

❖ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ❖ (PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης
Κωνσταντίνος Κουτρούμπας

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

ΜΙΑ ΣΥΜΒΑΣΗ: Προκειμένου να καταστήσουμε πιο συμπαγή το συμβολισμό

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

ορίζουμε

$$\mathbf{w}^* = [\mathbf{w}^T, w_0]^T \text{ and } \mathbf{x}^* = [\mathbf{x}^T, 1]^T.$$

Έτσι έχουμε

$$(H) : \mathbf{w}^{*T} \mathbf{x}^* = 0$$

ΣΗΜΕΙΩΣΗ: Στη συνέχεια, εκτός αν ορίζεται διαφορετικά, θα γράφουμε \mathbf{w} και \mathbf{x} , εννοώντας τα \mathbf{w}^* και \mathbf{x}^* , αντίστοιχα.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Υπόθεση: (Τα διανύσματα των) δύο κλάσεων ω_1 και ω_2 είναι γραμμικώς διαχωρίσιμα.

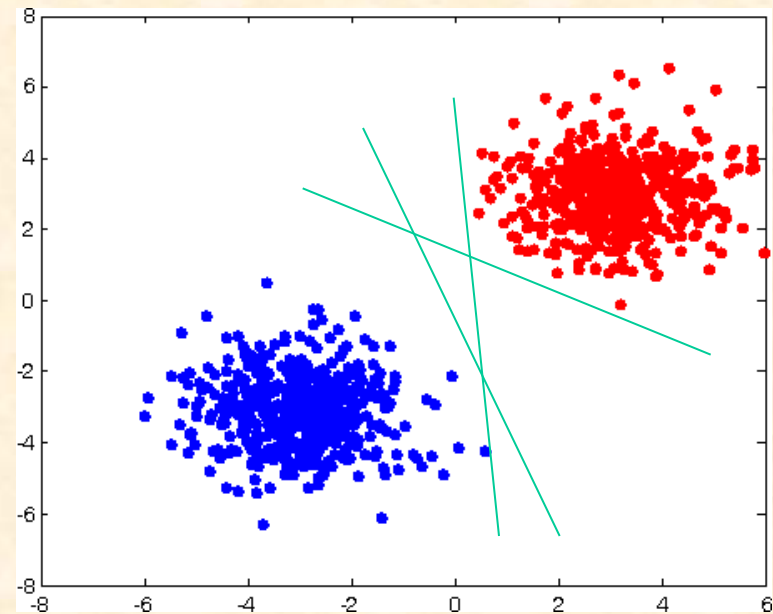
Έτσι, υπάρχει ένα υπερεπίπεδο (H^*): $\mathbf{w}^{*T}\mathbf{x}=0$, ώστε

$$\mathbf{w}^{*T} \mathbf{x} > 0, \quad \forall \mathbf{x} \in \omega_1$$

$$\mathbf{w}^{*T} \mathbf{x} < 0, \quad \forall \mathbf{x} \in \omega_2$$

Το κριτήριο στο perceptron: Προσδιόρισε ένα υπερεπίπεδο το οποίο διαχωρίζει τέλεια τα διανύσματα από τις δύο κλάσεις.

ΣΗΜΕΙΩΣΗ: Αν οι κλάσεις είναι γραμμικώς διαχωρίσιμες, υπάρχουν άπειρα υπερεπίπεδα που ικανοποιούν το κριτήριο perceptron.



ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Η συνάρτηση κόστους:

$$J(\underline{w}) = \sum_{x \in Y(\underline{w})} (\delta_x \underline{w}^T x)$$

όπου

- $Y(\underline{w})$ το σύνολο των **λανθασμένα ταξινομημένων** διανυσμάτων από το \underline{w} και

$$\begin{aligned} \delta_x &= -1 \text{ if } x \in \omega_1 \\ \delta_x &= +1 \text{ if } x \in \omega_2 \end{aligned}$$

Σημειώσεις:

- Είναι $J(\underline{w}) \geq 0$.

Πράγματι, αν $x \in Y(\underline{w})$ και $x \in \omega_1$ τότε $\underline{w}^T x < 0$ και $\delta_x = -1$. Επομένως $\delta_x \underline{w}^T x > 0$.

Επίσης, αν $x \in Y(\underline{w})$ και $x \in \omega_2$ τότε $\underline{w}^T x > 0$ και $\delta_x = +1$. Επομένως $\delta_x \underline{w}^T x > 0$.

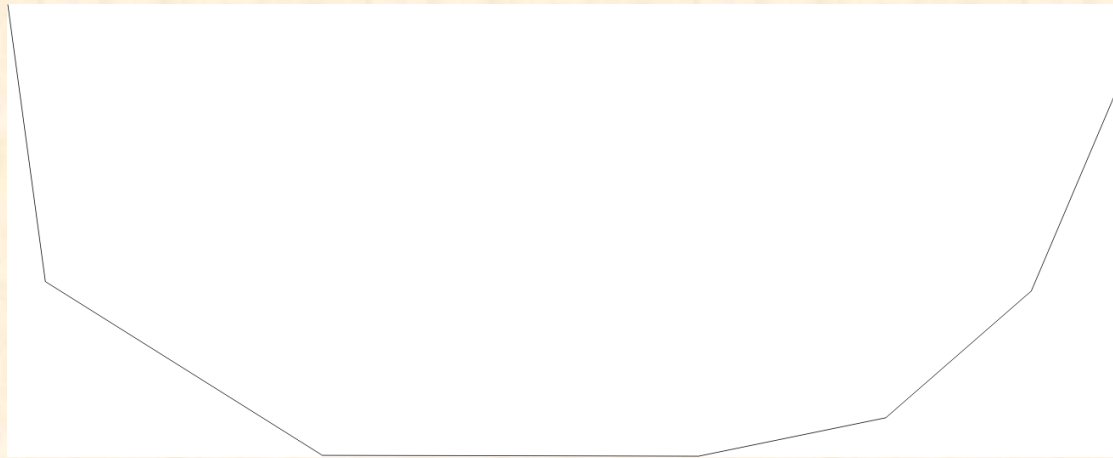
- Είναι $J(\underline{w}) = 0$ μόνον όταν $Y = \emptyset$, πράγμα που σημαίνει ότι προσδιορίστηκε **μία λύση** που ικανοποιεί το κριτήριο perceptron.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Η συνάρτηση κόστους:

Η $J(w)$ είναι τμηματικά γραμμική (piecewise linear) (γιατί;)



Η **ελαχιστοποίηση** της $J(w)$ επιτυγχάνεται μέσω διαδικασίας που ακολουθεί τη φιλοσοφία της **GD**.

Όπου είναι έγκυρο, ορίζουμε
$$\frac{\partial J(w)}{\partial w} = \frac{\partial}{\partial w} \left(\sum_{x \in Y(w)} \delta_x w^T x \right) = \sum_{x \in Y(w)} \delta_x x$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

- Αρχικοποίησε $w=w(0)$

- $t=0$

- Επανάλαβε

- Προσδιόρισε τα **λανθασμένα ταξινομημένα διανύσματα** του X από το $w(t)$ και συσσωρεύσέ τα στο $Y(w_t)$.

$$w(t+1) = w(t) - \rho_t \sum_{x \in Y(w_t)} \delta_x x$$

- $t=t+1$

- Έως ότου επιτευχθεί σύγκλιση

Σημειώσεις:

• Πρόκειται για αλγόριθμο επεξεργασίας **κατά δέσμες** (batch algorithm), όπου η ενημέρωση των παραμέτρων σε κάθε επανάληψη πραγματοποιείται μετά την επεξεργασία όλων των διανυσμάτων του X .

• Ο αλγόριθμος perceptron **συγκλίνει μόνον όταν** οι κλάσεις είναι **γραμμικώς διαχωρίσιμες**. Διαφορετικά, αποτυγχάνει να συγκλίνει.

