ε.δε.μ^2
ΕΠΙΣΤΗΜΗ ΔΕΔΟΜΕΝΩΝ & ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΕΡΓΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Αλγόριθμοι Πυρήνα (Kernel Methods) - Διαχωρισιμότητα Προτύπων

Θεώρημα του Cover & Εφαρμογές:

1. Radial-Basis Function (RBF) Networks

2. RBF Hybrid Learning

3. Support Vector Machines (SVM)

καθ. Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

www.netmode.ntua.gr

Video Conference μέσω Cisco Webex

Πέμπτη 28/5/2020

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

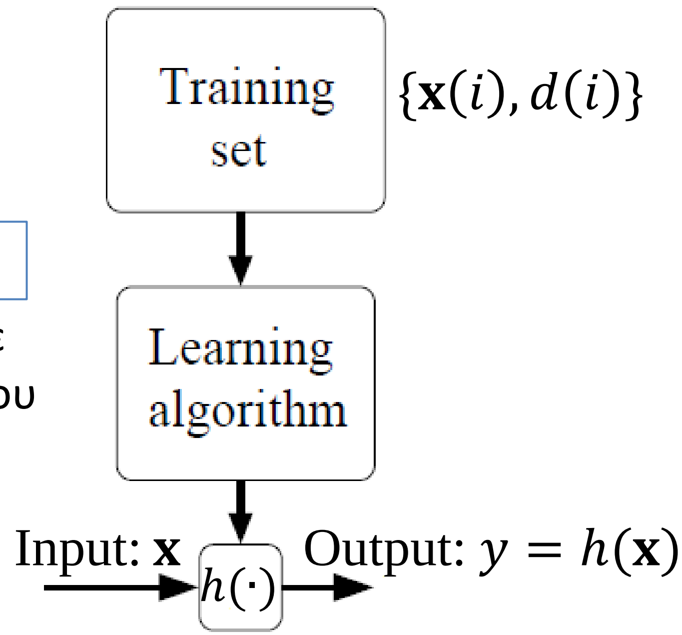
Γενικό Μοντέλο Επιβλεπόμενης Μάθησης - Supervised Learning (*επανάληψη*)

Βασισμένο στο Andrew Ng, “CS229 Lecture Notes”, Stanford University, Fall 2018

- Στόχος του συστήματος είναι η αντιστοίχηση ενός δειγματικού στοιχείου εισόδου (***input sample point***, ***example***) $\mathbf{x} = [x_1\ x_2\ \dots\ x_m]^{\mathrm{T}}$ σε τιμές εξόδου y που εκτιμούν επιθυμητή (***desired***) έξοδο d συμβατή με δεδομένη υπόθεση (π.χ. πρόβλεψη ή ταξινόμηση). Τα στοιχεία x_i είναι αριθμητικές τιμές που κωδικοποιούν m ειδοποιά χαρακτηριστικά (***features***) του $\mathbf{x}$

Ζητείται ο προσδιορισμός της συνάρτησης $y = h(\mathbf{x}) \cong d$

- Η σχεδίαση της $h(\cdot)$ προκύπτει από αλγόριθμο μάθησης, με προσαρμογή της μορφής και των παραμέτρων ενός μοντέλου σε εμπειρικά δεδομένα ενός συνόλου N χαρακτηρισμένων (***labeled***) του δείγματος μάθησης (***Training Set***) ζευγών, $\{\mathbf{x}(i), d(i)\}$, $i = 1,2, \dots, N$ ώστε να προσεγγίζεται ο στόχος της υπόθεσης $d(i) \cong y(i) = h(\mathbf{x}(i))$

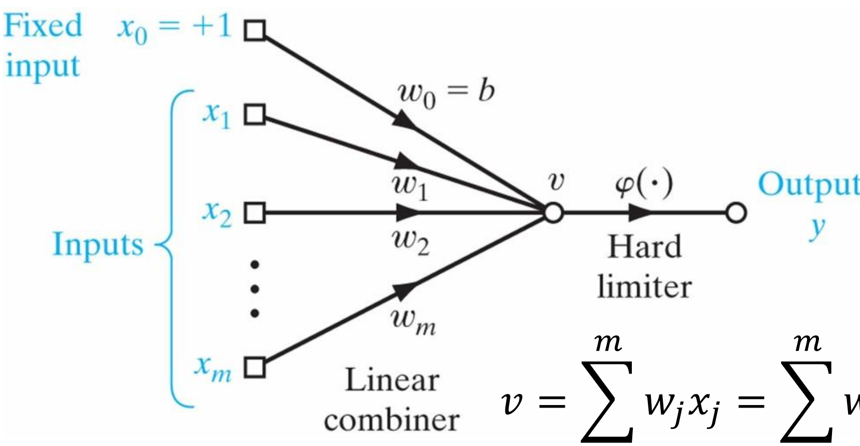


- Αν ο στόχος ικανοποιείται με μικρό αριθμό διακριτών επιλογών της y πρόκειται για πρόβλημα Ταξινόμησης, ***Classification*** (για δύο επιλογές έχουμε δυαδική ταξινόμηση)
- Αν η έξοδος y λαμβάνει συνεχείς τιμές, το πρόβλημα αναφέρεται σαν Παλινδρόμηση, ***Regression***

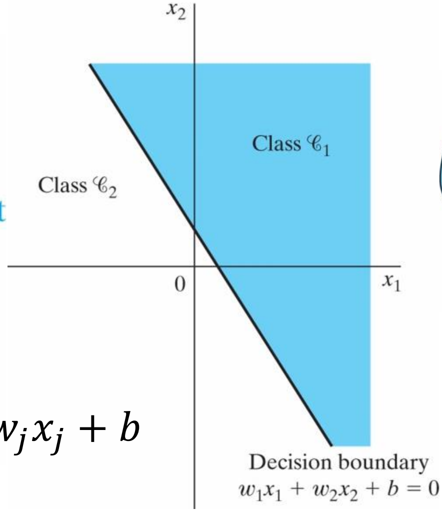
ΣΗΜΕΙΩΣΗ:
Στα επόμενα χρησιμοποιείται εναλλακτικά ο συμβολισμός $\mathbf{x}_i, d_i$ αντί $\mathbf{x}(i), d(i)$ για τα στοιχεία των δειγμάτων μάθησης

