ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ (PATTERN RECOGNITION)

Σέργιος Θεοδωρίδης

Κωνσταντίνος Κουτρούμπας

Version 2

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) $w_1, w_2, \ldots, w_0$

έτσι ώστε η διαφορά ανάμεσα

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

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$_{\boldsymbol{x}}$: Πίν. αυτοσυσχέτισης - **autocorrelation matrix**

E[***x****d*] : Διάν. Ετεροσυσχέτ.-**cross-correlation vector**

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

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

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

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

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 w που ελαχιστοποιεί τη συνάρτηση κόστους

- $J(w) = \sum_{i=1}^{N} (d_i - w^T x_i)^2$

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

Minimizing *J*(***w***) with respect to ***w*** results in:

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

$$\boxed{(\sum_{i=1}^{N} x_i x_i^{\,T}) w = \sum_{i=1}^{N} x_i d_i}$$

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ
Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

<u>Ένας εναλλακτικός τρόπος διατύπωσης:</u>

Ορίζουμε τα ακόλουθα

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \dots \\ x_N^T \end{bmatrix}$$ *N*x*l* πίνακας

$$\underline{\mathrm{d}} = \begin{bmatrix} d_1 \\ \dots \\ d_N \end{bmatrix}$$ Αντίστοιχες επιθυμητές αποκρίσεις

$$X^T = [x_1, x_2, \dots, x_N]$$ *l*x*N* 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

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

$$\hat{w} = (X^T X)^{-1} X^T d = X^{\neq} d$$

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

Ψευδοαντίστροφος (pseudoinverse) του *X*

Έστω *N=l*. Τότε ο *X* είναι τετραγωνικός και (γενικά) αντιστρέψιμος. Τότε έχουμε

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

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

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

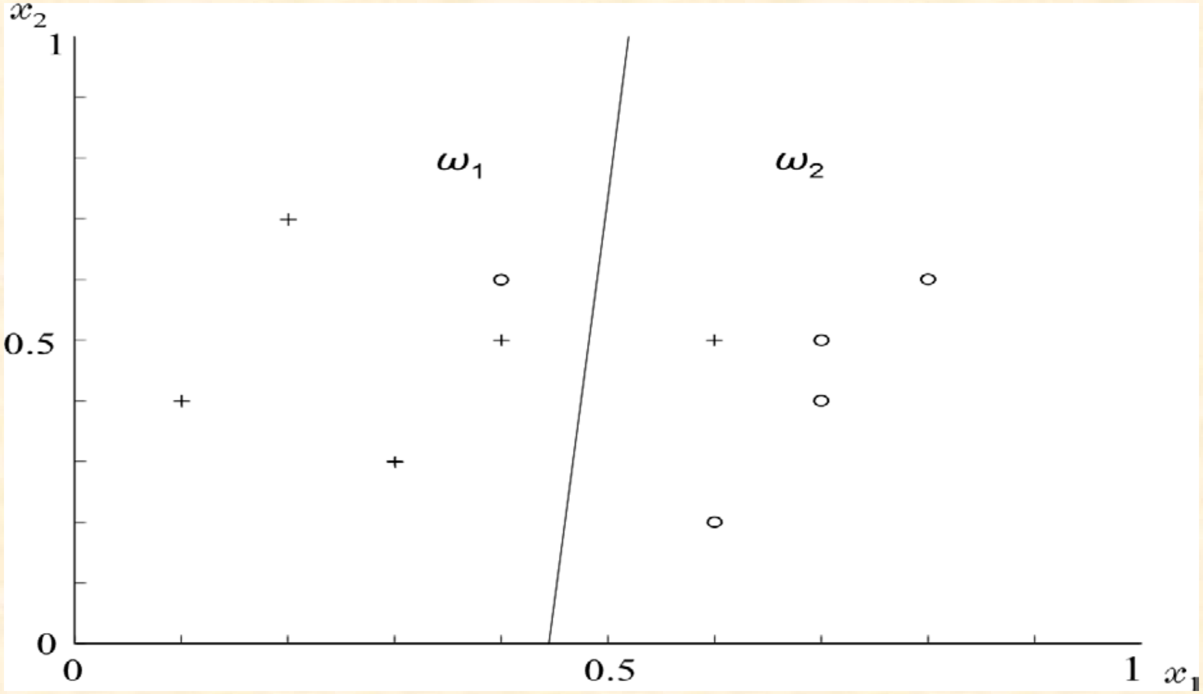
Σ' αυτή την περίπτωση, η "λύση" $w = X^{\neq} d$ ελαχιστοποιεί το άθροισμα των τετραγώνων των σφαλμάτων.

