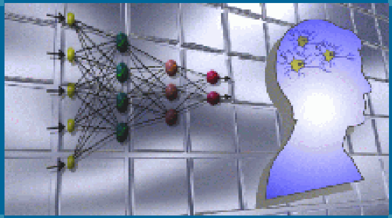
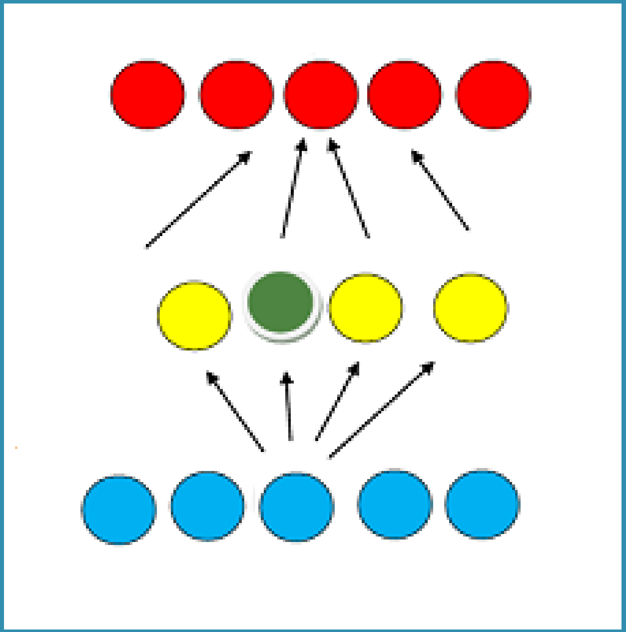


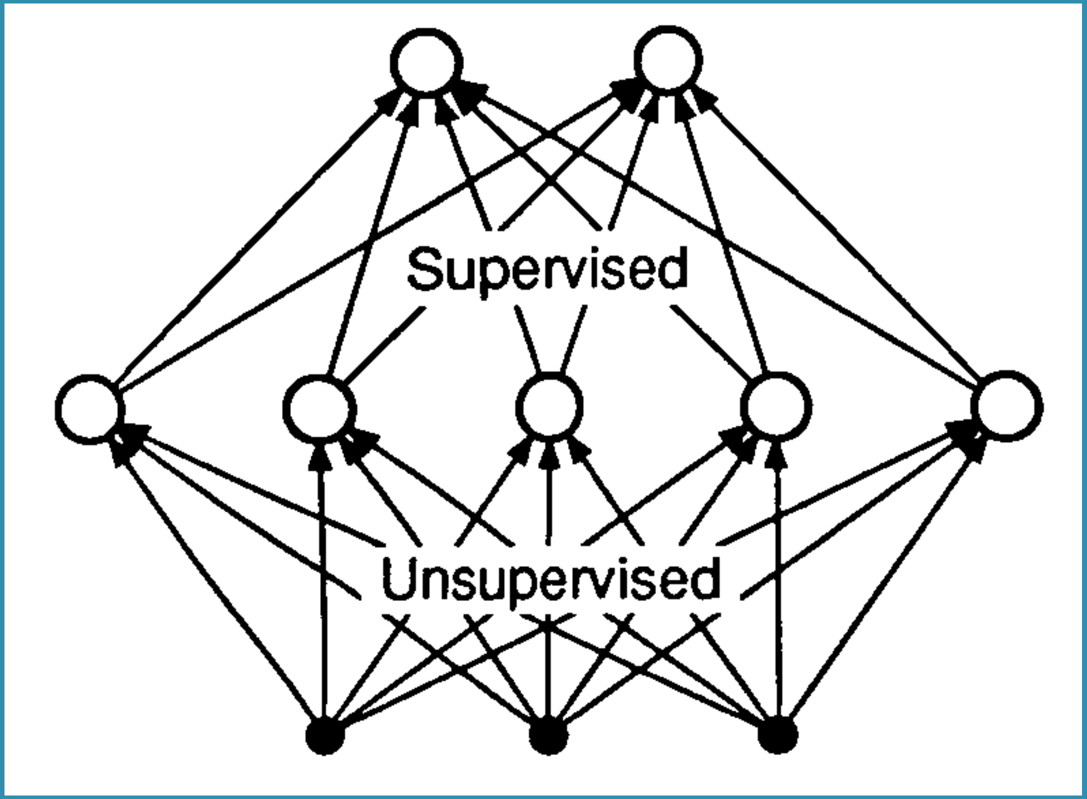
Υβριδική μάθηση

Δίκτυα RBF



Υβριδικά δίκτυα

- Στρώμα μη επιβλεπόμενης μάθησης
- Στρώμα επιβλεπόμενης μάθησης (πληροφορία κατηγορίας)