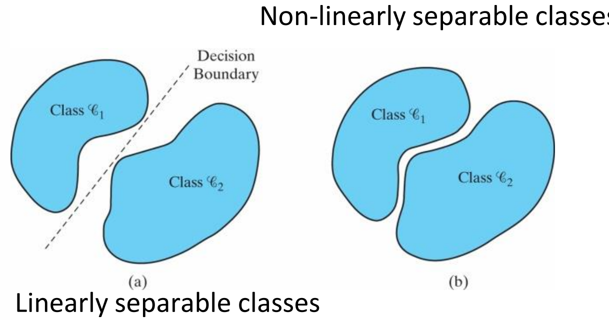
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Rosenblatt's Perceptron (*επανάληψη*)



$$v = \sum_{j=0}^{m} w_j x_j = \sum_{j=1}^{m} w_j x_j + b$$





Σύνοψη:

1. Ένας νευρώνας με γραμμικό induced local field v και συνάρτηση ενεργοποίησης $\varphi(v)$ προσήμου (***Hard Limiter***) για ταξινόμηση δειγματικών στοιχείων $\mathbf{x} = [x_0\ x_1 \ldots x_m]^{\mathrm{T}}$ σε δύο ***γραμμικά διαχωριζόμενες*** κλάσεις : C_1 αν $y = \varphi(v) = 1$, C_2 αν $y = \varphi(v) = -1$
2. Τα βάρη $\mathbf{w} = [w_0\ w_1 \ldots w_m]^{\mathrm{T}}$ ρυθμίζονται on-line (stochastic iterative method) με την διαδοχική εφαρμογή ***Error-correction Algorithm*** σε δεδομένα μάθησης $\{\mathbf{x}(n), d(n)\}, n = 1,2, \ldots, N$ σε περιβάλλον ***supervised learning*** προς ελαχιστοποίηση σφαλμάτων $[d(n) - y(n)]$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \eta[d(n) - y(n)]\mathbf{x}(n)$$

Η παράμετρος $\eta, 0 < \eta \leq 1$ (*learning-rate parameter*) αν είναι μικρή σταθεροποιεί την επαναληπτική διαδικασία, ενώ αν είναι μεγάλη την επιταχύνει π.χ. σε περιβάλλοντα με μεγάλες αποκλίσεις των δεδομένων $\mathbf{x}(n)$

Σε περιβάλλον με δείγματα $\mathbf{x}$ κατανεμημένα κατά Gauss, η ταξινόμησή τους σε δύο κλάσεις C_1 , C_2 μέσω **Bayes Classifier** (ελαχιστοποίηση ρίσκου σφάλματος με βάση a-priori πιθανότητες p_1, p_2) ταυτίζεται με το **Rosenblatt Perceptron**

Διαχωρισιμότητα Προτύπων (Separability of Patterns)

Θεώρημα του Cover (1965)

Ένα περίπλοκο πρόβλημα ταξινόμησης προτύπων μη γραμμικά ορισμένο σε χώρο πολλών διαστάσεων, είναι πιθανότερο να είναι γραμμικά διαχωρίσιμο (linearly separable) από ότι σε χώρο λίγων διαστάσεων, αν δεν υπάρχουν πυκνά σημεία (πρότυπα)

Για την εύκολη ταξινόμηση προτύπων προτιμάται η γραμμική διαχωρισιμότητα μέσω μη γραμμικού μετασχηματισμού συντεταγμένων, και ας απαιτούνται περισσότερες διαστάσεις

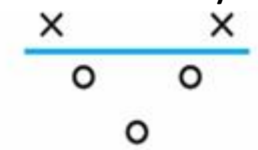
Κρυφές Συναρτήσεις (Hidden Functions)

Για ένα σύνολο $\mathcal{X}$ από N πρότυπα (patterns) $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ με $\mathbf{x}_i$ διανύσματα διαστάσεως m_0 τα οποία ταξινομούνται στις κλάσεις C_1 και C_2. Η δυαδική ταξινόμηση είναι διαχωρίσιμη (separable dichotomy) κατά $m_1 \geq m_0$ *κρυφές* (*μη γραμμικές*) *συναρτήσεις* $\varphi_j(\mathbf{x}) \in \mathbb{R},\ j = 1,2,\dots,m_1$ μετασχηματισμού διανυσμάτων $\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x}) = \left[\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_{m_1}(\mathbf{x})\right]^{\mathrm{T}}$ όταν υπάρχει διάνυσμα $\mathbf{w}$ διαστάσεως m_1 που να ορίζει δύο διακριτές περιοχές αντίστοιχες με τις κλάσεις C_1 και C_2 των $\mathbf{x}$:

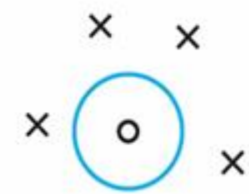
$$w^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}) > 0 \Rightarrow \mathbf{x} \in \mathrm{C}_1 \text{ και } w^{T}\boldsymbol{\varphi}(\mathbf{x}) < 0 \Rightarrow \mathbf{x} \in \mathrm{C}_2$$

φ-separable dichotomies

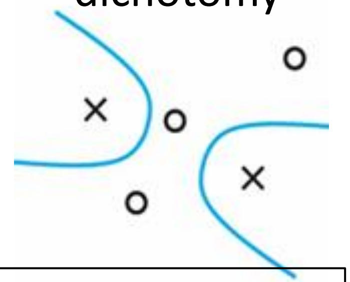
(a) Linearly separable dichotomy



(b) Spherically separable dichotomy



(c) Quadrically separable dichotomy



Συνήθης επιλογή για $\varphi_j(\mathbf{x})$: ***Gaussian Radial-Basis Function*** (***RBF***)

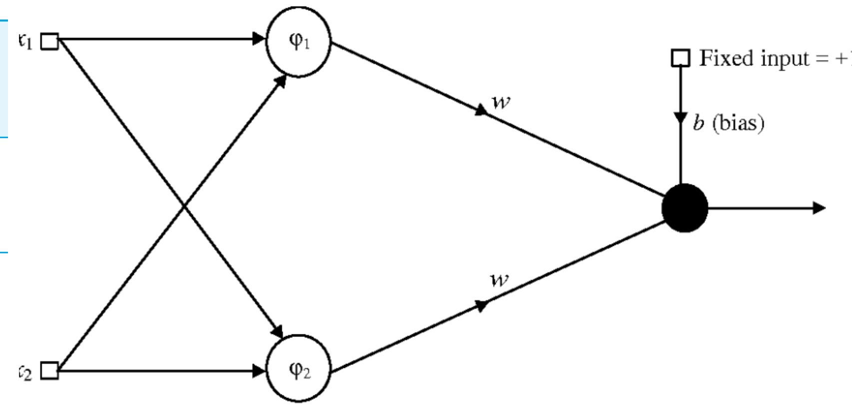
$\varphi_j(\mathbf{x}) = \exp\left(-\left\|\mathbf{x} - \boldsymbol{\mu}_j\right\|^2\right)$ όπου $\boldsymbol{\mu}_j$ διάνυσμα διαστάσεως m_0 των *μέσων τιμών* (κέντρων) της $\varphi_j(\mathbf{x})$ και $\left\|\mathbf{x} - \boldsymbol{\mu}_j\right\|$ η *Ευκλείδεια απόσταση* του σημείου $\mathbf{x}$ από το $\boldsymbol{\mu}_j$