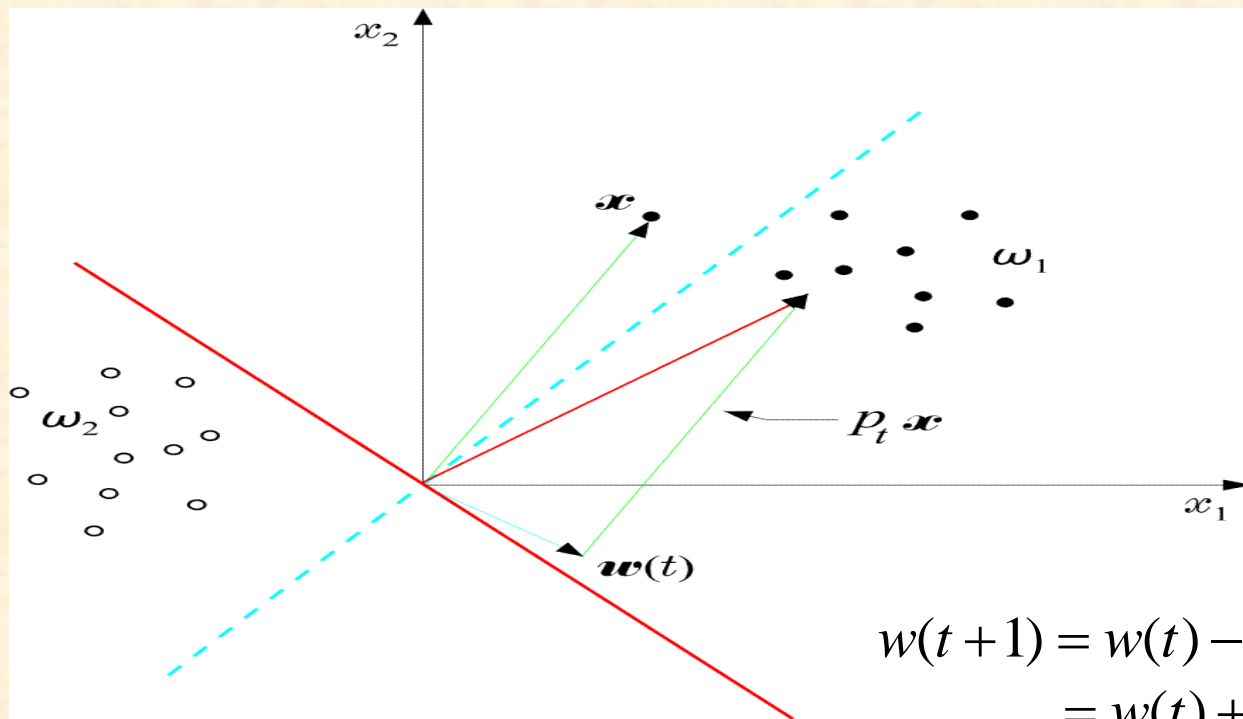
• Αν οι κλάσεις είναι **γραμμικώς διαχωρίσιμες**, ο αλγόριθμος **συγκλίνει σε πεπερασμένο αριθμό βημάτων**

$$\lim_{t \rightarrow \infty} \sum_{k=0}^t \rho_k \rightarrow \infty, \quad \lim_{t \rightarrow \infty} \sum_{k=0}^t \rho_k^2 < +\infty \text{ π.χ. } \rho_t = c/t$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Παράδειγμα 1:



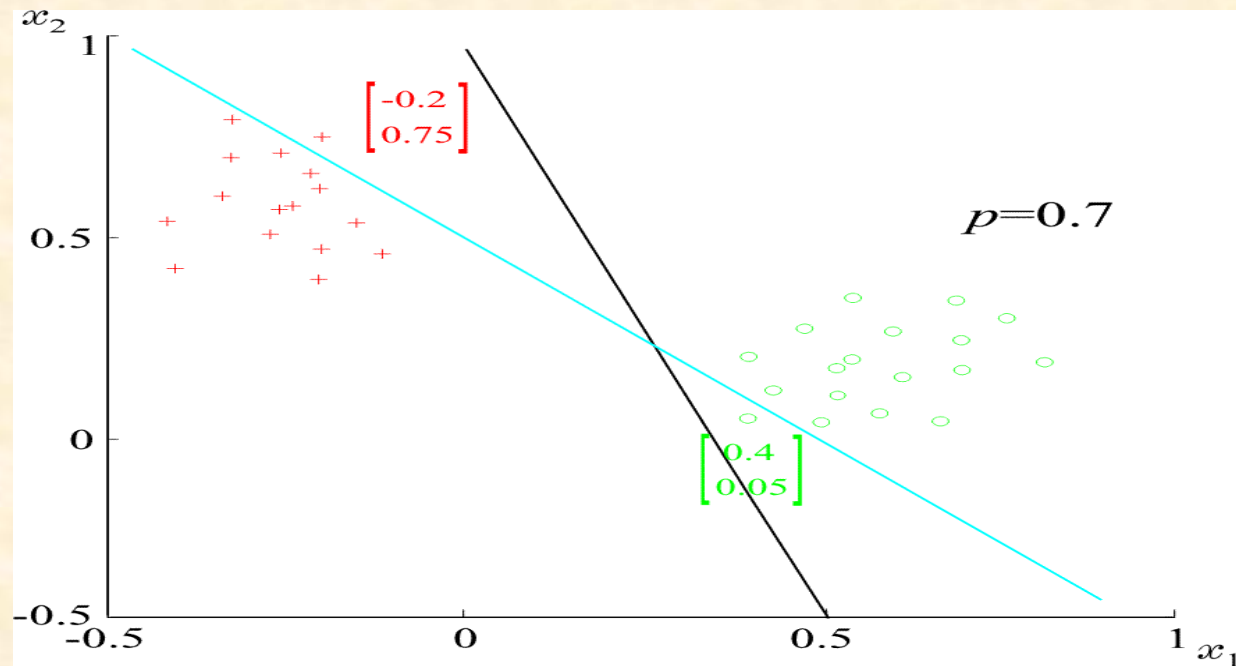
ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Ο αλγόριθμος perceptron

Παράδειγμα 2: Σε κάποια επανάληψη t ο αλγόριθμος perceptron δίνει

$w_1 = 1$, $w_2 = 1$, $w_0 = -0.5$ που αντιστοιχεί στην ευθεία $x_1 + x_2 - 0.5 = 0$

Το υπερέπιπεδο της επόμενης επανάληψης είναι



$$\underline{w}(t+1) = \begin{bmatrix} 1 \\ 1 \\ -0.5 \end{bmatrix} - 0.7(-1) \begin{bmatrix} 0.4 \\ 0.05 \\ 1 \end{bmatrix} - 0.7(+1) \begin{bmatrix} -0.2 \\ 0.75 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.42 \\ 0.51 \\ -0.5 \end{bmatrix}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Παραλλαγές του αλγόριθμου perceptron:

-Ο αλγόριθμος “τσέπης” (rocket algorithm).

- Κρατά την καλύτερη λύση που βρέθηκε κατά την εκτέλεση του αλγόριθμου perceptron για συγκεκριμένο αριθμό επαναλήψεων.
- Κατάλληλος για προβλήματα μη γραμμικώς διαχωρίσιμων κλάσεων.

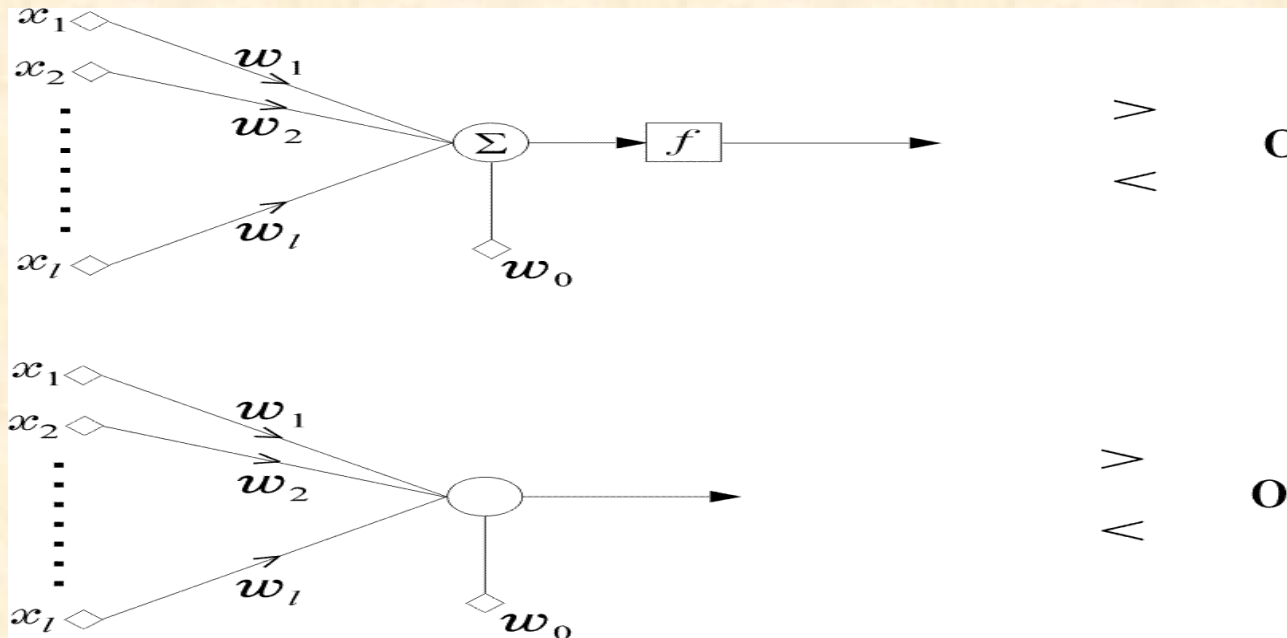
-Ο on-line αλγόριθμος perceptron:

- Η ενημέρωση των παραμέτρων λαμβάνει χώρα μετά την μεμονωμένη επεξεργασία κάθε διανύσματος του X .

$$\begin{aligned} w(t+1) &= w(t) + \rho x_{(t)}, & w^T(t)x_{(t)} &\leq 0 \\ & & x_{(t)} &\in \omega_1 \\ w(t+1) &= w(t) - \rho x_{(t)}, & w^T(t)x_{(t)} &\geq 0 \\ & & x_{(t)} &\in \omega_2 \\ w(t+1) &= w(t) & \text{otherwise} & \end{aligned}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Η δομή perceptron



w_i 's synapses or synaptic weights
 w_0 threshold

- Η δομή αυτή καλείται **perceptron** ή **νευρώνας (neuron)**
- Είναι μια **μηχανή** που μπορεί να **“μάθει”** (δηλ. να προσδιορίσει τις τιμές των εμπλεκόμενων παραμέτρων) από τα **διανύσματα εκπαίδευσης** μέσω του **αλγόριθμου perceptron**.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων (Least square methods)

- Αν οι κλάσεις είναι γραμμικώς διαχωρίσιμες, το perceptron θα δώσει σαν έξοδο ± 1
- Αν οι κλάσεις ΔΕΝ είναι γραμμικώς διαχωρίσιμες, θα υπολογίσουμε τα βάρη (παραμέτρους) w_1, w_2, \dots, w_0

έτσι ώστε η **διαφορά** ανάμεσα

- Στην **πραγματική απόκριση** του **ταξινομητή**, $w^T x$, και
- Την αντίστοιχη **επιθυμητή απόκριση**, δηλ.
$$\begin{cases} +1 & \text{if } x \in \omega_1 \\ -1 & \text{if } x \in \omega_2 \end{cases}$$

να είναι όσο το δυνατόν **ΜΙΚΡΟΤΕΡΗ** για όλα τα διανύσματα του X .

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το MSE

– **ΜΙΚΡΟΤΕΡΑ**, ως προς το κριτήριο του μέσου τετραγωνικού σφάλματος (**mean square error - MSE**), σημαίνει επιλογή του w ώστε η συνάρτηση κόστους

- $J(w) \equiv E[(d - w^T x)^2]$ να ελαχιστοποιηθεί
 $\hat{w} = \arg \min_w J(w)$

όπου d οι αντίστοιχες επιθυμητές αποκρίσεις

Ελαχιστοποιώντας την $J(w)$ ως προς το w έχουμε:

$$\frac{\partial J(w)}{\partial w} = \frac{\partial}{\partial w} E[(d - w^T x)^2] = 0$$
$$= 2E[x(d - x^T w)] \Rightarrow$$

R_x : Πίν. αυτοσυσχέτισης - **autocorrelation matrix**
 $E[xd]$: Διάν. Ετεροσυσχέτ.-**cross-correlation vector**

$$E[xx^T]w = E[xd] \Rightarrow \hat{w} = R_x^{-1}E[xd]$$

$$E[xd] = \begin{bmatrix} E[x_1 d] \\ \dots \\ E[x_l d] \end{bmatrix}$$

$$R_x \equiv E[xx^T] = \begin{bmatrix} E[x_1 x_1] & E[x_1 x_2] \dots & E[x_1 x_l] \\ \dots & \dots & \dots \\ E[x_l x_1] & E[x_l x_2] \dots & E[x_l x_l] \end{bmatrix}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το MSE: Περ. από 2 κλάσεις

Στην περίπτωση που το υπό μελέτη πρόβλημα περιλαμβάνει M κλάσεις, λύνουμε M προβλήματα σαν το παραπάνω. Συγκεκριμένα:

-Για την j -στή κλάση

- Θέσε $d_i=1$, αν $\mathbf{x}_i \in \omega_j$ και $d_i=0$, διαφορετικά.

- Λύσε το πρόβλημα δύο κλάσεων που προκύπτει και έστω \mathbf{w}^j το αντίστοιχο διάνυσμα παραμέτρων.

Μετά τον προσδιορισμό των \mathbf{w}^j 's:

-Για δεδομένο \mathbf{x}_i :

- Υπολόγισε τις ποσότητες $g^j(\mathbf{x}_i) = \mathbf{w}^{jT} \mathbf{x}_i$, $j=1, \dots, M$.

- Καταχώρησε το \mathbf{x}_i στην κλάση ω_q για την οποία $g^q(\mathbf{x}_i) = \max_{j=1, \dots, M} g^j(\mathbf{x}_i)$

Σημείωση: Το κριτήριο MSE ανήκει σε μια γενικότερη κλάση συναρτήσεων κόστους με την ακόλουθη σημαντική ιδιότητα:

Η τιμή $g^j(\mathbf{x}_i)$ είναι μια εκτίμηση, ως προς το κριτήριο MSE, της εκ των υστέρων πιθανότητας, $P(\omega_j | \mathbf{x})$, υπό την προϋπόθεση ότι οι επιθυμητές αποκρίσεις που χρησιμοποιούνται κατά την εκπαίδευση είναι $d_i=1$, αν $\mathbf{x}_i \in \omega_j$ και $d_i=0$, διαφορετικά.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

- **ΜΙΚΡΟΤΕΡΟ**, ως προς το κριτήριο του **αθροίσματος των τετραγώνων των σφαλμάτων**, σημαίνει επιλογή του \mathcal{W} που **ελαχιστοποιεί** τη συνάρτηση κόστους

$$\bullet J(\mathbf{w}) = \sum_{i=1}^N (d_i - \mathbf{w}^T \mathbf{x}_i)^2$$

όπου $d_i = +1$, αν $\mathbf{x}_i \in \omega_1$ και $d_i = -1$, αν $\mathbf{x}_i \in \omega_2$.

Minimizing $J(\mathbf{w})$ with respect to \mathbf{w} results in:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \sum_{i=1}^N (d_i - \mathbf{w}^T \mathbf{x}_i)^2 = 0 \Rightarrow$$

$$\left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T \right) \mathbf{w} = \sum_{i=1}^N \mathbf{x}_i d_i$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

Ένας εναλλακτικός τρόπος διατύπωσης:

Ορίζουμε τα ακόλουθα

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \dots \\ x_N^T \end{bmatrix} \text{ } N \times l \text{ πίνακας} \quad \underline{d} = \begin{bmatrix} d_1 \\ \dots \\ d_N \end{bmatrix} \text{ Αντίστοιχες επιθυμητές αποκρίσεις}$$

$$X^T = [x_1, x_2, \dots, x_N] \text{ } l \times N \text{ πίνακας}$$

$$X^T X = \sum_{i=1}^N x_i x_i^T$$

$$X^T d = \sum_{i=1}^N x_i d_i$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

Ένας εναλλακτικός τρόπος διατύπωσης:

Τότε

$$\left(\sum_{i=1}^N x_i^T x_i\right) \hat{w} = \left(\sum_{i=1}^N x_i d_i\right) \Leftrightarrow (X^T X) \hat{w} = X^T d \Rightarrow \hat{w} = (X^T X)^{-1} X^T d = X^\# d$$

$$X^\# \equiv (X^T X)^{-1} X^T \quad \text{Ψευδοαντίστροφος (pseudoinverse) του } X$$

Έστω $N=I$. Τότε ο X είναι τετραγωνικός και (γενικά) αντιστρέψιμος. Τότε έχουμε

$$(X^T X)^{-1} X^T = X^{-1} X^{-T} X^T = X^{-1} \Rightarrow X^\# = X^{-1}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

Ένας εναλλακτικός τρόπος διατύπωσης:

Έστω $N > l$. Τότε, γενικά, **δεν υπάρχει λύση** που να ικανοποιεί ταυτόχρονα όλους τους περιορισμούς

$$Xw = d : \begin{array}{l} x_1^T w = d_1 \\ x_2^T w = d_2 \\ \dots \\ x_N^T w = d_N \end{array} \quad N \text{ equations} > l \text{ unknowns}$$

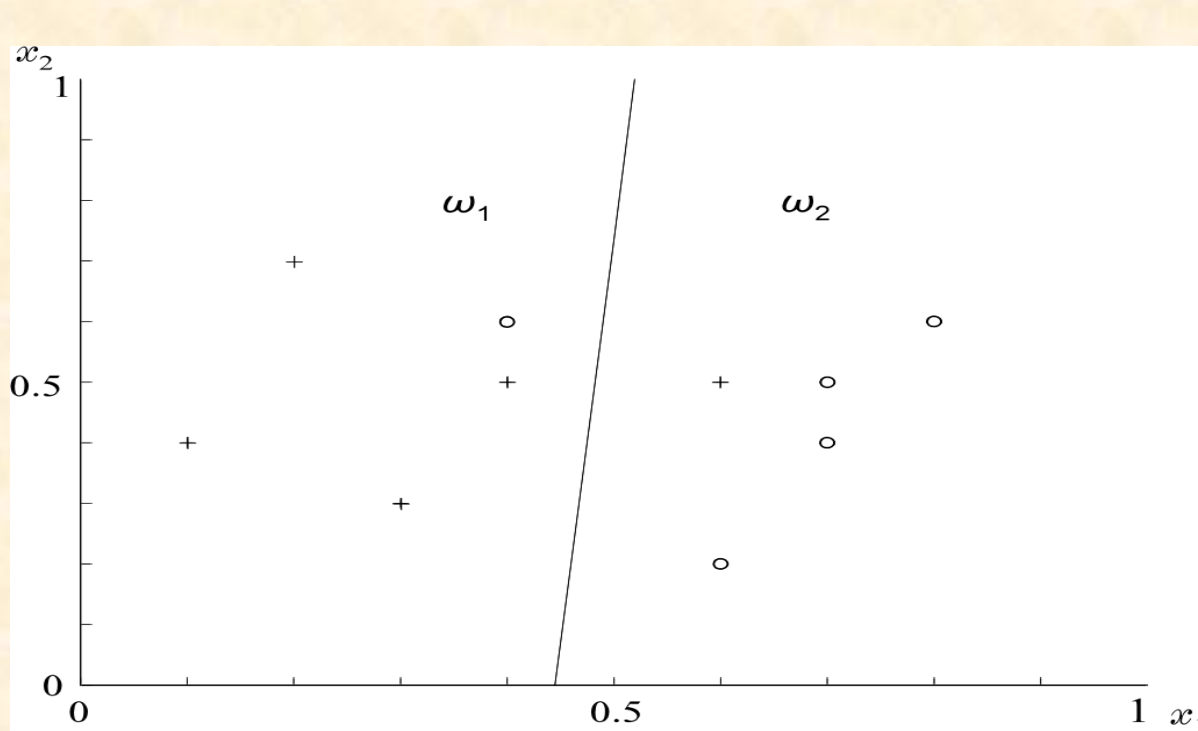
Σ' αυτή την περίπτωση, η **“λύση”** $w = X^\dagger d$ ελαχιστοποιεί το **άθροισμα των τετραγώνων των σφαλμάτων**.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

Ένα παράδειγμα:

$$\omega_1 : \begin{bmatrix} 0.4 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.6 \\ 0.5 \end{bmatrix}, \begin{bmatrix} 0.1 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 0.2 \\ 0.7 \end{bmatrix}, \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix} \quad \omega_2 : \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}, \begin{bmatrix} 0.6 \\ 0.2 \end{bmatrix}, \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}, \begin{bmatrix} 0.7 \\ 0.5 \end{bmatrix}$$



$$X = \begin{bmatrix} 0.4 & 0.5 & 1 \\ 0.6 & 0.5 & 1 \\ 0.1 & 0.4 & 1 \\ 0.2 & 0.7 & 1 \\ 0.3 & 0.3 & 1 \\ 0.4 & 0.6 & 1 \\ 0.6 & 0.2 & 1 \\ 0.7 & 0.4 & 1 \\ 0.8 & 0.6 & 1 \\ 0.7 & 0.5 & 1 \end{bmatrix} \quad \underline{d} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 2.8 & 2.24 & 4.8 \\ 2.24 & 2.41 & 4.7 \\ 4.8 & 4.7 & 10 \end{bmatrix}, \quad X^T \underline{d} = \begin{bmatrix} -1.6 \\ 0.1 \\ 0.0 \end{bmatrix} \quad \underline{w} = (X^T X)^{-1} X^T \underline{d} = \begin{bmatrix} -3.13 \\ 0.24 \\ 1.34 \end{bmatrix}$$

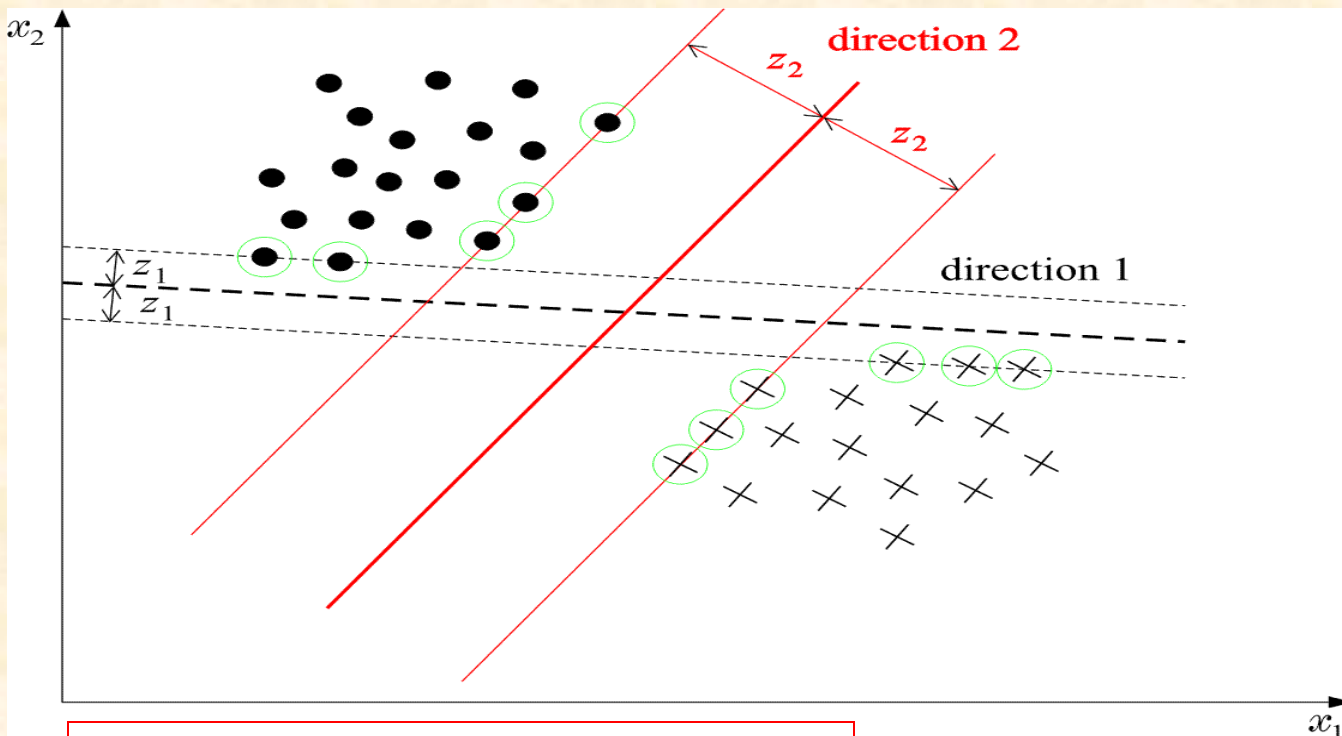
ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

Ο στόχος: Δοθέντων δύο **γραμμικώς διαχωρίσιμων** κλάσεων, προσδιόρισε τον ταξινομητή

$$g(x) = w^T x + w_0 = 0$$

που αφήνει το **μέγιστο περιθώριο** από τις δύο κλάσεις.