- ➢ Ομοιότητα εισόδων/εξόδων
- ➢ Μη βέλτιστο σχήμα

- Hierarchical feature maps
- Counterpropagation networks
- Radial Basis Function networks

Γενική περίπτωση υβριδικού ταξινομητή

1. Εκπαίδευση πρώτου (κρυμμένου) στρώματος με ανταγωνιστική μάθηση (είσοδοι x_k, βάρη w_{kj}, τιμές κόμβων z_j)
2. Εκπαίδευση στρώματος εξόδου (γραμμική απεικόνιση) με τον κανόνα δέλτα (βάρη v_{ji}, τιμές κόμβων εξόδου y_i)

Γενική περίπτωση υβριδικού ταξινομητή

Εκπαίδευση στρώματος εξόδου με τον κανόνα δέλτα:

$$\Delta v_{ji} = \eta (d_i - y_i) z_j = \eta (d_i - v_{ji}) z_j, \quad 0 < \eta < 1$$

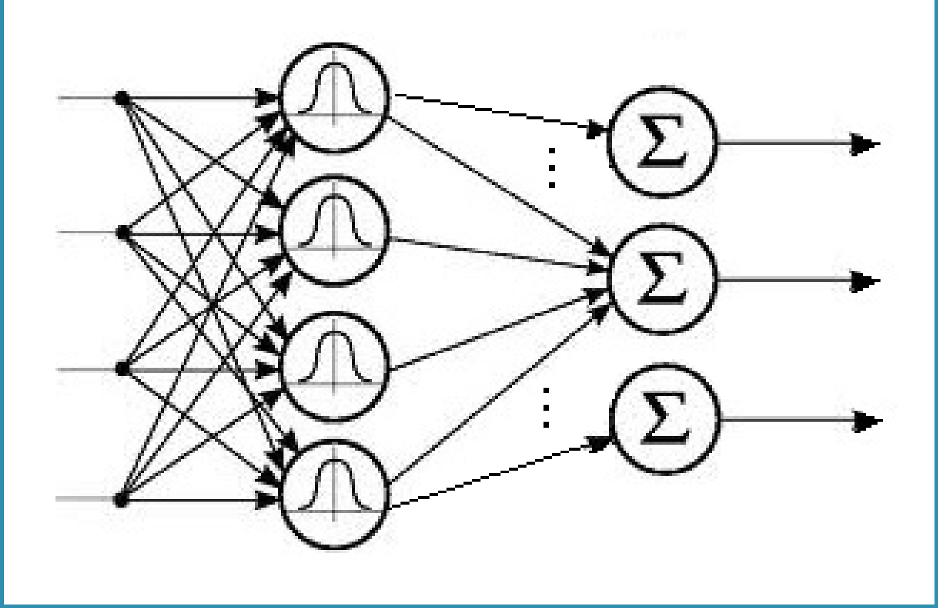
(z_j=1 για τον νικητή, 0 αλλιώς)

$$y_i = \sum_j v_{ji} z_j$$

- Nearest-match lookup table (key→value)
- Το διάνυσμα των βαρών v_{ji} (δεδομένο j) προσεγγίζει τον μέσο όρο των εξόδων y για την ομάδα εισόδων j.
- Δυνατότητα παρεμβολής (μετά την εκπαίδευση)

Ακτινικές συναρτήσεις βάσης
Radial Basis Functions - RBF

Broomhead & Lowe,
Moody & Darken, 1988



- Δεκτικά πεδία (receptive fields)
- Τοπικός συντονισμός (local tuning) κρυμμένων μονάδων
- Παρεμβολή (interpolation) μεταξύ προτύπων

Δίκτυα RBF

Ακτινικές συναρτήσεις βάσης (κρυφό επίπεδο)

Μη γραμμικές συναρτήσεις φ_j (χώρος χαρακτηριστικών)

$$y_i = f_i(\vec{x}) = \sum_j v_{ji}\varphi_j\left(\left\|\vec{x} - \vec{\mu}_j\right\|\right)$$

Η *j*-στη συνάρτηση βάσης φ_j εξαρτάται από την απόσταση μεταξύ του ***x*** και του κέντρου $\boldsymbol{\mu_j}$

Ακτινική συμμετρία:

$$\left\|\overrightarrow{z_1}\right\| = \left\|\overrightarrow{z_2}\right\| \Rightarrow \varphi\left(\left\|\overrightarrow{z_1}\right\|\right) = \varphi\left(\left\|\overrightarrow{z_2}\right\|\right)$$