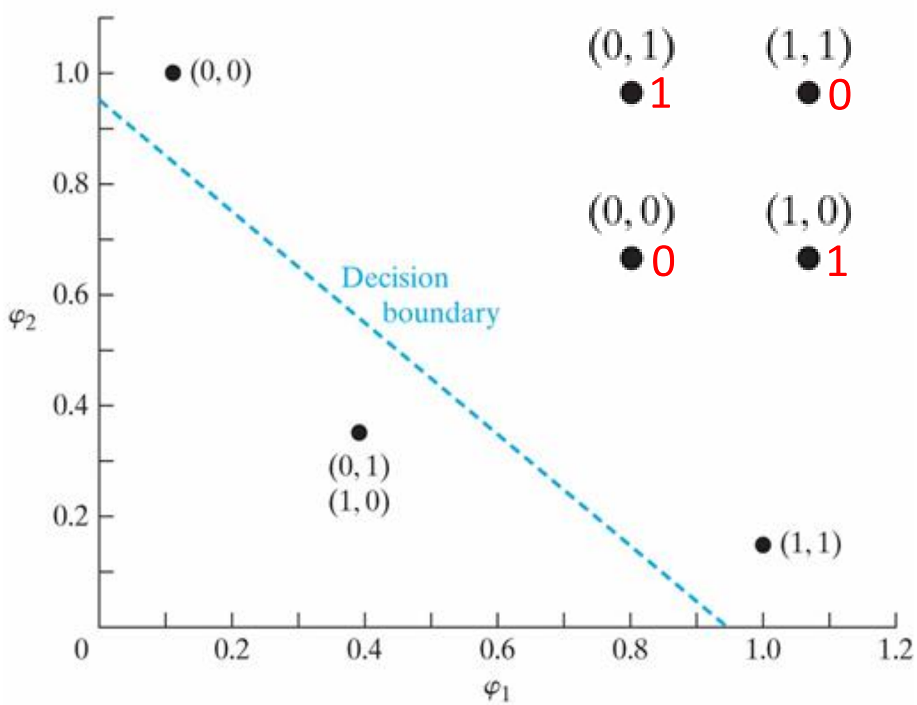
ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Διαχωρισιμότητα Προτύπων – Το πρόβλημα XOR

TABLE 5.1 Specification of the Hidden Functions for the XOR Problem of Example 1

Input Pattern $\mathbf{x}$	First Hidden Function $\varphi_1(\mathbf{x})$	Second Hidden Function $\varphi_2(\mathbf{x})$
(1,1)	1	0.1353
(0,1)	0.3678	0.3678
(0,0)	0.1353	1
(1,0)	0.3678	0.3678





Διαστάσεις: $(\mathbf{x}: m_0 = 2) \rightarrow (\boldsymbol{\varphi}(\mathbf{x}): m_1 = 2)$

$$\varphi_1(\mathbf{x}) = \exp(-\|\mathbf{x} - \boldsymbol{\mu}_1\|^2)$$
$$\varphi_2(\mathbf{x}) = \exp(-\|\mathbf{x} - \boldsymbol{\mu}_2\|^2)$$
$$\boldsymbol{\mu}_1 = [1,1]^{\mathrm{T}}, \boldsymbol{\mu}_2 = [0,0]^{\mathrm{T}}$$

Παράδειγμα: $\mathbf{x} = [1,1]$

$$\varphi_1(1,1) = \exp(-\|[1,1] - [1,1]\|^2) = 1$$
$$\varphi_2(1,1) = \exp(-\|[1,1] - [0,0]\|^2)$$
$$= \exp(-2) = 0.1353$$

Έξοδος: $y = w\varphi_1(\mathbf{x}) + w\varphi_2(\mathbf{x}) + b$

$$w + w \times 0.1353 + b = 0$$
$$w \times 0.3678 + w \times 0.3678 + b = 1$$
$$w \times 0.1353 + w + b = 0$$
$$w \times 0.3678 + w \times 0.3678 + b = 1$$

Αποτέλεσμα: $w = -2.502, \; b = 2.841$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ορισμοί Radial-Basis Function (RBF), Kernels, Hybrid Learning

- Radial- Based Function (RBF): $\mathbf{x} \in \mathbb{R}^{m_0} \rightarrow \varphi_j(\mathbf{x}) = \varphi(\|\mathbf{x} - \mathbf{x}_j\|) = \varphi(r) \in \mathbb{R}$

  $r = \|\mathbf{x} - \mathbf{x}_j\|$ μέτρο ακτινικής απόστασης διανυσματικών σημείων $\mathbf{x}$ και $\mathbf{x}_j$ (συνήθως *Ευκλείδεια*)

  Μετασχηματισμός: $\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x}) = [\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots, \varphi_{m_1}(\mathbf{x})]^{\mathrm{T}}, m_1 \geq m_0$

  Παράδειγμα: *Gaussian RBF* $\varphi_j(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma_j^2}\|\mathbf{x} - \mathbf{x}_j\|^2\right)$ που αφορά στην Ευκλείδεια απόσταση δειγματικού σημείου (προτύπου, *pattern*) $\mathbf{x}$ με m_0 χαρακτηριστικά (*features*) από N δειγματικά σημεία μάθησης (*patterns*) $\mathbf{x}_j, \quad j = 1,2, \dots, N$

  Ταξινόμηση Προτύπων: Τα $\varphi_j(\mathbf{x})$ απεικονίζουν N κρυφά χαρακτηριστικά (*hidden features*) του $\mathbf{x}$ σαν αποστάσεις από τα κέντρα $\mathbf{x}_j$ τα οποία προκύπτουν από το δείγμα μάθησης (*patterns*) σε *1η Φάση Υβριδικής Μάθησης* (***Hybrid Learning***) για ταξινόμηση προτύπων γύρω από κέντρα βάρους μέσω μη επιβλεπόμενης μάθησης (π.χ. με αλγόριθμο *K-Means*)