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

Μέθοδοι ελαχίστων τετραγώνων – μικρά ως προς το SSE

Ένα παράδειγμα:

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



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

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

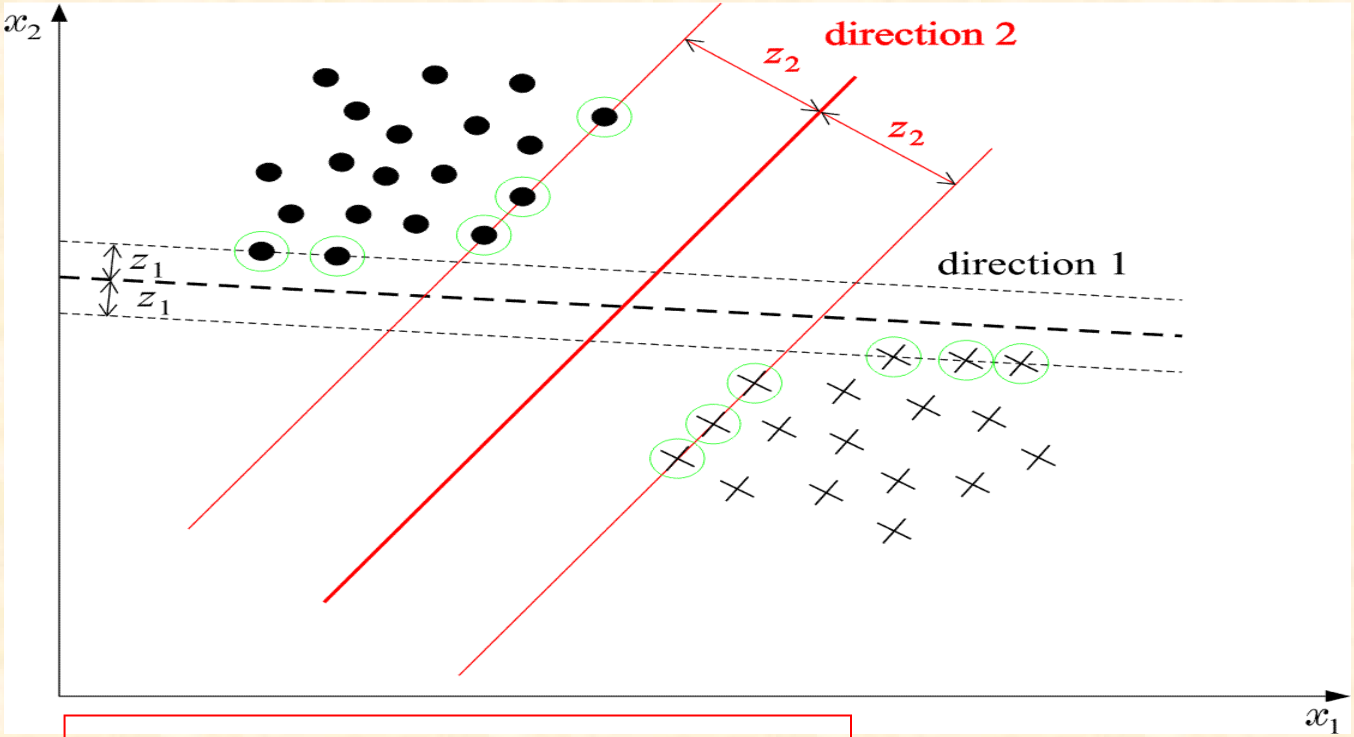
ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

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 $\boldsymbol{w}$
- Τη θέση του χώρο, i.e., w_0

- Για ΚΑΘΕ προσανατολισμό, $\boldsymbol{w}$, 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