Δίκτυα RBF

- Συνήθως χρησιμοποιούνται (κανονικοποιημένες) γκαουσιανές συναρτήσεις βάσης

$$\varphi_j\left(\left\|\vec{x}-\vec{\mu}_j\right\|\right)=R_j(\vec{x})\approx \exp\left(-\frac{\left\|\vec{x}-\vec{\mu}_j\right\|^2}{2\sigma_j^2}\right)$$

$$y_i = f_i(\vec{x}) = \sum_j v_{ji} R_j(\vec{x})$$

- μ_j: κέντρο
- σ_j: ακτίνα

Συναρτήσεις πυρήνα

(Kernel functions)

Ελαχιστοποίηση συνάρτησης κόστους

$$E(\vec{v}) = \tfrac{1}{2}\sum_{l}\sum_{i}\sum_{j}\left(d_i^l - v_{ji}R_j(\vec{x}^{\,l})\right)^2$$

$$y_i^l = \sum_j v_{ji} z_j^l = \sum_j v_{ji} R_j(\vec{x}^{\,l})$$

$\vec{x}^{\,l}$: πρότυπο εισόδου

[Υποθέτουμε αμελητέα επικάλυψη μεταξύ των δεκτικών πεδίων
⇨ Νικητής νευρώνας]

Συνήθως χρησιμοποιούμενες RBF

Gaussian $\phi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right)$ $\sigma > 0$

Multi-Quadric $\phi(r) = \left(r^2 + \sigma^2\right)^{1/2}$ $\sigma > 0$

Generalized Multi-Quadric $\phi(r) = \left(r^2 + \sigma^2\right)^{\beta}$ $\sigma > 0,\ 1 > \beta > 0$

Inverse Multi-Quadric $\phi(r) = \left(r^2 + \sigma^2\right)^{-1/2}$ $\sigma > 0$

Generalized Inverse Multi-Quadric $\phi(r) = \left(r^2 + \sigma^2\right)^{-\alpha}$ $\sigma > 0,\ \alpha > 0$

Thin Plate Spline $\phi(r) = r^2 \ln(r)$

Cubic $\phi(r) = r^3$

Linear $\phi(r) = r$

Δίκτυα RBF
Προσδιορισμός βαρών (γραμμικής) εξόδου

N: αριθμός εκπαιδευτικών προτύπων

n: διάσταση προτύπων εισόδου

M: αριθμός συναρτήσεων βάσης

C: αριθμός κόμβων εξόδου

- Δίκτυα παρεμβολής

$$M=N$$

- Δίκτυα προσέγγισης (ταξινόμησης)

$$M<N$$

Δίκτυα RBF Ακριβής παρεμβολή

$$y_i^l = f_i(\vec{x}^l) = \sum_j v_{ji}\varphi_j\left(\left\|\vec{x}^l - \vec{x}^j\right\|\right) = d_i^l, \quad l, j = 1, \ldots, N, i = 1, \ldots, C$$

M=N Τα κέντρα συμπίπτουν με τα πρότυπα εισόδου.

Κοινή μήτρα για τις μονάδες εξόδου

$$\Phi = \{\varphi_{lj}\} = \left\{\varphi_j\left(\left\|\vec{x}^l - \vec{x}^j\right\|\right)\right\} \quad l, j = 1, \ldots, N$$

Για κάθε μονάδα εξόδου *i*, τα βάρη v_{ji} προκύπτουν από τη λύση του γραμμικού συστήματος **Φv=d**

$$\boxed{\mathbf{v} = \mathbf{\Phi}^{-1}\mathbf{d}}$$

Ακριβής παρεμβολή:

υπερεκπαίδευση, χαμηλή γενίκευση, υψηλό υπολογιστικό κόστος

Βελτίωση: Δίκτυα Προσέγγισης

$$y_i^l = f_i(\vec{x}^l) = d_i^l,$$

$$l = 1,\ldots,N,\ j = 1,\ldots,M,\ i = 1,\ldots,C$$

$M<N$ Τα κέντρα δεν συμπίπτουν υποχρεωτικά με πρότυπα εισόδου.

Ο προσδιορισμός κατάλληλων κέντρων είναι μέρος της εκπαίδευσης.