- Kernel: $k(\mathbf{x}, \mathbf{x}_j) \in \mathbb{R}$ μέτρο ομοιότητας (~εσωτερικό γινόμενο) του διανύσματος $\mathbf{x} \in \mathbb{R}^{m_0}$ με διάνυσμα $\mathbf{x}_j \in \mathbb{R}^{m_0}$

  Σχέση με RBF: $k(\mathbf{x}, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x})^{\mathrm{T}} \boldsymbol{\varphi}(\mathbf{x}_j), j = 1,2, \dots, m_1$ (εσωτερικό γινόμενο). Λόγω του *Θεωρήματος του Cover* επιλέγεται συνήθως $m_1 > m_0$ για μετασχηματισμό $\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x})$ σε *linearly separable* περιοχές ταξινόμησης του προτύπου $\mathbf{x}$

  Ταξινόμηση Προτύπων: Τελική επιλογή κλάσης για δειγματικό σημείο (*pattern*) $\mathbf{x} \in \mathbb{R}^{m_0}$ σε *2η Φάση Υβριδικής Μάθησης* (***Hybrid Learning***) στο πλησιέστερο σημείο από τα κέντρα βάρους της 1ης Φάσης, μέσω επιβλεπόμενης μάθησης και με δίκτυο ***feed-forward***

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Δυαδική Ταξινόμηση - Kernel Perceptron

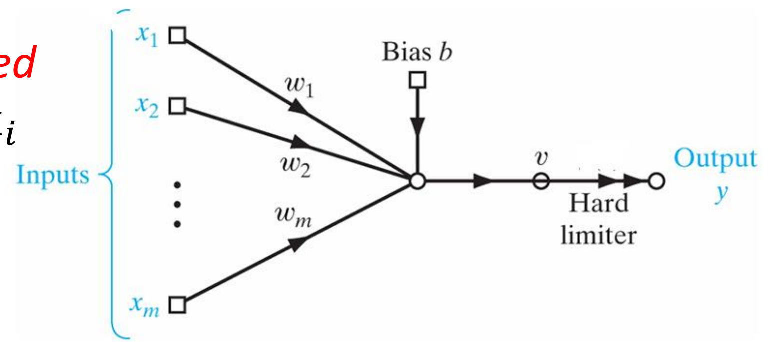
https://en.wikipedia.org/wiki/Kernel_perceptron

Γραμμική Δυαδική Ταξινόμηση, Αλγόριθμος Perceptron

$$y = \mathrm{sgn}(\mathbf{w}^{\mathrm{T}}\mathbf{x}) \in \{-1,1\}$$

Αλγόριθμος Μάθησης: Επαναληπτικός προσδιορισμός συναπτικών βαρών $\mathbf{w}$ με διαδοχική εφαρμογή του *labeled* δείγματος μάθησης $\{\mathbf{x}_i, d_i\}$, $i = 1,2,\ldots,N$ με εισόδους $\mathbf{x}_i$ που οδηγούν σε εξόδους $y_i = \mathrm{sgn}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) \in \{-1,1\}$

$$\mathbf{w} \leftarrow \begin{cases} \mathbf{w} & \text{αν } d_i = y_i \quad \text{(ορθή επιλογή)} \\ \mathbf{w} + d_i\mathbf{x}_i & \text{αν } d_i \neq y_i \quad \text{(λάθος επιλογή)} \end{cases}$$



Μη Γραμμική Δυαδική Ταξινόμηση, Kernel Perceptron

Η *Kernel Machine* αποθηκεύει ένα υποσύνολο από n σημεία $\mathbf{x}_i$ διαστάσεως m_0 του δείγματος μάθησης $\{\mathbf{x}_i, d_i\}$, ορίζει μετρητή α_i για ταξινομήσεις $\mathbf{x}_i \to y_i \in \{-1,1\}$ και επιλέγει δυαδική κλάση y για νέο σημείο $\mathbf{x}$ με βάση τον κανόνα

$$y = \mathrm{sgn}\sum_{i=1}^{n} \alpha_i d_i k(\mathbf{x}, \mathbf{x}_i) \in \{-1,1\}$$

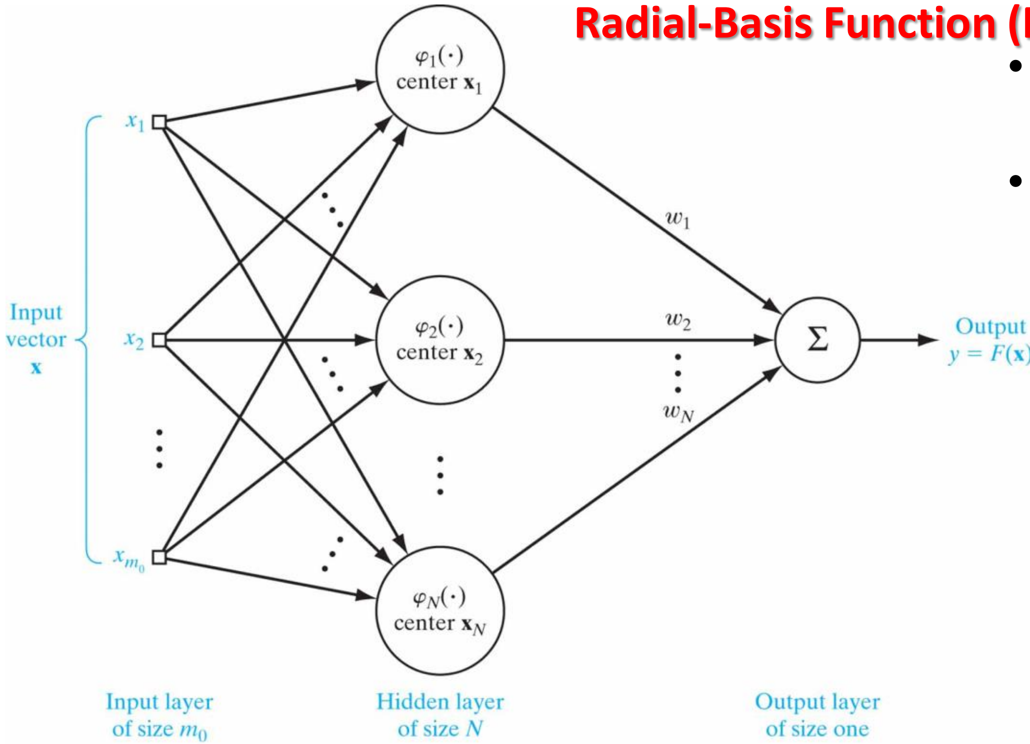
Ο πυρήνας (*Kernel*) $k(\mathbf{x}, \mathbf{x}_i) \in \mathbb{R}$ ορίζεται σαν εσωτερικό γινόμενο διανυσμάτων διαστάσεως $m_1 \geq m_0$ με στοιχεία μη γραμμικές *hidden functions*: $k(\mathbf{x}, \mathbf{x}_i) = \boldsymbol{\varphi}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_i)$