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

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

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

–Τότε το περιθώριο είναι

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

–Επίσης, ισχύουν και οι ακόλουθοι περιορισμοί

$$w^T x + w_0 \geq 1 \ \forall x \in \omega_1$$
$$w^T x + w_0 \leq -1 \ \forall x \in \omega_2$$

(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

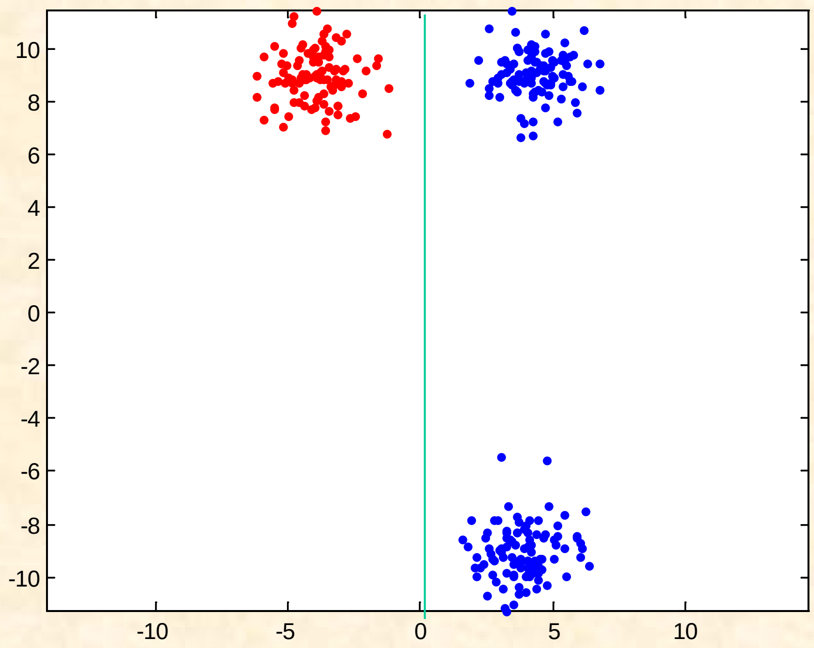
$$d_i(\underline{w}^T \underline{x}_i + w_0) \geq 1, \ i = 1,2,\ldots,N \qquad \begin{array}{l} d_i = 1, \text{ for } \underline{x}_i \in \omega_i, \\ d_i = -1, \text{ for } \underline{x}_i \in \omega_2 \end{array}$$

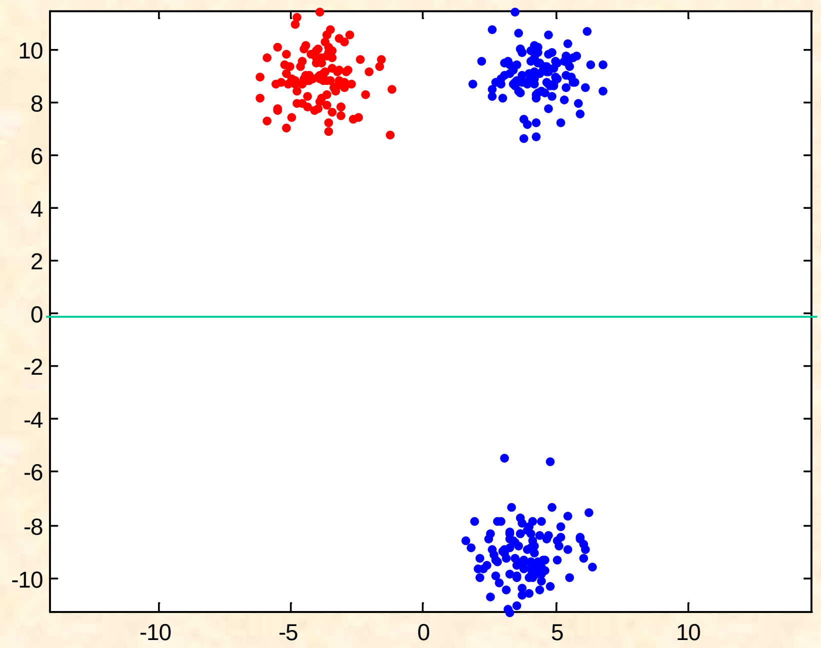
– Τα παραπάνω δικαιολογούνται από το γεγονός ότι ελαχιστοποιώντας το $\|\underline{w}\|^2$

μεγιστοποιείται το περιθώριο $\dfrac{2}{\|w\|}$.

