ΚΕΣ 03 – Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Support Vector Machines

Τμήμα Επιστήμης και Τεχνολογίας Τηλεπικοινωνιών

Πανεπιστήμιο Πελοποννήσου

Περιεχόμενα – Βιβλιογραφία

- Περιεχόμενα Ενότητας
  - Εισαγωγή
  - Γραμμικά διαχωρίσιμες κλάσεις
  - Μη γραμμικά διαχωρίσιμες κλάσεις
  - Χρήση συναρτήσεων πυρήνα (kernel functions)
- Βιβλιογραφία:
  - Duda [2004]: Chapter 5
  - Theodoridis [2002]: Chapter 3

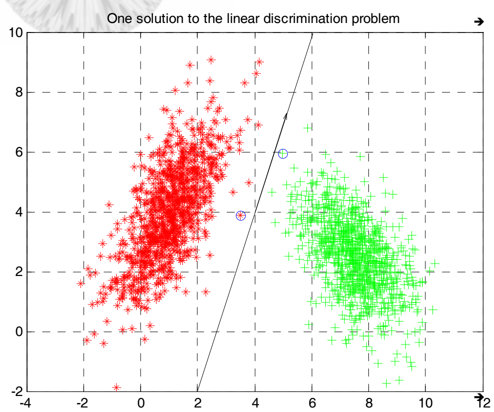
Εισαγωγή

- Έστω ένα πρόβλημα ταξινόμησης σε δύο κλάσεις ω_1, ω_2
  - Αν τα διανύσματα $\underline{x}_i \in R^l$ i=1,2,…,N είναι γραμμικά διαχωρίσιμα τότε υπάρχει μια γραμμική συνάρτηση των *l*-στοιχείων του $\underline{x}$ της μορφής
    $$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 = 0 =$$
    $$w_1 x_1 + w_2 x_2 + \ldots + w_l x_l + w_0$$
    η οποία διαχωρίζει τις δύο κλάσεις.
  - Το διάνυσμα παραμέτρων $\underline{w} = [w_1, w_2, \ldots, w_l]^T$ καθώς και το κατώφλι w_0 προσδιορίζονται από μια διαδικασία μάθησης με βάση τα διανύσματα εκπαίδευσης
  - Η συνάρτηση διαχωρισμού μπορεί να υπολογιστεί μέσω του αλγορίθμου Perceptron
  - Η ταξινόμηση πραγματοποιείται:
    $$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 > 0 \quad \underline{x} \in \omega_1$$
    $$g(\underline{x}) = \underline{w}^T \underline{x} + w_0 < 0 \quad \underline{x} \in \omega_2$$

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

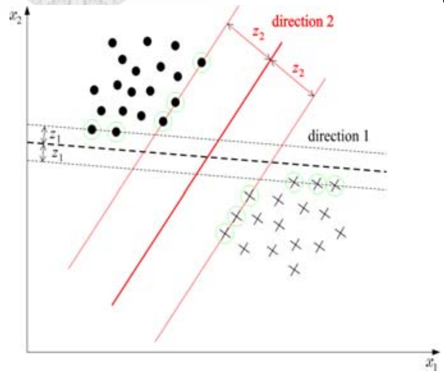
★ Εισαγωγή
☐ Γραμμικά διαχωρίσιμες κλάσεις
☐ Μη γραμμικά διαχωρίσιμες κλάσεις
☐ Kernel functions

Πολλαπλές λύσεις



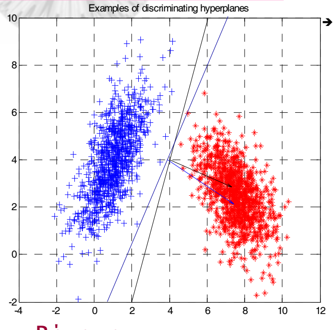
- Ο αλγόριθμος Perceptron δεν εγγυάται μοναδική λύση επειδή:
  - Χρησιμοποιείται τυχαία αρχικοποίηση των βαρών
  - Το κριτήριο κόστους ελαχιστοποιεί τα σφάλματα κόστους (και όχι κάποια γεωμετρική απόσταση όπως π.χ. συμβαίνει με τον αλγόριθμο των ελαχίστων τετραγώνων)
- Αν θέλαμε να επιλέξουμε την καλύτερη από τις λύσεις τι θα κάναμε;

