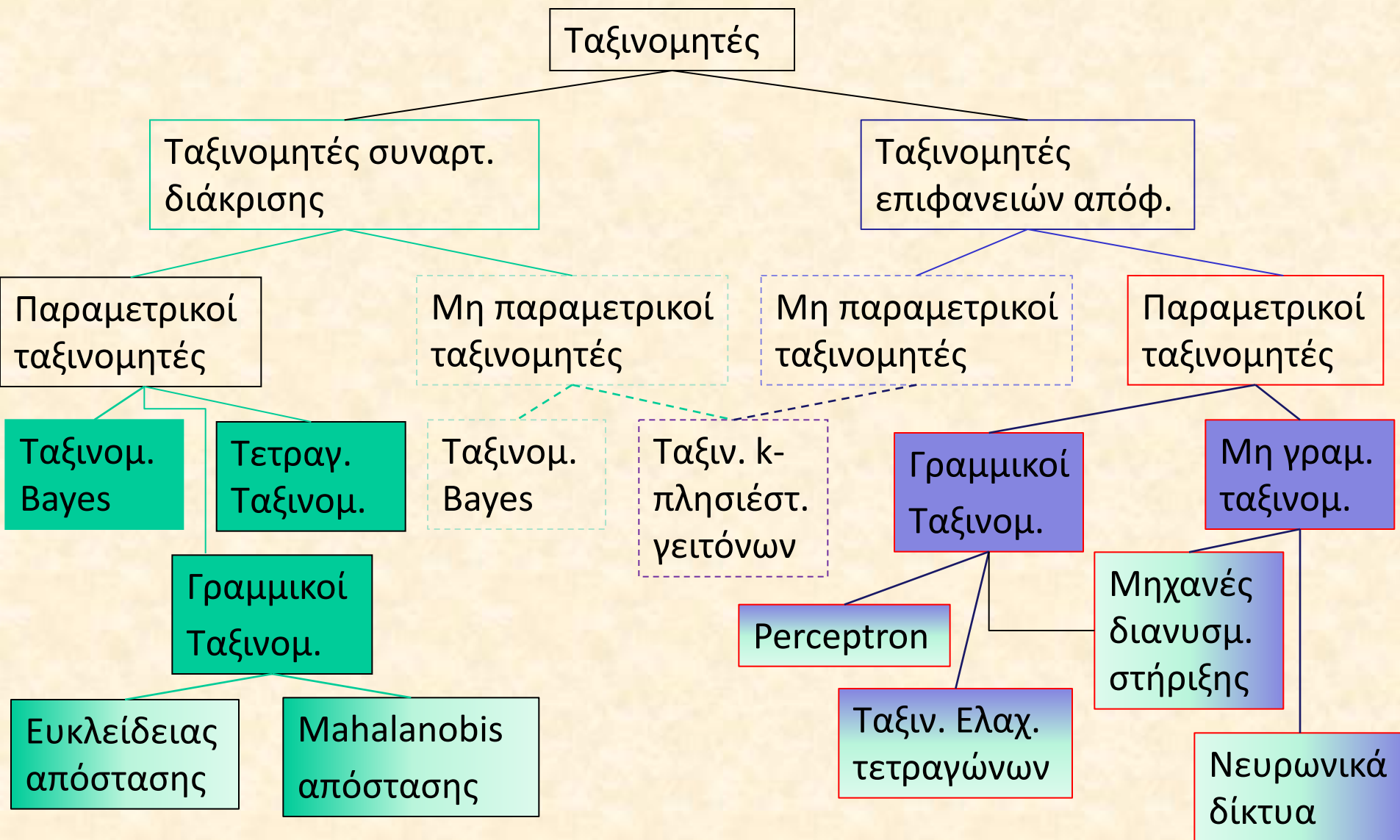


ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

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 ω_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. θ_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. μ_j, Σ_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. μ_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. μ_j, Σ_j με βάση το X_j
(ML): $X_j \rightarrow \hat{\mu}_j, \hat{\Sigma}_j$
- Εκτίμ. Σ 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 ω_j ,

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

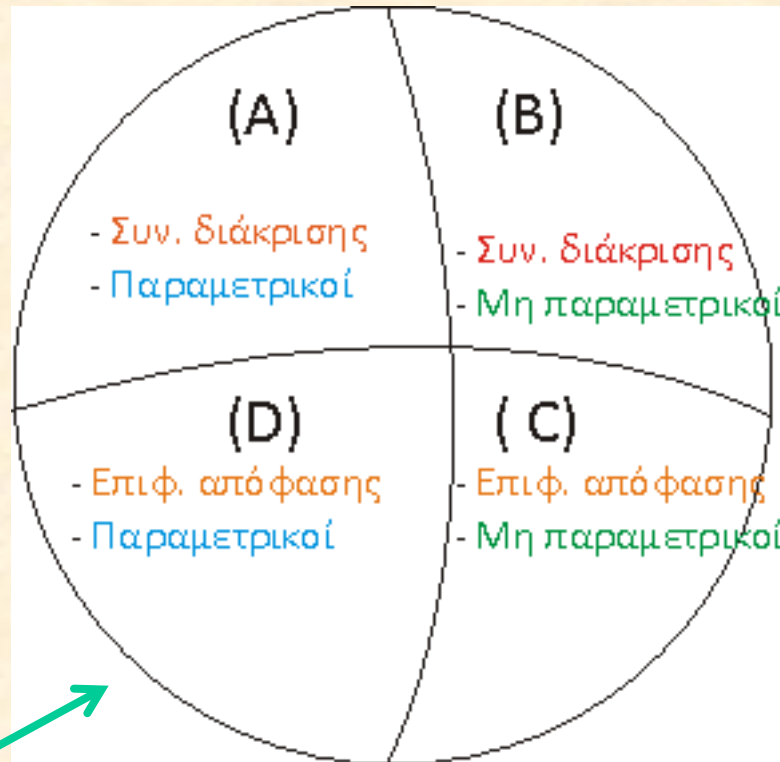
Μη παραμετρικοί ταξινομητές με βάση τις συναρτήσεις διάκρισης

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 ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Η παραπάνω ανάλυση ήταν χρήσιμη προκειμένου να κατανοήσουμε τη **λογική** των δικτύων perceptrons πολλών επιπέδων καθώς επίσης και τις **δυνατότητές** τους.

Ωστόσο, **στην πράξη** το μόνο που έχουμε στη διάθεσή μας είναι ένα σύνολο δεδομένων X και το πρόβλημά μας είναι το ακόλουθο:

Ορισμός προβλήματος:

1. **Υιοθέτησε/καθόρισε** δίκτυο perceptron πολλαπλών επιπέδων συγκεκριμένης δομής
2. **Εκτίμησε** το διάνυσμα παραμέτρων του w (το οποίο περιέχει όλες τις παραμέτρους των νευρώνων του δικτύου), ώστε να επιτευχθεί ο **“βέλτιστος” δυνατός διαχωρισμός των κλάσεων**, με βάση το X .