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

Min $J(\vartheta)$

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

Ορίζουμε τη συνάρτηση Lagrange

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

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

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

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

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

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΣΥΝΑΡΤΗΣΗΣ ΥΠΟ ΓΡΑΜ. ΑΝΙΣΟΤΙΚΟΥΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

Έστω

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

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

Πρωτογενές (primal) πρόβλημα

Δυϊκό (dual) πρόβλημα

$$L(\vartheta^*,\lambda^*) = \underbrace{\min_\vartheta \max_{\lambda \geq 0} L(\vartheta,\lambda)}_{\text{primal}} = \underbrace{\max_{\lambda \geq 0} \min_\vartheta L(\vartheta,\lambda)}_{\text{dual}}$$

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 πλαίσιο του **SVM** λύνουμε το **δυϊκό πρόβλημα:** θα **ελαχιστοποιήσουμε** την $L(.)$ πρώτα ως προς ϑ και μετά θα την **μεγιστοποιήσουμε** 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) $$\frac{\partial}{\partial w}\mathrm{L}(w, w_0, \lambda) = 0$$
- (2) $$\frac{\partial}{\partial w_0} L(w, w_0, \lambda) = 0$$
- (3) $$\lambda_i \geq 0, i = 1,2,\ldots, N$$
- (4) $$\lambda_i\left[d_i(w^T x_i + w_0) - 1\right] = 0, i = 1,2,\ldots, N$$
- Όπου $L(\bullet,\bullet,\bullet)$ είναι η συνάρτηση Lagrange

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

ΓΡΑΜΜΙΚΟΙ ΠΑΡΑΜΕΤΡΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ ΕΠΙΦΑΝΕΙΩΝ ΑΠΟΦΑΣΗΣ

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

– Η λύση: από τα παραπάνω προκύπτει ότι είναι

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

Αντικαθιστώντας τα παραπάνω στην $L(.)$, τα λ_i's εκτιμώνται ως οι λύσεις του ακόλουθου προβλήματος:

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 $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_N]^T$

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

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

ΣΗΜ.: Το πρόβλημα λύνεται συνήθως με χρήση τεχνικών τετραγωνικού προγραμματισμού (quadratic programming – QP)

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_is έχουν τα διανύσματα που ικανοποιούν τη συνθήκη $w^T x_i + w_0 = \pm 1$ λόγω των περιορισμών $\lambda_i [d_i (\underline{w}^T \underline{x}_i + w_0) - 1] = 0, \quad i = 1,2,\ldots, N$

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***.

