ΓΡΑΜΜΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ (LINEAR CLASSIFIERS)

❖ **Πρόβλημα**: Θεωρείστε ένα πρόβλημα δύο κλάσεων ω_1, ω_2

➢ $$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 = 0 =$$
$$w_1 x_1 + w_2 x_2 + \ldots + w_l x_l + w_0$$

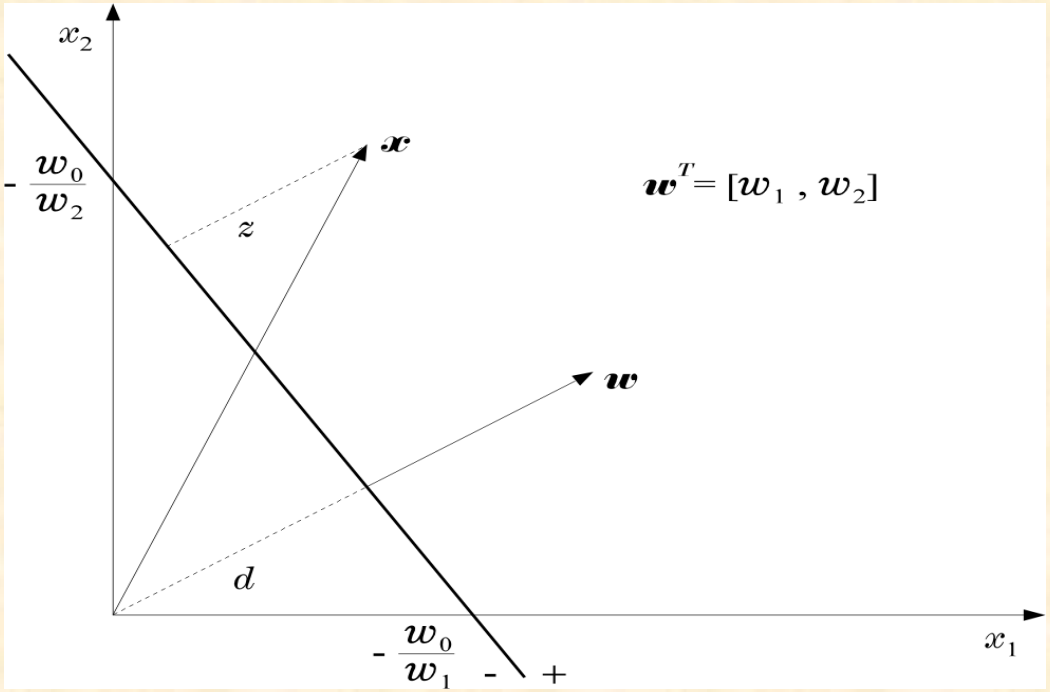
➢ Έστω $\underline{x}_1, \underline{x}_2$ πάνω στο υπερεπίπεδο απόφασης:

$$0 = \underline{w}^T \underline{x}_1 + w_0 = \underline{w}^T \underline{x}_2 + w_0 \Rightarrow$$
$$\underline{w}^T (\underline{x}_1 - \underline{x}_2) = 0 \quad \forall \underline{x}_1, \underline{x}_2$$

- Έτσι:

$\underline{w} \perp$ στο υπερεπίπεδο

$$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 = 0$$



$$d = \frac{|w_0|}{\sqrt{w_1^2 + w_2^2}}, \quad z = \frac{|g(\underline{x})|}{\sqrt{w_1^2 + w_2^2}}$$

❖ Ο αλγόριθμος Perceptron

- Έστω ότι οι κλάσεις είναι γραμμικώς διαχωρίσιμες, δηλ.,

$$\exists\ \underline{w}^*:\ w^{*T}\ \underline{x} > 0\ \ \forall \underline{x} \in \omega_1$$

$$\underline{w}^{*T}\ \underline{x} < 0\ \ \forall \underline{x} \in \omega_2$$

- Η περίπτωση $\underline{w}^{*T}\ \underline{x} + w_0^*$ εμπίπτει στην παραπάνω διατύπωση, αφού

  - $\underline{w}' \equiv \begin{bmatrix} \underline{w}^* \\ w_0^* \end{bmatrix},\ \underline{x}' = \begin{bmatrix} \underline{x} \\ 1 \end{bmatrix}$

  - $\underline{w}^{*T}\ \underline{x} + w_0^* = \underline{w}'^T\ \underline{x}' = 0$

- **Στόχος**: Υπολογισμός λύσης, δηλ., προσδιορισμός υπερεπιπέδου $\underline{w}$, έτσι ώστε

$$\underline{w}^T \underline{x}(><)0 \quad x \in \begin{cases} \omega_1 \\ \omega_2 \end{cases}$$

  - **Βήματα**
    - Ορισμός συνάρτησης κόστους που θα ελαχιστοποιηθεί
    - Επιλογή αλγορίθμου για την ελαχιστοποίηση της συνάρτησης κόστους
    - Το ελάχιστο αντιστοιχεί σε μία λύση.

- Η συνάρτηση κόστους

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

- όπου Y το υποσύνολο των λανθασμένα ταξινομημένων διανυσμάτων από το $\underline{w}$. Όταν $Y=\varnothing$ έχουμε μία λύση και είναι $J(\underline{w}) = 0$