Support vector machines:
Μηχανές διανυσματικής στήριξης

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

- Περιθώριο (Margin): Κάθε υπερεπίπεδο χαρακτηρίζεται από
 - Τον **προσανατολισμό** του στο χώρο, δηλ. το διάνυσμα w
 - Τη **θέση** του χώρο, i.e., w_0
- Για **ΚΑΘΕ** προσανατολισμό, w , επέλεξε το υπερεπίπεδο που **αφήνει την ΙΔΙΑ απόσταση** από τα **εγγύτερα** σημεία από κάθε κλάση. Το περιθώριο είναι το διπλάσιο αυτής της απόστασης.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

– Η απόσταση ενός σημείου \hat{x} από ένα υπερεπίπεδο είναι

$$z_{\hat{x}} = \frac{g(\hat{x})}{\|w\|}$$

– Διαβάθμισε τα w , w_0 , ώστε για τα **εγγύτερα σημεία** από κάθε κλάση να είναι:

$$|g(x)| = 1 \quad \{g(x) = +1 \text{ for } \omega_1 \text{ and } g(x) = -1 \text{ for } \omega_2\}$$

– Τότε το **περιθώριο** είναι

$$\frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|}$$

– Επίσης, ισχύουν και οι ακόλουθοι περιορισμοί

$$\begin{aligned} w^T x + w_0 &\geq 1 \quad \forall x \in \omega_1 \\ w^T x + w_0 &\leq -1 \quad \forall x \in \omega_2 \end{aligned}$$

(δηλ. όλα τα στοιχεία της κλάσης +1 (-1) βρίσκονται στη θετική (αρνητική) πλευρά του υπερεπιπέδου)

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

– SVM (γραμμικός) ταξινομητής

$$g(\underline{x}) = \underline{w}^T \underline{x} + w_0$$

– (Πρόβλημα SVM) Ελαχιστοποίησε την

$$J(\underline{w}) = \frac{1}{2} \|\underline{w}\|^2$$

– Υπό τους περιορισμούς

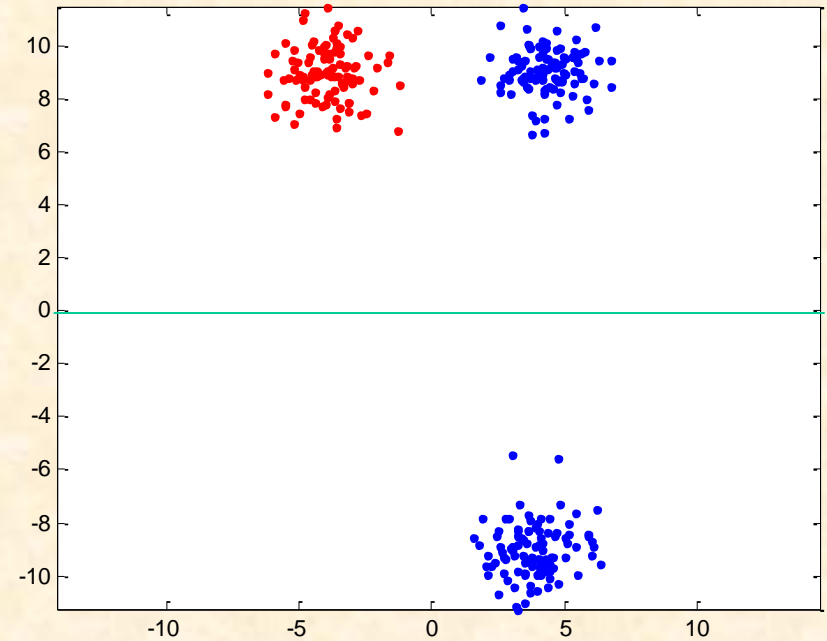
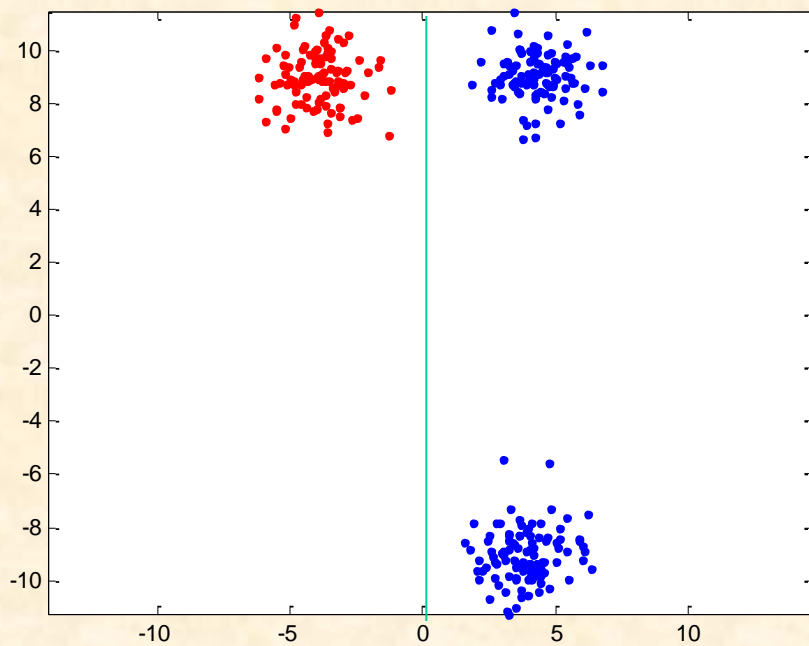
$$d_i(\underline{w}^T \underline{x}_i + w_0) \geq 1, \quad i = 1, 2, \dots, N \quad \begin{array}{l} d_i = 1, \text{ for } \underline{x}_i \in \omega_1, \\ d_i = -1, \text{ for } \underline{x}_i \in \omega_2 \end{array}$$

– Τα παραπάνω δικαιολογούνται από το γεγονός ότι ελαχιστοποιώντας το $\|\underline{w}\|^2$

μεγιστοποιείται το περιθώριο $\frac{2}{\|\underline{w}\|}$.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις



Οι επιπλέον περιορισμοί ισότητας αποθαρρύνουν λύσεις σαν αυτή του 2^{ου} σεναρίου πιο πάνω.

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΥΝΑΡΤΗΣΗΣ ΥΠΟ ΓΡΑΜ. ΑΝΙΣΟΤΙΚΟΥΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

Έστω το πρόβλημα

$$\text{Min } J(\vartheta)$$

Υπό τους περιορισμούς $f_i(\vartheta) \geq 0, i=1, \dots, m$.

Ορίζουμε τη **συνάρτηση Lagrange**

$$L(\vartheta, \lambda) = J(\vartheta) - \sum \lambda_i f_i(\vartheta)$$

Συνθήκες KKT (Karush-Kuhn-Tacker): Για τη θέση του **ελαχίστου** ισχύουν τα ακόλουθα

$$\frac{\partial}{\partial \vartheta} L(\vartheta, \lambda) = 0$$

$$\lambda_i \geq 0, i = 1, 2, \dots, m$$

$$\lambda_i f_i(\vartheta) = 0, i = 1, 2, \dots, m$$

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΥΝΑΡΤΗΣΗΣ ΥΠΟ ΓΡΑΜ. ΑΝΙΣΟΤΙΚΟΥΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

Έστω

$$L^*(\vartheta) = \max_{\lambda} L(\vartheta, \lambda) = \max_{\lambda} (J(\vartheta) - \sum_i \lambda_i f_i(\vartheta))$$

Αφού $\lambda_i \geq 0$, $f_i(\vartheta) \geq 0$ έχουμε

$$L^*(\vartheta) = J(\vartheta)$$

Άρα

$$\min_{\vartheta} J(\vartheta) = \min_{\vartheta} L^*(\vartheta) = \min_{\vartheta} \max_{\lambda \geq 0} L(\vartheta, \lambda)$$

Θεώρημα:

- Έστω (α) η $J(\vartheta)$ είναι **κυρτή** και (β) οι $f_i(\vartheta)$ είναι **γραμμικές**.
- Έστω ϑ^* μία θέση ελαχίστου για το πρόβλημα ελαχιστοποίησης και λ^* το αντίστοιχο διάνυσμα των πολ/στών Lagrange.
- Τότε το (ϑ^*, λ^*) είναι ένα σαγματικό σημείο (saddle point) της συνάρτησης Lagrange, για το οποίο ισχύει

$$L(\vartheta^*, \lambda^*) = \min_{\vartheta} \max_{\lambda \geq 0} L(\vartheta, \lambda) = \max_{\lambda \geq 0} \min_{\vartheta} L(\vartheta, \lambda)$$

Συνεπώς, σύμφωνα με το θεώρημα, για τον προσδιορισμό του ελαχίστου **δεν παίζει ρόλο η σειρά με την οποία βελτιστοποιούμε την $L(\vartheta, \lambda)$ ως προς τα ϑ και λ .**

Στο SVM πρόβλημα θα βελτιστοποιήσουμε την $L(\cdot)$ πρώτα ως προς ϑ και μετά ως προς λ .