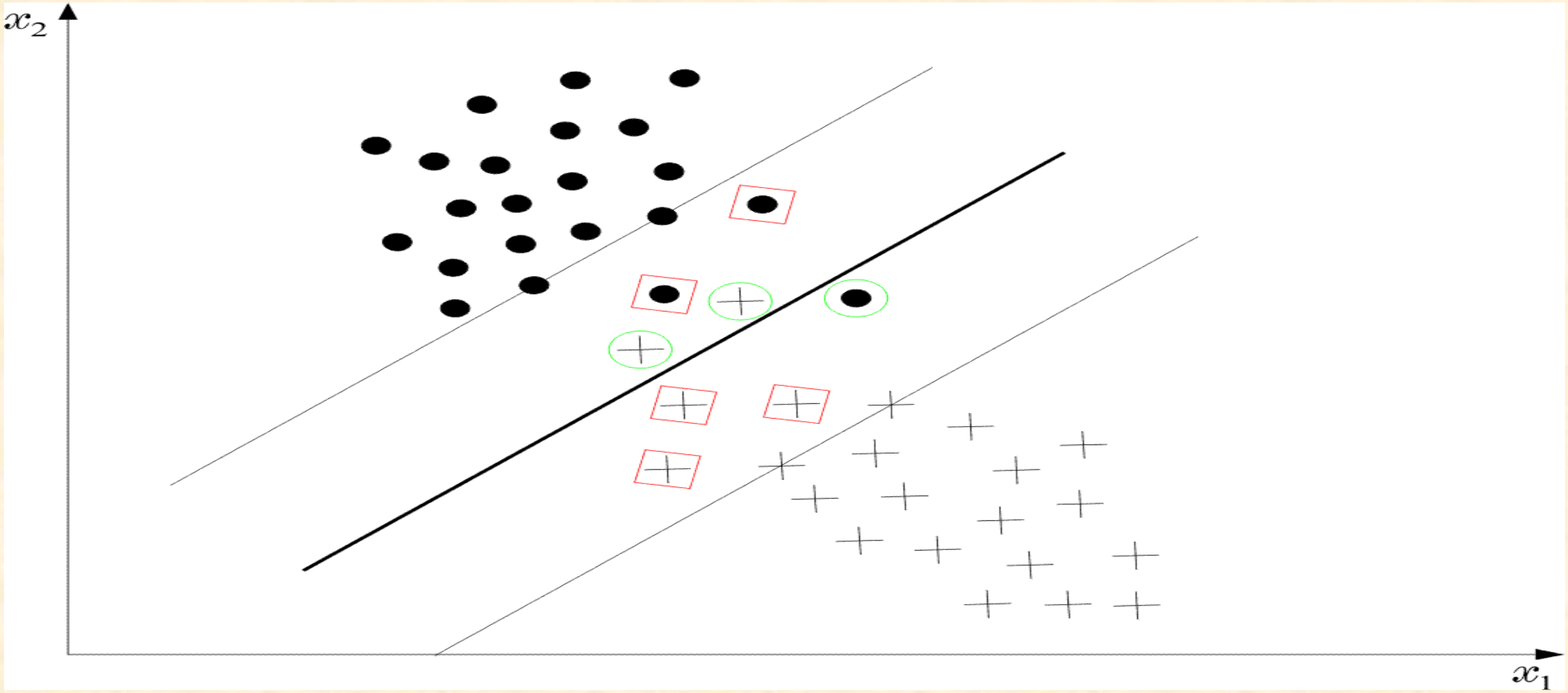
- Μετά τον προσδιορισμό του ***w***, το w_0 προσδιορίζεται από τις συνθήκες (4).

- 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

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

- 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

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

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

1ο σενάριο $\rightarrow \xi_i = 0$

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

3ο 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

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

όπου *C* είναι μια σταθερά και

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

•Η συνάρτηση *I*(.) 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 KKT συνθήκες είναι:

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

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

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

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

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

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

- Αντικαθιστώντας τα παραπάνω στην L(.), οι πολ/στές Lagrange προσδιορίζονται ως οι λύσεις του ακόλουθου δυϊκού SVM προβλήματος

Μεγιστοποίησε $_{\lambda \geq 0}(\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{ij} \lambda_i \lambda_j d_i d_j x_i^T x_j)$ ως προς το $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_N]^T$

υπό τις προϋποθέσεις

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

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

- **<u>Σχόλιο</u>**:

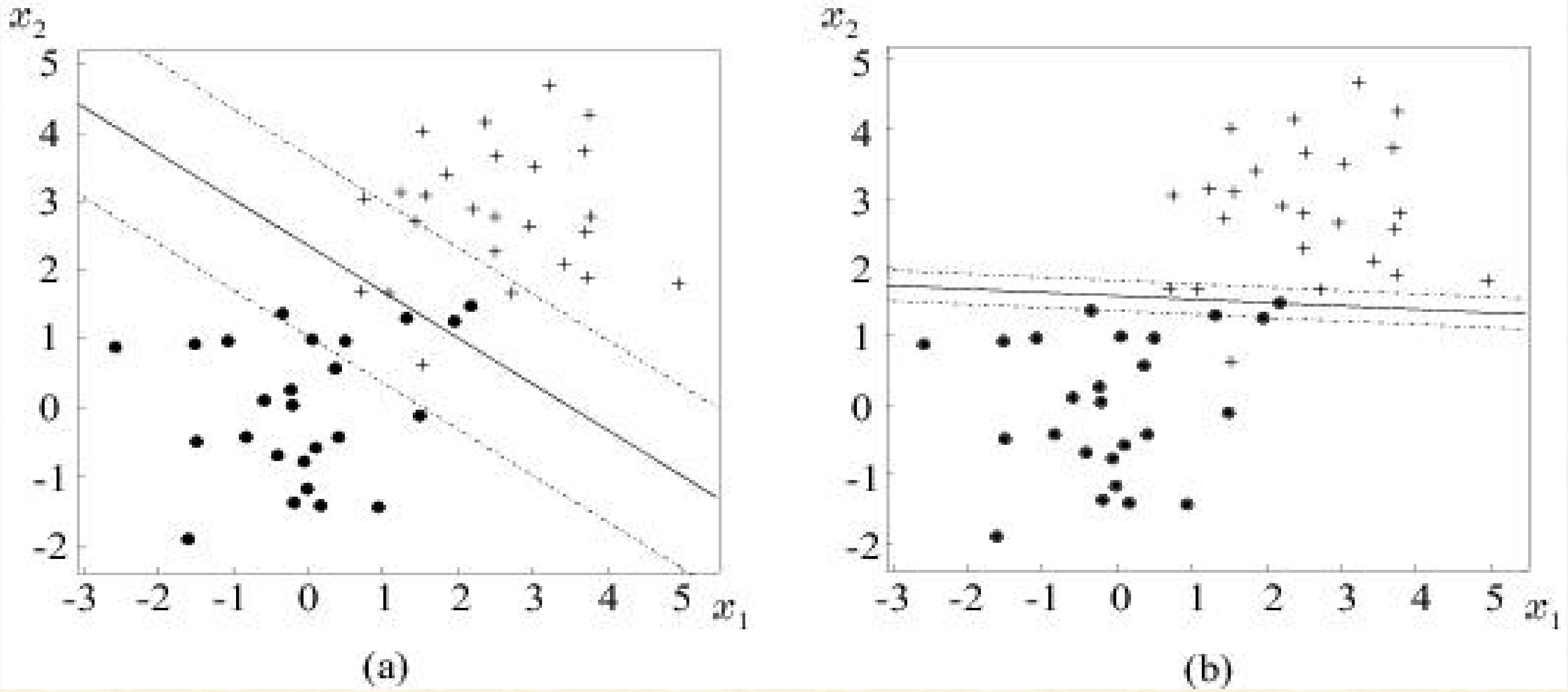
Η μόνη διαφορά με την περίπτωση των διαχωρίσιμων κλάσεων είναι η ύπαρξη του *C* στους περιορισμούς.

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.