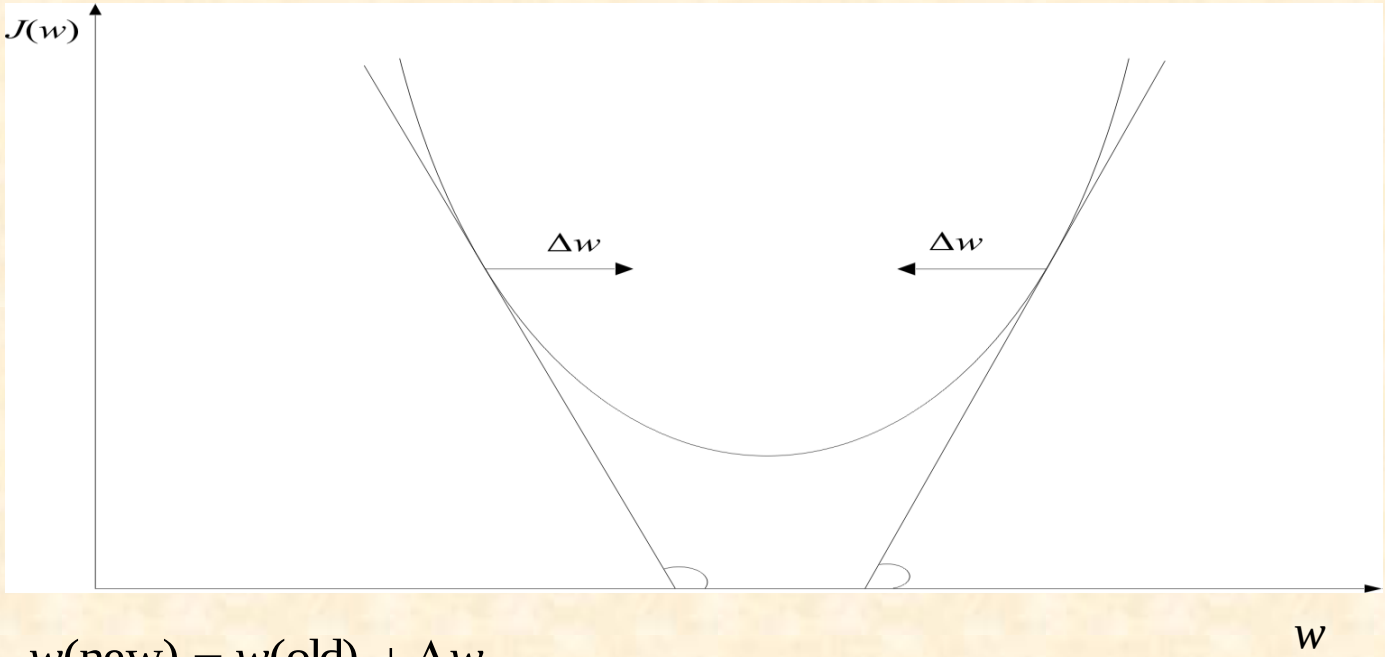
- $\delta_x = -1$ if $\underline{x} \in Y$ and $\underline{x} \in \omega_1$

  $\delta_x = +1$ if $\underline{x} \in Y$ and $\underline{x} \in \omega_2$

- Γενικά, $J(\underline{w}) \geq 0$

- Η $J(\underline{w})$ είναι τμηματικά γραμμική (Γιατί?)

- Ο Αλγόριθμος
  - Υιοθετείται η φιλοσοφία της σταδιακής καθόδου κατά την κλίση (gradient descent).



$$\underline{w}(\text{new}) = \underline{w}(\text{old}) + \Delta\underline{w}$$

$$\Delta\underline{w} = -\mu \frac{\partial J(\underline{w})}{\partial \underline{w}} \Big| \underline{w} = \underline{w}(\text{old})$$

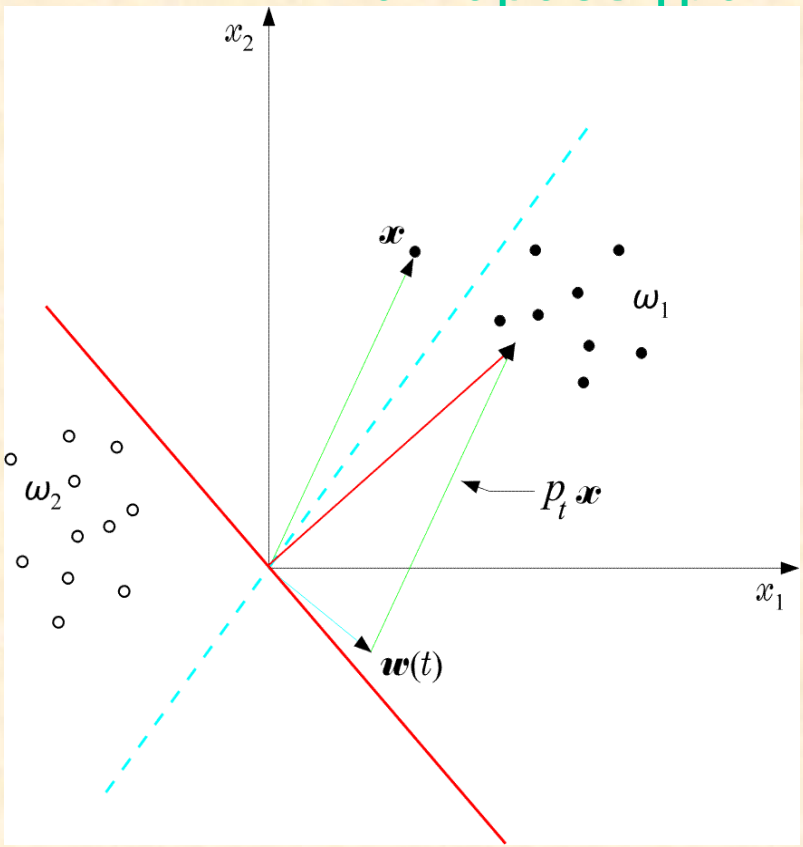
- Όπου είναι νόμιμο (valid)

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

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

Αυτός είναι ο γνωστός αλγόριθμος Perceptron.