Δίκτυα RBF

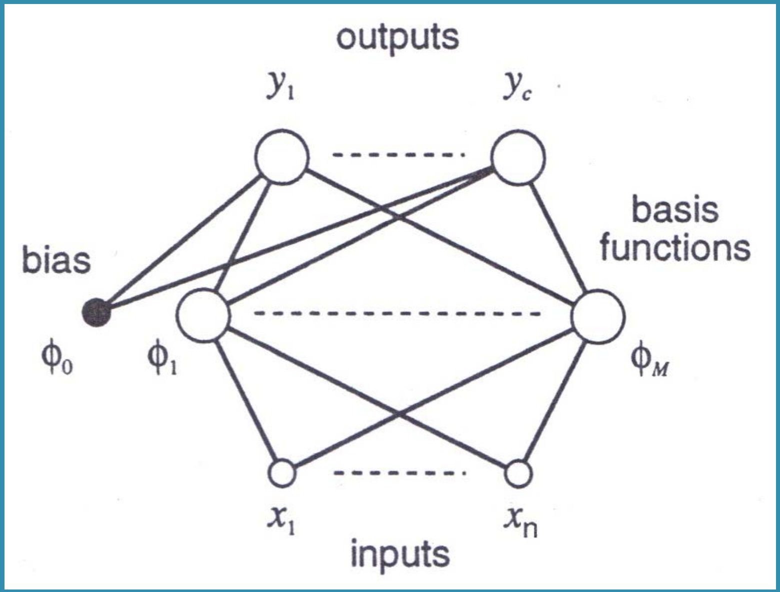
Προσέγγιση

$$y_i^l = f_i(\vec{x}^l) = \sum_j v_{ji}\varphi_j\left(\left\|\vec{x}^l - \vec{\mu}_j\right\|\right) + v_{0i}$$

$$\boxed{\varphi_0 = 1}$$

$$= \sum_{j=0}^{M} v_{ji}\varphi_j\left(\left\|\vec{x}^l - \vec{\mu}_j\right\|\right), \quad l = 1,\ldots,N, \; i = 1,\ldots,C$$

Η γραμμική συνάρτηση εξόδου περιλαμβάνει σταθερό όρο εξισορρόπησης (πόλωση)



Δίκτυα RBF Προσέγγιση

Προσδιορισμός βαρών

Βελτιστοποίηση του γραμμικού επιπέδου ως προς τα βάρη

Ελαχιστοποίηση της συνάρτησης σφάλματος (sum-of-squares)

$$E = \frac{1}{2} \sum_{i=1}^{C} \sum_{l=1}^{N} \left\{ d_i^l - y_i^l \right\}^2$$

- Επίλυση του γραμμικού συστήματος **Φv=d (**περισσότερες εξισώσεις από αγνώστους) με χρήση της ψευδοαντίστροφης μήτρας $\mathbf{\Phi^{+} = (\Phi^{T}\Phi)^{-1}\Phi^{T}}$

$$\boxed{\mathbf{v^{T} = \Phi^{+} d}}$$

- Επαναληπτική επίλυση με κατάβαση κλίσης

Δίκτυα RBF

Θεωρία ομαλοποίησης (Regularization theory)

Αντί να περιορίσουμε τον αριθμό των κρυμμένων νευρώνων, για να αποφύγουμε την υπερεκπαίδευση, μπορούμε να διατηρήσουμε μια συνάρτηση βάσης ανά εκπαιδευτικό πρότυπο, προσθέτοντας όμως έναν επιπλέον όρο στη συνάρτηση σφάλματος, ο οποίος αποτρέπει τις μη ομαλές απεικονίσεις (όρος ομαλοποίησης).

$$E = \frac{1}{2}\sum_{i=1}^{C}\sum_{l=1}^{N}\left\{d_i^l - y_i^l\right\}^2 + \lambda\Omega$$

- λ: παράμετρος ομαλοποίησης
- Ω: διάφορες επιλογές (κυρίως δευτεροβάθμιες ως προς τα βάρη εξόδου)

Δίκτυα RBF   Ταξινόμηση

Θεώρημα του Cover

"A complex pattern classification problem, cast in a high dimensional space non-linearly, is more likely to be linearly separable than in a low dimensional space".

T.M. Cover, IEEE Transactions on Electronic Computers, 1965

1. Μη γραμμικοί κρυμμένοι νευρώνες
2. (Χώρος κρυμμένων νευρώνων) > (χώρος εισόδου)

