)$$

Πολυωνυμικές συναρτήσεις (Polynomial functions) (για κατάλληλες τιμές του  $q$ )

$$K(\underline{x}, \underline{z}) = (\underline{x}^T \underline{z} + 1)^q, \quad q > 0$$

Συνάρτηση υπερβολικής εφαπτομένης (Hyperbolic tangent functions) (για κατάλληλες τιμές των  $\beta$  και  $\gamma$ )

$$K(\underline{x}, \underline{z}) = \tanh(\beta \underline{x}^T \underline{z} + \gamma)$$

## ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

Μια σημαντική παρατήρηση: Στο παρόν πλαίσιο, υπονοείται μια μη γραμμική απεικόνιση

$$f: \mathbf{x} \in R^l \rightarrow \mathbf{y} \in R^k$$

η οποία, ωστόσο, είναι **άγνωστη**. Έτσι, για δεδομένο διάνυσμα  $\mathbf{x}$ , **δεν είναι γνωστή** η εικόνα του  $\mathbf{y}=f(\mathbf{x})$  στο μετασχηματισμένο χώρο (με άλλα λόγια, ο μετασχηματισμένος χώρος δεν μας είναι ρητά γνωστός).

Αυτό που μας είναι **γνωστό** είναι η συνάρτηση  $K(.,.)$  (η οποία καλείται **συνάρτηση πυρήνα – kernel function**) η οποία εκφράζει το εσωτερικό γινόμενο δύο διανυσμάτων στο μετασχηματισμένο (υψηλής διάστασης) χώρο, σε συνάρτηση με το αντίστοιχο εσωτερικό γινόμενο στον αρχικό χώρο, δηλ.,

$$\mathbf{y}_i^T \mathbf{y}_j = K(\mathbf{x}_i, \mathbf{x}_j)$$

Μ' άλλα λόγια, για δύο δεδομένα διανύσματα, μας είναι γνωστό **το εσωτερικό γινόμενό τους στο μετασχηματισμένο χώρο**.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

Τα βήματα για τη λύση του SVM προβλήματος

- Βήμα 1: **Επέλεξε κατάλληλη συνάρτηση πυρήνα**. Αυτό υπονοεί απεικόνιση σε έναν (άγνωστο) χώρο υψηλότερης διάστασης.

- Βήμα 2: 
$$\max_{\lambda} \left( \sum_i \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j d_i d_j K(x_i, x_j) \right)$$
 subject to:  $0 \leq \lambda_i \leq C, i = 1, 2, \dots, N$   
$$\sum_i \lambda_i d_i = 0$$

Αυτό καταλήγει σε έναν **υπονοούμενο** συνδυασμό

$$w = \sum_{i=1}^{N_s} \lambda_i d_i f(x_i)$$

- Βήμα 3: **Καταχώρησε** ένα δεδομένο  $x$  στην κλάση **+1** (**-1**), ανάλογα με το αν

$$g(x) = \sum_{i=1}^{N_s} \lambda_i d_i K(x_i, x) + w_0 > (<) 0$$

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

Η λογική πίσω από το βήμα 3:

- Έστω  $\mathbf{x}$  δεδομένο διάνυσμα και  $\mathbf{y}=f(\mathbf{x})$  η εικόνα του στο μετασχηματισμένο χώρο.
- Προκειμένου να ταξινομήσουμε το  $\mathbf{x}$  σε μία από τις κλάσεις  $+1$  ή  $-1$ , πρέπει να ελέγξουμε αν η ποσότητα

$$\mathbf{w}^T \mathbf{y} + w_0$$

είναι **θετική** ή **αρνητική**.

- Ωστόσο, αφού η απεικόνιση  $f(\cdot)$  είναι άγνωστη, δεν γνωρίζουμε ούτε το  $\mathbf{y}$  ούτε τα  $\mathbf{y}_i=f(\mathbf{x}_i)$  που εμπλέκονται στον τύπο του  $\mathbf{w}$  (που επαναλαμβάνεται παρακάτω).

$$\mathbf{w} = \sum_{i=1}^{N_s} \lambda_i d_i f(\mathbf{x}_i)$$

- Παρ' όλες τις παραπάνω δυσχέρειες, η ποσότητα  $\mathbf{w}^T \mathbf{y} + w_0$  μπορεί να υπολογιστεί ως