Γραμμικά διαχωρίσιμες κλάσεις

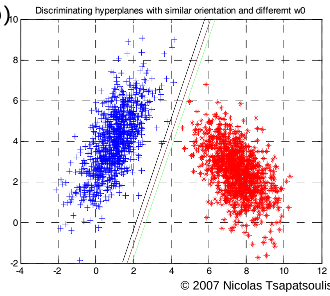


- Η μεθοδολογία SVM (Support Vector Machines) επιλέγει εκείνη την υπερεπιφάνεια διαχωρισμού η οποία:
  - Ισαπέχει από τα πλησιέστερα (σε αυτήν) διανύσματα των δύο κλάσεων. Τα διανύσματα αυτά είναι γνωστά ως διανύσματα υποστήριξης (support vectors)
  - Μεγιστοποιεί την απόσταση των διανυσμάτων υποστήριξης από την υπερεπιφάνεια διαχωρισμού. Το διπλάσιο της απόστασης αυτής ονομάζεται περιθώριο (margin)

Περιθώριο



- Είδαμε στους γραμμικούς ταξινομητές ότι κάθε υπερεπιφάνεια διαχωρισμού χαρακτηρίζεται από:
  - Την κατεύθυνση της όπως προσδιορίζεται από το κάθετο στην επιφάνεια διάνυσμα w (βλέπε σχήμα αριστερά)
  - Την απόσταση της από την αρχή των αξόνων όπως προσδιορίζεται από το κατώφλι w_0. (βλέπε σχήμα κάτω)



Βήματα:

- Για κάθε κατεύθυνση τοποθετούμε την επιφάνεια έτσι ώστε η ελάχιστη απόσταση από τις δύο κλάσεις να είναι ίδια
- Ανάμεσα σε όλες τις υπερεπιφάνειες που διαχωρίζουν τα δεδομένα μας επιλέγουμε εκείνη που μεγιστοποιεί το περιθώριο

Γραμμικός ταξινομητής SVM

- Η απόσταση z_x ενός τυχαίου σημείου **x** από την υπερεπιφάνεια διαχωρισμού δίνεται από τη σχέση (αρνητική απόσταση σημαίνει ότι το **x** ανήκει στην κλάση ω_2): $z_x = \frac{g(\mathbf{x})}{\|\underline{w}\|}, \quad \|\underline{w}\| = \sqrt{\underline{w}^T \underline{w}}$
- Μπορούμε πάντοτε να τροποποιήσουμε τα βάρη $\underline{w}$, w_0 ώστε η απόσταση των πλησιέστερων στην επιφάνεια σημείων (εκατέρωθεν αυτής) να είναι ίση με ± 1:

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

  - Το περιθώριο σε αυτή την περίπτωση δίνεται από τη σχέση:

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

  - Ισχύει επίσης:

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

Γραμμικός ταξινομητής SVM (II)

- Για την εύρεση της βέλτιστης υπερεπιφάνειας διαχωρισμού σύμφωνα με την μεθοδολογία SVM ελαχιστοποιούμε:
  - Κριτήριο:

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

  - Υποκείμενο στους περιορισμούς:

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

- Ελαχιστοποίηση της ποσότητας $\|\underline{w}\|$ μεγιστοποιεί το περιθώριο $\frac{2}{\|w\|}$.
- Οι περιορισμοί διασφαλίζουν ότι δεν υπάρχει σφάλμα ταξινόμησης (υπό την προϋπόθεση ότι οι κλάσεις είναι γραμμικά διαχωρίσιμες)

Γραμμικός ταξινομητής SVM (III)