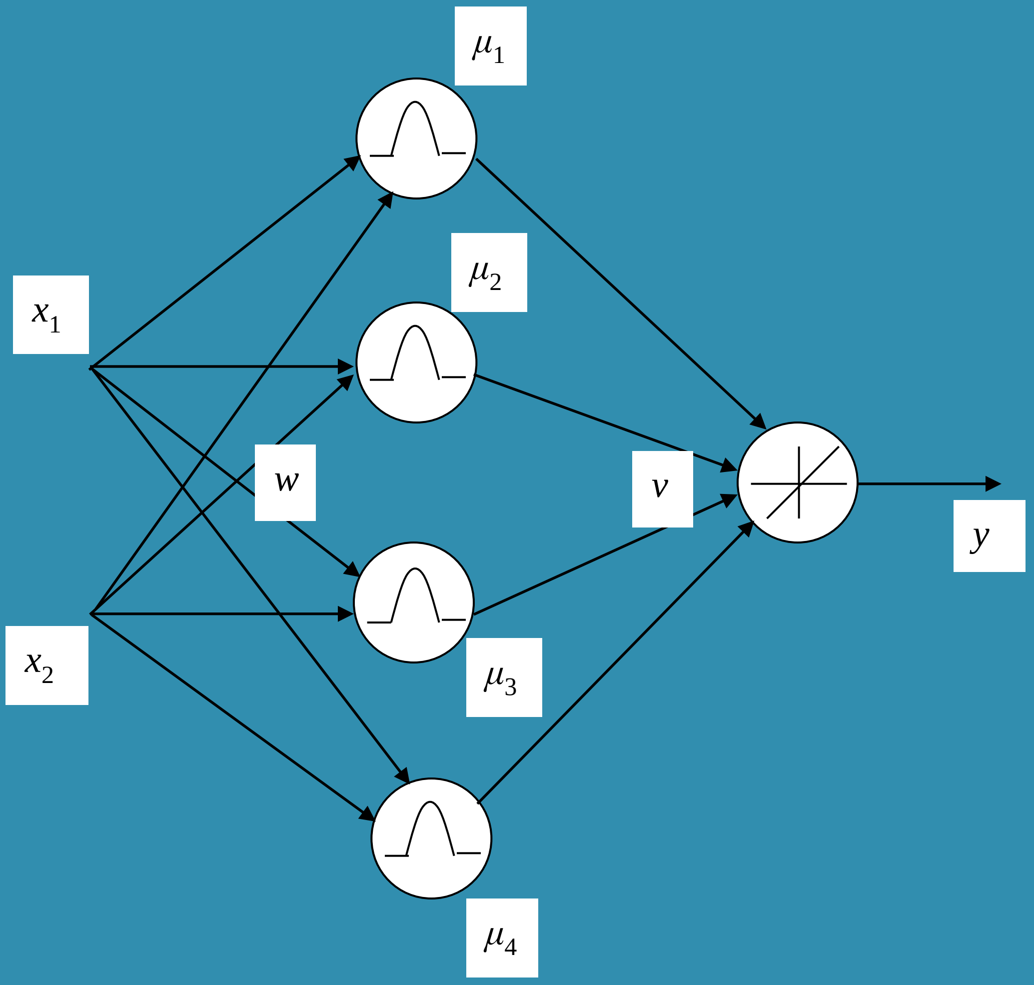
$M > n$

Συχνά το 1. είναι αρκετό ($M = n$) - Παράδειγμα XOR

Το πρόβλημα XOR

- Πρόβλημα XOR
  2 εισόδων
  ($x_k \in \{0, 1\}$, $k = 1, 2$)

- Μοντέλο RBF
  με 4 κρυμμένους κόμβους
  (ακριβής παρεμβολή)



Υλοποίηση γραμμικά διαχωρίσιμου προβλήματος σε μεγαλύτερη διάσταση

Συναρτήσεις βάσης: $\varphi_j(\boldsymbol{x}) = \exp (- | \boldsymbol{x} - \boldsymbol{\mu}_j |^2),\ j = 1, 2, 3, 4$

Κέντρα: $\boldsymbol{\mu}_1 = [1, 1], \boldsymbol{\mu}_2 = [0, 0], \boldsymbol{\mu}_3 = [0, 1], \boldsymbol{\mu}_4 = [1, 0]$

Έξοδος δικτύου: $y = v_1\, \varphi_1 (\boldsymbol{x}) + v_2\, \varphi_2 (\boldsymbol{x}) + v_3\, \varphi_3 (\boldsymbol{x}) + v_4\, \varphi_4 (\boldsymbol{x})$