εξής

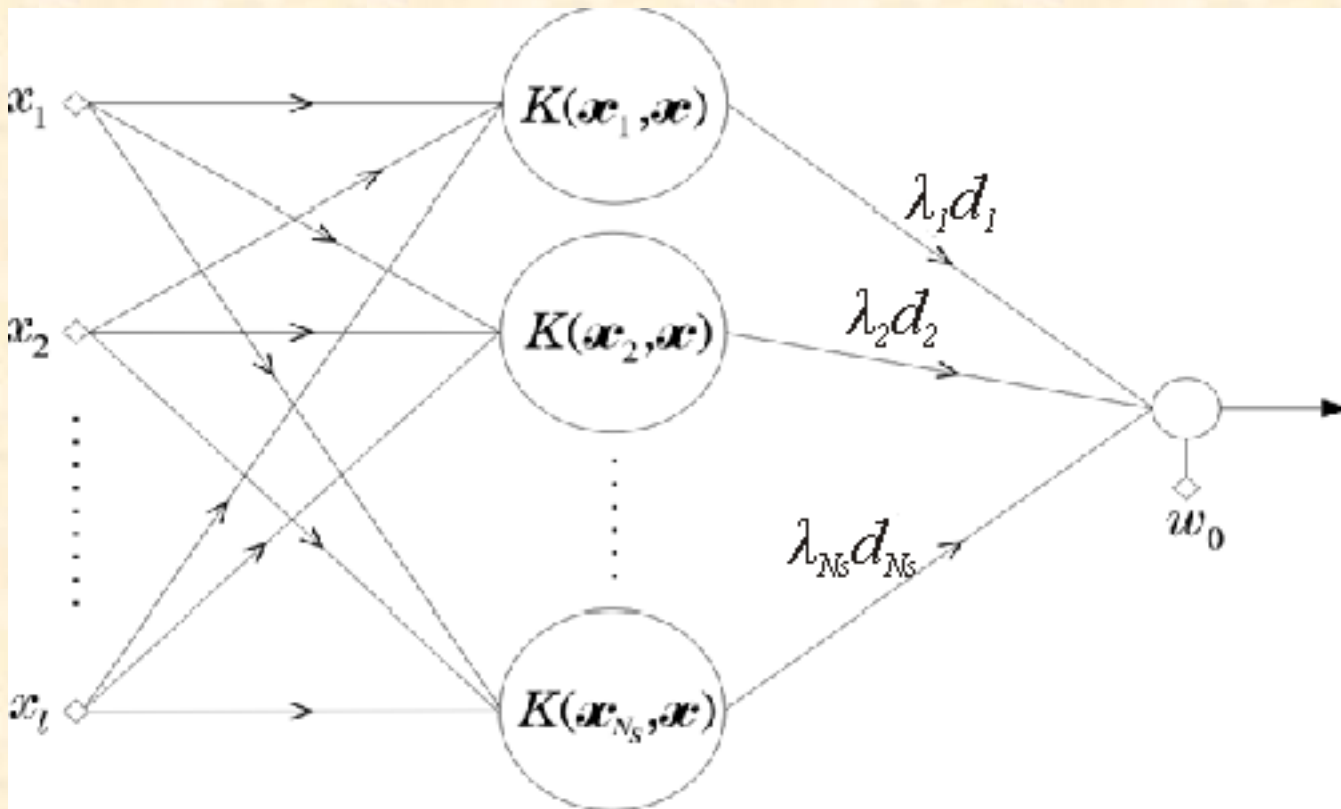
$$\begin{aligned} \mathbf{w}^T \mathbf{y} + w_0 &= \sum_{i=1}^{N_s} \lambda_i d_i \mathbf{y}_i^T \mathbf{y} + w_0 = \sum_{i=1}^{N_s} \lambda_i d_i f(\mathbf{x}_i)^T f(\mathbf{x}) + w_0 = \\ &= \sum_{i=1}^{N_s} \lambda_i d_i K(\mathbf{x}_i, \mathbf{x}) + w_0 \end{aligned}$$



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

## Η αρχιτεκτονική SVM

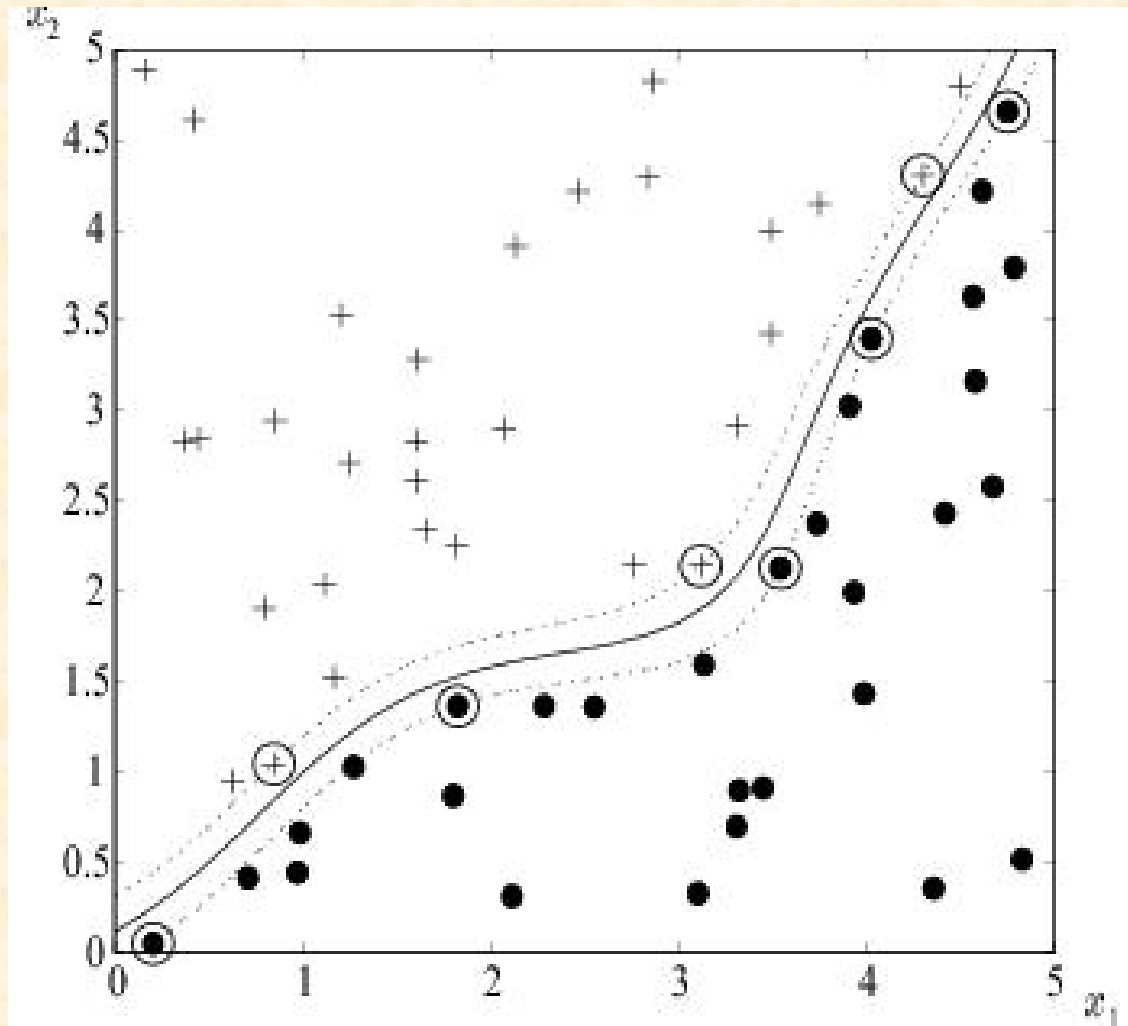


**ΣΗΜ.:** Οι μη γραμμικές SVM μπορεί να θεωρηθούν ως γενικευμένοι γραμμικοί ταξινομητές.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Μηχανές διανυσματικής στήριξης (SVM): Η μη γραμμική περίπτωση