Δύο κύριες κατευθύνσεις:

- **Ανέπτυξε** μια δομή η οποία ταξινομεί **σωστά όλα** τα διανύσματα εκπαίδευσης.
- **Υιοθέτησε** μια σταθερή δομή και υπολόγισε τις τιμές των παραμέτρων της μέσω της **βελτιστοποίησης μιας κατάλληλα επιλεγμένης συνάρτησης κόστους**.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

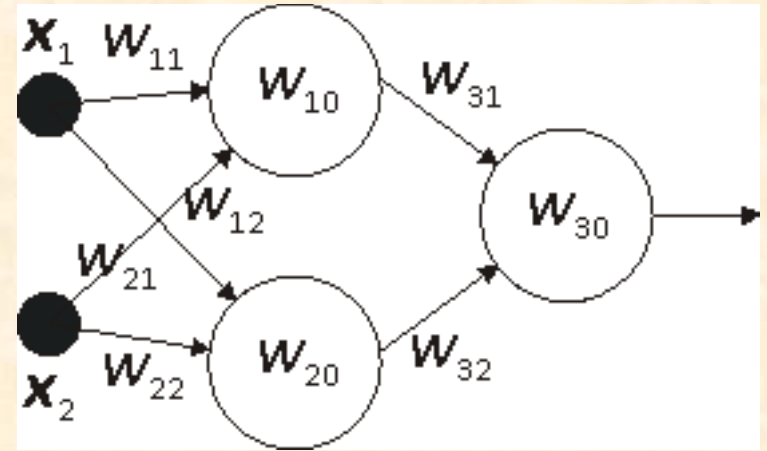
Θεωρείστε το ακόλουθο δίκτυο δύο επιπέδων με

-2 **νευρώνες εισόδου** (διάσταση χώρου εισόδου)

-2 νευρώνες **πρώτου επιπέδου**

-1 νευρώνας (εξόδου) **δευτέρου επιπέδου**.

Το δίκτυο αυτό περιγράφεται από τη συνάρτηση



$$\hat{y} = f(w_{31}f(w_{11}x_1 + w_{12}x_2 + w_{10}) + w_{32}f(w_{21}x_1 + w_{22}x_2 + w_{20}) + w_{30})$$
$$= f\left(\sum_{i=1}^2 w_{3i}f\left(\sum_{j=1}^2 w_{ij}x_j + w_{i0}\right) + w_{30}\right) = f\left(\sum_{i=1}^2 w_{3i}f\left(\sum_{j=1}^2 w_i^T x\right) + w_{30}\right)$$

όπου $\mathbf{w}_1 = [w_{11}, w_{12}, w_{10}]^T$, $\mathbf{w}_2 = [w_{21}, w_{22}, w_{20}]^T$, $\mathbf{w}_3 = [w_{31}, w_{32}, w_{30}]^T$, $\mathbf{x} = [x_1, x_2, 1]^T$ και $f(\cdot)$ η **συνάρτηση μοναδιαίου βήματος**.

Το διάνυσμα παραμέτρων \mathbf{w} του δικτύου είναι

$$\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \mathbf{w}_3^T] = [w_{11}, w_{12}, w_{10}, w_{21}, w_{22}, w_{20}, w_{31}, w_{32}, w_{30}]^T$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (Backpropagation (BP) Algorithm)

- Είναι μια αλγοριθμική διαδικασία στη λογική της **οξύτερης καθόδου (gradient descent)** που υπολογίζει **αναδρομικά** τα συναπτικά βάρη (παράμετροι) του δικτύου, έτσι ώστε να επιτυγχάνεται η **ελαχιστοποίηση** της επιλεγμένης **συνάρτησης κόστους**.
- Σε ένα μεγάλο αριθμό διαδικασιών βελτιστοποίησης, απαιτείται ο υπολογισμός **παραγώγων**. Έτσι, **ασυνεχείς συναρτήσεις ενεργοποίησης** δημιουργούν προβλήματα δηλ.,

$$\cancel{f(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}}$$

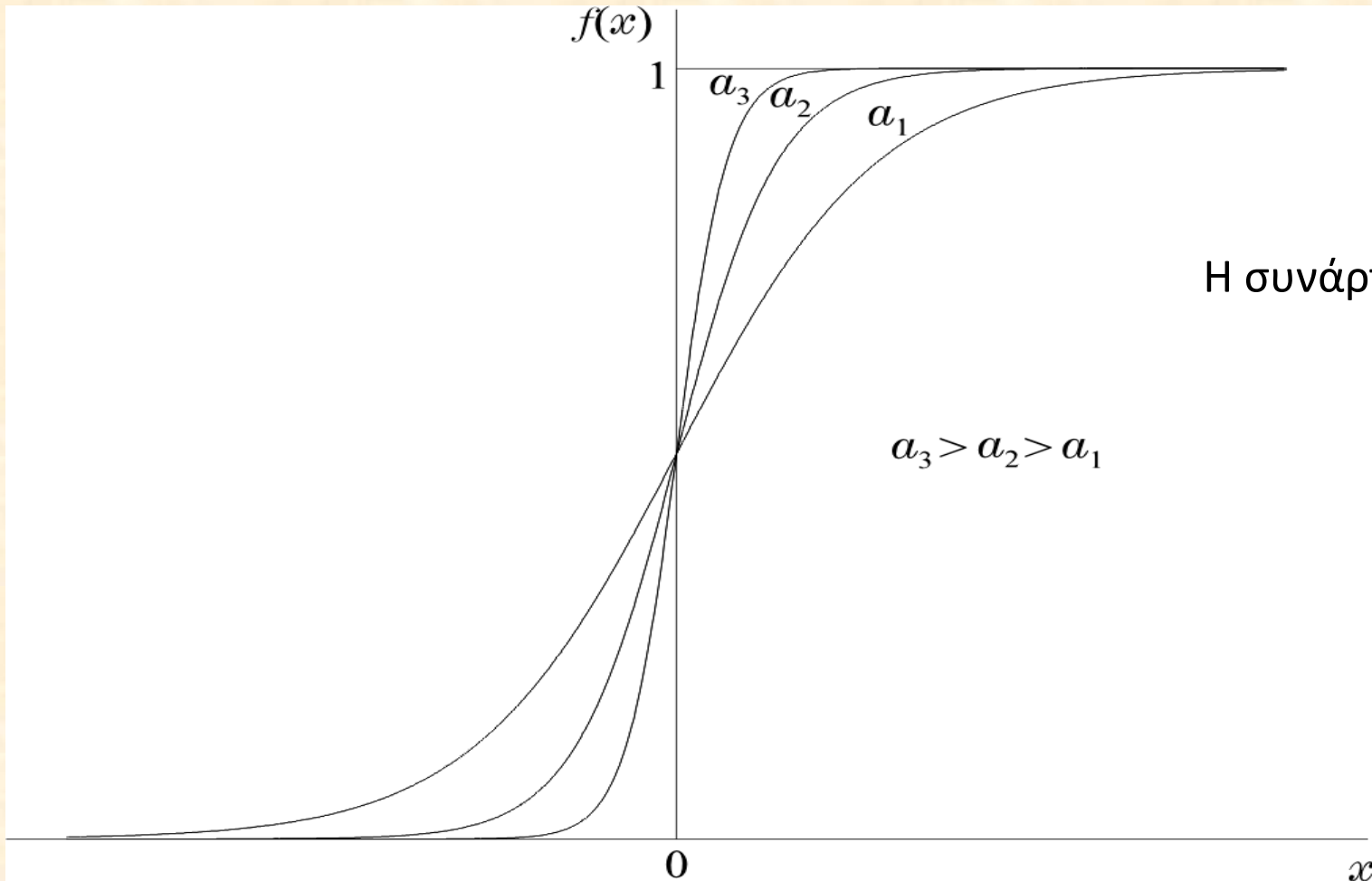
- Παρόλα αυτά, υπάρχουν δίοδοι διαφυγής!!! Η **logistic** συνάρτηση και η συνάρτηση **υπερβολικής εφαπτομένης** αποτελούν τέτοια παραδείγματα.

$$f(x) = \frac{1}{1 + \exp(-ax)}$$

$$f(x) = \frac{1 - \exp(-ax)}{1 + \exp(-ax)}$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