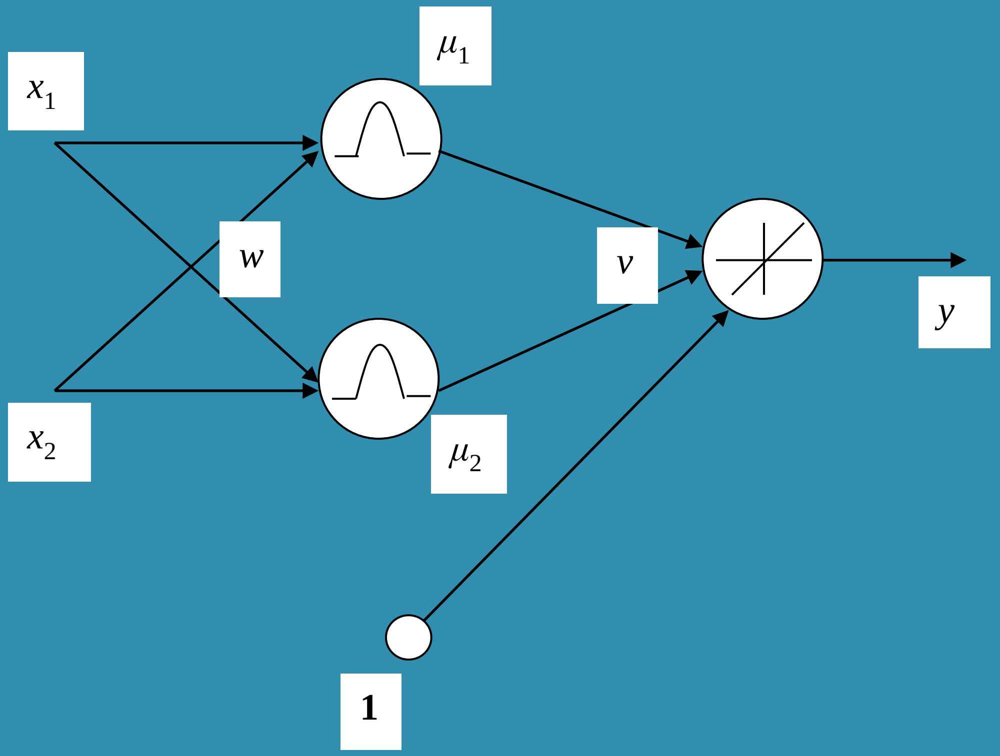
Πρότυπο εισόδου	$\boldsymbol{x} = [x_1, x_2]$	$\varphi_1(\boldsymbol{x})$	$\varphi_2(\boldsymbol{x})$	$\varphi_3(\boldsymbol{x})$	$\varphi_4(\boldsymbol{x})$	Επιθυμητή έξοδος
1	[1, 1]	1	0,1353	0,3678	0,3678	0
2	[0,0]	0,1353	1	0,3678	0,3678	0
3	[0, 1]	0,3678	0,3678	1	0,1353	1
4	[1, 0]	0,3678	0,3678	0,1353	1	1

$$v_1 . 1 + v_2 . 0{,}1353 + v_3 . 0{,}3678 + v_4 . 0{,}3678 = 0$$
$$v_1 . 0{,}1353 + v_2 . 1 + v_3 . 0{,}3678 + v_4 . 0{,}3678 = 0$$
$$v_1 . 0{,}3678 + v_2 . 0{,}3678 + v_3 . 1 + v_4 . 0{,}1353 = 1$$
$$v_1 . 0{,}3678 + v_2 . 0{,}3678 + v_3 . 0{,}1353 + v_4 . 1 = 1$$

Τελικά: $v_1 = v_2 = -0{,}9843$

$v_3 = v_4 = 1{,}5188$

Το πρόβλημα XOR



- Πρόβλημα XOR
  δύο εισόδων
  ($x_k \in \{0, 1\}$, $k = 1, 2$)

- Μοντέλο RBF
  με 2 κρυμμένους κόμβους
  (προσέγγιση)

- Επιλέγουμε κέντρα [1, 1] και [0, 0]

- Υλοποίηση γραμμικά διαχωρίσιμου προβλήματος στην ίδια διάσταση

Συναρτήσεις βάσης: $\varphi_j(\boldsymbol{x}) = \exp(-|\boldsymbol{x} - \boldsymbol{\mu}_j|^2),\ j = 1, 2$

Κέντρα: $\boldsymbol{\mu}_1 = [1, 1], \boldsymbol{\mu}_2 = [0, 0]$

Έξοδος δικτύου: $y = v_1\,\varphi_1(\boldsymbol{x}) + v_2\,\varphi_2(\boldsymbol{x}) + v_0$

Πρότυπο εισόδου	$\boldsymbol{x} = [x_1, x_2]$	$\varphi_1(\boldsymbol{x})$	$\varphi_2(\boldsymbol{x})$	Επιθυμητή έξοδος
1	[1, 1]	1	0,1353	0
2	[0,0]	0,1353	1	0
3	[0, 1]	0,3678	0,3678	1
4	[1, 0]	0,3678	0,3678	1

Το πρόβλημα γίνεται γραμμικά διαχωρίσιμο στο επίπεδο $\varphi 1$-$\varphi 2$

$$v_1 \,.\, 1 + v_2 \,.\, 0{,}1353 + v_0 = 0$$
$$v_1 \,.\, 0{,}1353 + v_2 \,.\, 1 + v_0 = 0$$
$$v_1 \,.\, 0{,}3678 + v_2 \,.\, 0{,}3678 + v_0 = 1$$
$$v_1 \,.\, 0{,}3678 + v_2 \,.\, 0{,}3678 + v_0 = 1$$