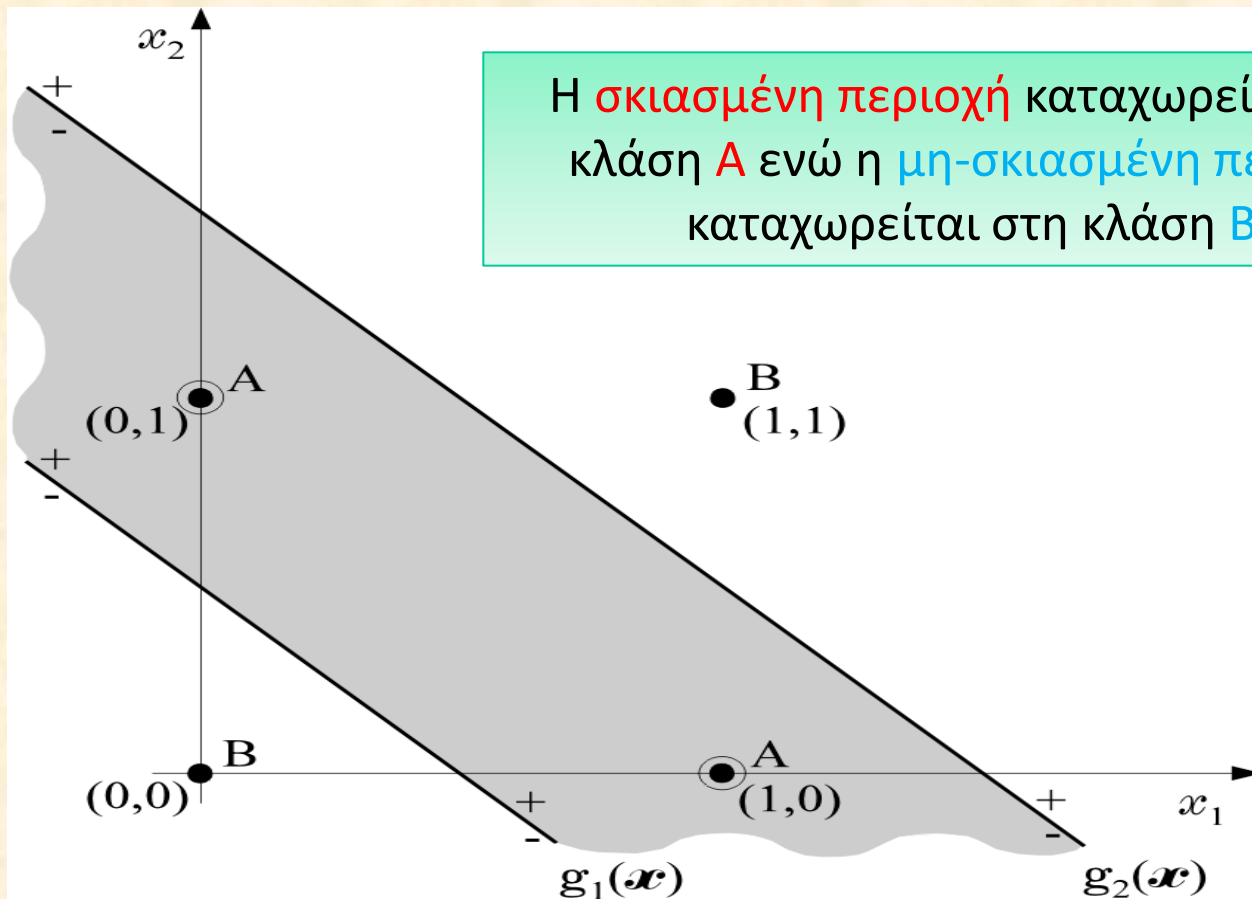
Παράδειγμα



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Η δομή δικτύου Perceptron δύο επιπέδων

Για το πρόβλημα **XOR**, σχεδίασε **δύο**, αντί για μία γραμμές.



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Η δομή δικτύου Perceptron δύο επιπέδων

Έστω ότι το  $\mathbf{x}=[x_1, x_2]^T$  απεικονίζεται στο διάνυσμα  $\mathbf{y}=[y_1, y_2]^T$ , όπου το  $y_i$  ισούται με 1(0), αν το  $\mathbf{x}$  ανήκει στη θετική (αρνητική) πλευρά της  $i$ -στής γραμμής,  $i=1,2$ . Τότε