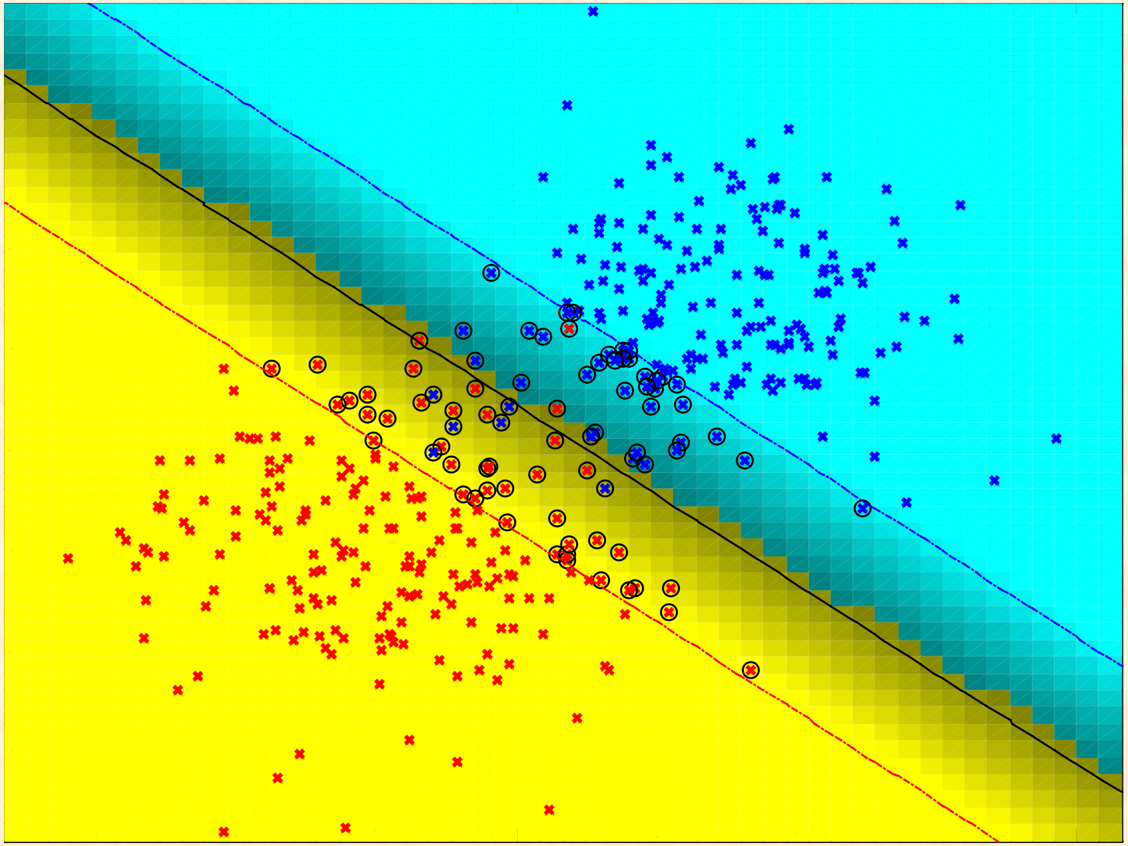
Ωστόσο, στην παρούσα περίπτωση, η παράμετρος *C* επηρεάζει την επιλογή της τελικής λύσης.