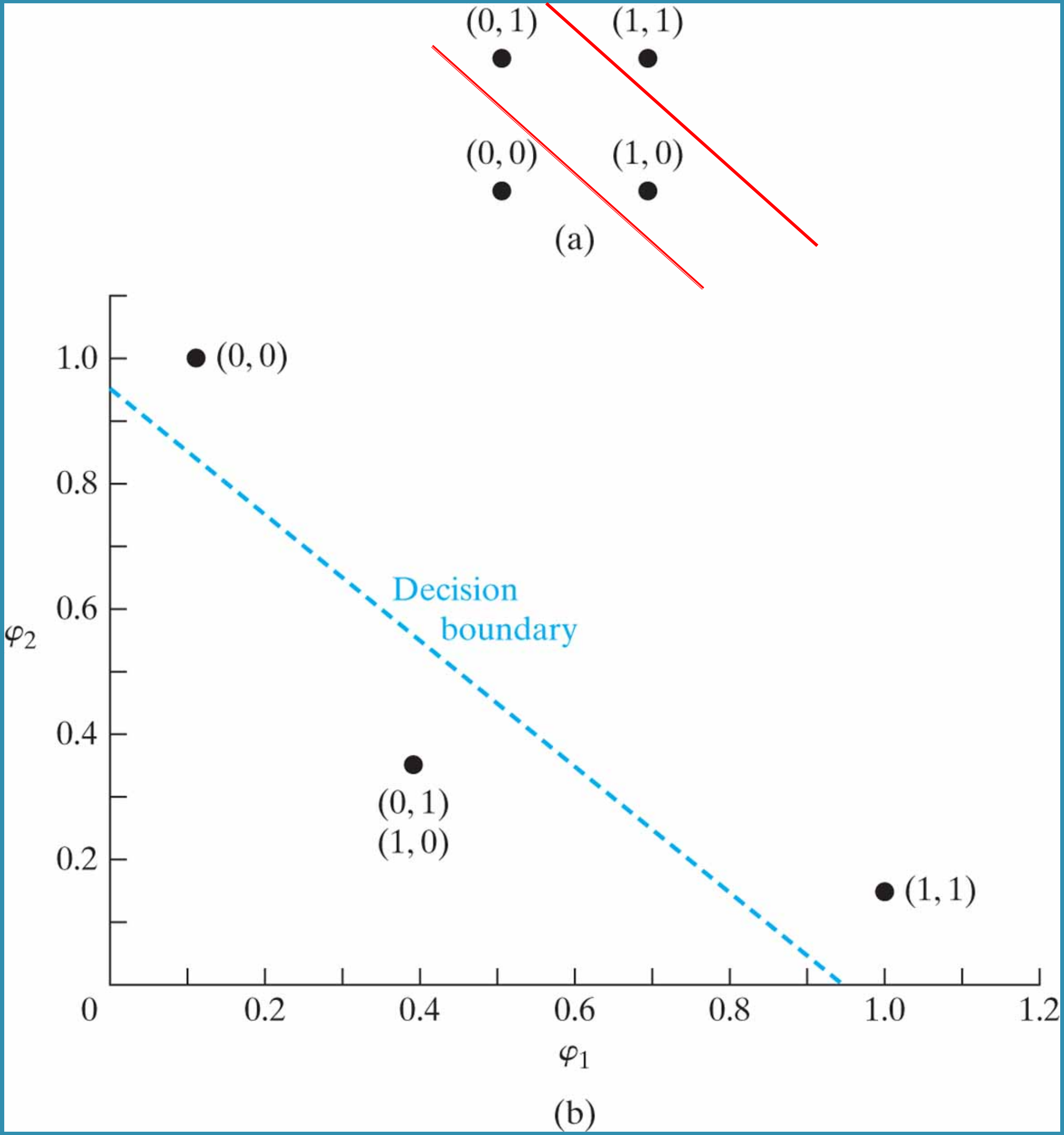
Τελικά: $v_1 = v_2 = -2{,}502$

$v_0 = 2{,}841$

Γραμμικά διαχωρίσιμο πρόβλημα

(a) Χώρος εισόδου

(b) Χώρος χαρακτηριστικών



Δίκτυα RBF

- Εκπαίδευση δύο σταδίων
  - Κρυμμένο επίπεδο

    ⇒ Καθορισμός κέντρων: ανταγωνιστική μάθηση (π.χ. αλγόριθμος *k*-μέσων)

    ⇒ Καθορισμός ακτίνων: ευρετικά (π.χ. μέση τετραγωνική απόσταση από τους *p*-πλησιέστερους γείτονες)
  - Επίπεδο εξόδου

