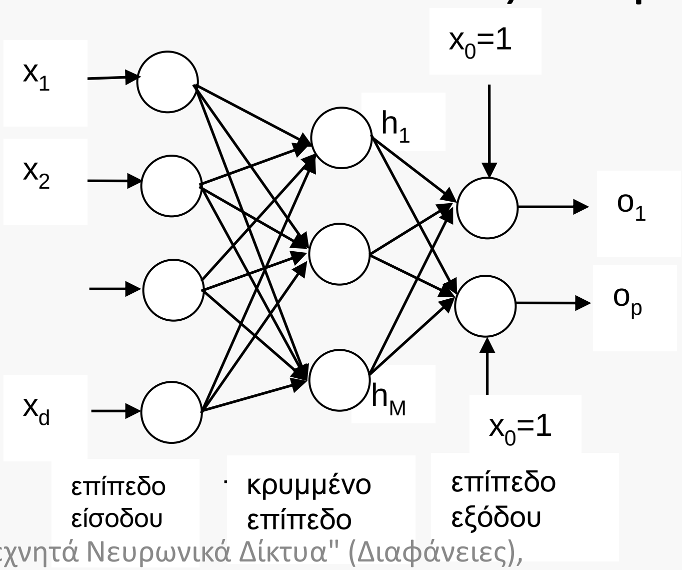
# ΔΙΚΤΥΟ RBF

# Αρχιτεκτονική δικτύου RBF

- Δίκτυα RBF: δίκτυα συναρτήσεων πυρήνα (radial basis function networks). Πρόσθιας τροφοδότησης (**feedforward**) για **προβλήματα μάθησης με επίβλεψη**. Εναλλακτικό του MLP.
- Υλοποιούν απεικονίσεις από το χώρο εισόδων στο χώρο εξόδων (όπως το MLP), ωστόσο η λειτουργία των νευρώνων του κρυμμένου επιπέδου είναι εντελώς διαφορετική.



# Αρχιτεκτονική δικτύου RBF

- Τα δίκτυα RBF έχουν **ένα μόνο κρυμμένο επίπεδο**, του οποίου οι κρυμμένοι νευρώνες j υπολογίζουν μια ειδική συνάρτηση $\mathbf{h_j(x)}$ του διανύσματος εισόδου x.
- Εστω ένα δίκτυο RBF με d εισόδους, M κρυμμένους νευρώνες και p εξόδους. Για είσοδο $x=(x_1,\ldots,x_d)^T$ , οι έξοδοι $o_i$ $(i=1,\ldots,p)$ του δικτύου υπολογίζονται ως:

$$o_i(x)=v_{i0}+\sum_{j=1}^{M} v_{ij} h_j(x)$$

  $v_{ij}$ είναι το βάρος της σύνδεσης από τον κρυμμένο νευρώνα j προς το νευρώνα εξόδου i και $v_{i0}$ είναι η πόλωση του νευρώνα εξόδου i.
- **Οι νευρώνες εξόδου** είναι τυπικοί **νευρώνες εσωτερικού γινομένου** (όπως και στο MLP) με **γραμμική** συνάρτηση ενεργοποίησης.