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

–Το πρόβλημα SVM είναι μια διαδικασία (κυρτής) τετραγωνικής βελτιστοποίησης (quadratic optimization task), με γραμμικούς περιορισμούς. Οι συνθήκες Karush-Kuhn-Tucker ορίζουν ότι για τη θέση του ελαχίστου ισχύουν τα ακόλουθα:

- (1) $\frac{\partial}{\partial w} L(w, w_0, \lambda) = 0$
- (2) $\frac{\partial}{\partial w_0} L(w, w_0, \lambda) = 0$
- (3) $\lambda_i \geq 0, i = 1, 2, \dots, N$
- (4) $\lambda_i [d_i (w^T x_i + w_0) - 1] = 0, i = 1, 2, \dots, N$
- Όπου $L(\bullet, \bullet, \bullet)$ είναι η συνάρτηση Lagrange

$$L(w, w_0, \lambda) \equiv \frac{1}{2} w^T w - \sum_{i=1}^N \lambda_i [d_i (w^T x_i + w_0)]$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

– Η λύση: από τα παραπάνω προκύπτει ότι είναι

$$w = \sum_{i=1}^N \lambda_i d_i x_i \quad \sum_{i=1}^N \lambda_i d_i = 0$$

Αντικαθιστώντας τα παραπάνω στην $L(\cdot)$, τα λ_i 's εκτιμώνται ως οι λύσεις του δυϊκού SVM προβλήματος:

(Δυϊκό SVM)

Μεγιστοποίησε την $\left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j d_i d_j x_i^T x_j \right)$ ως προς το $\lambda = [\lambda_1, \dots, \lambda_N]^T$

Έτσι ώστε $\sum_{i=1}^N \lambda_i d_i = 0$
 $\underline{\lambda} \geq \underline{0}$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – γραμ. διαχ. κλάσεις

Παρατηρήσεις:

- Οι πολ/στές Lagrange λ_i είναι είτε **θετικοί** είτε **μηδέν**. Έτσι

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

όπου N_s είναι ο αριθμός των διανυσμάτων με θετικά λ_i 's.

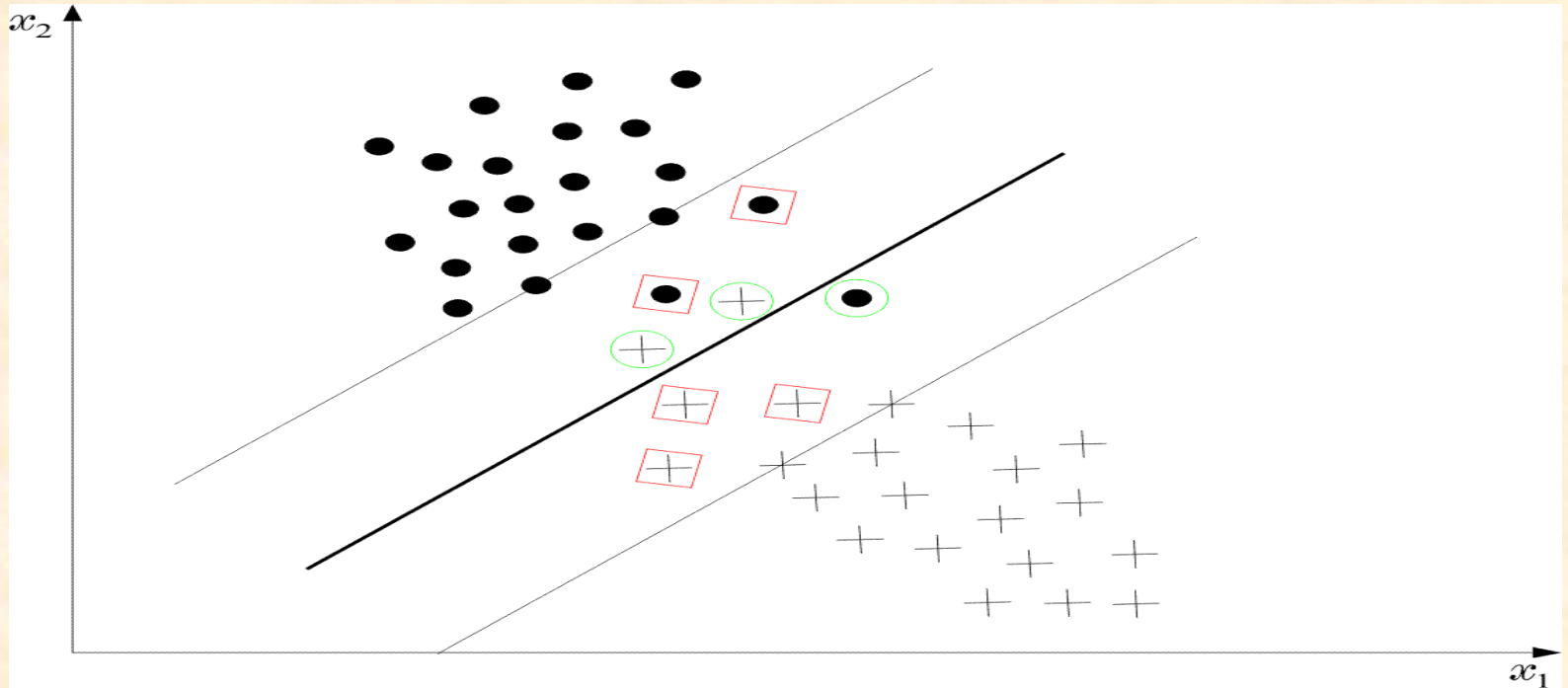
- Θετικά λ_i έχουν τα διανύσματα που ικανοποιούν τη συνθήκη $w^T x_i + w_0 = \pm 1$ λόγω των περιορισμών $\lambda_i [d_i (w^T x_i + w_0) - 1] = 0, \quad i = 1, 2, \dots, N$

Τα διανύσματα αυτά ονομάζονται **ΔΙΑΝΥΣΜΑΤΑ ΣΤΗΡΙΞΗΣ (SUPPORT VECTORS)** και είναι τα εγγύτερα προς το υπερεπίπεδο του ταξινομητή διανύσματα από κάθε κλάση. Αυτά είναι που καθορίζουν το w .

- Μετά τον προσδιορισμό του w , το w_0 προσδιορίζεται από τις συνθήκες (4).
- Το βέλτιστο υπερεπίπεδο ως προς το κριτήριο της μεγιστοποίησης του περιθωρίου είναι **ΜΟΝΑΔΙΚΟ**.
- Παρότι η λύση είναι μοναδική, οι πολ. Lagrange **δεν είναι απαραίτητα** μοναδικοί.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις



ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις

- Στην περίπτωση αυτή, δεν υπάρχει υπερεπίπεδο έτσι ώστε

$$w^T x + w_0 (><)1, \forall x$$

- Υπενθυμίζεται ότι το **περιθώριο** ορίζεται ως το **διπλάσιο της απόστασης** μεταξύ των ακόλουθων δύο υπερεπιπέδων

$$w^T x + w_0 = 1 \text{ and } w^T x + w_0 = -1$$

– Για τα **διανύσματα εκπαίδευσης** έχουμε **ένα** από τα ακόλουθα **τρία** δυνατά σενάρια