Ένα παράδειγμα



$$\underline{w}(t+1) = \underline{w}(t) + \rho_t \underline{x}$$
$$= \underline{w}(t) - \rho_t \delta_x \underline{x} \quad (\delta_x = -1)$$

- Ο αλγόριθμος perceptron **συγκλινει** σε **πεπερασμένο** αριθμό επαναλήψεων σε λύση αν

$$\lim_{t\to\infty} \sum_{k=0}^{t} \rho_k \to \infty \qquad \lim_{t\to\infty} \sum_{k=0}^{t} \rho_k^{\,2} < +\infty$$

$$\text{e.g.: } \rho_t = \frac{c}{t}$$

❖ Μία χρήσιμη παραλλαγή του αλγορίθμου perceptron

$$\underline{w}(t+1) = \underline{w}(t) + \rho \underline{x}_{(t)} \ , \quad \begin{matrix} \underline{w}^T(t)\underline{x}_{(t)} \le 0 \\ \underline{x}_{(t)} \in \omega_1 \end{matrix}$$

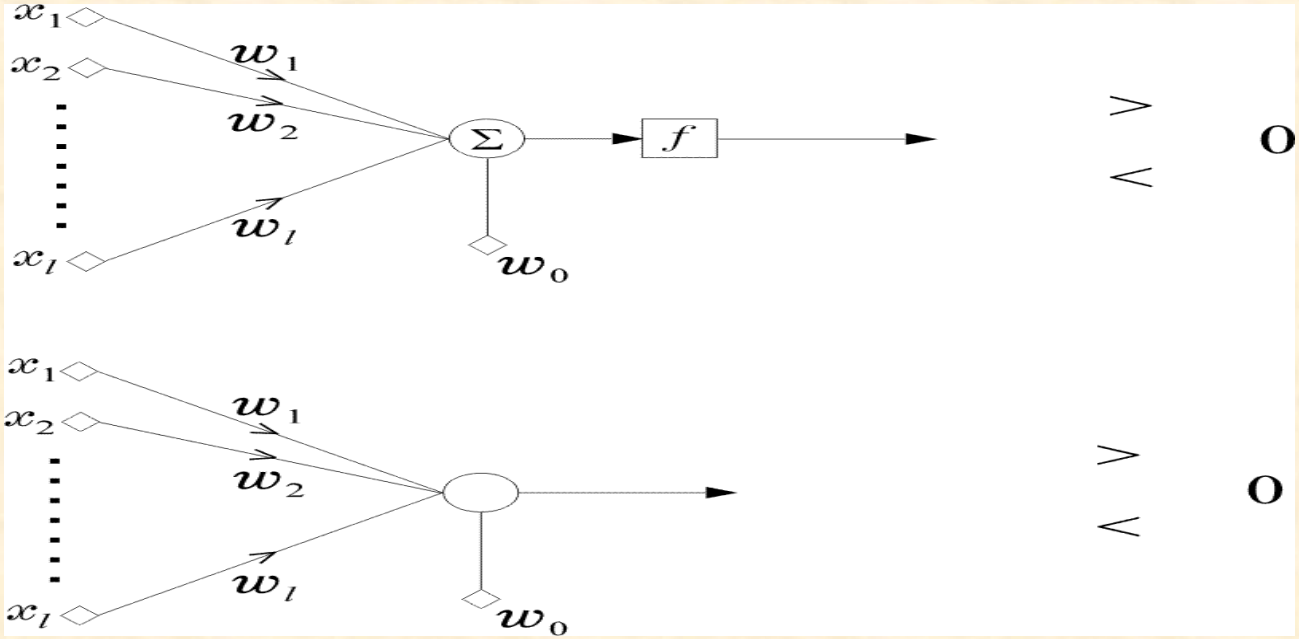
$$\underline{w}(t+1) = \underline{w}(t) - \rho \underline{x}_{(t)}, \quad \begin{matrix} \underline{w}^T(t)\underline{x}_{(t)} \ge 0 \\ \underline{x}_{(t)} \in \omega_2 \end{matrix}$$

$$\underline{w}(t+1) = \underline{w}(t) \quad \text{διαφορετικά}$$

➢ Είναι ένας αλγόριθμος του τύπου

επιβράβευσης και τιμωρίας

❖ Το perceptron



w_i 's συνάψεις ή συναπτικά βάρη

w_0 κατώφλι

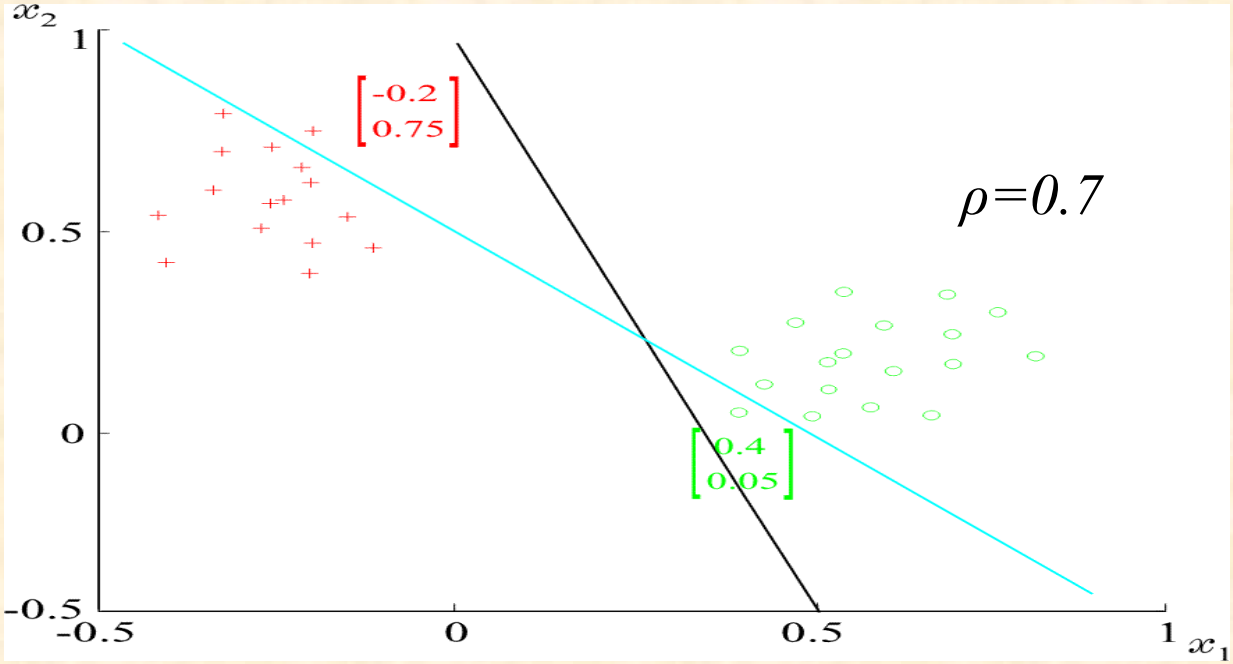
- Το δίκτυο καλείται perceptron ή νευρώνας (neuron)
- Είναι μία μηχανή μάθησης που μαθαίνει από τα διανύσματα εκπαίδευσης μέσω του αλγορίθμου perceptron.