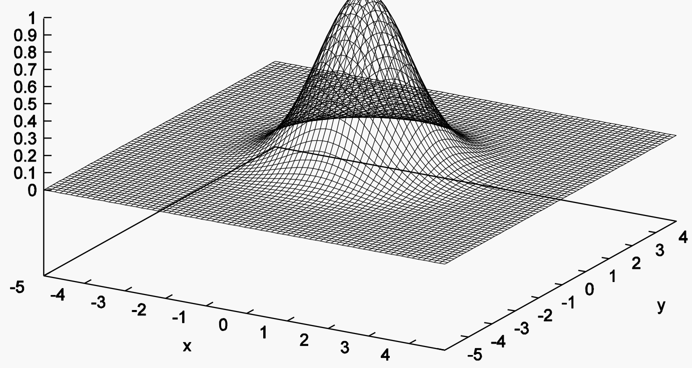
# Συναρτήσεις RBF

- Οι κρυμμένοι νευρώνες υπολογίζουν τις συναρτήσεις h(x) (με όρισμα το διάνυσμα εισόδου x) που ονομάζονται **ακτινικές συναρτήσεις βάσης** (radial basis functions).
- Συνήθως έχουν την ακόλουθη μορφή: υπάρχει κάποιο σημείο (που ονομάζεται **κέντρο**) για το οποίο παρέχουν μέγιστη τιμή και καθώς απομακρυνόμαστε **ακτινικά** από το κέντρο η τιμή της συνάρτησης μειώνεται και σχεδόν εκμηδενίζεται για σημεία x που είναι μακριά από το κέντρο.
- Χαρακτηριστικό παράδειγμα είναι η συνάρτηση Gauss, που είναι γνωστή και ως **πυρήνας RBF (RBF kernel)**:

$$h_j(x)=\exp\left(-\frac{\|x-w_j\|^2}{2\sigma_j^2}\right)=\exp\left(-\frac{\sum_{l=1}^{d}(x_l-w_{jl})^2}{2\sigma_j^2}\right)$$

# Συναρτήσεις RBF

- Το διάνυσμα βαρών $w_j=(w_{j1},\ldots, w_{jd})^T$ καθορίζει το **κέντρο** της συνάρτησης, ενώ η παράμετρος $σ_j$ καθορίζει την **ακτίνα** της συνάρτησης.
- Παίρνει τη μέγιστη τιμή στο κέντρο (για $x=w_j$) και η τιμή ελαττώνεται εκθετικά καθώς απομακρυνόμαστε ακτινικά από το κέντρο.
- Ο ρυθμός μείωσης καθορίζεται από την τιμή της ακτίνας $σ_j$. Για μικρές τιμές της ακτίνας ο ρυθμός μείωσης είναι μεγάλος, ενώ το αντίθετο συμβαίνει για μεγάλες τιμές της ακτίνας.




d=2

# Συναρτήσεις RBF

- Εντοπισμένες συναρτήσεις (local functions): χαρακτηρίζονται από μια **περιοχή εμβέλειας**, έξω από την οποία η συνάρτηση δίνει αμελητέες τιμές.
- Στην περίπτωση RBF, η περιοχή εμβέλειας είναι σφαιρική στις δύο διαστάσεις και γενικότερα **υπερσφαίρα** στις d διαστάσεις.
- Επομένως όταν ένα διάνυσμα εισόδου x εμφανίζεται ως είσοδος στο δίκτυο RBF, θα δώσουν μεγάλη τιμή εξόδου οι κρυμμένοι νευρώνες οι οποίοι περιέχουν το x στην περιοχή εμβέλειάς τους.
- Μπορούμε λοιπόν να θεωρούμε ότι **οι κρυμμένοι νευρώνες του δικτύου RBF ορίζουν σφαιρικές περιοχές επιρροής στον d-διάστατο χώρο των εισόδων.**

# Δίκτυο RBF (ευθύ πέρασμα)

- Εστω ένα δίκτυο RBF με d εισόδους, M κρυμμένους νευρώνες τύπου RBF, και p εξόδους. Για κάποια είσοδο $x=(x_1,\ldots,x_d)^T$, οι έξοδοι $o_i$ προκύπτουν ως εξής:
  - Επίπεδο εισόδου: $y_i^{(0)}=x_i,\ i=1,\ldots,d$
  - Κρυμμένο επίπεδο:

$$u_i^{(1)}=\sum_{j=1}^{d}(y_j^{(0)}-w_{ij})^2=\sum_{j=1}^{d}(x_j-w_{ij})^2,\ \ i=1,\ldots,M$$

$$h_i=e^{-u_i^{(1)}/\sigma_i^2},\ \ i=1,\ldots,M$$

  - Επίπεδο εξόδου:

$$u_i^{(2)}=v_{i0}+\sum_{j=1}^{M}v_{ij}h_j,\ \ \ i=1,\ldots p$$

$$y_i^{(2)}=u_i^{(2)}\qquad i=1,\ldots,p$$

  - Έξοδος του δικτύου: $o_i=y_i^{(2)}\qquad i=1,\ldots,p$

# Δίκτυο RBF

<u>Ερώτηση</u>: Δίνονται τρεις συναρτήσεις RBF $h_1(x)$, $h_2(x)$, $h_3(x)$ στο χώρο $R^2$ με κέντρα (0,0),(5,5), (-5,-5) αντίστοιχα και σ=1 και για τις τρεις συναρτήσεις. Χωρίς να κάνετε υπολογισμούς, να εκτιμήσετε ποια από τις τρεις συναρτήσεις θα δώσει μεγάλη έξοδο για τα ακόλουθα δεδομένα εισόδου:

α) x=(0.2,-0.2)

β) x=(-4.6,-5.2)

γ) x=(5.5,4.8)

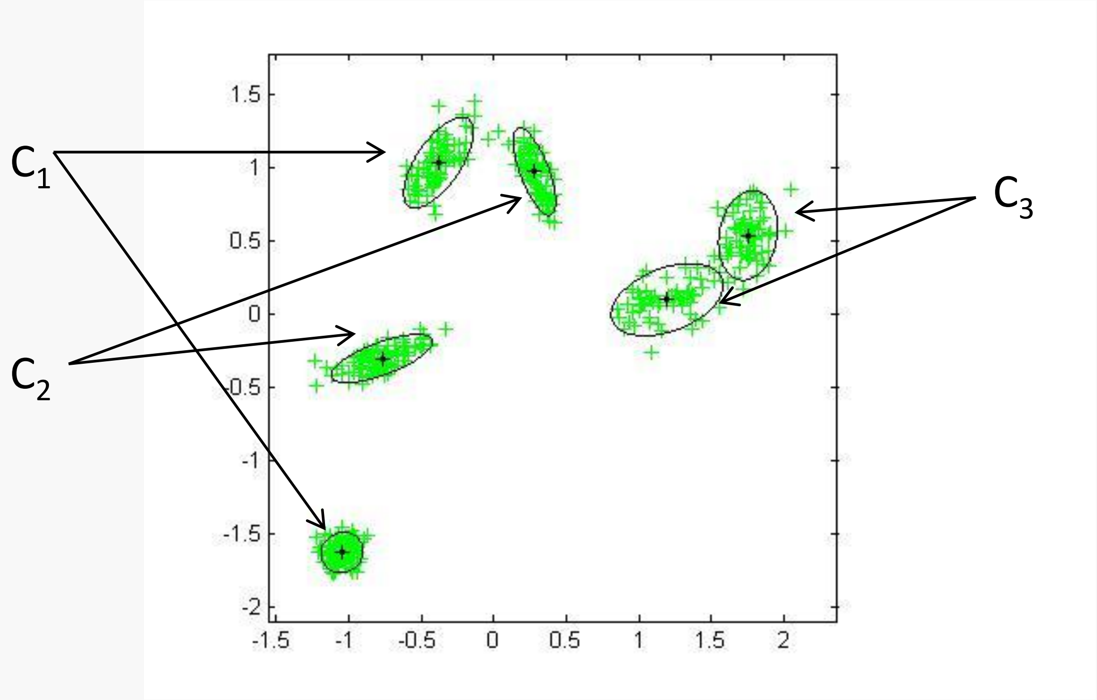
δ) x=(-5,5)

# Δυνατότητες δικτύου RBF

- Το δίκτυο RBF (όπως και το MLP) χαρακτηρίζεται από την ιδιότητα της **παγκόσμιας προσέγγισης** (universal approximation).
- Η ιδιότητα αυτή μας εξασφαλίζει ότι ένα δίκτυο RBF μπορεί να προσεγγίσει οποιαδήποτε συνάρτηση με οποιαδήποτε ακρίβεια, αυξάνοντας επαρκώς τον αριθμό των νευρώνων RBF.
- Ωστόσο αυτή η ιδιότητα δεν είναι πρακτικά εκμεταλλεύσιμη, διότι δεν μας προτείνει πόσους RBF νευρώνες να χρησιμοποιήσουμε δοθέντος ενός συνόλου παραδειγμάτων εκπαίδευσης.
- **Για πρόβλημα της επιλογής του αριθμού των νευρώνων RBF χρησιμοποιούνται οι μέθοδοι που αναφέρθηκαν για το MLP**.