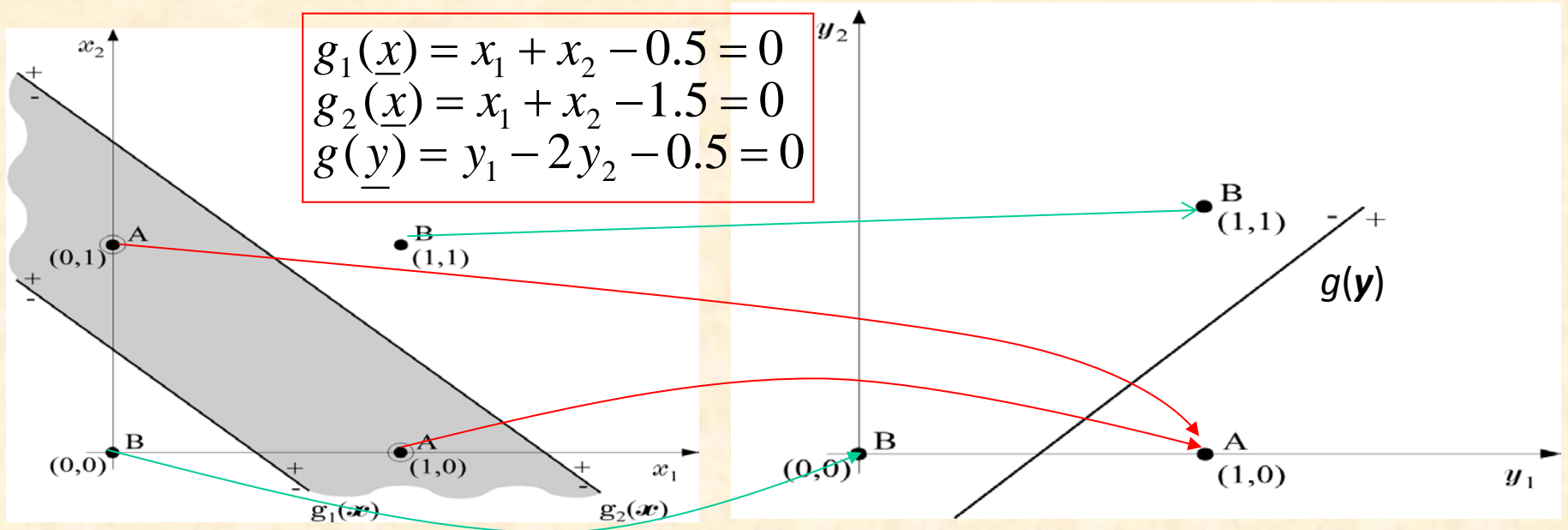
Απεικόνιση				Διαχωρισμός
$x_1$	$x_2$	$y_1$	$y_2$	
0	0	0(-)	0(-)	B(0)
0	1	1(+)	0(-)	A(1)
1	0	1(+)	0(-)	A(1)
1	1	1(+)	1(+)	B(0)

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## Η δομή δικτύου Perceptron δύο επιπέδων

Έστω ότι το  $\underline{x}=[x_1, x_2]^T$  απεικονίζεται στο διάνυσμα  $\underline{y}=[y_1, y_2]^T$ , όπου το  $y_i$  ισούται με  $1(0)$ , αν το  $\underline{x}$  ανήκει στη **θετική** (**αρνητική**) πλευρά της  $i$ -στής γραμμής,  $i=1,2$ . Τότε,

- κάθε σημείο της **σκιασμένης περιοχής** απεικονίζεται στο σημείο **(1,0)** στο νέο χώρο
- κάθε σημείο της **μη σκιασμένης περιοχής** απεικονίζεται είτε στο **(0,0)** είτε στο **(1,1)**.



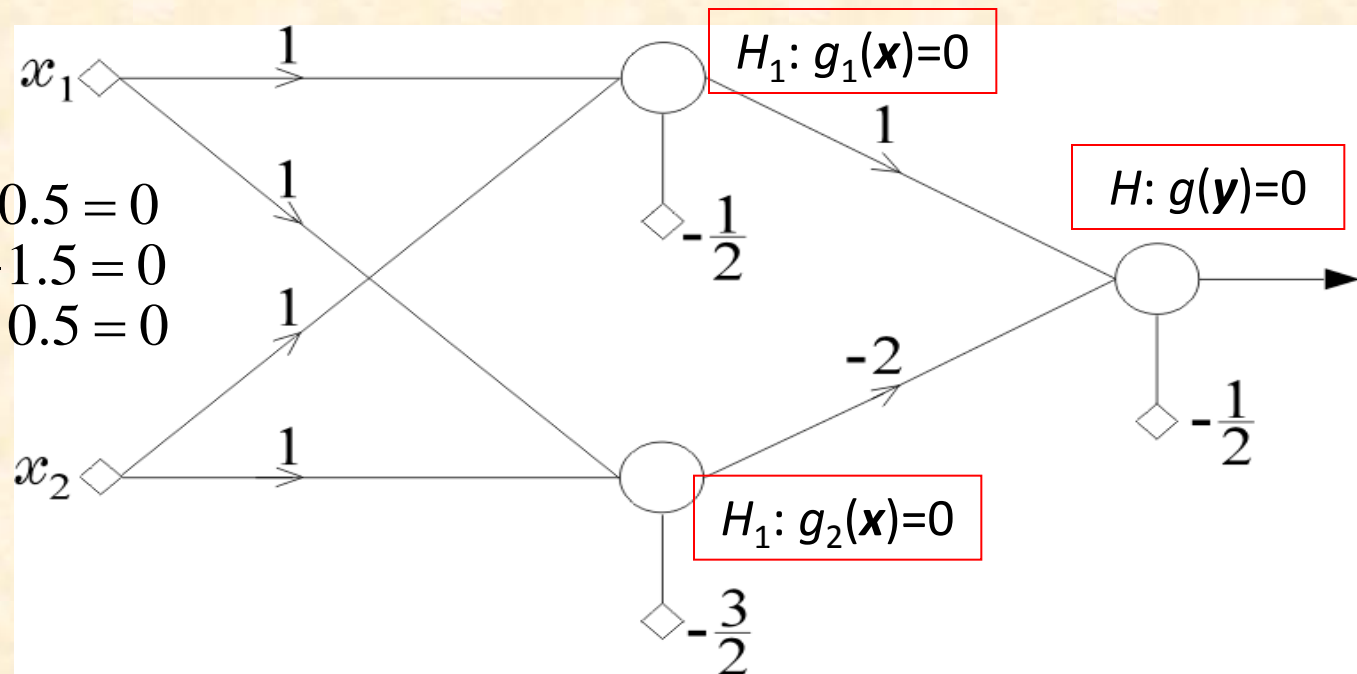
**ΣΗΜ.:** Το πρόβλημα γίνεται **γραμμικώς διαχωρίσιμο** στο **μετασχηματισμένο χώρο**.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## Η δομή δικτύου Perceptron δύο επιπέδων - Υλοποίηση