➢ Παράδειγμα: Σε κάποια επανάληψη t ο αλγόριθμος perceptron βρίσκεται στην κατάσταση

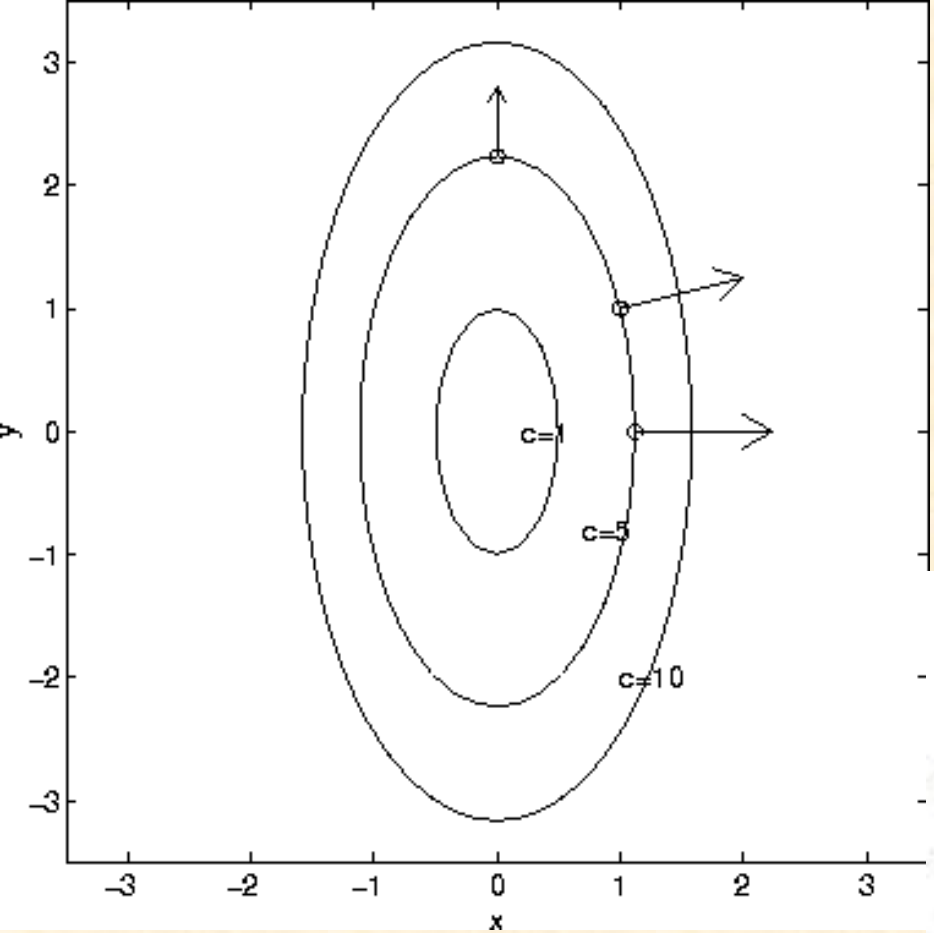
$$w_1 = 1,\ w_2 = 1,\ w_0 = -0.5$$

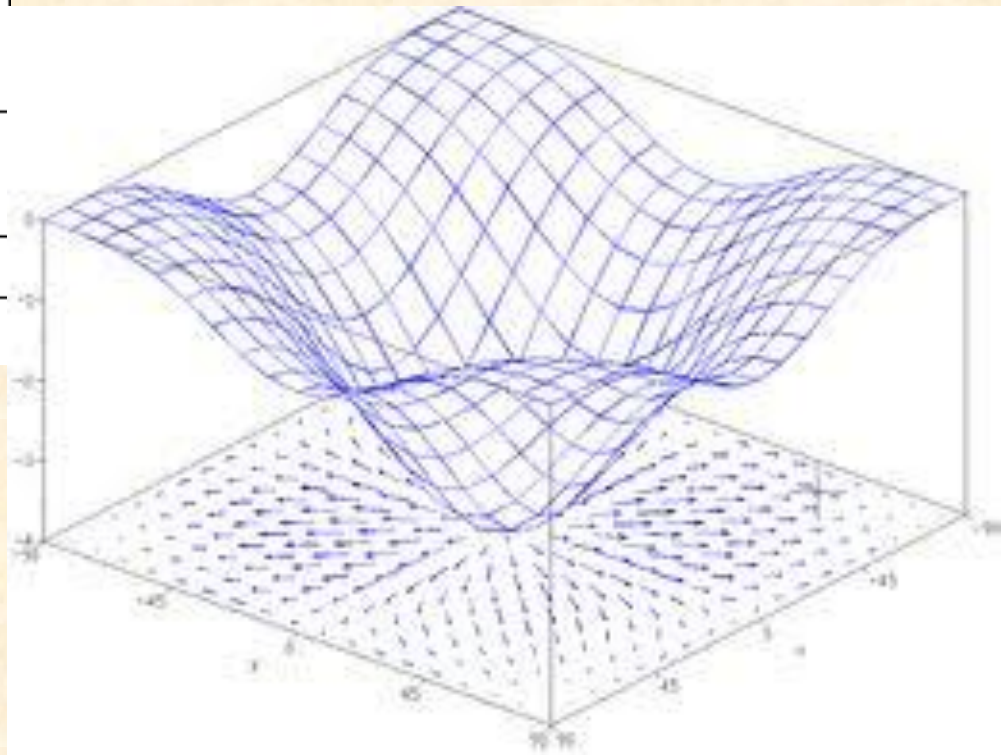
$$x_1 + x_2 - 0.5 = 0$$

Το αντίστοιχο υπερεπίπεδο είναι



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


c=1
c=5
c=10
x
y



❖ Μέθοδοι ελαχίστων τετραγώνων

- Αν οι κλάσεις είναι γραμμικώς διαχωρίσιμες, η έξοδος του perceptron είναι ± 1

- Αν οι κλάσεις ΔΕΝ είναι γραμμικώς διαχωρίσιμες, θα υπολογίσουμε τα βάρη $w_1, w_2, \ldots, w_0$

  έτσι ώστε η διαφορά μεταξύ

  - Της πραγματικής εξόδου του ταξινομητή, $\underline{w}^T \underline{x}$, και
  - Των επιθυμητών εξόδων, π.χ.