# Κατασκευή δικτύου RBF

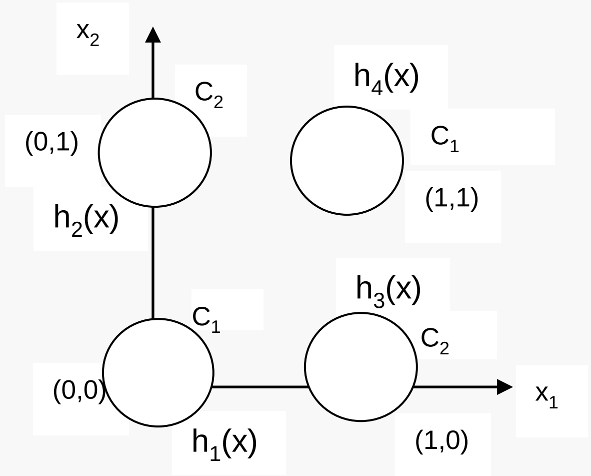
- Έστω ότι γνωρίζουμε πως τα δεδομένα εκπαίδευσης ενός προβλήματος ταξινόμησης σχηματίζουν σφαιρικές (κατά προσέγγιση) και **διακριτές** ομάδες, και τα δεδομένα κάθε ομάδας είναι της ίδιας κατηγορίας.



# Κατασκευή δικτύου RBF

- **Για κάθε ομάδα δεδομένων**, μπορούμε να ορίσουμε **ένα νευρώνα RBF** με κέντρο το κέντρο της αντίστοιχης ομάδας και ακτίνα τη διασπορά της ομάδας.
- Για κάποια είσοδο x η τιμή της εξόδου $h_j(x)$ του νευρώνα j δηλώνει εάν το x ανήκει στην ομάδα j. Ετσι το δεδομένο εισόδου x απεικονίζεται στο διάνυσμα $h(x)=(h_1(x),\ldots,h_M(x))^T$, που δείχνει το βαθμό στον οποίο το x ανήκει σε κάποια από τις ομάδες.
- Στη συνέχεια είναι πολύ εύκολο να καθορίσουμε τα βάρη στο επίπεδο εξόδου, ώστε ο κρυμμένος νευρώνας που αντιστοιχεί στην ομάδα κατηγορίας $C_k$ να 'διεγείρει' μόνο τον αντίστοιχο νευρώνα εξόδου k.

# Κατασκευή δικτύου RBF για το XOR




- Θεωρούμε **ένα κρυμμένο νευρώνα για κάθε παράδειγμα εκπαίδευσης**.
- Κάθε νευρώνας έχει για κέντρο ένα από τα τέσσερα παραδείγματα (0,0), (0,1), (1,0) και (1,1) και μια μικρή ακτίνα σ (π.χ σ=0.1).
- Κατά συνέπεια όταν κάποιο από τα παραδείγματα εμφανιστεί ως είσοδος θα δώσει μεγάλη έξοδο μόνο ο αντίστοιχος κρυμμένος νευρώνας.

# Κατασκευή δικτύου RBF για το XOR

- Στη συνέχεια θα πρέπει να καθορίσουμε τα **βάρη από το κρυμμένο επίπεδο προς το επίπεδο εξόδου**.
- Δεδομένου ότι έχουμε δύο κατηγορίες $C_1$ και $C_2$ θα θεωρήσουμε δύο γραμμικές εξόδους $o_1$ και $o_2$ μία για κάθε κατηγορία.
- Το διάνυσμα των εξόδων θέλουμε να είναι (0,1) για τα διανύσματα h=(0,1,0,0) και h=(0,0,1,0) και (1,0) για τα διανύσματα h=(1,0,0,0) και h=(0,0,0,1).
- Αυτή η απαίτηση επιτυγχάνεται εύκολα εάν θέσουμε:

$$
\begin{aligned}
&v_{11} = 1,\ v_{21} = 0, \qquad && v_{14} = 1,\ v_{24} = 0 \\
&v_{12} = 0,\ v_{22} = 1, \qquad && v_{13} = 0,\ v_{23} = 1
\end{aligned}
$$