- Τοποθέτησε δύο νευρώνες (perceptrons) στο ίδιο (πρώτο) επίπεδο. Ο πρώτος (δεύτερος) υλοποιεί το διαχωρισμό που ορίζεται από τη γραμμή  $g_1(\mathbf{x})=0$  ( $g_2(\mathbf{x})=0$ ).
- Τοποθέτησε έναν επιπλέον νευρώνα στο δεύτερο επίπεδο, που παίρνει σαν είσοδο τις εξόδους των προηγούμενων δύο νευρώνων και υλοποιεί την ταξινόμηση στο μετασχηματισμένο χώρο.

$$H_1 : g_1(\underline{x}) = x_1 + x_2 - 0.5 = 0$$
$$H_2 : g_2(\underline{x}) = x_1 + x_2 - 1.5 = 0$$
$$H : g(\underline{y}) = y_1 - 2y_2 - 0.5 = 0$$

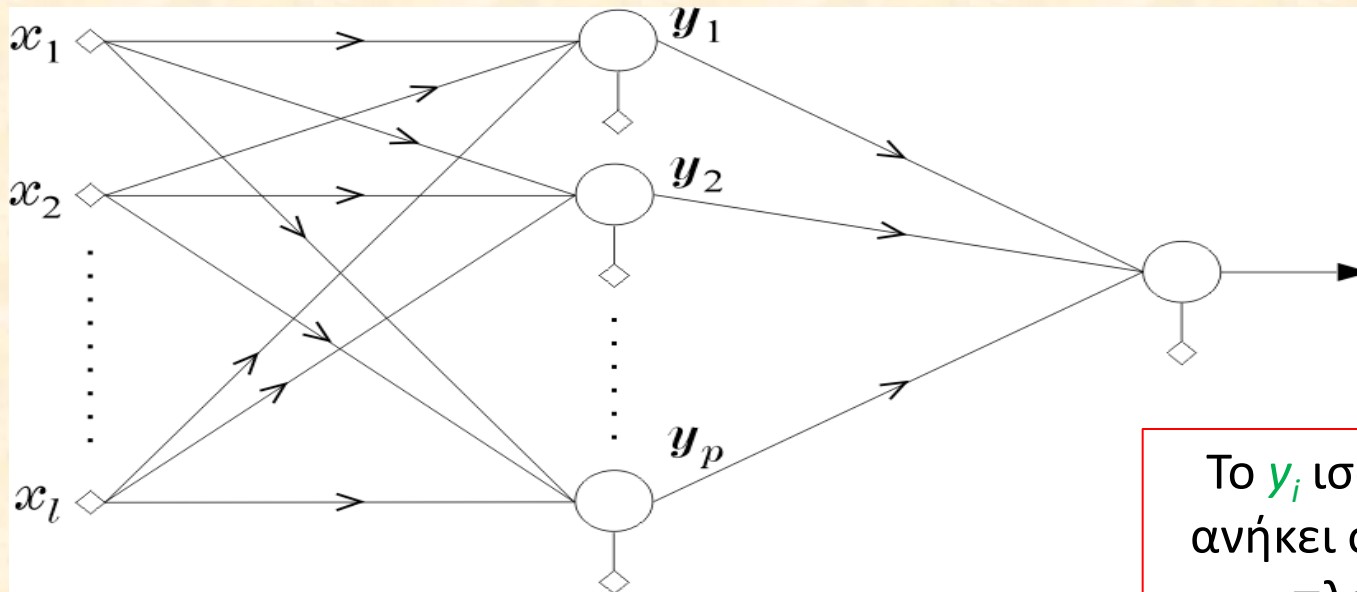


# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Δυνατότητες ταξινόμησης των δικτύων perceptrons δύο επιπέδων

- Στην περίπτωση το προβλήματος **XOR**, οι νευρώνες του πρώτου επιπέδου απεικονίζουν τα διανύσματα εισόδου ( $\mathbf{x}$ ) στις κορυφές του κύβου με πλευρά ίση με 1, δηλ., στις (0, 0), (0, 1), (1, 0), (1, 1).
- Στη γενικότερη περίπτωση όπου  $\mathbf{x} \in \mathbb{R}^l$  και χρησιμοποιούνται  $p$  (πρώτου επιπέδου) νευρώνες, η απεικόνιση γίνεται στις κορυφές του  $H_p$  υπερκύβου, δηλ.,

$$\mathbf{x} \rightarrow \mathbf{y} = [y_1, \dots, y_p]^T, \quad y_i \in \{0, 1\} \quad i = 1, 2, \dots, p$$