- Το προηγούμενο είναι ένα πρόβλημα τετραγωνικής βελτιστοποίησης υποκείμενο σε γραμμικούς περιορισμούς υπό τη μορφή ανισοτήτων.
- Για τέτοιου είδους προβλήματα οι συνθήκες ΚΚΤ (Karush – Kuhh – Tucker) ορίζουν ότι η βέλτιστη λύση ικανοποιεί τις συνθήκες:
  - (1) $\frac{\partial}{\partial \underline{w}} \mathrm{L}(\underline{w}, w_0, \underline{\lambda}) = \underline{0}$
  - (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) $\lambda_i \left(y_i (\underline{w}^T \underline{x}_i + w_0) - 1 \right) = 0, i = 1,2,\ldots, N$
- Η συνάρτηση L(•,•,•) είναι μια συνάρτηση Langrange:

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

Γραμμικός ταξινομητής SVM (IV)

- Η λύση του ανωτέρου προβλήματος ικανοποιεί τις σχέσεις (εφαρμογή των σχέσεων (1), (2) στη συνάρτηση Langrange):

$$\underline{w} = \sum_{i=1}^{N} \lambda_i y_i \underline{x}_i \qquad \sum_{i=1}^{N} \lambda_i y_i = 0$$

και υπόκειται στους περιορισμούς:

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

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

- Στις παραπάνω σχέσεις προκύπτει μια μοναδική λύση για το διάνυσμα βαρών $\underline{w}$ η οποία όμως μπορεί να αντιστοιχεί σε διάφορες τιμές των πολλαπλασιαστών Langrange (λ_i)

Παρατηρήσεις σχετικά με τη βέλτιστη λύση

- Οι πολλαπλασιαστές Langrange (λ_i) μπορεί να είναι είτε θετικοί είτε μηδενικοί. Επομένως:

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

όπου Ns≤N είναι ο αριθμός των μη μηδενικών πολλαπλασιαστών Langrange.

- Από τη σχέση: $\lambda_i \left(y_i (\underline{w}^T \underline{x}_i + w_0) - 1 \right) = 0,\, i = 1,2,\ldots,\, N$

προκύπτει ότι τα διανύσματα τα οποία συνεισφέρουν στον υπολογισμό του w σύμφωνα με τη σχέση (1) ανωτέρω πληρούν τη σχέση:

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

είναι δηλαδή διανύσματα υποστήριξης (support vectors)

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
★ Γραμμικά διαχωρίσιμες κλάσεις
☐ Μη γραμμικά διαχωρίσιμες κλάσεις
☐ Kernel functions

Παρατηρήσεις σχετικά με τη βέλτιστη λύση (ΙΙ)

- Εφόσον υπολογιστεί το διάνυσμα w το κατώφλι w_0 υπολογίζεται από τις σχέσεις:

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

- Οι επίλυση του προηγούμενου προβλήματος τετραγωνικής βελτιστοποίησης πραγματοποιείται με επαναληπτικές μεθόδους βελτιστοποίησης διότι δεν υπάρχει μοναδική λύση για τους πολλαπλασιαστές Langrange

- Μια πιο απλή μορφή του προβλήματος μπορεί να διατυπωθεί με τη βοήθεια της αρχής της δυαδικότητας (duality principle)

Δυαδικό πρόβλημα εύρεσης της βέλτιστης λύσης

- Με τη μεθοδολογία SVM που περιγράφηκε νωρίτερα ορίζεται ένα πρόβλημα βελτιστοποίησης με:
  - Κυρτή συνάρτηση κόστους (μοναδικό ελάχιστο)
  - Κυρτή περιοχή πιθανών λύσεων
- Οι δύο προηγούμενες συνθήκες μας εξασφαλίζουν ότι το πρόβλημα μπορεί να λυθεί με δυαδικό τρόπο (μεγιστοποίηση της συνάρτησης Langrange ως προς λ αντί ελαχιστοποίηση ως προς w, w_0):

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

  - Μεγιστοποίηση της $L(\underline{w}, w_0, \underline{\lambda})$ ως προς λ $\underline{\lambda}^* = \arg\max_{\underline{\lambda}} \left(L(\underline{w}, w_0, \underline{\lambda}) \right)$