# Εκπαίδευση δικτύου RBF

- Οι παράμετροι (βάρη) του δικτύου RBF που πρέπει να καθοριστούν κατά τη διάρκεια της εκπαίδευσης:
  - **τα κέντρα $\mathbf{w_j=(w_{j1},\ldots,w_{jd})^T}$** (M·d παράμετροι) και οι **ακτίνες $\boldsymbol{\sigma_j}$** (M παράμετροι) **των νευρώνων RBF**
  - τα **βάρη $\mathbf{v_{ij}}$** ((M+1)·p παράμετροι συμπεριλαμβανομένων των p πολώσεων $v_{i0}$).
- Διαφορετική ‘σημασία-ερμηνεία’:
  - τα **βάρη w** αναπαριστούν τις **συντεταγμένες των κέντρων** των κρυμμένων νευρώνων
  - τα **βάρη v** έχουν την τυπική σημασία των νευρώνων εσωτερικού γινομένου (όπως και στο MLP) και **ορίζουν την εξίσωση υπερεπιπέδου για τους νευρώνες εξόδου**.
- Μπορεί να χρησιμοποιηθεί διαφορετική εκπαίδευση για τα δύο είδη βαρών.

# Εκπαίδευση δικτύου RBF

- Όπως και στο MLP γίνεται μέσω της ελαχιστοποίησης του τετραγωνικού σφάλματος μεταξύ πραγματικών και επιθυμητών εξόδων του δικτύου.
- Έστω σύνολο παραδειγμάτων εκπαίδευσης $D=\{(x^n,t^n)\}$, $n=1,\ldots,N$, $x^n=(x_{n1,\ldots,}x_{nd})^T$ και $t^n=(t_{n1,\ldots,}t_{np})^T$
- Το δίκτυο RBF θα πρέπει να έχει d νευρώνες στο επίπεδο εισόδου και p νευρώνες στο επίπεδο εξόδου. Εστω M ο αριθμός των κρυμμένων νευρώνων RBF (δίνεται από τον χρήστη).
- $\theta=(\theta_{1,}\ldots,\theta_L)^T$ $(L=(M+1)(d+p))$ ένα διάνυσμα στο οποίο συγκεντρώνουμε όλες τις παραμέτρους του δικτύου
- $o(x^n;\theta)$ το διάνυσμα εξόδου του δικτύου RBF όταν το διάνυσμα εισόδου είναι το $x^n$. Επομένως μπορούμε να ορίσουμε την **τετραγωνική συνάρτηση σφάλματος**:

# Εκπαίδευση δικτύου RBF

- **Τετραγωνική συνάρτηση σφάλματος:**

$$E(\theta)=\sum_{n=1}^{N} E^n(\theta), \quad E^n(\theta) = \frac{1}{2}\|t^n\text{-}o(x^n;\theta)\|^2 = \frac{1}{2}\sum_{m=1}^{p}(t_{nm}\text{-}o_m(x^n;\theta))^2$$

  – $E^n(\theta)$ το τετραγωνικό σφάλμα ανά παράδειγμα ($x^n$, $t^n$)

- Δύο μεθοδολογίες εκπαίδευσης:
  – Ενιαία εκπαίδευση
  – Εκπαίδευση δύο σταδίων

# Ενίαια Εκπαίδευση δικτύου RBF

- Μέθοδος **gradient descent** για ελαχιστοποίηση του Ε(θ) ως προς τις παραμέτρους $\theta_i$ του δικτύου.

$$\theta_i(t+1)=\theta_i(t)-\eta\frac{\partial E}{\partial \theta_i} \text{ ---> } \theta_i(t+1)=\theta_i(t)-\eta\sum_{n=1}^{N}\frac{\partial E^n}{\partial \theta_i}, \quad i=1,..,L$$

- Για το παράδειγμα $(x^n,t^n)$: $E^n=\frac{1}{2}\sum_{i=1}^{p}(o_i-t_{ni})^2$

$$h_j(x)=\exp\left(-\frac{\|x^n-w_j\|^2}{2\sigma_j^2}\right)=\exp\left(-\frac{\sum_{l=1}^{d}(x_l^n-w_{jl})^2}{2\sigma_j^2}\right) \qquad o_i=v_{i0}+\sum_{j=1}^{M}v_{ij}h_j(x^n), \quad i=1,...p$$