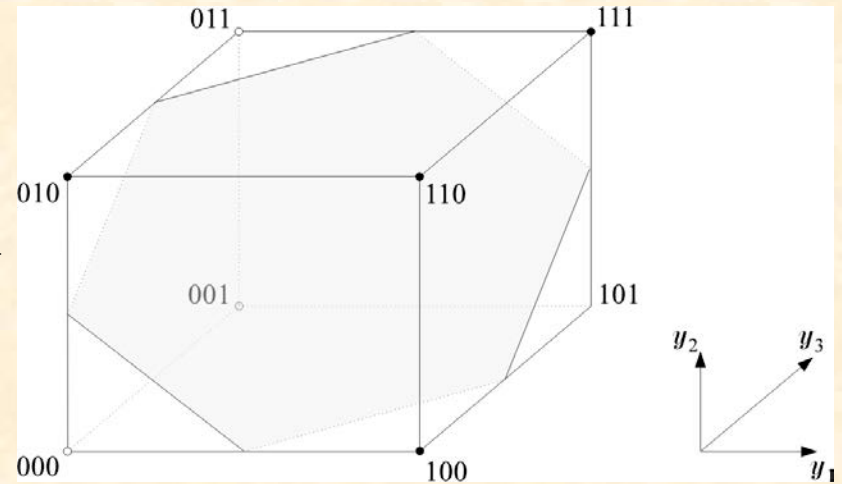
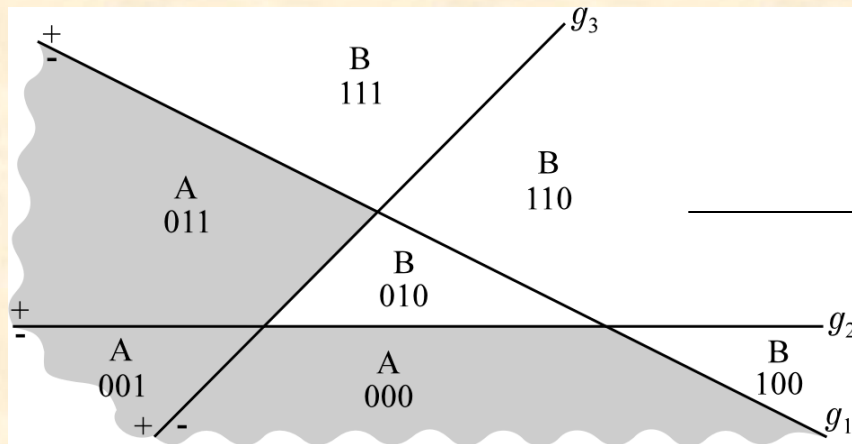
Το  $y_i$  ισούται με 1(0) αν το  $\mathbf{x}$  ανήκει στη θετική (αρνητική) πλευρά του  $g_i(\mathbf{x})=0$ .

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Δυνατότητες ταξινόμησης των δικτύων perceptrons δύο επιπέδων

➤ Τεμνόμενα **υπερεπίπεδα** στο χώρο που αντιστοιχούν στους  $p$  νευρώνες του πρώτου επιπέδου, ορίζουν **περιοχές** με την ακόλουθη ιδιότητα:

“**ΟΛΑ** τα σημεία μιας περιοχής έχουν την **ίδια σχετική θέση** ως προς τα  $p$  υπερεπίπεδα. Επομένως, **απεικονίζονται στην ίδια κορυφή του  $H_p$  υπερκύβου.**”



**Παράδειγμα:** Η κορυφή **001** αντιστοιχεί στην περιοχή που βρίσκεται στην (-) πλευρά της  $g_1(\underline{x})=0$ , στην (-) πλευρά της  $g_2(\underline{x})=0$ , στη (+) πλευρά της  $g_3(\underline{x})=0$ .

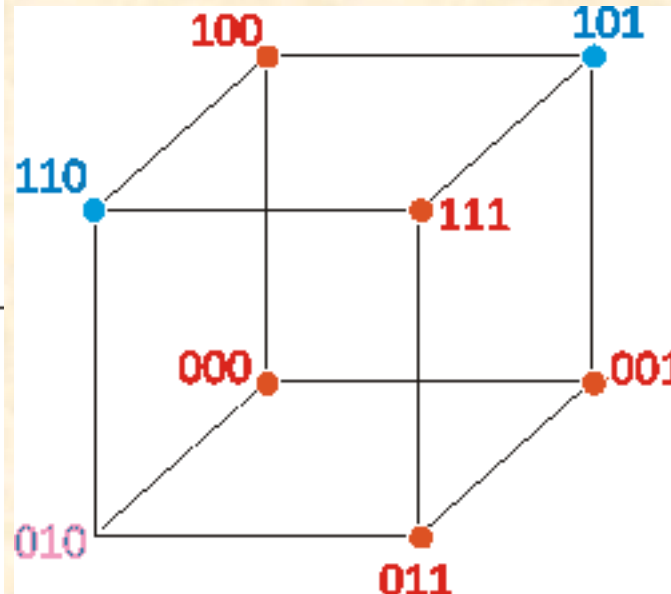
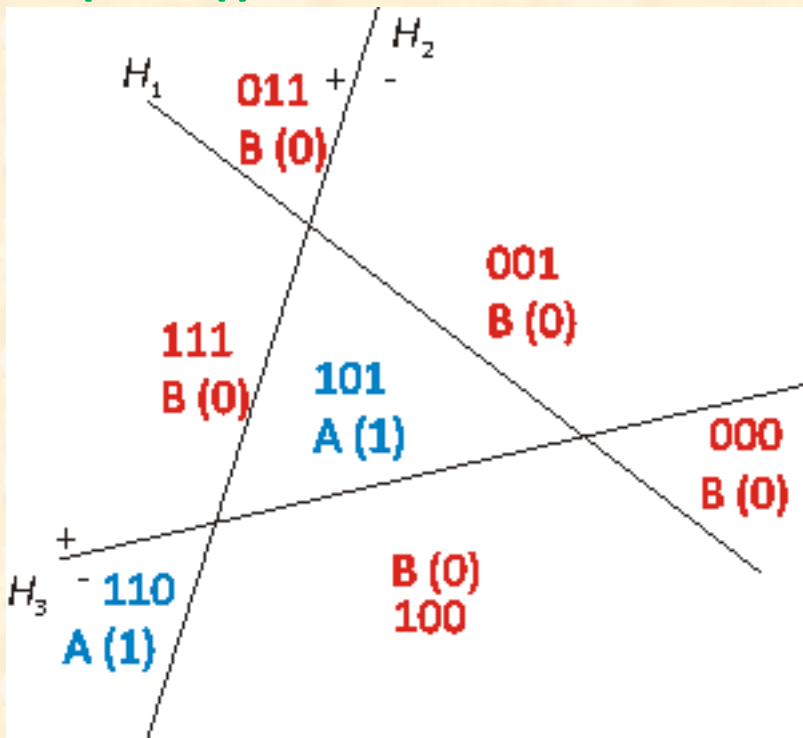


# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## Δυνατότητες ταξινόμησης των δικτύων perceptrons δύο επιπέδων

- Ο νευρώνας εξόδου αντιστοιχεί σε ένα υπερεπίπεδο στο μετασχηματισμένο χώρο το οποίο χωρίζει κάποιες κορυφές του  $H_p$  υπερκύβου από τις υπόλοιπες. Συνεπώς, ένα perceptron δύο επιπέδων έχει τη δυνατότητα να διαχωρίσει **κλάσεις που αποτελούνται από ενώσεις πολυεδρικών περιοχών**.
- Αλλά **ΌΧΙ ΟΠΟΙΕΣΔΗΠΟΤΕ** ενώσεις. Αυτό εξαρτάται από τη σχετική θέση των αντίστοιχων κορυφών του υπερκύβου.