Αλγόριθμος Μάθησης: Ανάλογη του Αλγορίθμου Perceptron με $\mathbf{w} = \sum_{i=1}^{N} \alpha_i\, d_i, \mathbf{x}_i$ όπου α_i μετρητής λανθασμένων επιλογών $\mathbf{x}_i \to y_i \neq d_i$

Για κάθε σημείο $\{\mathbf{x}_i, d_i\}, i = 1,2,\ldots,n$ υπολογίζω $y_i = \mathrm{sgn}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right) = \mathrm{sgn}\sum_{j=1}^{n} \alpha_j d_j k(\mathbf{x}_i, \mathbf{x}_j)$ και αν $y_i \neq d_i$ (λάθος επιλογή) αυξάνεται ο μετρητής $\alpha_i \leftarrow \alpha_i + 1$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Radial-Basis Function (RBF) Networks



- **Input Layer**: Είσοδος διανυσμάτων $\mathbf{x}$ με m_0 χαρακτηριστικά (*features*)
- **Hidden Layer**: Για *Δείγμα Μάθησης* με N στοιχεία (*πρότυπα*), ορίζονται κόμβοι επεξεργασίας *RBF* με *Gaussian* συναρτήσεις βάσης διαστάσεως $N \geq m_0$:

$$\varphi_j(\mathbf{x}) = \varphi(\mathbf{x}, \mathbf{x}_j) = \varphi(\|\mathbf{x} - \mathbf{x}_j\|)$$
$$= \exp(-\frac{1}{2\sigma_j^2}\|\mathbf{x} - \mathbf{x}_j\|^2)$$

(εκτίμηση της απόστασης $\|\mathbf{x} - \mathbf{x}_j\|$ από κέντρα $\mathbf{x}_j$ που προκύπτουν στη διαδικασία μάθησης – συνήθως $\sigma_j = \sigma$)

- **Output Layer**: Γραμμικός συνδυασμός συναρτήσεων βάσης $\boldsymbol{\varphi}(\mathbf{x}) = [\varphi_j(\mathbf{x}), \varphi_j(\mathbf{x}), \ldots \varphi_j(\mathbf{x})]^{\mathrm{T}}$

$$y = F(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}) = \sum_{j=1}^{N} w_j \varphi(\|\mathbf{x} - \mathbf{x}_j\|)$$

Για δυαδική ταξινόμηση $y \in \{-1,1\}$ ή $y \in \{0,1\}$

- **Training**:
  a. Επίλυση γραμμικού συστήματος N εξισώσεων από τα *labeled* στοιχεία $\{\mathbf{x}_i, d_i\}, i = 1,2, \ldots, N$ του δείγματος μάθησης $F(\mathbf{x}_i) = \sum_j w_j \varphi(\|\mathbf{x}_i - \mathbf{x}_j\|) = d_i$ με N αγνώστους w_j
  b. Οι σχέσεις $F(\mathbf{x}_i) = d_i, i = 1,2, \ldots, N$ ορίζουν *υπερ-επιφάνεια δυαδικού διαχωρισμού* κλάσεων για το δείγμα μάθησης. Η *γενίκευση* για σημεία με ανακριβή στοιχεία ή νέα δειγματικά σημεία προκύπτει σαν *interpolation* επί της υπερ-επιφανείας διαχωρισμού
  c. Το σύστημα έχει πάντα λύση για διακριτά σημεία $\mathbf{x}_i$ (*Θεώρημα Michellis*)

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Radial-Basis Function (RBF) Network για XOR

(βασισμένο στη παρουσίαση «**Υβριδική Μάθηση – RBF**», ***Α. Σταφυλοπάτη***, ΣΗΜΜΥ Ε.Μ.Π.
http://mycourses.ntua.gr/courses/ECE1080/document/%C4%E9%E1%EB%DD%EE%E5%E9%F2_2019-2020/rbf.pdf)

$$m_0 = 2, N = 4$$

Συναρτήσεις Βάσης: $\varphi_j(\mathbf{x}) = \exp(-\frac{1}{2\sigma_j^2}\left\|\mathbf{x} - \boldsymbol{\mu}_j\right\|^2, \quad j = 1,2,3,4$

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

$$y = F(\mathbf{x}) = w_1\varphi_1(x) + w_2\varphi_2(x) + w_3\varphi_3(x) + w_4\varphi_4(x)$$

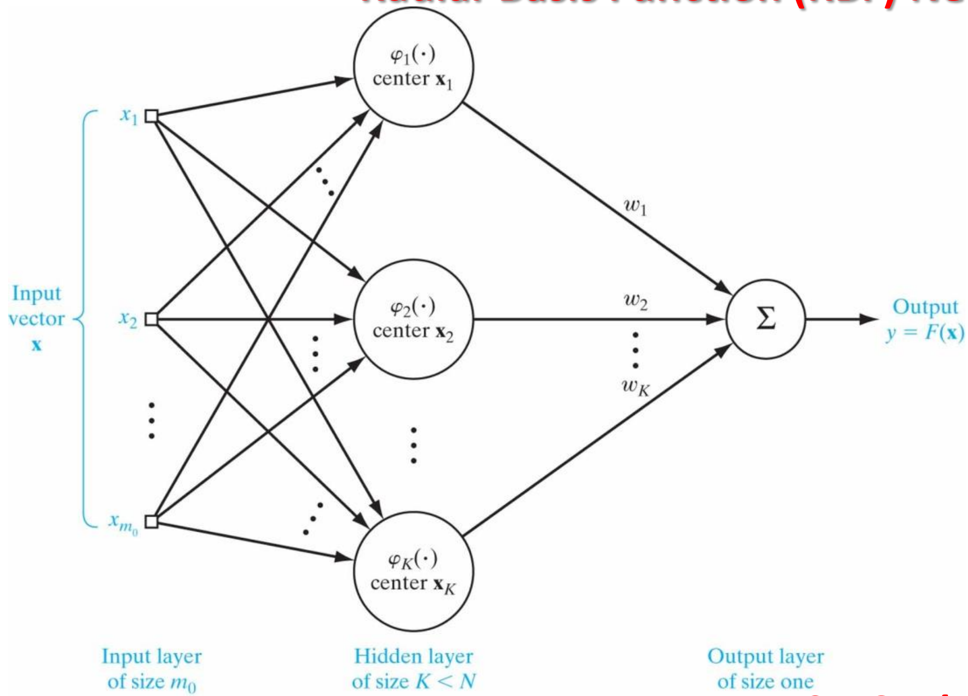
$\mathbf{x}$	$\varphi_1(\mathbf{x})$	$\varphi_2(\mathbf{x})$	$\varphi_3(\mathbf{x})$	$\varphi_4(\mathbf{x})$	y
(1,1)	1	0.1353	0.3678	0.3678	0
(0,0)	0.1353	1	0.3678	0.3678	0
(0,1)	0.3678	0.3678	1	0.1353	1
(1,0)	0.3678	0.3678	0.1353	1	1