    Εκπαίδευση γραμμικού ταξινομητή (κανόνας δέλτα)

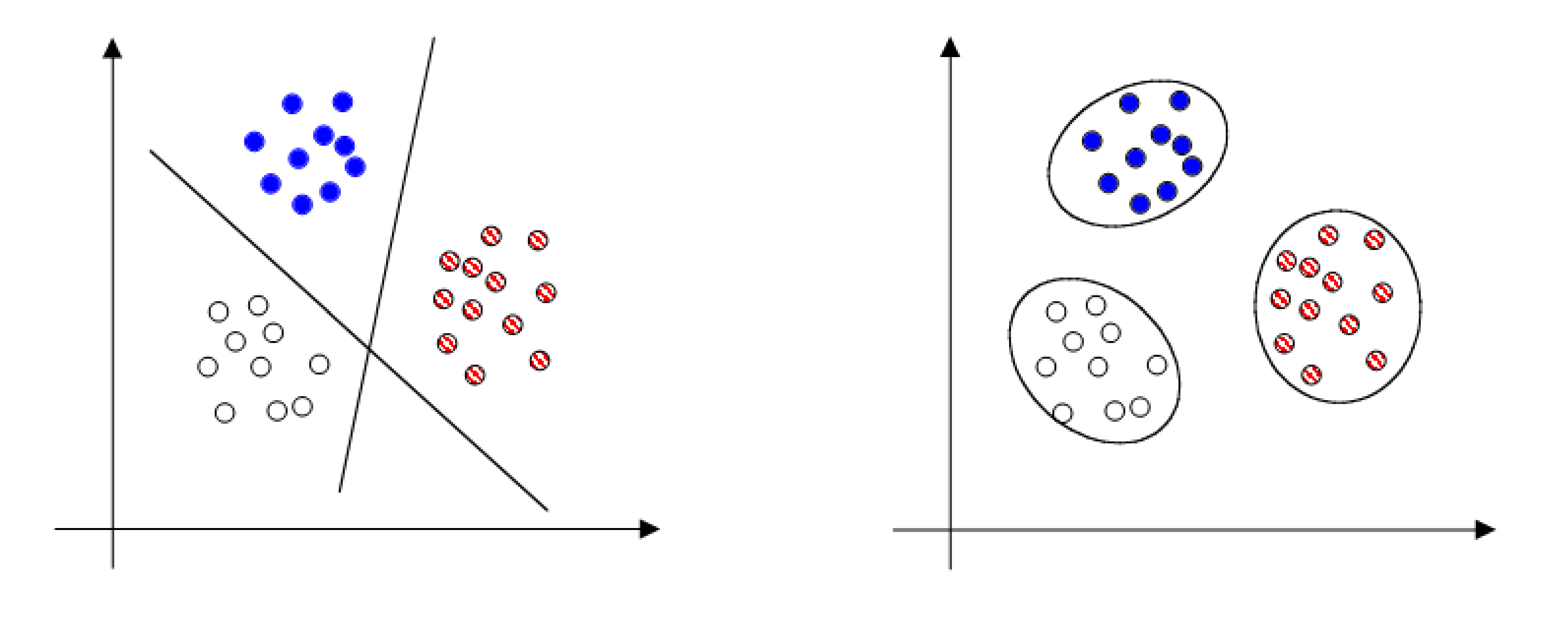
- Ενιαία εκπαίδευση
  - Ταυτόχρονη ενημέρωση όλων των παραμέτρων του δικτύου (κάθοδος κλίσης)
  - Χρήση ανταγωνιστικής μάθησης για αρχικοποίηση κρυμμένου επιπέδου

Σύγκριση MLP και RBF

- MLP
  - Κρυμμένες μονάδες: υπερεπίπεδα
  - Κατανεμημένη αναπαράσταση
  - Ανοχή σε σφάλματα

- RBF
  - Κρυμμένες μονάδες: σφαιρικά δεκτικά πεδία
  - Τοπική αναπαράσταση (διαμερισμός χώρου)
  - Υψηλή ταχύτητα εκπαίδευσης

Πρόβλημα ταξινόμησης



MLP vs RBF

Σύγκριση MLP και RBF

Ομοιότητες

- Μη γραμμικά, στρωματικά δίκτυα πρόσθιας τροφοδότησης
- Καθολικοί προσεγγιστές (universal approximators): *Οποιαδήποτε συνάρτηση μπορεί να προσεγγιστεί με αυθαίρετα μικρό σφάλμα εφόσον υπάρχουν αρκετοί κρυμμένοι νευρώνες.*
- Παρόμοιες περιοχές εφαρμογών