### Παράδειγμα:

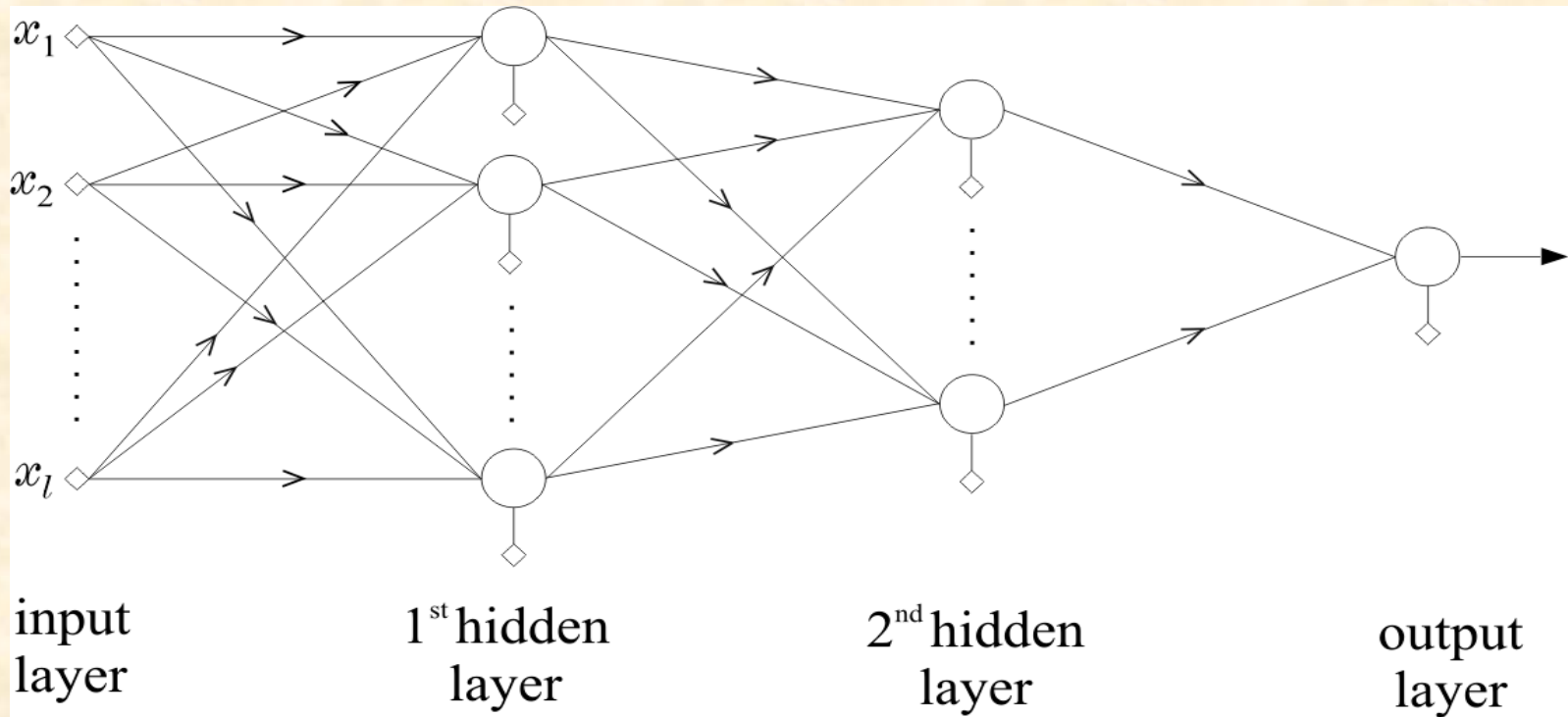


Δεν υπάρχει υπερεπίπεδο που να διαχωρίζει τις **μπλε** κορυφές από τις υπόλοιπες

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## Δίκτυα Perceptrons τριών επιπέδων