Αποτέλεσμα

$$w_1 = w_2 = -0.9843$$
$$w_3 = w_4 = 1.5188$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Radial-Basis Function (RBF) Networks – Πρακτική Υλοποίηση



Η ταξινόμηση με δίκτυα RBF απαιτεί μεγάλο αριθμό κρυφών κόμβων N (ίσων με τον αριθμό στοιχείων μάθησης) και ακριβείς μετρήσεις των $\{\mathbf{x}_i, d_i\}, i = 1,2,\ldots,N$

Προσεγγιστική Υλοποίηση

Μικρότερος αριθμός κρυφών κόμβων $K < N$ που ορίζει χώρο K διαστάσεων:

$$y = F(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}) = \sum_{j=1}^{K} w_j \varphi\left(\|\mathbf{x} - \boldsymbol{\mu}_j\|\right)$$

Υβριδική Μάθηση

Προσδιορισμός των Κέντρων $\mathbf{x}_j$ και Συναπτικών Βαρών w_j, $j = 1,2,\ldots,K$

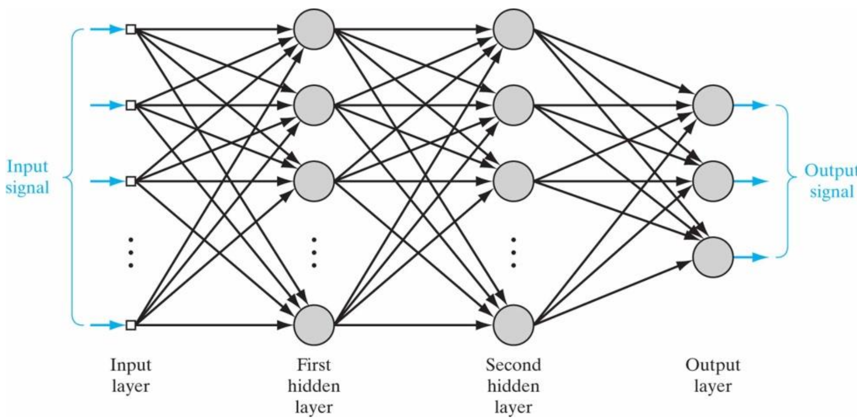
- **Input Layer**: Διάνυσμα $\mathbf{x}$ διαστάσεως m_0 (αριθμός *features*)
- **Hidden Layer**: K κρυφοί κόμβοι $\varphi\left(\|\mathbf{x} - \boldsymbol{\mu}_j\|\right)$ που έχουν κέντρα $\boldsymbol{\mu}_j$. Τα K κέντρα $\boldsymbol{\mu}_j$ προκύπτουν σαν *cluster heads* των $\mathbf{x}$ από αλγόριθμο *μη επιβλεπόμενης μάθησης K-Means Clustering* με Ευκλείδειο μέτρο απόστασης $\|\mathbf{x} - \boldsymbol{\mu}_j\|^2$ (ο K ορίζεται από τον αναλυτή)
- **Output Layer**: Γραμμικός συνδυασμός των K συναρτήσεων βάσης $\varphi_j(\mathbf{x})$:

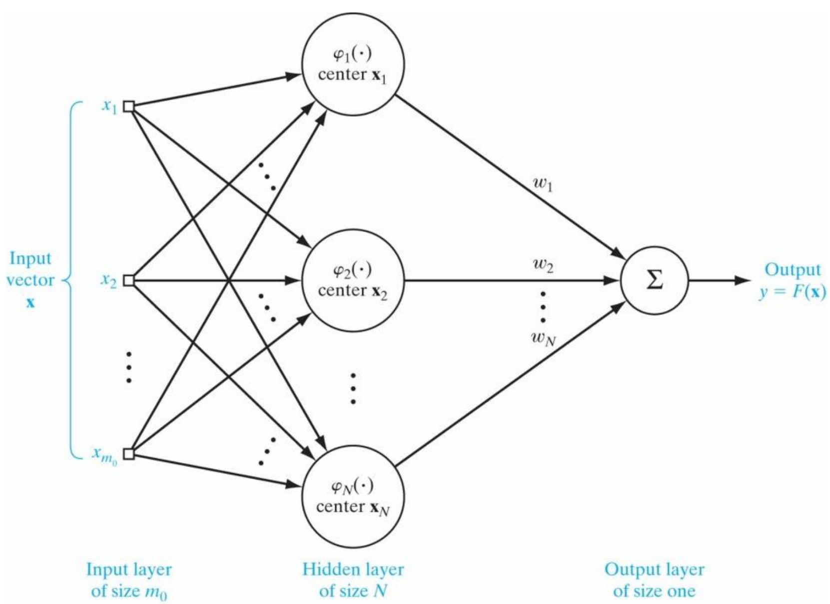
$$y = F(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}) = \sum_{j=1}^{K} w_j \varphi\left(\|\mathbf{x} - \boldsymbol{\mu}_j\|\right)$$

Εκτίμηση των w_j από στοιχεία του δείγματος μάθησης $\{\boldsymbol{\varphi}(\mathbf{x}_i), d_i\}$ με *επιβλεπόμενη μάθηση* κατά προσέγγιση *ελαχίστων τετραγώνων*: N εξισώσεις $d_i = F(\mathbf{x}_i)$, K άγνωστοι w_j $(K < N)$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Multi-layer Perceptron (MLP) vs. RBF