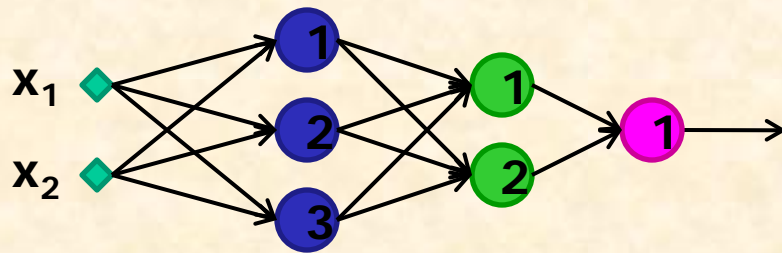
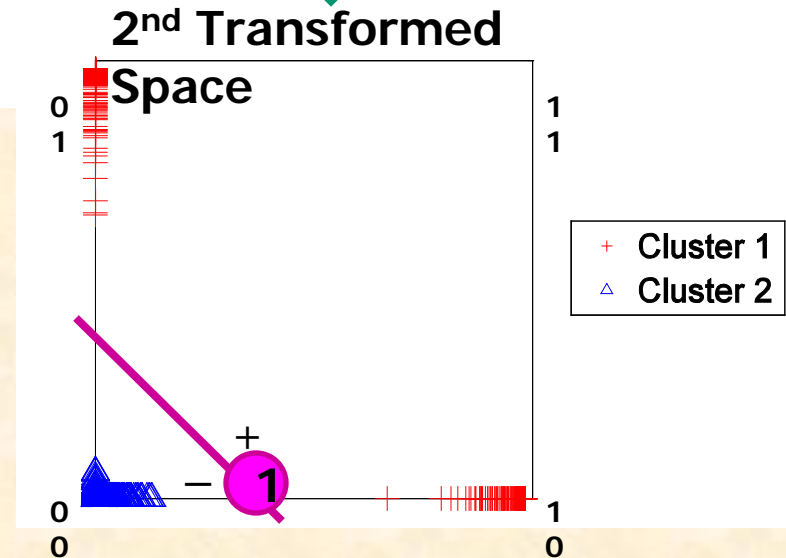
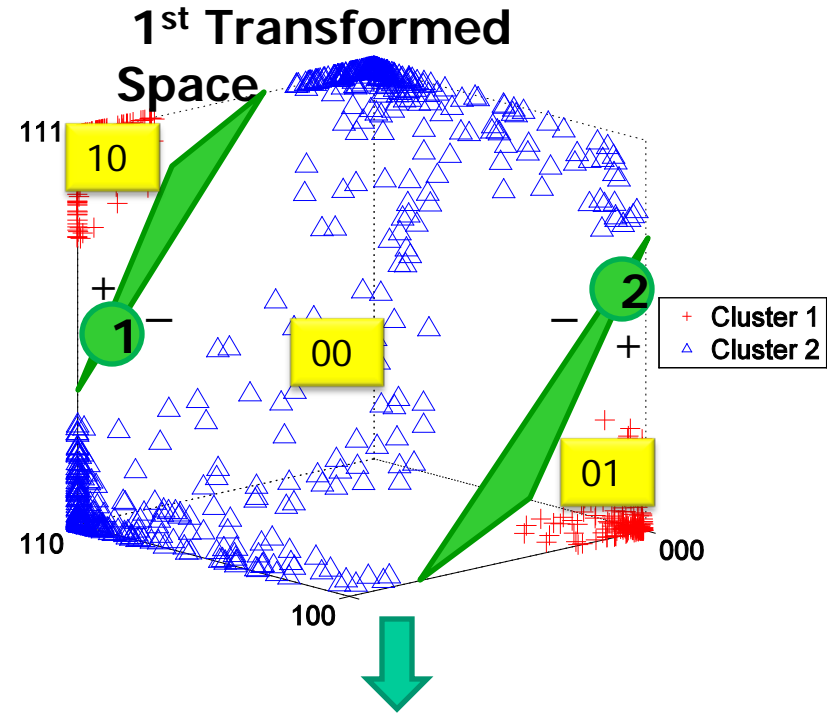
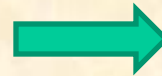
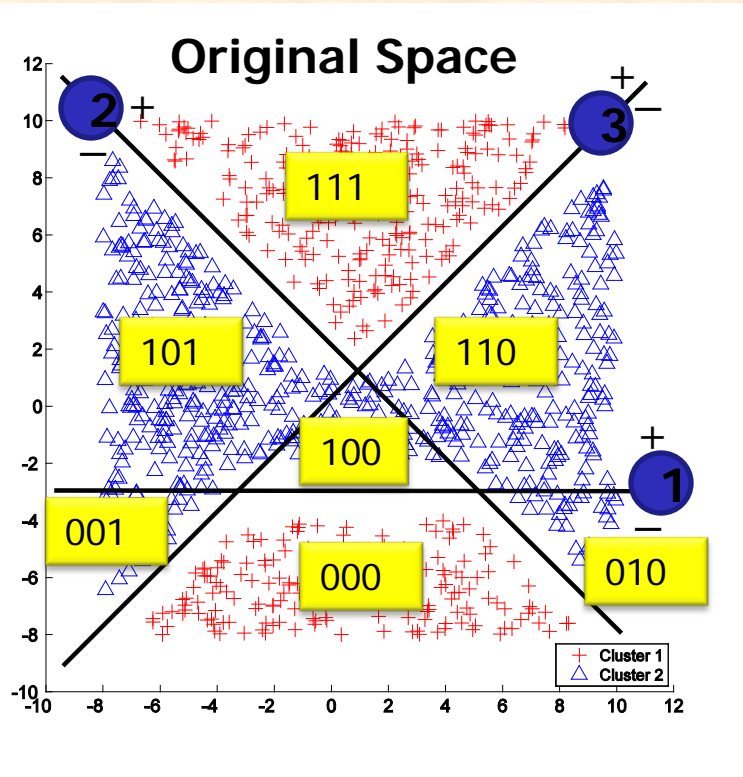
Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)



Η συνάρτηση **logistic**

$$a_3 > a_2 > a_1$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ



Sigmoid output functions

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)

➤ Τα βήματα:

- Υιοθέτησε μία κατάλληλη **συνάρτηση κόστους $J(\mathbf{w})$** , π.χ.,
Σφάλμα ελαχίστων τετραγώνων (Least Squares Error),
Σχετική εντροπία (Relative Entropy)

που ορίζεται λαμβάνοντας υπ' όψιν τις **επιθυμητές αποκρίσεις** και τις **πραγματικές αποκρίσεις του δικτύου για τα διαθέσιμα διανύσματα εκπαίδευσης**. Αυτό υπονοεί ότι από εδώ και στο εξής αποδεχόμαστε την ύπαρξη λαθών. Προσπαθούμε μόνο να τα ελαχιστοποιήσουμε, χρησιμοποιώντας συγκεκριμένα κριτήρια.

- Υιοθέτησε έναν **αλγόριθμο** για τη **βελτιστοποίηση** της **συνάρτησης κόστους ως προς τα συναπτικά βάρη** π.χ.,
 - Αλγόριθμος οξύτερης καθόδου (Gradient descent)
 - Αλγόριθμος Newton
 - Αλγόριθμος συζυγών διευθύνσεων (Conjugate gradient)

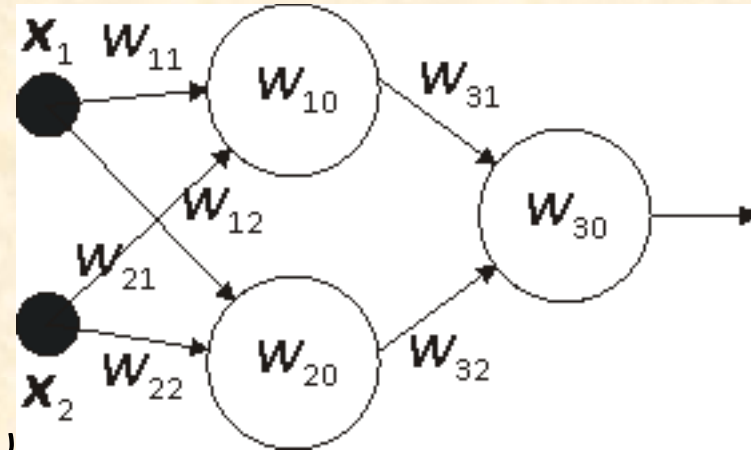
ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)

Για την περίπτωση της **οξύτερης καθόδου** έχουμε

$$w_i(t+1) = w_i(t) + \Delta w_i(t)$$

$$\Delta w_i(t) = -\mu \frac{\partial J}{\partial w_i} \Big|_{w_i=w_i(t)}$$



Για τα διανύσματα όλων των νευρώνων του δικτύου.