- Οπότε: $\frac{\partial E^n}{\partial v_{ij}}=(o_i-t_{ni})h_j(x^n) \qquad \frac{\partial E^n}{\partial v_{i0}}=(o_i-t_{ni})$

$$\frac{\partial E^n}{\partial w_{jl}}=\sum_{i=1}^{p}(o_i-t_{ni})v_{ij}h_j(x^n)\frac{(x_{nl}-w_{jl})}{\sigma_j^2} \qquad \frac{\partial E^n}{\partial \sigma_j}=\sum_{i=1}^{p}(o_i-t_{ni})v_{ij}h_j(x^n)\frac{\|x_n-w_j\|^2}{2\sigma_j^3}$$

# Ενίαια Εκπαίδευση δικτύου RBF

- Μπορεί να εφαρμοστεί gradient descent με ομαδική ή σειριακή ενημέρωση. Προτιμάται η ομαδική ενημέρωση.
- Τα βάρη $v_{ij}$ αρχικοποιούνται όπως στο MLP: τυχαία στο (-1,1).
- Αν τα κέντρα και οι ακτίνες των συναρτήσεων βάσης αρχικοποιούνται εντελώς τυχαία, είναι πολύ δύσκολο για τη μέθοδο gradient descent να αποφύγει **ρηχά** τοπικά ελάχιστα.
- Στην περίπτωση όμως, που τα κέντρα και οι ακτίνες έχουν αρχικοποιηθεί σε καλές τιμές, τότε βρίσκουμε καλές λύσεις.
- Μια πιο αποδοτική προσέγγιση αρχικού καθορισμού των κέντρων βασίζεται στην παρατήρηση ότι **επιθυμούμε τα κέντρα των νευρώνων RBF να τοποθετούνται σε κέντρα ομάδων** που υπάρχουν στα δεδομένα εκπαίδευσης. Έτσι προέκυψε η **μέθοδος εκπαίδευσης των δύο σταδίων.**

# Εκπαίδευση δύο σταδίων δικτύου RBF

**Πρώτο στάδιο**:

- Χρησιμοποιείται το σύνολο των δεδομένων εκπαίδευσης $X=\{x^n\}$ για τον καθορισμό νευρώνων RBF.

- Αγνοείται η πληροφορία σχετικά με την κατηγορία κάθε προτύπου εκπαίδευσης.

- **Εφαρμόζεται κάποια μέθοδος ομαδοποίησης**, π.χ. k-means ή LVQ. Προκύπτουν άμεσα τα κέντρα των ομάδων $w_j$ ενώ ο καθορισμός της ακτίνας γίνεται υπολογίζοντας τη διασπορά των δεδομένων κάθε ομάδας:

$$\sigma_j^2 = \frac{1}{N_j} \sum_{x^n \in O_j} \|x^n - w_j\|^2$$

# Εκπαίδευση δύο σταδίων δικτύου RBF

**Δεύτερο στάδιο**:

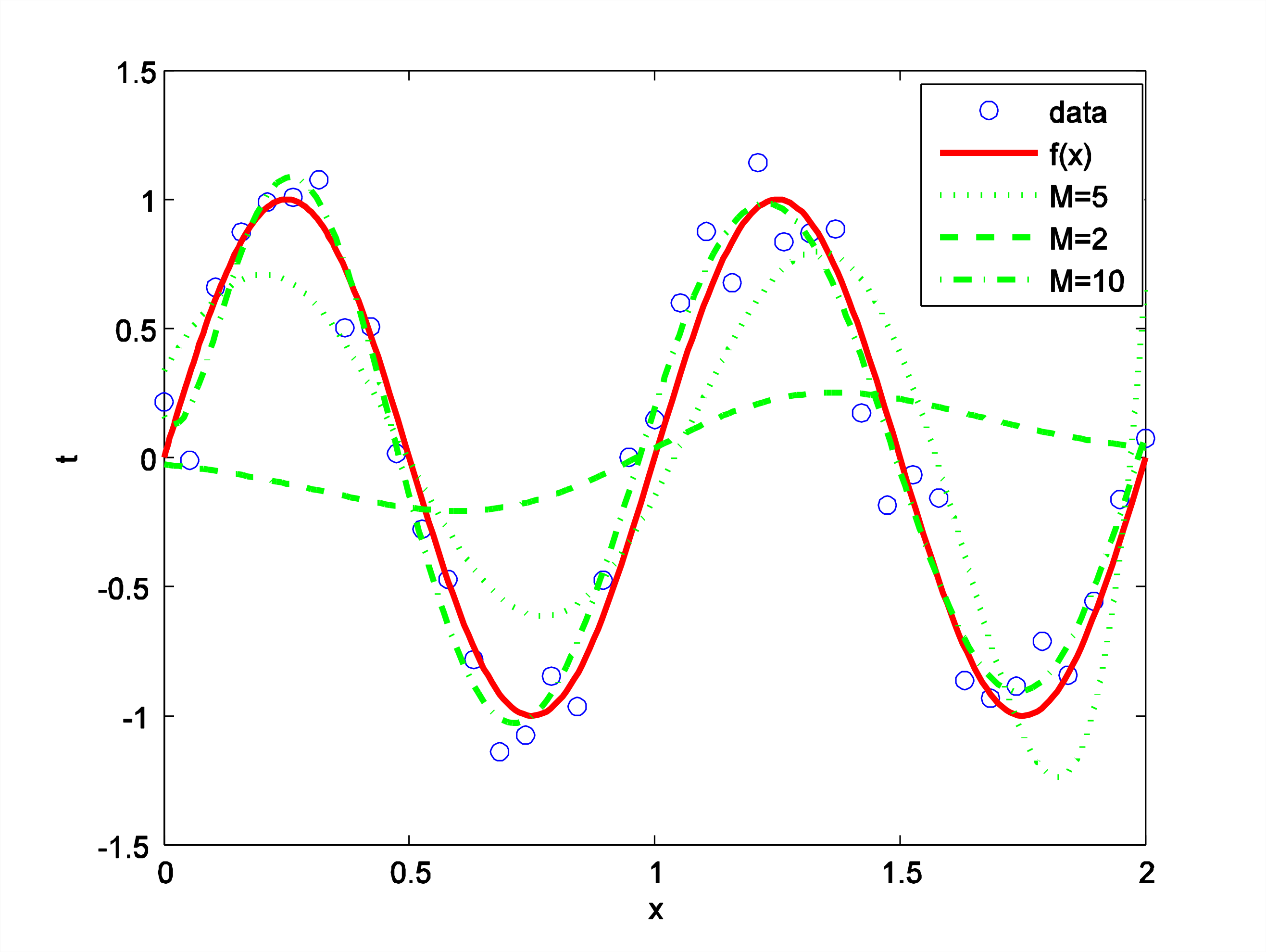
- $1^η$ επιλογή
  - Τα $w_{ij}$ και $\sigma_j$ παραμένουν σταθερά στις τιμές που καθορίστηκαν από το πρώτο στάδιο. Απομένει ο η εύρεση των βαρών $v_{ij}$ και των πολώσεων $v_{i0}$ του επιπέδου εξόδου.
  - Το πρόβλημα ανάγεται στην εκπαίδευση με το **νέο σύνολο εκπαίδευσης** $\{(h^n,t^n)\}$ $h^n=(h_1(x^n),\ldots,h_M(x^n))^T$
  - Μέθοδος gradient descent με μερικές παραγώγους

$$\frac{\partial E^n}{\partial v_{i0}}=(o_i-t_{ni}) \qquad \frac{\partial E^n}{\partial v_{ij}}=(o_i-t_{ni})h_j(x^n)$$

- $2^η$ επιλογή
  - Εφαρμογή gradient descent (όπως ακριβώς στην ενιαία εκπαίδευση), αρχικοποιώντας τα κέντρα και τις ακτίνες στις τιμές που προκύπτουν από το πρώτο στάδιο.