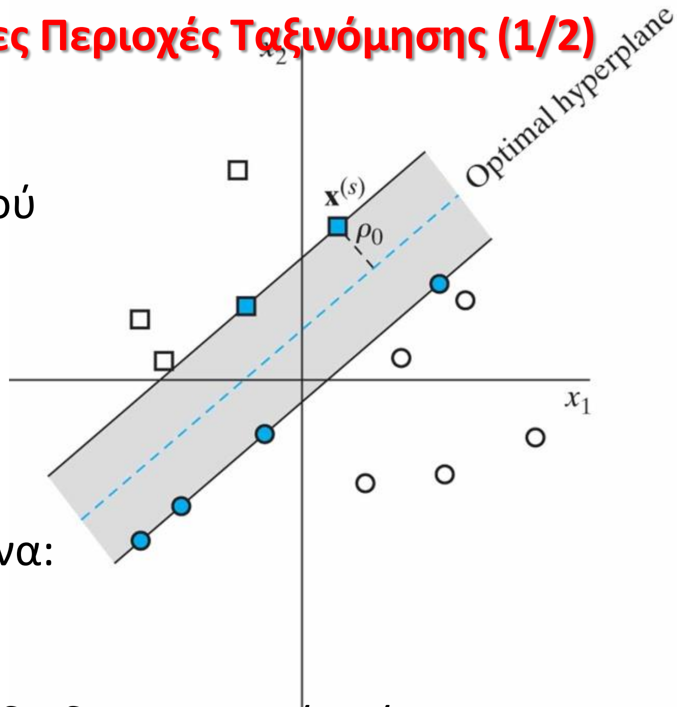
MLP

- Κρυφά επίπεδα νευρώνων
- Επιβλεπόμενη Μάθηση (Supervised Learning)
- Batch ή On-line (Stochastic) Learning
- Back-propagation Algorithm
- Μη γραμμική συνάρτηση ενεργοποίησης
- Βραδεία εκπαίδευση
- Ανοχή σε ανακρίβειες μετρήσεων δειγματικών σημείων

RBF

- Κρυφό επίπεδο νευρώνων
- Υβριδική Μάθηση (Hybrid Learning)
- Μη γραμμικός μετασχηματισμός διανυσματικών σημείων μέσω Radial-Basis Functions (Gaussian)
- Ευελιξία στη διαχωρισιμότητα περιοχών κατάταξης διανυσματικών σημείων (pattern vectors)
- Γρήγορη εκπαίδευση
- Ευαισθησία σε ανακρίβειες μετρήσεων δειγματικών σημείων

Support Vector Machines (SVM) – Γραμμικά Διαχωριζόμενες Περιοχές Ταξινόμησης (1/2)



- Για *labeled* δείγμα μάθησης με στοιχεία $\{\mathbf{x}_i, d_i\}$, $d_i \in \{-1, +1\}, i = 1,2, \dots, N$ η **SVM** ορίζει βέλτιστες περιοχές δυαδικής ταξινόμησης με το *μέγιστο περιθώριο* διαχωρισμού (*margin of separation*) μεταξύ τους
- Για περιπτώσεις γραμμικά διαχωριζόμενων διανυσματικών στοιχείων $\mathbf{x}$ (*patterns*) με m διαστάσεις (*features*) το υπερ-επίπεδο διαχωρισμού ορίζεται από την εξίσωση

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0$$

- Η ταξινόμηση του σημείου μάθησης $\mathbf{x}_i$ ακολουθεί τον κανόνα:

$$\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b \geq 0 \text{ αν } d_i = +1$$
$$\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b < 0 \text{ αν } d_i = -1$$

- Η απόσταση του πλησιέστερου σημείου από το υπερ-επίπεδο διαχωρισμού ορίζει το περιθώριο ρ που πρέπει να μεγιστοποιηθεί για *βέλτιστο διαχωρισμό*: $\mathbf{w}_o^{\mathrm{T}}\mathbf{x} + b_o = 0$
- Γεωμετρικά προκύπτει πως $\rho = \frac{2}{\|\mathbf{w}_o\|}$ όπου $\|\mathbf{w}_o\|$ το Ευκλείδειο μέτρο του διανύσματος $\mathbf{w}_o$
- Για τα στοιχεία του δείγματος μάθησης$\{\mathbf{x}_i, d_i\}$ σε κανονικό υπερ-επίπεδο διαχωρισμού ισχύει:

$$\mathbf{w}_o^{\mathrm{T}}\mathbf{x}_i + b_o \geq 1 \text{ αν } d_i = +1$$
$$\mathbf{w}_o^{\mathrm{T}}\mathbf{x}_i + b_o \leq 1 \text{ αν } d_i = -1$$

- Τα διανύσματα $\mathbf{x}_i$ για τα οποία ισχύει η ισότητα είναι τα ***Support Vectors*** (*Διανύσματα Υποστήριξης*) $\mathbf{x}_i^S$
- Οι ανισότητες ενοποιούνται σαν περιορισμοί (*constraints*) για το δείγμα μάθησης:

$$d_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1, i = 1,2, \dots, N$$

ΣΤΟΧΑΣΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ & ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Support Vector Machines (SVM) – Γραμμικά Διαχωριζόμενες Περιοχές Ταξινόμησης (2/2)

(βασισμένο στη παρουσίαση «**Μηχανές Διανυσμάτων Υποστήριξης**», ***Γ. Στάμου***, ΣΗΜΜΥ Ε.Μ.Π.
http://mycourses.ntua.gr/courses/ECE1078/document/%D5%EB%E9%EA%FC_%C4%E9%E1%EB%DD%EE%E5%F9%ED_2019-2020/NN-SVM-handouts.pdf)

Διατύπωση σαν Πρόβλημα Μη Γραμμικού Προγραμματισμού

Μεγιστοποίηση περιθωρίου διαχωρισμού $\rho = \frac{2}{\|\mathbf{w}_o\|} \Leftrightarrow$ Ελαχιστοποίηση $\|\mathbf{w}_o\|^2 = \mathbf{w}_o^{\mathrm{T}}\mathbf{w}_o$

Πρόβλημα βελτιστοποίησης για προσδιορισμό των παραμέτρων $\mathbf{w}$ και b με περιορισμούς:

$$\min_{\mathbf{w}} \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \quad \text{όταν} \quad d_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1, i = 1,2, \dots, N$$

Η συνάρτηση κόστους είναι άθροισμα τετραγώνων και οι περιορισμοί γραμμικοί. Το βέλτιστο $\mathbf{w}$ μπορεί να προσδιορισθεί με κλασσική μέθοδο μη γραμμικού προγραμματισμού, π.χ. με χρήση ***Lagrange Multipliers*** λ_i για τους περιορισμούς $d_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1$