Παράδειγμα: Η **συν. κόστους ελαχίστων τετραγώνων** για το παραπάνω δίκτυο είναι

$$J(w) = \sum_{n=1}^N (d_n - \hat{y}_n)^2 = \sum_{n=1}^N (d_n - f(\sum_{i=1}^2 w_{3i} f(w_i^T x_n) + w_{30}))^2 =$$

$$\sum_{n=1}^N (d_n - f(\sum_{i=1}^2 w_{3i} f(\sum_{j=1}^2 w_{ij} x_{nj} + w_{i0}) + w_{30}))^2$$

Batch mode

$$J(w) = (d_n - \hat{y}_n)^2 = (d_n - f(\sum_{i=1}^2 w_{3i} f(\sum_{j=1}^2 w_{ij}^T x_n) + w_{30}))^2$$

Pattern (online) mode

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)

- **Εμπροσθοδρομικοί υπολογισμοί (Forward Computations):** Για δεδομένο διάνυσμα δεδομένων, \mathbf{x}_n , χρησιμοποίησε τις τρέχουσες εκτιμήσεις των παραμέτρων και υπολόγισε την **έξοδο του δικτύου**, έστω \hat{y}_n , η οποία εξαρτάται από τις τρέχουσες εκτιμήσεις.
- **Οπισθοδρομικοί υπολογισμοί (Backward Computations):** Χρησιμοποιώντας την επιθυμητή απόκριση, y_n , και την εκτιμηθείσα, \hat{y}_n , υπολόγισε τους αντίστοιχους **όρους βαθμίδας (gradients)** της **συνάρτησης κόστους** ως προς όλες τις παραμέτρους. Για το σκοπό αυτό, οι υπολογισμοί **διαδίδονται προς τα πίσω**:
 - Υπολόγισε τους όρους βαθμίδας των παραμέτρων των νευρώνων του **τελευταίου επιπέδου**, L .
 - Χρησιμοποιώντας τους παραπάνω όρους βαθμίδας και τον **κανόνα αλυσίδας** για την παραγωγή, υπολόγισε τους όρους βαθμίδας των παραμέτρων των νευρώνων στο $L - 1$ επίπεδο.
 - Η **προς τα πίσω διάδοση** συνεχίζεται έως ότου όλοι οι όροι βαθμίδας (έως αυτούς των νευρώνων του 1^{ου} επιπέδου) έχουν υπολογιστεί.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)

Η διαδικασία:

1. **Αρχικοποίησε** τα άγνωστα βάρη σε τυχαίες μικρές τιμές.
2. **Υπολόγισε** τους **όρους βαθμίδας (gradient terms)** **προς τα πίσω**, ξεκινώντας με τα βάρη των νευρώνων του τελευταίου επιπέδου και κινούμενος προς αυτά των νευρώνων του πρώτου
3. **Ενημέρωσε** τα βάρη
4. **Επανάλαβε** τα βήματα 2 και 3 μέχρις ότου ικανοποιηθεί ένα κριτήριο τερματισμού.

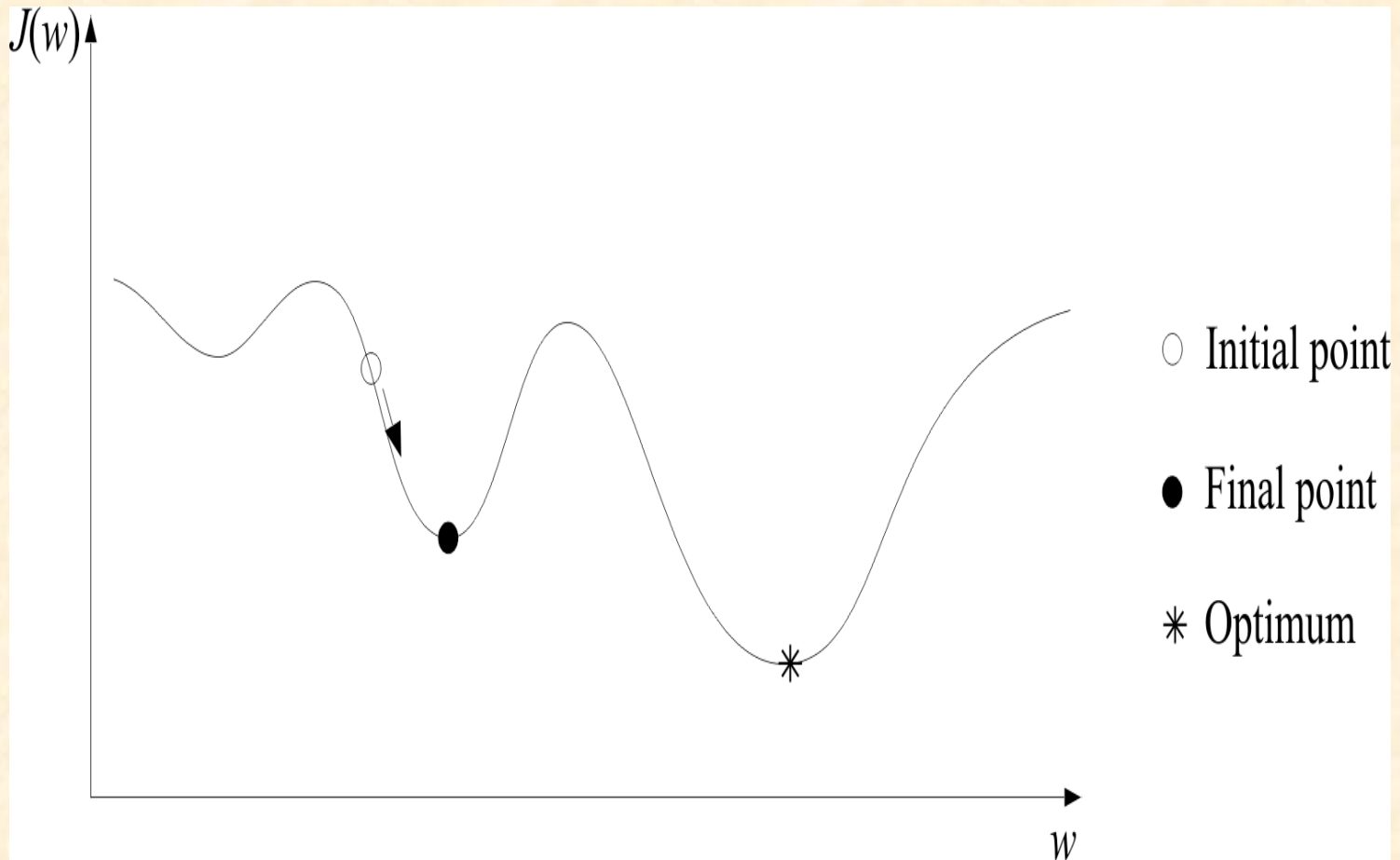
– **Δύο** κύριες φιλοσοφίες:

- **Επεξεργασία κατά συρροή (Batch mode)**: Οι όροι βαθμίδας του τελευταίου επιπέδου υπολογίζονται αφού πρώτα υποστούν επεξεργασία **ΌΛΑ** τα **διανύσματα εκπαίδευσης** δηλ., θεωρώντας το άθροισμα των λαθών για όλα τα διανύσματα.
- **Επεξεργασία κατά μόνας (Pattern mode)**: Οι όροι βαθμίδα υπολογίζονται για κάθε **νέο διάνυσμα εκπαίδευσης**. Έτσι, οι όροι αυτοί βασίζονται σε διαδοχικά μεμονωμένα λάθη.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP)

Ένα σημαντικό πρόβλημα: Ο αλγόριθμος μπορεί να συγκλίνει σε ένα **τοπικό ελάχιστο**.



ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP) – Επιλογή συνάρτησης κόστους

Παραδείγματα:

- Η συνάρτηση ελαχίστων τετραγώνων (Least Squares)

$$J = \sum_{i=1}^N E(i)$$
$$E(i) = \sum_{m=1}^k e_m^2(i) = \sum_{m=1}^k (y_m(i) - \hat{y}_m(i))^2$$
$$i = 1, 2, \dots, N$$

$y_m(i) \rightarrow$ Επιθυμητή απόκριση του m -στου νευρώνα εξόδου (1 or 0) για το $x(i)$

$\hat{y}_m(i) \rightarrow$ Πραγματική απόκριση του m -στου νευρώνα εξόδου, στο διάστημα $[0, 1]$, για είσοδο $x(i)$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP) – Επιλογή συνάρτησης κόστους

- Η συνάρτηση cross-entropy

$$J = \sum_{i=1}^N E(i)$$

$$E(i) = -\sum_{m=1}^k \{y_m(i) \ln \hat{y}_m(i) + (1 - y_m(i)) \ln(1 - \hat{y}_m(i))\}$$

Αυτή προϋποθέτει την ερμηνεία των y and \hat{y} ως **πιθανότητες**.

- Σφάλμα ταξινόμησης (Classification error rate).

Οι περισσότερες από τις σχετικές τεχνικές χρησιμοποιούν μία εξομαλυσμένη έκδοση του σφάλματος ταξινόμησης και όλες μαζί αποτελούν την κατηγορία αλγορίθμων **διακριτικής μάθησης** (discriminative learning).

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος οπισθοδρομικής διάδοσης (BP) – Επιλογή συνάρτησης κόστους

➤ **Σχόλιο 1:** Ένα κοινό χαρακτηριστικό των παραπάνω συναρτήσεων είναι ο κίνδυνος σύγκλισης σε κάποιο τοπικό ελάχιστο. Οι **“καλώς ορισμένες”** (**“Well formed”**) **συναρτήσεις κόστους** εγγυώνται σύγκλιση σε μία **“καλή”** λύση, δηλαδή σε μία λύση που να ταξινομεί σωστά **ΟΛΑ** τα δεδομένα εκπαίδευσης, υπό την προϋπόθεση ότι υπάρχει μία τέτοια λύση. Η **cross-entropy** συνάρτηση κόστους **είναι καλώς ορισμένη**. Η συνάρτηση ελαχίστων τετραγώνων **δεν είναι**.

➤ **Σχόλιο 2:** Αμφότερες οι συναρτήσεις κόστους **ελαχίστων τετραγώνων** και **cross entropy** οδηγούν σε τιμές εξόδου $\hat{y}_m(i)$ που προσεγγίζουν **κατά βέλτιστο** τρόπο τις εκ των υστέρων πιθανότητες για κάθε κλάση (Optimally class a-posteriori probabilities)!!!

$$\hat{y}_m(i) \cong P(\omega_m | \underline{x}(i))$$

Δηλ., την πιθανότητα της κλάσης ω_m δοθέντος του $x(i)$.

Πρόκειται για ένα πολύ ενδιαφέρον αποτέλεσμα. Δεν εξαρτάται από τις κατανομές των κλάσεων. Είναι ένα χαρακτηριστικό **ορισμένων** συναρτήσεων κόστους. Η ποιότητα της προσέγγισης εξαρτάται από το μοντέλο που υιοθετήθηκε. Επιπλέον, ισχύει **μόνο** στο **ολικό ελάχιστο**.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Παραλλαγές του αλγόριθμου οπισθοδρομικής διάδοσης (BP)

- Αλγόριθμος backpropagation με όρο ορμής (momentum term)

- Προστατεύει τον αλγόριθμο από περιπτώσεις ταλάντωσης γύρω από το ελάχιστο και, κατά συνέπεια, αργής σύγκλισης

- Εξισώσεις

$$w_i^r(t+1) = w_i^r(t) + \Delta w_i^r(t+1)$$

$$\Delta w_i^r(t+1) = a \Delta w_i^r(t) - \mu \frac{\partial J}{\partial w_i^r} \Big|_{w_i^r = w_i^r(t)}$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Παραλλαγές του αλγόριθμου οπισθοδρομικής διάδοσης (BP)

- Προσαρμοστικός (adaptive) αλγόριθμος backpropagation

- Επιταχύνει ή επιβραδύνει ανάλογα με το είδος της περιοχής του landscape της συνάρτησης κόστους που βρίσκεται η τρέχουσα εκτίμηση

- Εξισώσεις

$$\frac{J(t)}{J(t-1)} < 1, \quad \mu(t) = r_i \mu(t-1)$$

$$\frac{J(t)}{J(t-1)} > c, \quad \mu(t) = r_d \mu(t-1)$$

$$1 \leq \frac{J(t)}{J(t-1)} \leq c, \quad \mu(t) = \mu(t-1)$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Μέγεθος – αρχιτεκτονική δικτύου

Ένα σημαντικό ζήτημα: Πώς αποφασίζουμε για το μέγεθος και τη δομή του δικτύου;

Δύο κύριες κατευθύνσεις:

- **Σταθερή δομή δικτύου:** Υιοθέτησε ένα δίκτυο συγκεκριμένης σταθερής δομής και εφάρμοσε το αλγόριθμο BP. Αν η απόδοση του δικτύου μετά την εκπαίδευση δεν είναι ικανοποιητική, υιοθέτησε μια άλλη αρχιτεκτονική και επανέλαβε την εκπαίδευση.
- **Μεταβλητή δομή δικτύου:** Στην περίπτωση αυτή η δομή του δικτύου μεταβάλλεται κατά την διάρκεια της εκπαίδευσης του δικτύου. Υπάρχουν δύο βασικές φιλοσοφίες:
 - **Φιλοσοφία κλαδέματος (pruning):** Ξεκίνα με ένα δίκτυο μεγάλου μεγέθους και «κλάδεύε το» σταδιακά (αφαιρώντας συναπτικά βάρη και/ή νευρώνες) σύμφωνα με κάποια κριτήρια.
 - **Φιλοσοφία κατασκευής (Constructive philosophy):** Ξεκίνα με ένα δίκτυο μικρού μεγέθους (ανίκανο να λύσει το υπό μελέτη πρόβλημα) και σταδιακά πρόσθετε νευρώνες έως ότου το δίκτυο μάθει τη διαδικασία ταξινόμησης.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Μέγεθος – αρχιτεκτονική δικτύου

Παράδειγμα φιλοσοφίας κλαδέματος:

Μέθοδοι που βασίζονται στην κανονικοποίηση συνάρτησης
(function regularization)

$$J = \sum_{i=1}^N E(i) + aE_p(\underline{w})$$

Ο δεύτερος όρος ευνοεί μικρές τιμές για τα βάρη, π.χ.,

$$E_p(\underline{w}) = \sum_k h(w_k^2)$$

$$h(w_k^2) = \frac{w_k^2}{w_0^2 + w_k^2}$$

όπου $w_0 \cong 1$

Μετά από μερικά βήματα εκπαίδευσης, τα βάρη με μικρές τιμές απομακρύνονται.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος Backpropagation – θέματα ικανότητας γενίκευσης (**Generalization**)

Γιατί να μην ξεκινήσουμε με ένα δίκτυο μεγάλου μεγέθους και να αφήσουμε τον αλγόριθμο να αποφασίσει ποια βάρη είναι μικρά;;

Η προσέγγιση αυτή είναι απλοϊκή.