# Παράδειγμα

Εκπαιδευμένο RBF για διάφορες τιμές του M

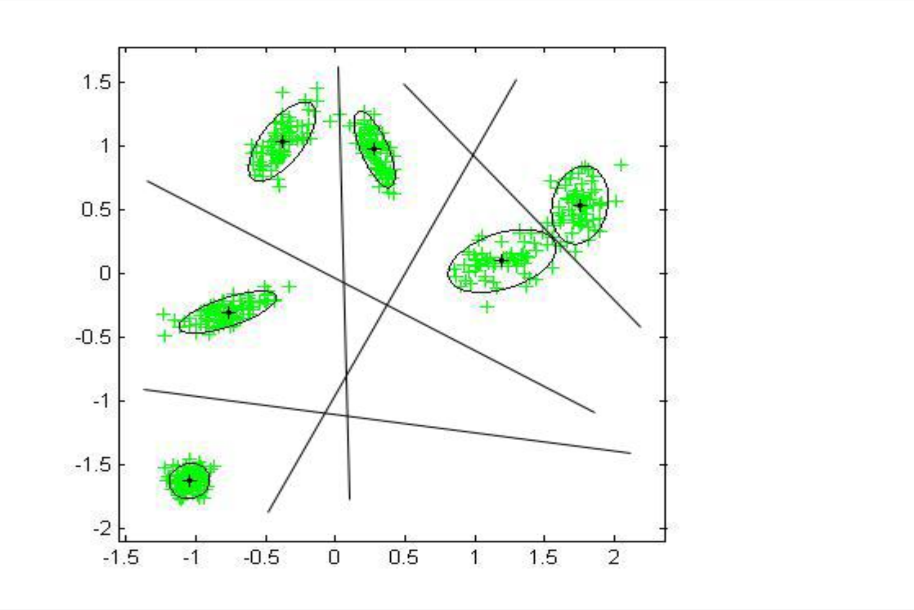


# Σύγκριση MLP και RBF

- Υλοποιούν μη γραμμικές απεικονίσεις από το χώρο των εισόδων $R^d$ στο χώρο των εξόδων $R^p$. Χαρακτηρίζονται από την ιδιότητα της **παγκόσμιας προσέγγισης.**
- **Θεμελιώδης διαφορά: στη μορφή των συναρτήσεων $h_j(x)$ του κρυμμένου επιπέδου**.
- Κάθε κρυμμένος νευρώνας του MLP ορίζει την εξίσωση ενός υπερεπιπέδου: το MLP υλοποιεί συναρτήσεις ταξινόμησης **διαχωρίζοντας το χώρο των προτύπων με υπερεπίπεδα** και ορίζοντας τις **περιοχές απόφασης ως τομές των υπερεπιπέδων**.
- Στον υπολογισμό μιας εξόδου συμμετέχουν όλες οι κρυμμένες μονάδες: **κατανεμημένη αναπαράσταση**, δηλαδή η γνώση σχετικά με την έξοδο που αντιστοιχεί σε κάποια είσοδο κατανέμεται στις τιμές των βαρών όλων των κρυμμένων

# Σύγκριση MLP και RBF

- Τα δίκτυα RBF δημιουργούν **τοπικές αναπαραστάσεις**, δηλαδή η γνώση σχετικά με την έξοδο που αντιστοιχεί σε κάποια είσοδο x ενσωματώνεται στα βάρη εκείνων των κρυμμένων νευρώνων οι οποίες περιλαμβάνουν το πρότυπο x στη σφαίρα επιρροής τους.
- Δίκτυα RBF: **ο χώρος παραδειγμάτων διαιρείται σε σφαιρικές περιοχές**, καθεμιά από τις οποίες αντιστοιχεί σε μια μονάδα του κρυμμένου επίπεδου.



# Σύγκριση MLP και RBF

- Άλλες διαφορές:
  - Στο MLP μπορούμε να έχουμε περισσότερα του ενός κρυμμένα επίπεδα, ενώ το RBF έχει ακριβώς ένα κρυμμένο επίπεδο.
  - Το MLP εκπαιδεύεται με ενιαίο τρόπο (δηλαδή ενημερώνονται όλα τα βάρη με τον ίδιο αλγόριθμο), ενώ στο RBF έχουμε συνήθως εκπαίδευση δύο σταδίων με διαφορετικούς αλγορίθμους σε κάθε στάδιο.