    $+1$ αν $\underline{x} \in \omega_1$

    -1 αν $\underline{x} \in \omega_2$

    να είναι ΜΙΚΡΗ

- ΜΙΚΡΗ, ως προς το μέσο τετραγωνικό σφάλμα (mean square error), σημαίνει επιλογή του $\underline{w}$ έτσι ώστε η συνάρτηση κόστους

  - $J(\underline{w}) \equiv E[(y - \underline{w}^T \underline{x})^2]$ να ελαχιστοποιηθεί
  - $\hat{\underline{w}} = \arg\min_{\underline{w}} J(\underline{w})$
  - y οι αντίστοιχες επιθυμητές απόκρισεις

- Ελαχιστοποίηση της

$J(\underline{w})$ ως προς $\underline{w}$ δίνει :

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

$$= 2E[\underline{x}(y - \underline{x}^T \underline{w})] \Rightarrow$$

$$E[\underline{x}\underline{x}^T]\underline{w} = E[\underline{x}y] \Rightarrow$$

$$\boxed{\hat{\underline{w}} = R_x^{-1} E[\underline{x}y]}$$

όπου R_x είναι ο πίνακας αυτοσυσχέτισης (autocorrelation matrix)

$$R_x \equiv E[\underline{x}\underline{x}^T] = \begin{bmatrix} E[x_1x_1] & E[x_1x_2]... & E[x_1x_l] \\ ..... & ..... & ..... \\ E[x_lx_1] & E[x_lx_2]... & E[x_lx_l] \end{bmatrix}$$

και $E[\underline{x}y] = \begin{bmatrix} E[x_1y] \\ ... \\ E[x_ly] \end{bmatrix}$ το διάνυσμα ετεροσυσχέτισης (crosscorrelation vector)

- ΜΙΚΡΗ, ως προς το άθροισμα των τετραγωνικών σφαλμάτων, σημαίνει
  - $$J(\underline{w}) = \sum_{i=1}^{N} (y_i - \underline{w}^T \underline{x}_i)^2$$

    $(y_i, \underline{x}_i)$: ζεύγη εκπ/σης δηλαδή, το διαν. εισ. x_i και η αντίστοιχη ετικέτα κλάσης y_i (±1).

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

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

- Ψευδοαντίστροφος πίνακας
  - Ορίζουμε

$$X = \begin{bmatrix} \underline{x}_1^T \\ \underline{x}_2^T \\ \dots \\ \underline{x}_N^T \end{bmatrix} \quad (\text{ένας } Nxl \text{ πίνακας})$$

$$\underline{y} = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix} \quad \text{αντίστοιχες επιθυμητές αποκρίσεις}$$

- $X^T = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_N]$ (ένας lxN πίνακας)
- $X^T X = \sum_{i=1}^{N} \underline{x}_i \underline{x}_i^T$
- $X^T \underline{y} = \sum_{i=1}^{N} \underline{x}_i y_i$

Έτσι $$(\sum_{i=1}^{N} \underline{x}_i^T \underline{x}_i)\underline{\hat{w}} = (\sum_{i=1}^{N} \underline{x}_i y_i)$$

$$(X^T X)\underline{\hat{w}} = X^T \underline{y} \Rightarrow$$

$$\underline{\hat{w}} = (X^T X)^{-1} X^T \underline{y}$$

$$= X^{\neq} \underline{y}$$

$$X^{\neq} \equiv (X^T X)^{-1} X^T$$ Ψευδοαντίστροφος του X

- Έστω $N=l$ $\Rightarrow$ ο X είναι τετραγωνικός και (γενικά) αντιστρέψιμος. Τότε

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

$$X^{\neq} = X^{-1}$$

- Έστω $N>l$. Τότε, γενικά, δεν υπάρχει λύση που να ικανοποιεί όλες τις εξισώσεις ταυτόχρονα:

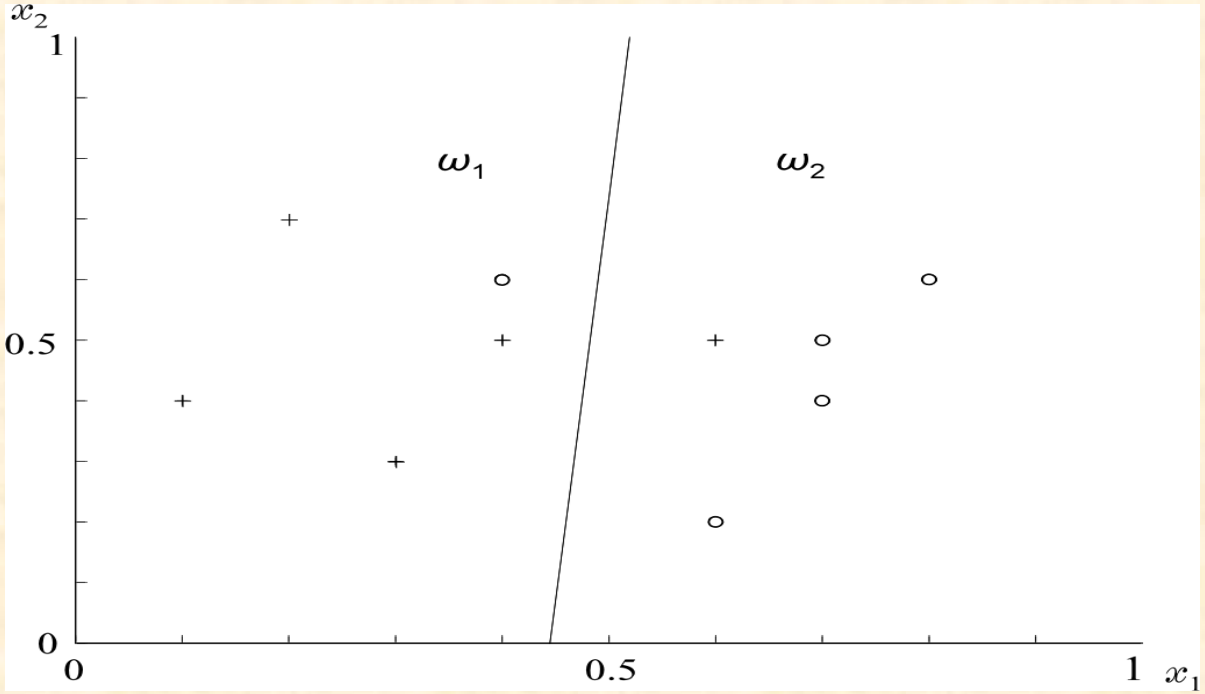
$$X\,\underline{w}=\underline{y}:\quad \begin{array}{l} \underline{x}_1^T\,\underline{w}=y_1 \\ \underline{x}_2^T\,\underline{w}=y_2 \\ \ldots. \\ \underline{x}_N^T\,\underline{w}=y_N \end{array} \qquad \text{N εξισώσεις } >\text{l άγνωστοι}$$

- Η "λύση" $\underline{w}=X^{\neq}\,\underline{y}$ αντιστοιχεί στην ελάχιστη τιμή του αθροίσματος τετραγώνων (minimum sum of squares solution).

- Παράδειγμα:

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

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



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

- $X^T X = \begin{bmatrix} 2.8 & 2.24 & 4.8 \\ 2.24 & 2.41 & 4.7 \\ 4.8 & 4.7 & 10 \end{bmatrix}, X^T \underline{y} = \begin{bmatrix} -1.6 \\ 0.1 \\ 0.0 \end{bmatrix}$

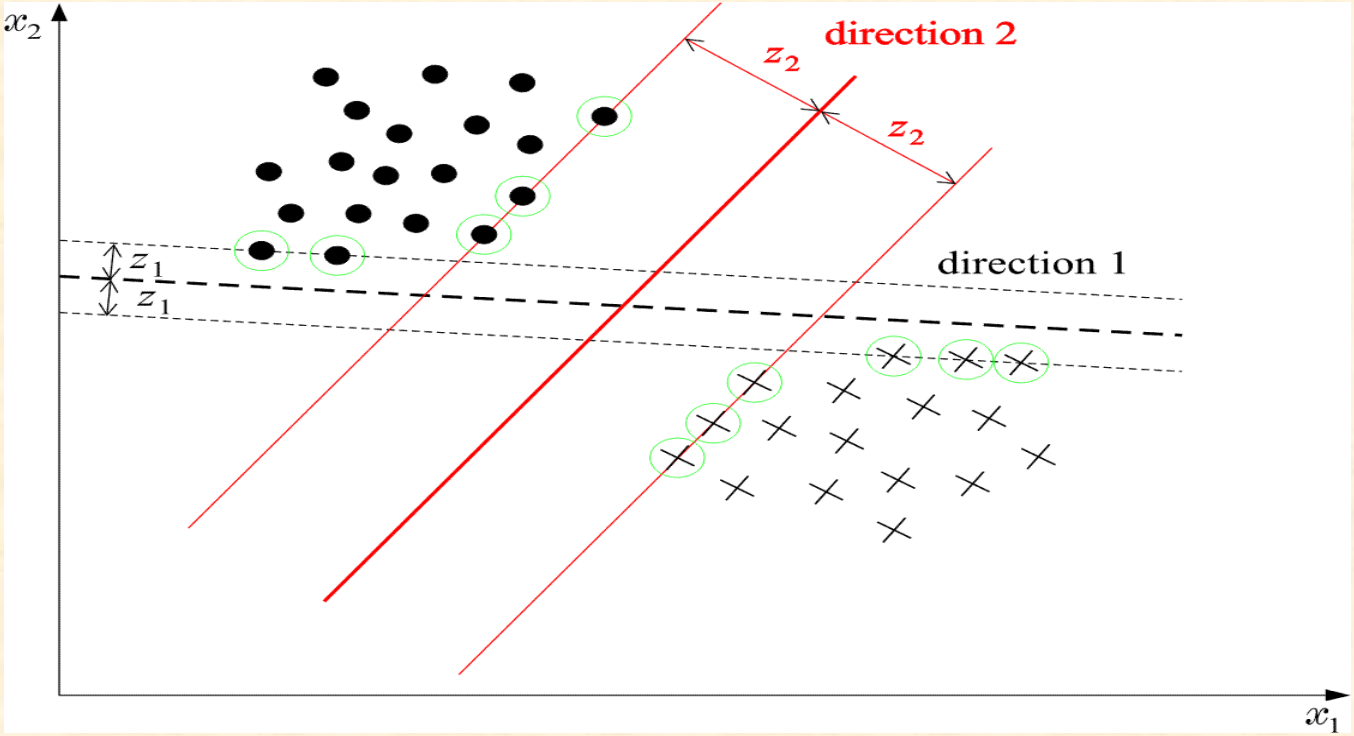
$$\underline{w} = (X^T X)^{-1} X^T \underline{y} = \begin{bmatrix} -3.13 \\ 0.24 \\ 1.34 \end{bmatrix}$$

❖ ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΙΚΗΣ ΣΤΗΡΙΞΗΣ (SUPPORT VECTOR MACHINES)

- Ο στόχος: Δοθέντων δύο γραμμικώς διαχωρίσιμων κλάσεων, να σχεδιαστεί ο ταξινομητής

$$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 = 0$$

που αφήνει το μέγιστο περιθώριο από τις δύο κλάσεις



- Περιθώριο: Κάθε υπερεπίπεδο χαρακτηρίζεται από
  - Την κατεύθυνση στο χώρο, δηλ., $\underline{w}$
  - Τη θέση του στο χώρο, δηλ., w_0
  - Για ΚΑΘΕ κατεύθυνση, $\underline{w}$ επέλεξε το υπερεπίπεδο που αφήνει την ΙΔΙΑ απόσταση από τα πλησιέστερα σημεία από κάθε κλάση. Το περιθώριο είναι διπλάσιο αυτής της απόστασης.