  Υποκείμενη στους περιορισμούς:

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

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

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

Δυαδικό πρόβλημα εύρεσης της βέλτιστης λύσης (ΙΙ)

- Με συνδυασμό των πρώτων δύο σχέσεων:

$$\underline{\lambda}^* = \arg\max_{\underline{\lambda}} \left(L(\underline{w}, w_0, \underline{\lambda}) \right)$$

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

Καταλήγουμε στο πρόβλημα: $\underline{\lambda}^* = \arg\max_{\underline{\lambda}} (\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j)$

υποκείμενο στους περιορισμούς: $\sum_{i=1}^{N} \lambda_i y_i = 0$

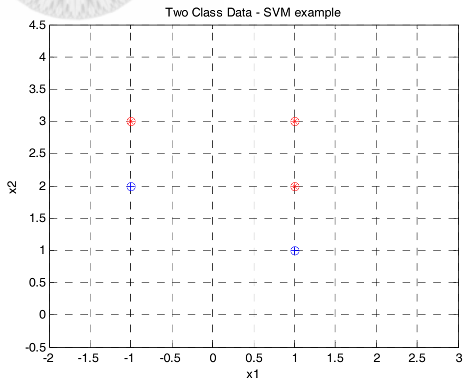
$$\underline{\lambda} \geq \underline{0}$$

το οποίο είναι εμφανώς απλούστερο στη λύση από το αρχικό

- Παρατηρήστε ότι στη σχέση $\underline{\lambda}^* = \arg\max_{\underline{\lambda}} (\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j)$

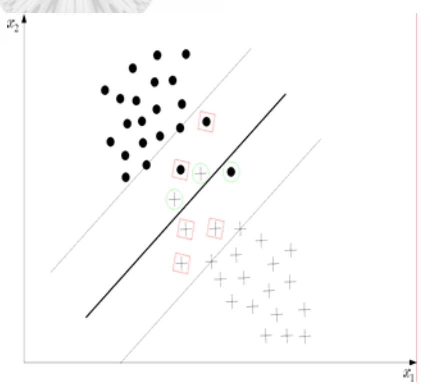
  τα διανύσματα υποστήριξης υπεισέρχονται ανά ζεύγη

Παράδειγμα



- Να βρεθεί η συνάρτηση διαχωρισμού των κλάσεων του σχήματος (ω_1 => κόκκινα πρότυπα) με τη μεθοδολογία SVM.
  - Θεωρήστε ότι είναι γνωστό ότι η γραμμή διαχωρισμού περνά από το σημείο [0 2]
  - Προχωρήστε στην διατύπωση και επίλυση των συνθηκών ΚΚΤ
  - Τι θα συνέβαινε αν δεν γνωρίζαμε ότι το σημείο ικανοποιεί την εξίσωση της συνάρτησης διαχωρισμού;

Μη γραμμικά διαχωρίσιμες κλάσεις

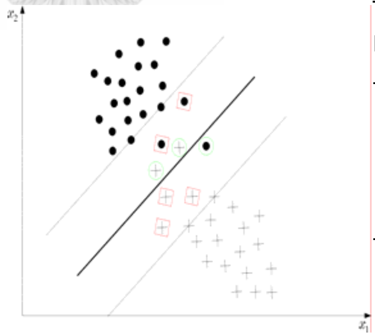


- Μπορεί η μεθοδολογία SVM να εφαρμοστεί σε μη γραμμικά διαχωρίσιμες κλάσεις ω_1, ω_2 (όπως αυτές του σχήματος);
  - Προφανώς σε αυτή την περίπτωση δεν υπάρχει υπερεπιφάνεια για την οποία να ισχύει:

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

☑ Εισαγωγή
☑ Γραμμικά διαχωρίσιμες κλάσεις
★ Μη γραμμικά διαχωρίσιμες κλάσεις
☐ Kernel functions