Παραβλέπει το γεγονός ότι οι ταξινομητές πρέπει να έχουν καλή δυνατότητα **γενίκευσης (generalization)**. Ένα μεγάλο δίκτυο μπορεί να δώσει μικρά σφάλματα για το σύνολο εκπαίδευσης, αφού μπορεί να μάθει τις συγκεκριμένες λεπτομέρειές του. Από την άλλη μεριά, δεν αναμένεται να παρουσιάζει καλή απόδοση για δεδομένα στα οποία δεν εκπαιδεύτηκε.

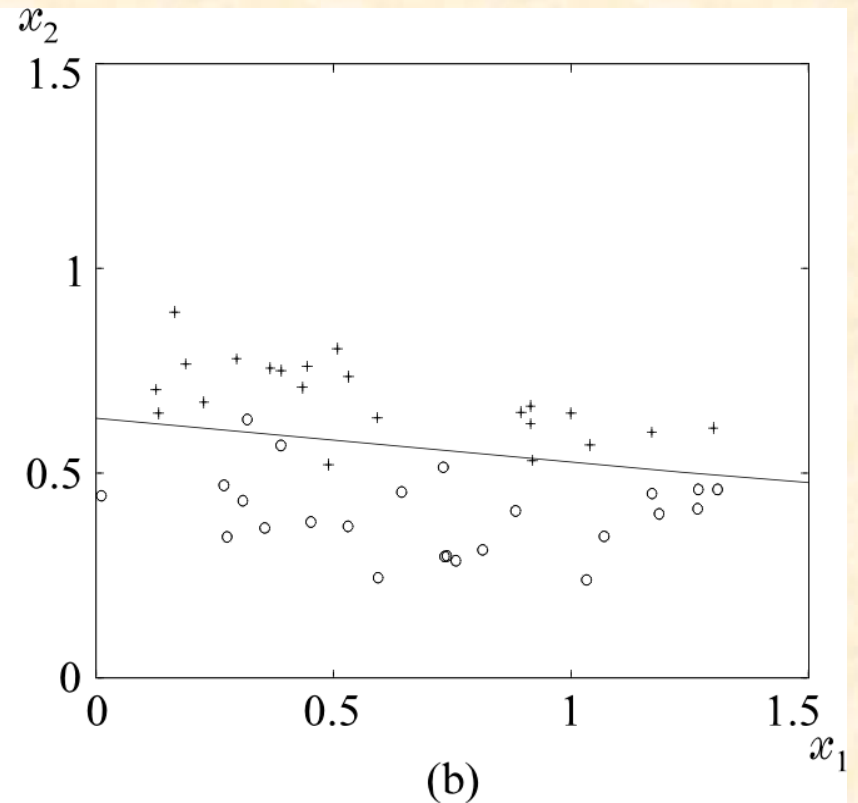
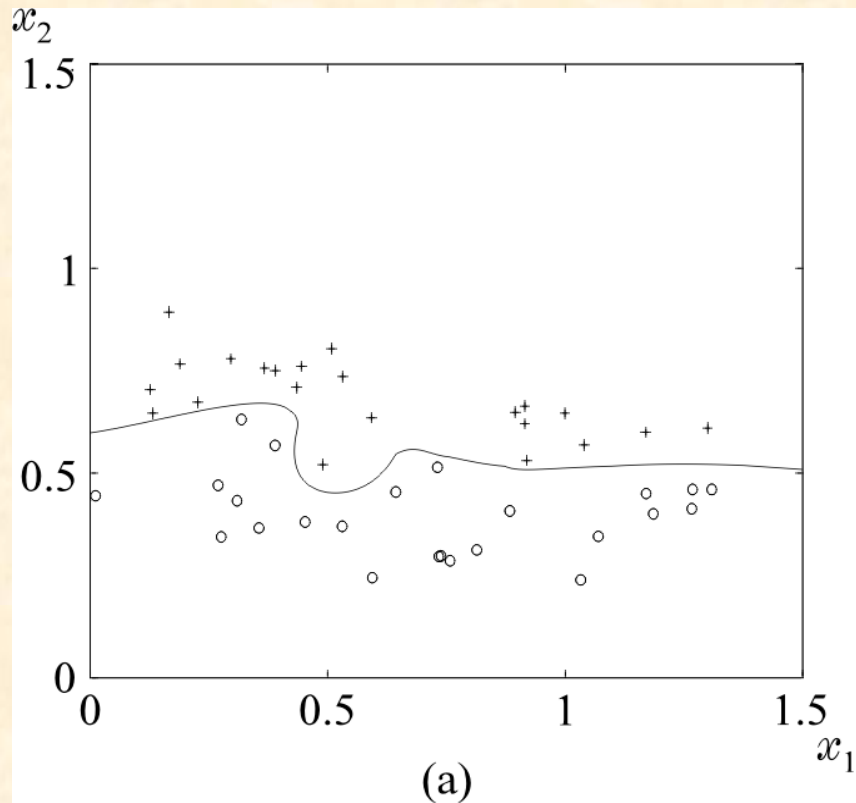
Γενικά, το μέγεθος του δικτύου πρέπει να είναι:

- Αρκούντως μεγάλο** για να μπορεί να μάθει τι κάνει **όμοια** τα δεδομένα της ίδιας κλάσης και τι κάνει **ανόμοια** τα δεδομένα διαφορετικών κλάσεων.
- Αρκούντως μικρό** για να μην μπορεί να μάθει τις διαφορές μεταξύ δεδομένων της ίδιας κλάσης. Το τελευταίο οδηγεί στο λεγόμενο **υπερ-ταίριασμα (overfitting)**.

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος Backpropagation – θέματα ικανότητας γενίκευσης

Παράδειγμα:



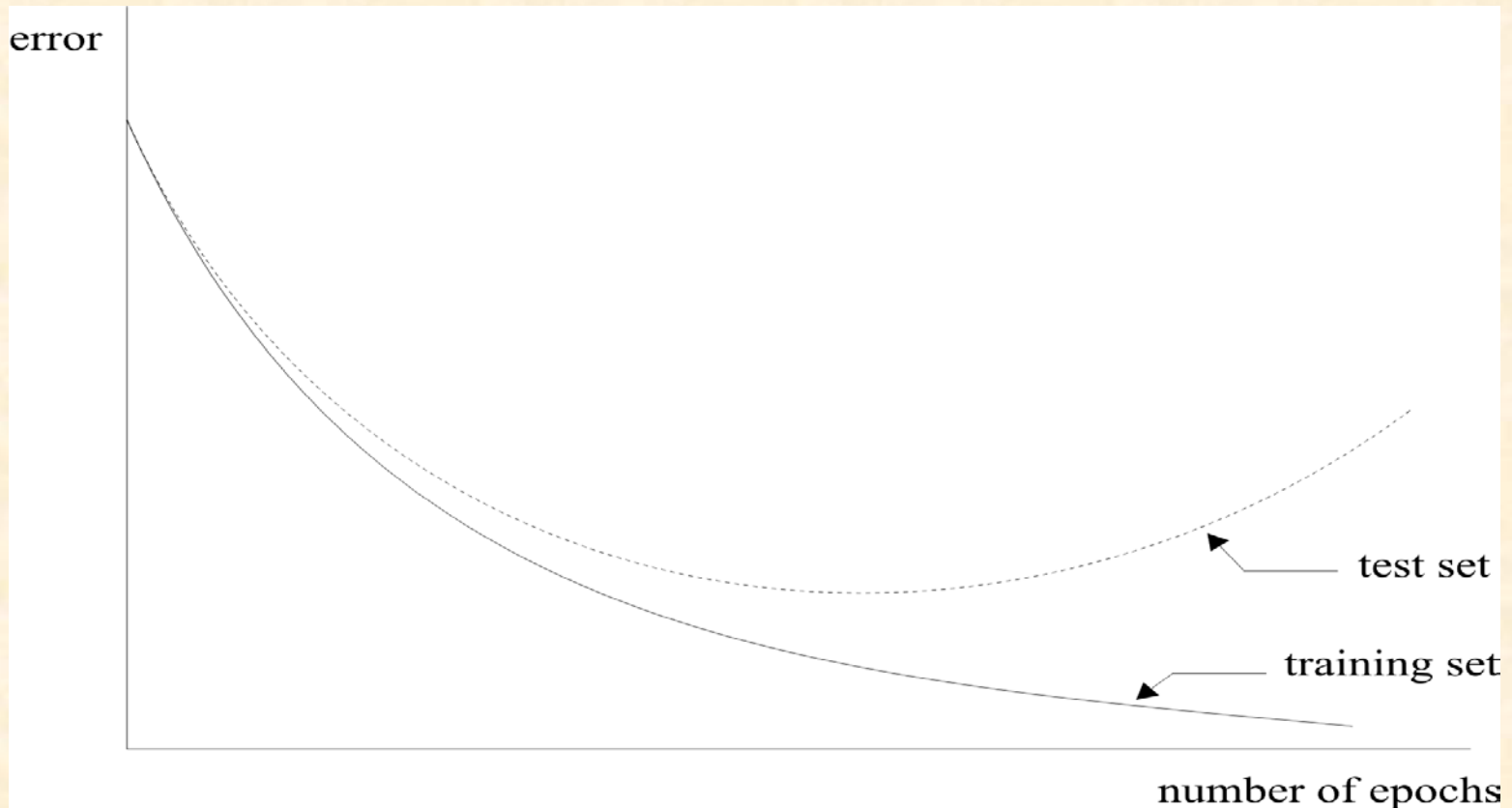
ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ο αλγόριθμος Backpropagation