- Η απόσταση ενός σημείου $\hat{\underline{x}}$ από ένα υπερεπίπεδο είναι

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

- Κλιμάκωσε τα, $\underline{w}$, $\underline{w}_0$, έτσι ώστε στα εγγύτερα από κάθε κλάση η συνάρτηση διάκρισης είναι ±1:

$$|g(\underline{x})| = 1 \quad \{g(\underline{x}) = +1 \text{ για } \omega_1 \text{ και } g(\underline{x}) = -1 \text{ για } \omega_2\}$$

- Έτσι το περιθώριο δίνεται από τη σχέση

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

- Επίσης, ισχύουν τα ακόλουθα

$$\underline{w}^T \underline{x} + w_0 \geq 1 \ \forall \underline{x} \in \omega_1$$

$$\underline{w}^T \underline{x} + w_0 \leq -1 \ \forall \underline{x} \in \omega_2$$

- SVM (γραμμικός) ταξινομητής

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

- Ελαχιστοποίησε

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

- Υπό τους περιορισμούς

$$y_i(\underline{w}^T \underline{x}_i + w_0) \geq 1, \; i = 1,2,\ldots,N$$

$$y_i = 1, \text{ for } \underline{x}_i \in \omega_1,$$

$$y_i = -1, \text{ for } \underline{x}_i \in \omega_2$$

- Αυτό γίνεται διότι ελαχιστοποιώντας το $\|\underline{w}\|$

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

- Η παραπάνω είναι μία διαδικασία τετραγωνικής βελτιστοποίησης (quadratic optimization task) υπό τον περιορισμό ενός συνόλου γραμμικών ανισοτήτων. Οι Karush-Kuhh-Tucker συνθήκες, δηλώνουν ότι το ελάχιστο ικανοποιεί τις συνθήκες:
  - (1) $\frac{\partial}{\partial \underline{w}} \mathrm{L}(\underline{w}, w_0, \underline{\lambda}) = 0$
  - (2) $\frac{\partial}{\partial w_0} L(\underline{w}, w_0, \underline{\lambda}) = 0$
  - (3) $\lambda_i \geq 0, i = 1,2,\ldots,N$
  - (4) $\lambda_i \left[y_i (\underline{w}^T \underline{x}_i + w_0) - 1 \right] = 0, i = 1,2,\ldots,N$
  - Όπου $L(.,.,.)$ είναι η συνάρτηση Lagrange

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

- Η λύση: από τα παραπάνω, προκύπτει ότι
  - $\underline{w} = \sum_{i=1}^{N} \lambda_i y_i \underline{x}_i$
  - $\sum_{i=1}^{N} \lambda_i y_i = 0$

- Σχόλια:
  - Οι πολ/στές Lagrange μπορεί να είναι είτε μηδέν είτε θετικοί. Έτσι,
    - $$\underline{w} = \sum_{i=1}^{N_s} \lambda_i y_i \underline{x}_i$$

      όπου $N_s \leq N$ αντιστοιχεί στους θετικούς πολ/στές Lagrange

    - Από τον περιορισμό (4) πιο πάνω, δηλ.,

      $$\lambda_i [y_i (\underline{w}^T \underline{x}_i + w_0) - 1] = 0, \quad i = 1, 2, \ldots, N$$

      τα διανύσματα που συνεισφέρουν στο $\underline{w}$ ικανοποιούν

      $$\underline{w}^T \underline{x}_i + w_0 = \pm 1$$

– Τα διανύσματα αυτά καλούνται ΔΙΑΝΥΣΜΑΤΑ ΣΤΗΡΙΞΗΣ και είναι τα κοντινότερα διανύσματα, από κάθε κλάση, προς τον ταξινομητή.

– Μετά τον υπολογισμό του $\underline{w}$ το w_0 προσδιορίζεται από τις συνθήκες (4).

– Ο βέλτιστος ταξινομητής-υπερεπίπεδο μίας μηχανής διανυσματικής υποστήριξης είναι ΜΟΝΑΔΙΚΟΣ.

- Σχηματισμός Δυϊκού προβλήματος
  - Το πρόβλημα των SVM είναι διατυπωμένο ως ένα κυρτό πρόβλημα προγραμματισμού (convex programming problem), με
    - Κυρτή συνάρτηση κόστους
    - Κυρτή περιοχή εφικτών (feasible) λύσεων
  - Έτσι, η λύση μπορεί να προέλθει από το δυϊκό του πρόβλημα, δηλ.,
    - Μεγιστοποίησε $\underset{\underline{\lambda}}{} L(\underline{w}, w_0, \underline{\lambda})$
    - Υπό τους περιορισμούς $\underline{w} = \sum_{i=1}^{N} \lambda_i y_i \underline{x}_i$