Κατάταξη διανυσμάτων εκπαίδευσης ως προς το περιθώριο



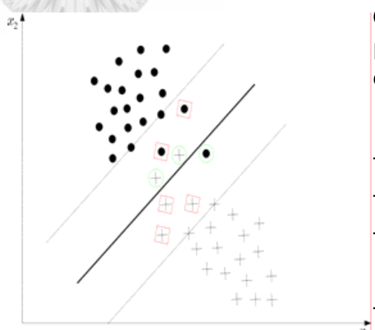
Τα διανύσματα εκπαίδευσης υπάγονται σε μια από τις τρεις επόμενες κατηγορίες:

- Διανύσματα εκτός της περιοχής που ορίζεται από το περιθώριο τα οποία έχουν ταξινομηθεί ορθά. Για τα διανύσματα αυτά ισχύει η σχέση:
$$y_i(\underline{w}^T \underline{x} + w_0) > 1$$
- Διανύσματα εντός της περιοχής που ορίζεται από το περιθώριο τα οποία έχουν ταξινομηθεί ορθά. Για τα διανύσματα αυτά ισχύει η σχέση:
$$0 \le y_i(\underline{w}^T \underline{x} + w_0) < 1$$
- Εσφαλμένα ταξινομημένα διανύσματα. Σε αυτή την περίπτωση ισχύει η σχέση:
$$y_i(\underline{w}^T \underline{x} + w_0) < 0$$

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
☑ Γραμμικά διαχωρίσιμες κλάσεις
★ Μη γραμμικά διαχωρίσιμες κλάσεις
☐ Kernel functions

Κατάταξη διανυσμάτων εκπαίδευσης ως προς το περιθώριο (II)



Οι τρεις προηγούμενες περιπτώσεις μπορούν να περιγραφούν από μια ενιαία σχέση:
$$y_i(\underline{w}^T \underline{x} + w_0) \ge 1 - \xi_i$$

- (1) $\xi_i = 0$
- (2) $0 < \xi_i \le 1$
- (3) $1 < \xi_i$

- Οι μεταβλητές ξ_i είναι γνωστές ως «χαλαρές» μεταβλητές (slack variables)

Εύρεση βέλτιστης λύσης

- Η βελτιστοποίηση τώρα αφορά δύο επίπεδα
  - Μεγιστοποίηση του περιθωρίου
  - Ελαχιστοποίηση του αριθμού των προτύπων για τα οποία ισχύει $\xi_i > 0$ (δηλαδή πρότυπα είτε εσφαλμένα ταξινομημένα είτε εντός της περιοχής του περιθωρίου)
- Σύμφωνα με τα προηγούμενα μια κατάλληλη συνάρτηση κόστους είναι:

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

όπου *C* είναι μια σταθερά που καθορίζει τη βαρύτητα της εσφαλμένης ταξινόμησης στη συνολική βελτιστοποίηση και

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

Εύρεση βέλτιστης λύσης (II)

- Στη προηγούμενη συνάρτηση κόστους η μεταβλητή *I(·)* δεν είναι παραγωγίσιμη με αποτέλεσμα να μην είναι εφικτή εύρεση της βέλτιστης λύσης μέσω των συνθηκών ΚΚΤ.
  - Για να αντιμετωπιστεί το προηγούμενο προσεγγίζουμε τη προηγούμενη συνάρτηση κόστους με τη συνάρτηση:

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

η οποία είναι παραγωγίσιμη

- Εφαρμόζοντας όμοια μεθοδολογία ορίζουμε τη συνάρτηση Langrange με δύο κατηγορίες πολλαπλασιαστών (μ_i, λ_i):

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

Εύρεση βέλτιστης λύσης (III)

- Οπότε οι συνθήκες ΚΚΤ υπολογίζονται με παρόμοιο τρόπο όπως προηγουμένως :

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

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

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

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

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

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

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

☑ Εισαγωγή
☑ Γραμμικά διαχωρίσιμες κλάσεις
★ Μη γραμμικά διαχωρίσιμες κλάσεις
☐ Kernel functions