Υπερεκπαίδευση (Overtraining) είναι μία άλλη όψη του ίδιου νομίσματος.

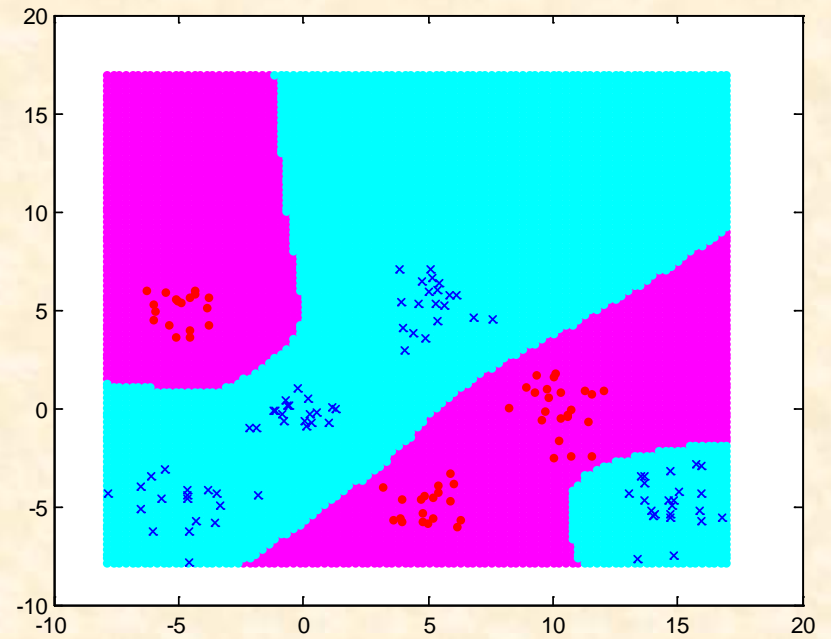
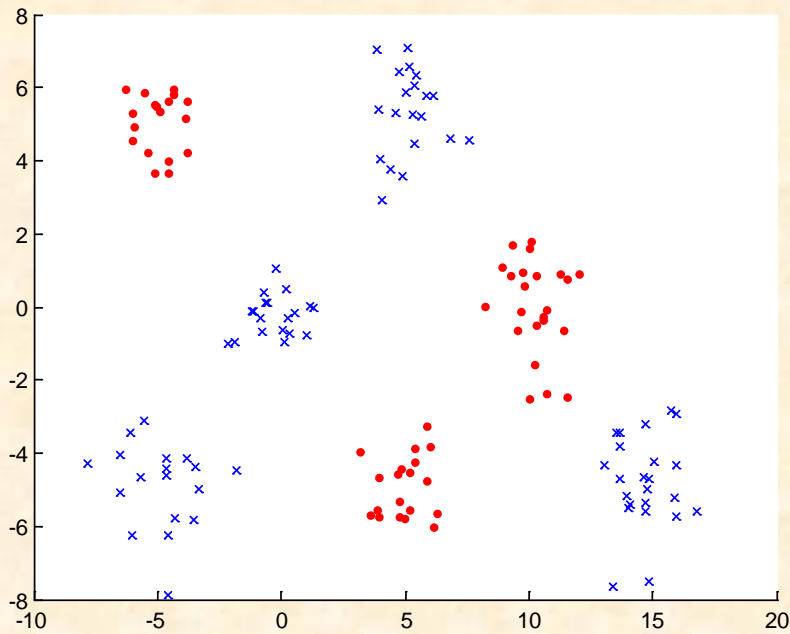
Εμφανίζεται όταν ένα **(μεγάλου μεγέθους)** δίκτυο αρχίζει να προσαρμόζεται στις ιδιαιτερότητες του συνόλου δεδομένων.

Αντιμετώπιση: Διακοπή εκπαίδευσης όταν το λάθος ταξινόμησης αρχίζει να αυξάνει για το σύνολο δοκιμής.



ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ

Ένα παράδειγμα:



Όπως αναμενόταν, η επιφάνεια απόφασης δεν ορίζεται μέσω υπερεπιπέδων.

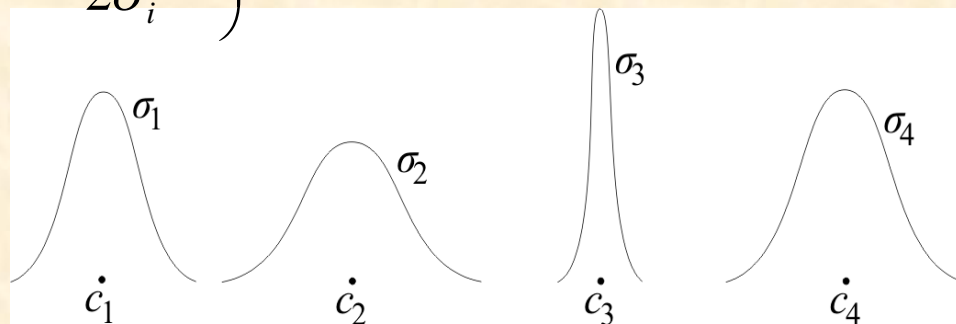
ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Συναρτήσεις ακτινωτής βάσης (Radial basis functions -RBFs)

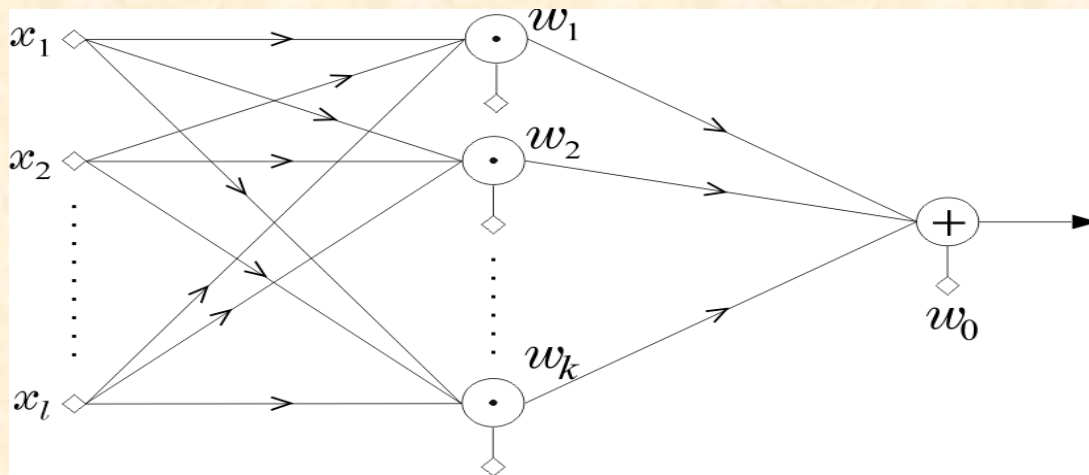
➤ Στην περίπτωση αυτή, οι $f_i(\cdot)$ δε σχετίζονται με υπερεπίπεδο αλλά με σημείο στο χώρο και επιλέγονται ως

$$f_i(\underline{x}) = \exp\left(-\frac{\|\underline{x} - \underline{c}_i\|^2}{2\sigma_i^2}\right)$$

➤ Έχουν σχήμα «**καμπάνας**», είναι κεντραρισμένες στο \underline{c}_i και το πλάτος τους εξαρτάται από το σ_i .