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

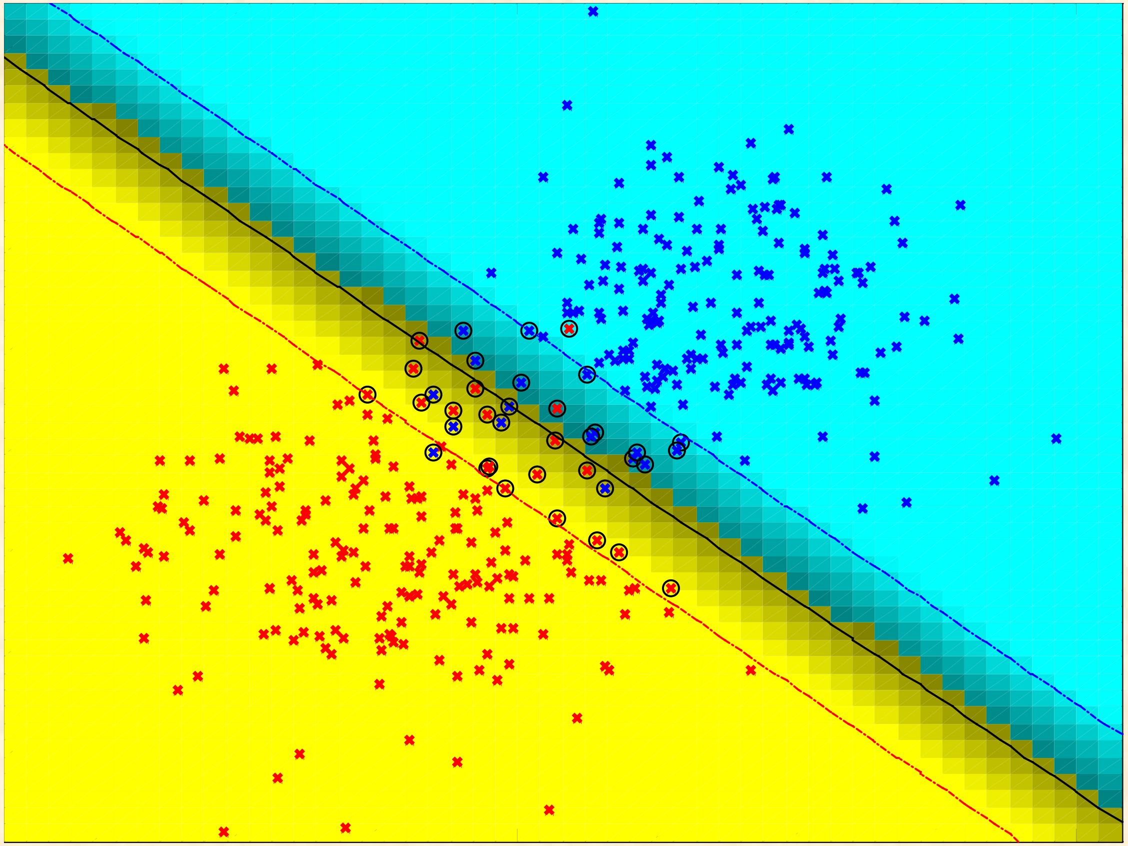
Pe_te =0.0325

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

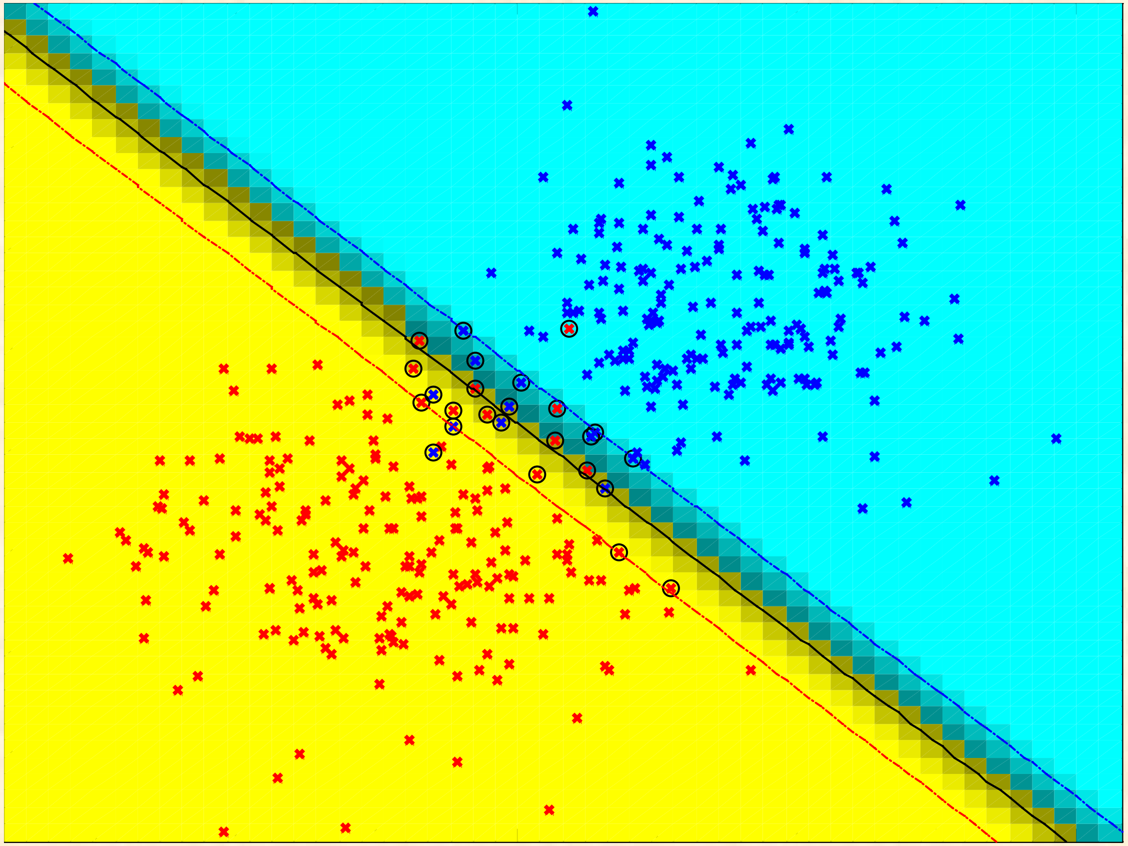
Pe_te =0.0325

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