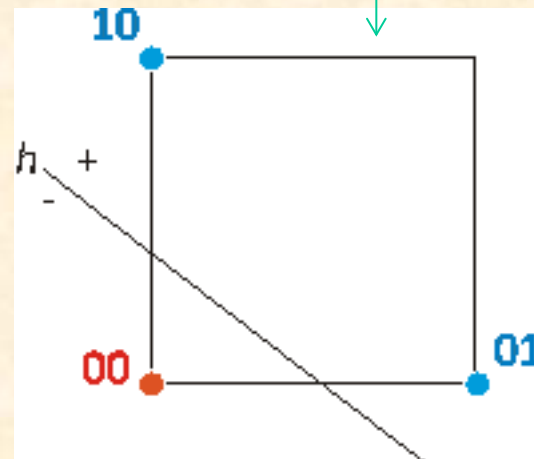
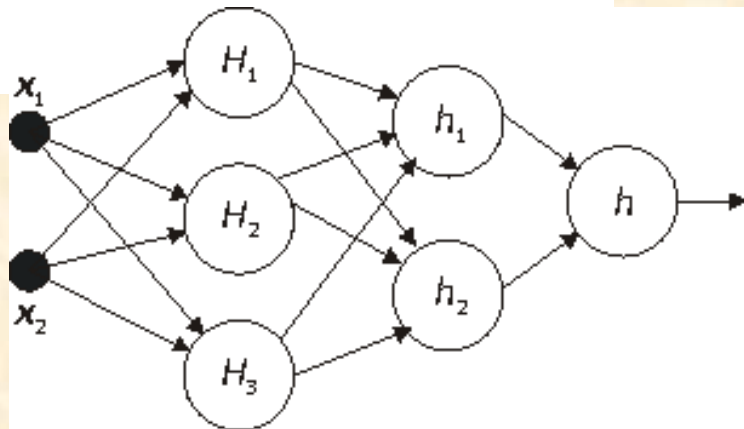
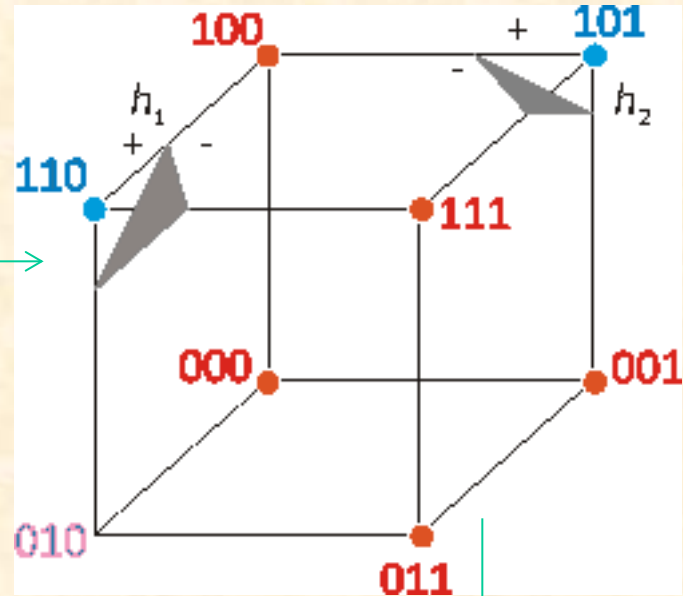
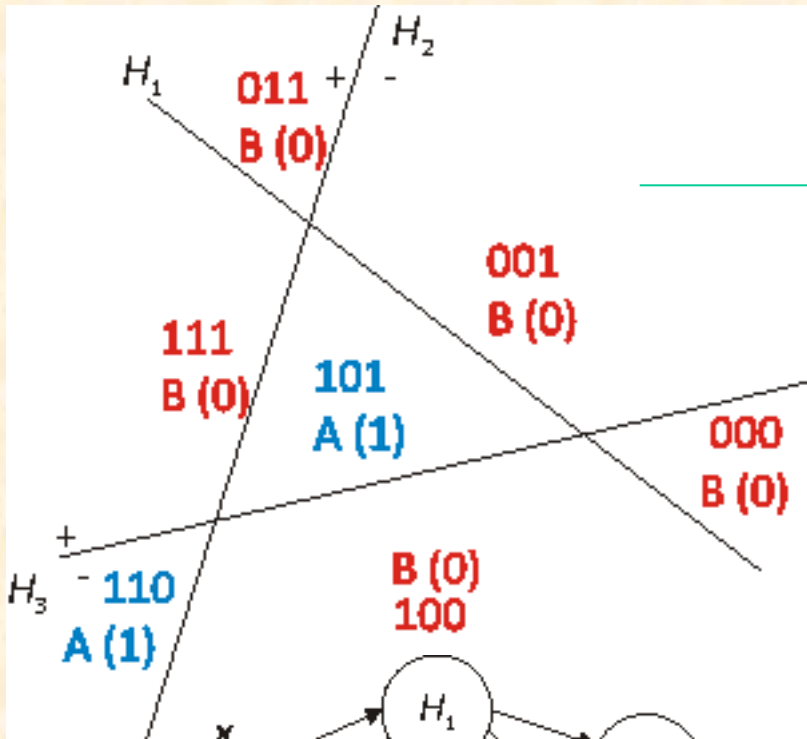
Η αρχιτεκτονική



Τέτοιες δομές είναι ικανές να διαχωρίσουν κλάσεις που ορίζονται από **οποιαδήποτε** ένωση πολυεδρικών περιοχών.

# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Δίκτυα Perceptrons τριών επιπέδων– παράδειγμα



# ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

## Δίκτυα Perceptrons τριών επιπέδων

Έστω  $A=R_1 \cup \dots \cup R_k$  (**1**) και  $B=R_1' \cup \dots \cup R_m'$  (**0**) ( $A \cup B=R'$ ), όπου οι περιοχές  $R_i$  και  $R_j'$  ορίζονται από την τομή  $p$  υπερεπιπέδων,  $H_1, \dots, H_p$ .

Πώς ένα **perceptron τριών επιπέδων** μπορεί να υλοποιήσει ΟΠΟΙΟΝΔΗΠΟΤΕ διαχωρισμό που ορίζεται από υπερεπίπεδα:

- Τοποθέτησε  $p$  **νευρώνες** στο **πρώτο επίπεδο**, καθένας από τους οποίους αντιστοιχεί σε ένα από τα υπερεπίπεδα  $H_1, \dots, H_p$ .
- Τοποθέτησε  $k$  **nodes** στο **δεύτερο επίπεδο**, καθένας από τους οποίους διαχωρίζει μια κορυφή του  $H_p$  υπερκύβου που αντιστοιχεί σε μια περιοχή  $R_i$ , από όλες τις υπόλοιπες κορυφές.
- Τοποθέτησε **έναν OR νευρώνα** στο  **τρίτο επίπεδο** (αυτός επιστρέφει **0** για την κορυφή  $00\dots 0$  του  $H_k$  υπερκύβου και **1** για όλες τις υπόλοιπες).

**ΣΗΜ.:** Το **πρώτο επίπεδο** του δικτύου ορίζει τα **υπερεπίπεδα**, το **δεύτερο επίπεδο** ορίζει τις **περιοχές** και ο νευρώνας εξόδου ορίζει τις κλάσεις.