Ορίζω συνάρτηση κόστους $J(\mathbf{w}, b, \lambda_1, \lambda_2, \dots, \lambda_N) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} - \sum_{i=1}^{N} \lambda_i [\, d_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b)]$. Στο βέλτιστο σημείο ισχύουν οι συνθήκες ***Kuhn-Tucker***:

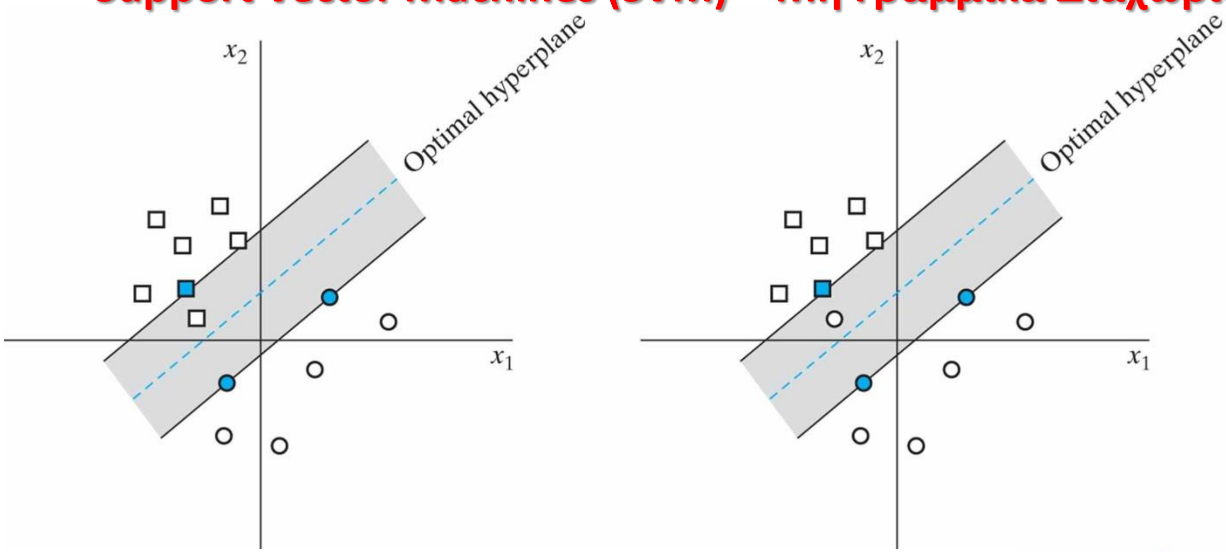
$$\frac{\partial J}{\partial \mathbf{w}} = 0 \;\rightarrow\; \mathbf{w} = \sum_{i=1}^{N} \lambda_i d_i \mathbf{x}_i$$

$$\frac{\partial J}{\partial b} = 0 \;\rightarrow\; \sum_{i=1}^{N} \lambda_i d_i = 0$$

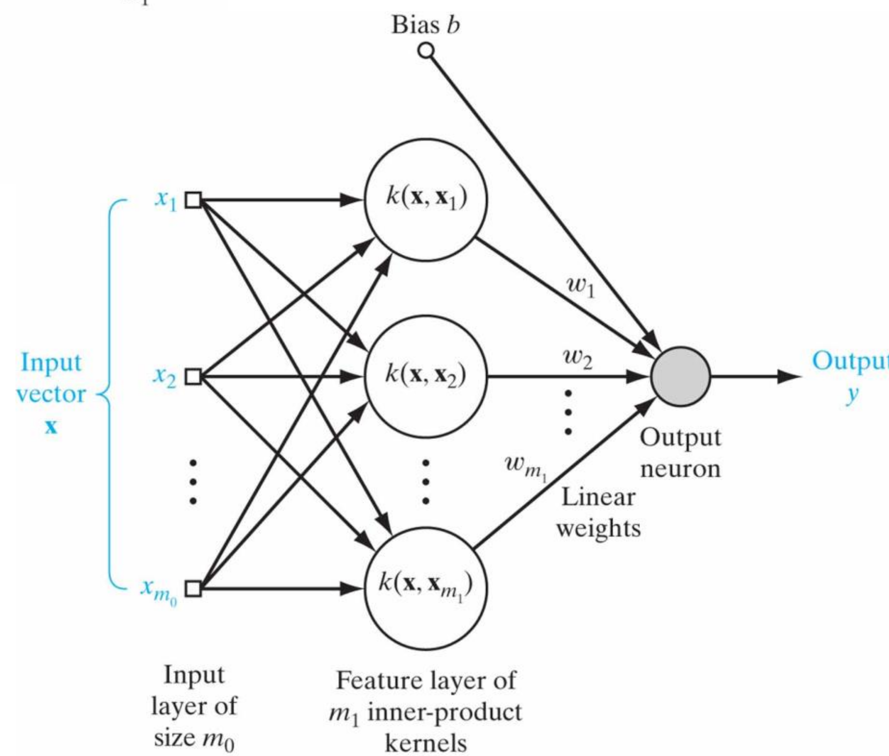
Τα $\mathbf{w}, b$ προσδιορίζουν το βέλτιστο υπερ-επίπεδο διαχωρισμού $\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0$

Τα ***Support Vectors*** $\mathbf{x}_i^S$ αντιστοιχούν σε $\lambda_i > 0$. Τα υπόλοιπα $\mathbf{x}_i$ του δείγματος μάθησης αντιστοιχούν σε $\lambda_i = 0$

Support Vector Machines (SVM) – Μη Γραμμικά Διαχωριζόμενες Περιοχές Ταξινόμησης



Παραβάσεις Γραμμικής Διαχωρισιμότητας:
- $\{\mathbf{x}_i, d_i\}$ εντός της διαχωριστικής ζώνης από την σωστή πλευρά του βέλτιστου υπερ-επιπέδου
- $\{\mathbf{x}_i, d_i\}$ εντός της διαχωριστικής ζώνης από την λάθος πλευρά του βέλτιστου υπερ-επιπέδου



Αρχιτεκτονική SVM με χρήση Δικτύου RBF

Χρήση μεγάλου αριθμού *hidden features* m_1 (π.χ. ίση με τον αριθμό στοιχείων του δείγματος μάθησης N) που μετασχηματίζουν μη γραμμικά διαχωρίσιμες περιοχές των διανυσμάτων εισόδου $\mathbf{x}$ διαστάσεως $m_0 \ll m_1$ σε γραμμικά διαχωρίσιμες περιοχές