1) Διανύσματα **εκτός** της ζώνης των δύο επιπέδων που είναι **ορθά** ταξινομημένα

$$d_i(w^T x + w_0) > 1$$

2) Διανύσματα **εντός** της ζώνης των δύο επιπέδων που είναι **ορθά** ταξινομημένα

$$0 \leq d_i(w^T x + w_0) < 1$$

3) Διανύσματα τα οποία είναι **μη ορθά ταξινομημένα**

$$d_i(w^T x + w_0) < 0$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις

– Οι παραπάνω περιπτώσεις μπορούν να παρασταθούν με συμπαγή τρόπο ως εξής

$$d_i(\underline{w}^T \underline{x} + w_0) \geq 1 - \xi_i$$

Όπου για

1^ο σενάριο $\rightarrow \xi_i = 0$

2^ο σενάριο $\rightarrow 0 < \xi_i \leq 1$

3^ο σενάριο $\rightarrow 1 < \xi_i$

Οι μεταβλητές ξ_i είναι γνωστές ως **μεταβλητές χαλαρότητας (slack variables)**

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις

–Ο στόχος της βελτιστοποίησης είναι τώρα διττός

•Μεγιστοποίηση του περιθωρίου

•Ελαχιστοποίηση του αριθμού των διανυσμάτων με $\xi_i > 0$,

•Ένας τρόπος να επιτύχουμε τον παραπάνω στόχο είναι μέσω της ακόλουθης συνάρτησης κόστους

$$J(\underline{w}, w_0, \underline{\xi}) = \frac{1}{2} \|\underline{w}\|^2 + C \sum_{i=1}^N I(\xi_i)$$

όπου C είναι μια σταθερά και

$$I(\xi_i) = \begin{cases} 1 & \xi_i > 0 \\ 0 & \xi_i = 0 \end{cases}$$

•Η συνάρτηση $I(\cdot)$ ΔΕΝ είναι διαφορίσιμη. Στην πράξη, μπορούμε να

χρησιμοποιήσουμε μια προσέγγιση

$$J(\underline{w}, w_0, \underline{\xi}) = \frac{1}{2} \|\underline{w}\|^2 + C \sum_{i=1}^N \xi_i$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις

- Η **συνάρτηση Lagrange** για την περίπτωση αυτή γίνεται

$$L(w, w_0, \xi, \lambda, \mu) \equiv \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \lambda_i [d_i (w^T x_i + w_0) - 1 + \xi_i]$$

(τα λ_i και μ_i είναι πολ/στές Lagrange)

- Οι αντίστοιχες **ΚΚΤ συνθήκες** είναι:

$$(1) \quad w = \sum_{i=1}^N \lambda_i d_i x_i$$

$$(2) \quad \sum_{i=1}^N \lambda_i d_i = 0$$

$$(3) \quad C - \mu_i - \lambda_i = 0, \quad i = 1, 2, \dots, N$$

$$(4) \quad \lambda_i [y_i (w^T x_i + w_0) - 1 + \xi_i] = 0, \quad i = 1, 2, \dots, N$$

$$(5) \quad \mu_i \xi_i = 0, \quad i = 1, 2, \dots, N$$

$$(6) \quad \mu_i, \lambda_i \geq 0, \quad i = 1, 2, \dots, N$$

➤ Αντικαθιστώντας τα παραπάνω στην $L(\cdot)$, οι πολ/στές Lagrange προσδιορίζονται ως οι λύσεις του ακόλουθου δυϊκού SVM προβλήματος

Μεγιστοποίηση $\lambda_{\geq 0} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j d_i d_j x_i^T x_j \right)$ ως προς το $\lambda = [\lambda_1, \dots, \lambda_N]^T$

υπό τις προϋποθέσεις

$$0 \leq \lambda_i \leq C, \quad i = 1, 2, \dots, N$$

$$\sum_{i=1}^N \lambda_i d_i = 0$$

➤ Σχόλιο:

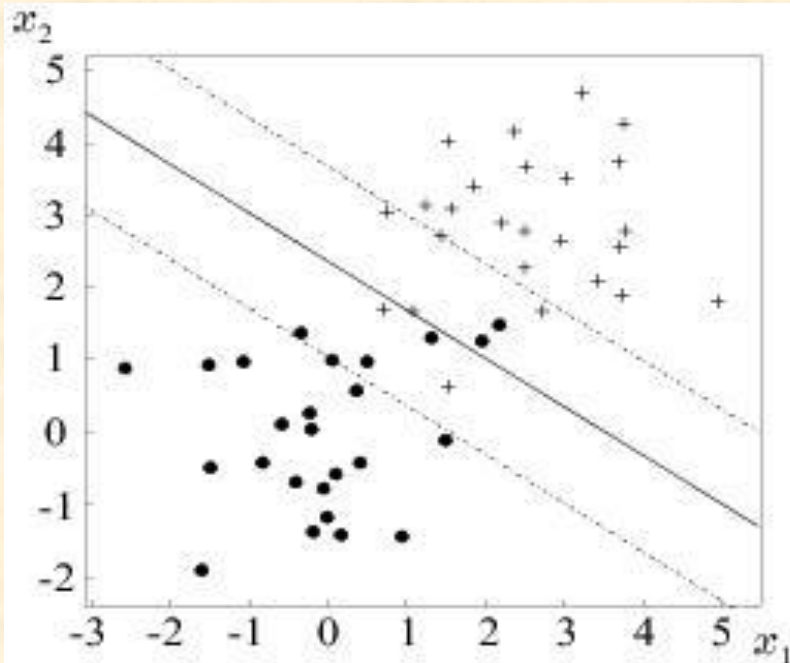
Η μόνη διαφορά με την περίπτωση των διαχωρίσιμων κλάσεων είναι η ύπαρξη του C στους περιορισμούς.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

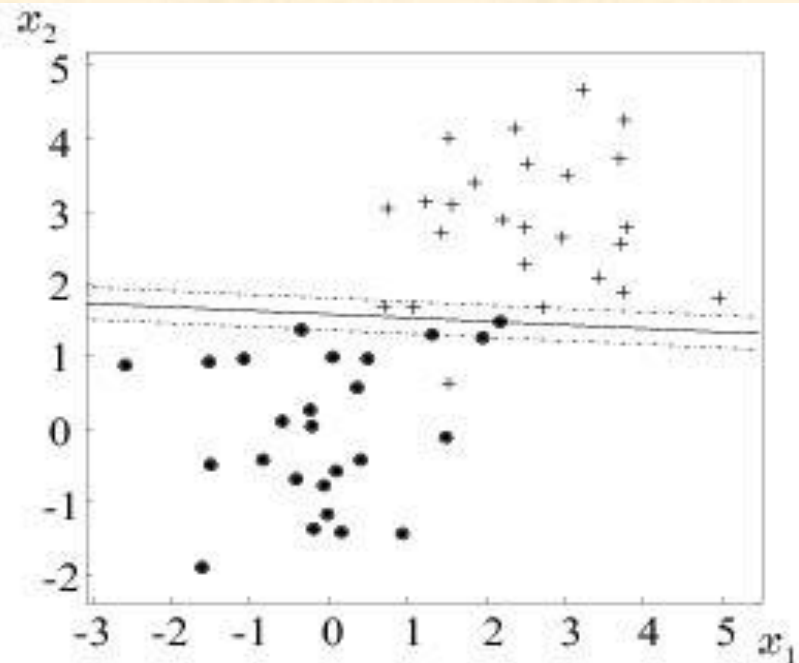
Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις

Η λύση του SVM προβλήματος ακολουθεί βήματα ανάλογα με αυτά της προηγούμενης περίπτωσης.

Ωστόσο, στην παρούσα περίπτωση, η παράμετρος C επηρεάζει την επιλογή της τελικής λύσης.



(a)