➤ Η παραπάνω απεικόνιση $\mathbf{x} \rightarrow \mathbf{y}$ μπορεί να υλοποιηθεί από ένα **δίκτυο δύο επιπέδων**, όπου οι νευρώνες του πρώτου επιπέδου υλοποιούν τις συναρτήσεις $f_i(\cdot)$ ενώ ο νευρώνας του δεύτερου επιπέδου πραγματοποιεί γραμμικό διαχωρισμό στο μετασχηματισμένο χώρο.



ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Συναρτήσεις ακτινωτής βάσης (Radial basis functions-RBFs)

Παράδειγμα: Το πρόβλημα **XOR**.

Έστω

$$c_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, c_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}$$

Χρησιμοποιώντας συναρτήσεις RBF με τις παραπάνω παραμέτρους επιτυγχάνουμε την απεικόνιση του **αρχικού 2-διάστατου** χώρου σε ένα νέο **μετασχηματισμένο** χώρο ίδιας **διάστασης**.

$$x \rightarrow y = \begin{bmatrix} \exp(-\|x - c_1\|^2) \\ \exp(-\|x - c_2\|^2) \end{bmatrix}$$

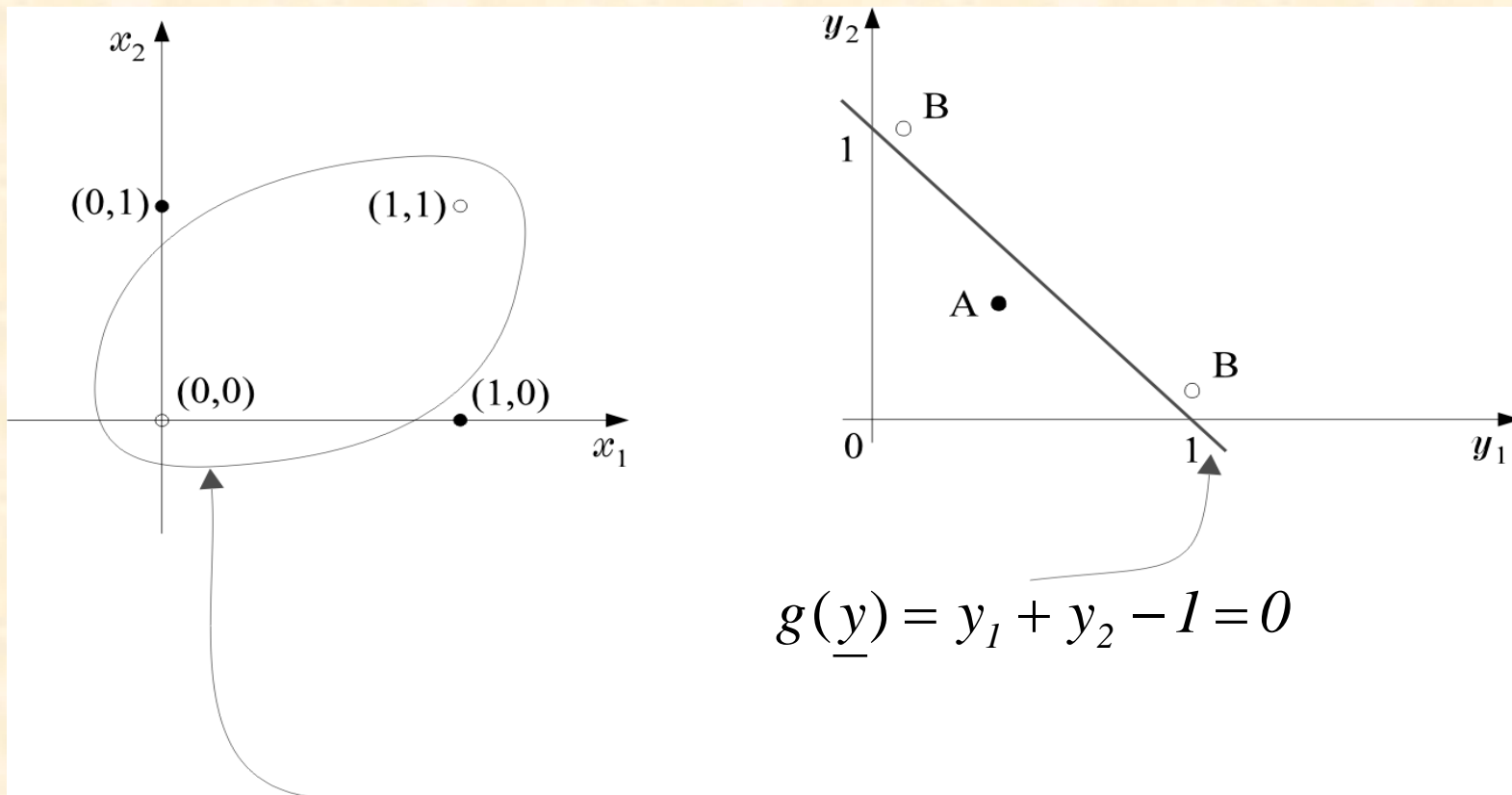
Μέσω αυτού του μετασχηματισμού έχουμε

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0.135 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ 0.135 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0.368 \\ 0.368 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0.368 \\ 0.368 \end{bmatrix}$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Συναρτήσεις ακτινωτής βάσης (Radial basis functions-RBFs)

Παράδειγμα: Το πρόβλημα **XOR**.



$$g(\underline{x}) = \exp(-\|\underline{x} - \underline{c}_1\|^2) + \exp(-\|\underline{x} - \underline{c}_2\|^2) - 1 = 0$$

ΜΗ ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ – ΓΕΝΙΚΕΥΜΕΝΟΙ ΓΡΑΜ. ΤΑΞΙΝΟΜΗΤΕΣ

Συναρτήσεις ακτινωτής βάσης - ζητήματα εκπαίδευσης

Οι εμπλεκόμενες παράμετροι είναι τα c_i , τα σ_i και τα w_i .

- σ_i : Συνήθως τίθενται εκ των προτέρων σε σταθερές τιμές (υπάρχουν όμως και άλλες εναλλακτικές).

- c_i : Μπορεί

- είτε να τεθούν ίσα με **τυχαία επιλεγμένα διανύσματα εκπαίδευσης** ή
- είτε να τεθούν ίσα με τα **κέντρα των ομάδων (clusters)** που παράγονται μετά την εφαρμογή ενός αλγόριθμου ομαδοποίησης στα σημεία κάθε κλάσης.

- w_i : Υπό την προϋπόθεση ότι τα c_i και σ_i έχουν εκτιμηθεί, η εκτίμηση των w_i είναι μία **γραμμική διαδικασία** (μπορεί π.χ. να χρησιμοποιηθεί μια μέθοδος ελαχίστων τετραγώνων).