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

$$\underline{\lambda} \geq \underline{0}$$

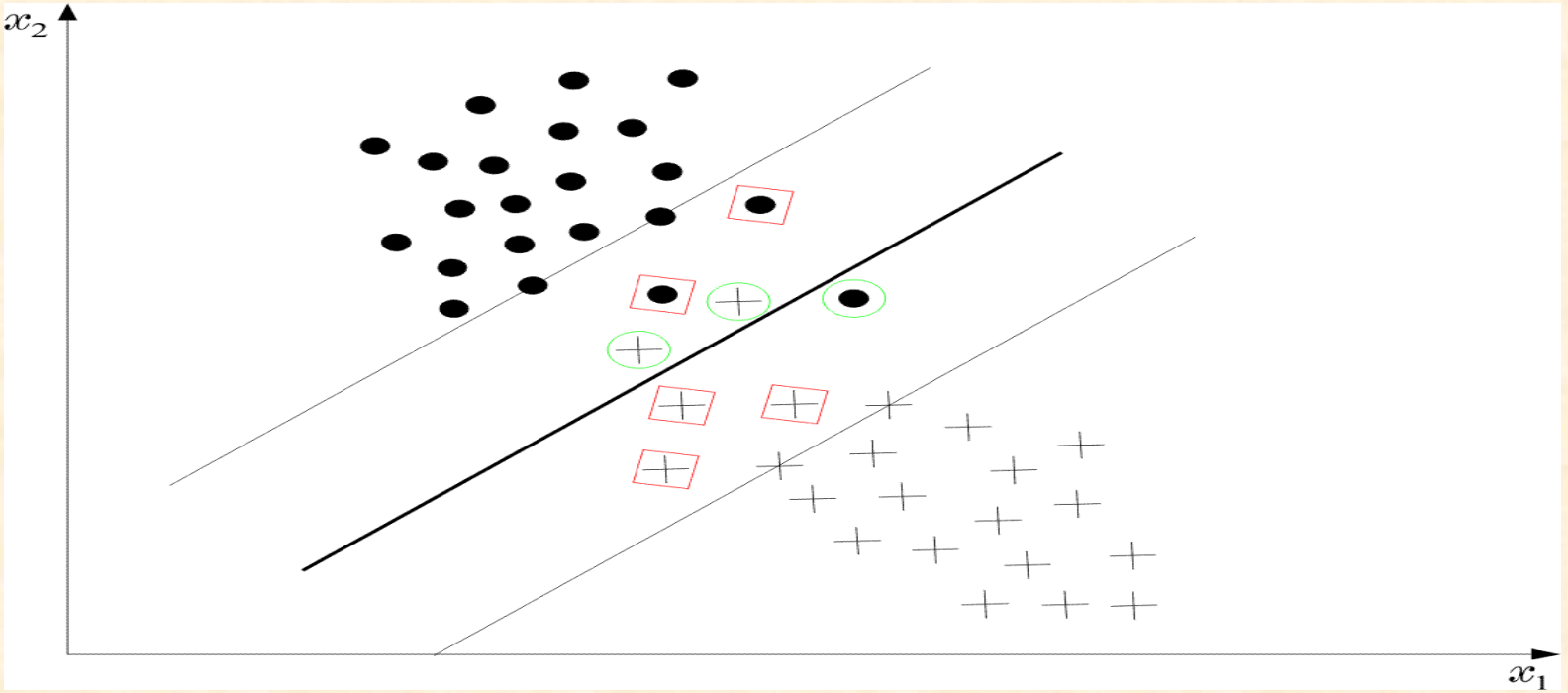
- Συνδυάζοντας τα παραπάνω παίρνουμε
  - Μεγιστοποίησε την $(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{ij} \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j)$
  - Υπό τους περιορισμούς

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

$$\underline{\lambda} \geq \underline{0}$$

- Σχόλια:
  - Τα διανύσματα στήριξης εισέρχονται μέσω εσωτερικών γινομένων
  - Παρότι η λύση, w, είναι μοναδική, οι πολ/στές Lagrange ΔΕΝ ΕΙΝΑΙ.

- Non-Separable classes



Στην περίπτωση αυτή δεν υπάρχει υπερεπίπεδο, ώστε

$$\underline{w}^T \underline{x} + w_0 (>)1, \ \forall \underline{x}$$

- Θυμηθείτε ότι το περιθώριο ορίζεται ως η απόσταση μεταξύ των δύο ακόλουθων υπερεπιπέδων

$$\underline{w}^T \underline{x} + w_0 = 1$$

και

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

- Τα διανύσματα εκπαίδευσης ανήκουν σε μία από τρεις δυνατές κατηγορίες

  - Α. Διανύσματα εκτός περιθωρίου που ταξινομούνται σωστά, δηλ.,

$$y_i(\underline{w}^T \underline{x} + w_0) > 1$$

  - Β. Διανύσματα εντός περιθωρίου, που είναι σωστά ταξινομημένα, δηλ.,

$$0 \le y_i(\underline{w}^T \underline{x} + w_0) < 1$$

  - C. Διανύσματα λανθασμένα ταξινομημένα, i.e.

$$y_i(\underline{w}^T \underline{x} + w_0) < 0$$

- Όλες οι παραπάνω περιπτώσεις μπορούν να αναπαρασταθούν από

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

A. $\rightarrow \xi_i = 0$

B. $\rightarrow 0 < \xi_i \leq 1$

C. $\rightarrow 1 < \xi_i$

Οι ξ_i είναι γνωστές ως μεταβλητές χαλαρότητας (slack variables)

- Ο στόχος της βελτιστοποίησης είναι τώρα διττός
  - Μεγιστοποίηση περιθωρίου
  - Ελαχιστοποίηση του αριθμού των διανυσμάτων με $\xi_i > 0$ Δηλαδή,

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

όπου C σταθερά και

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

  - I(.) δεν είναι παραγωγίσιμη. Στην πράξη, χρησιμοποιούμε μία προσέγγιση
  - $J(\underline{w}, w_0, \underline{\xi}) = \frac{1}{2}\|\underline{w}\|^2 + C\sum_{i=1}^{N} \xi_i$
  - Εργαζόμενοι όπως και προηγουμένως παίρνουμε

- ΚΚΤ συνθήκες

$$(1)\ \ \underline{w} = \sum_{i=1}^{N} \lambda_i y_i \underline{x}_i$$

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

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

$$(4)\ \ \lambda_i [y_i (\underline{w}^T \underline{x}_i + w_0) - 1 + \xi_i] = 0, \quad i = 1,2,...,N$$

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

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

- Το σχετικό δυϊκό πρόβλημα

Μεγιστοποίησε $\lambda - (\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j)$

υπό τις προϋποθέσεις

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

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

- **<u>Σχόλιο</u>**:

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