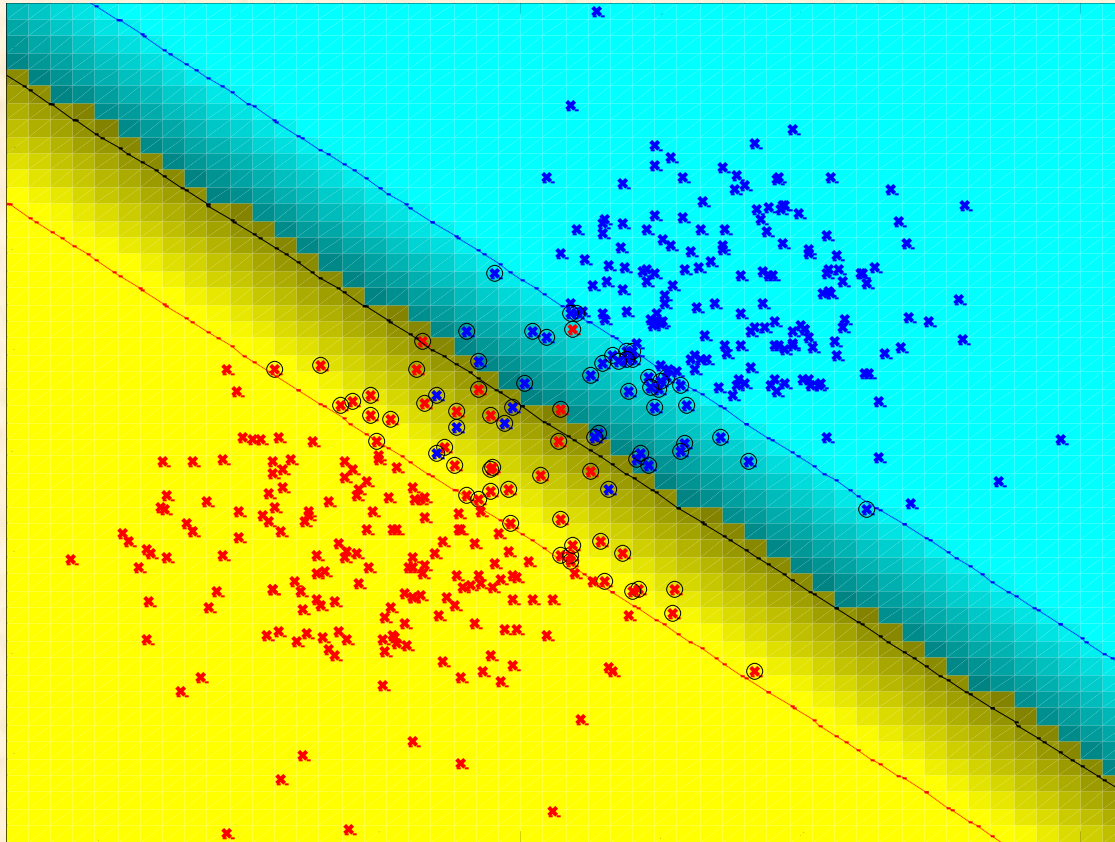


(b)

Στο παραπάνω παράδειγμα η C έχει “μικρότερη τιμή” για το σχήμα (a) και “μεγαλύτερη τιμή” για το σχήμα (b).

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις



C = 0.1

Pe_tr = 0.0225

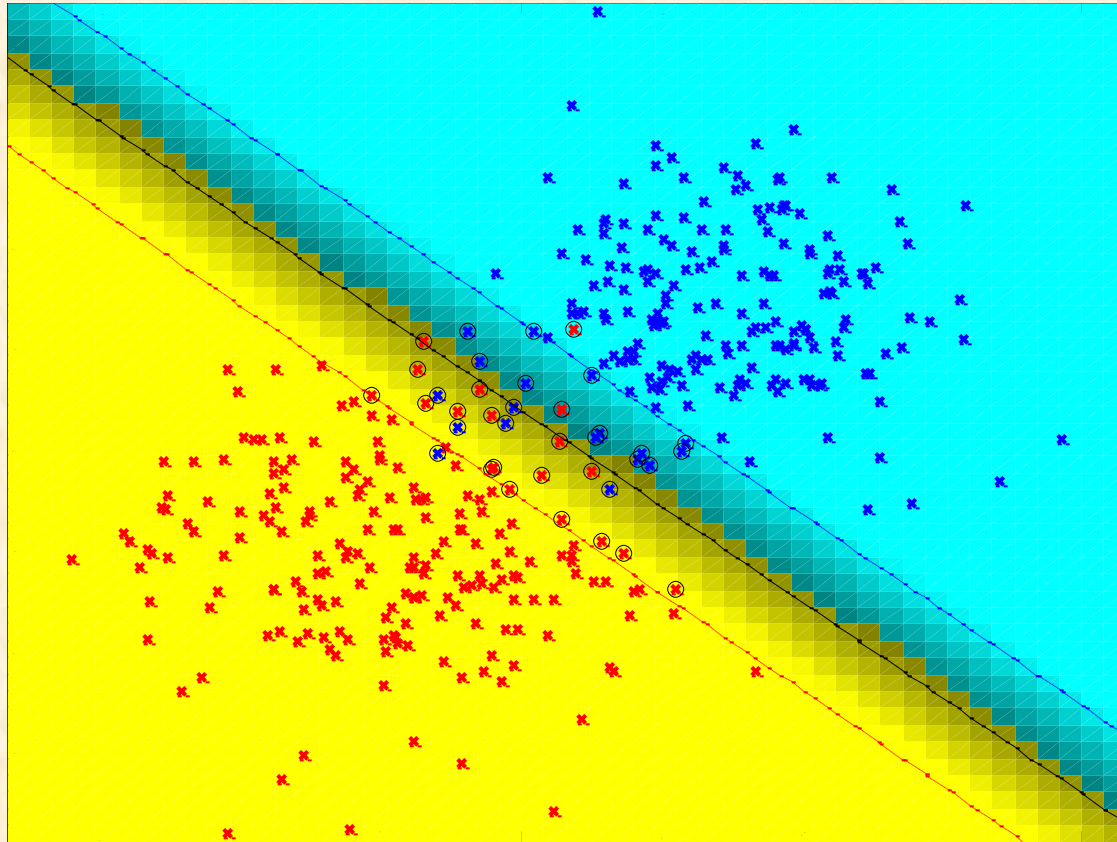
Pe_te = 0.0325

sup_vec = 82

marg = 0.9412

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις



C = 1

Pe_tr = 0.0225

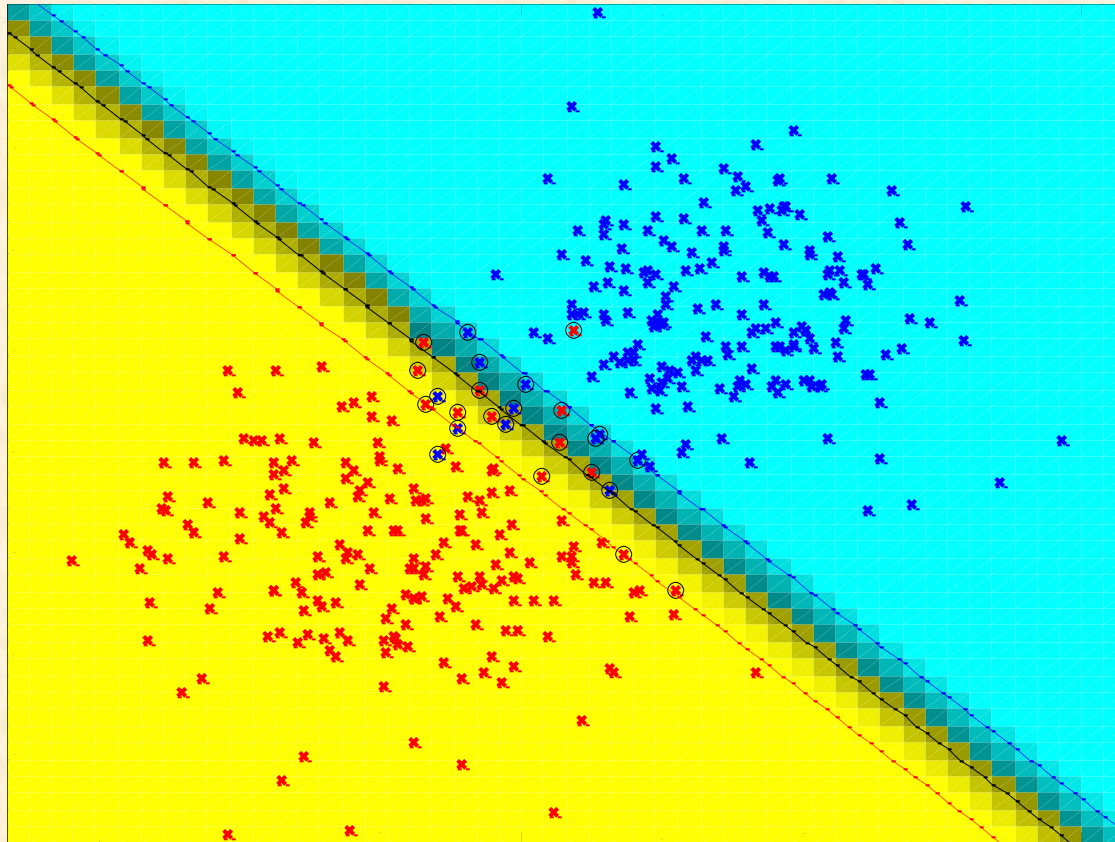
Pe_te = 0.0325

sup_vec = 37

marg = 0.6317

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Support vector machines – η γραμμική περίπτωση – μη γραμ. διαχ. κλάσεις



C =20

Pe_tr =0.0225

Pe_te =0.0350

sup_vec =25

marg = 0.3573