Δυαδικό πρόβλημα

- Το δυαδικό πρόβλημα ορίζεται όπως στην περίπτωση των γραμμικά διαχωρίσιμων κλάσεων και καταλήγει στις σχέσεις:

$$\underline{\lambda}^* = \arg\max_{\underline{\lambda}}\left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \underline{x}_i^T \underline{x}_j\right)$$

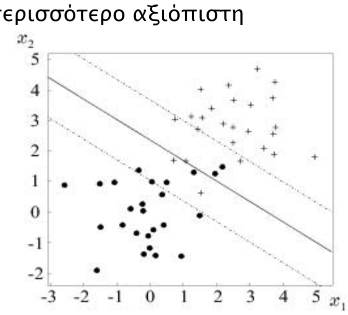
$$0 \leq \lambda_i \leq C,\ i = 1,2,\ldots,N$$

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

το οποίο διαφέρει από το αντίστοιχο των γραμμικά διαχωρίσιμων κλάσεων μόνο ως προς την παρουσία της σταθεράς C στους περιορισμούς

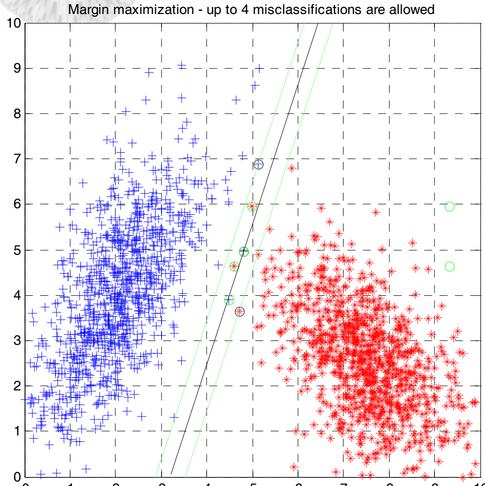
Παράδειγμα

- Στο παρακάτω σχήμα φαίνεται η επίδραση της σταθεράς C στον ορισμό του περιθωρίου ανάμεσα σε δύο μη διαχωρίσιμες κλάσεις (στο σχήμα (a) έχουμε μικρή τιμή για το C (C = 0.2) ενώ στη περίπτωση (b) σαφώς μεγαλύτερη τιμή (C = 1000))
  - Για σκοπούς βελτίωσης της γενικευτικής ικανότητας (ικανότητα σωστής ταξινόμησης άγνωστων διανυσμάτων) η περίπτωση (a) είναι γενικά περισσότερο αξιόπιστη



(a) (b)

Παράδειγμα (ΙΙ)



- Στο παράδειγμα του σχήματος:
  - Να βρείτε τα διανύσματα υποστήριξης (support vectors)
  - Να βρείτε τις εσφαλμένες ταξινομήσεις
  - Να βρείτε τα διανύσματα που έχουν ταξινομηθεί σωστά και ανήκουν στην περιοχή του περιθωρίου

Γενίκευση σε πολλαπλές κλάσεις

- Η μεθοδολογία SVM (είτε για περίπτωση γραμμικά διαχωρίσιμων κλάσεων είτε για μη γραμμικά διαχωρίσιμες) μπορεί να εφαρμοστεί σε περίπτωση που τα πρότυπα μας ανήκουν σε Μ κλάσεις (ω_1, ω_2, …, ω_M).
  - Ο απλούστερος τρόπος είναι να θεωρήσουμε Μ προβλήματα δύο κλάσεων (πρότυπα στη κλάση ω_i, l = 1,2,…M, σε σχέση με όλα τα υπόλοιπα)
  - Εναλλακτική, και πιο πολύπλοκη μέθοδος, υπολογίζει τα περιθώρια των κλάσεων ανά δύο σχηματίζοντας M*(M-1)/2 συναρτήσεις διαχωρισμού

ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

SVM με συναρτήσεις πυρήνα

- Ένας τρόπος αντιμετώπισης μη γραμμικά διαχωρίσιμων κλάσεων είναι με μετασχηματισμό των διανυσμάτων εισόδου σε ένα χώρο μεγαλύτερης διάστασης
  - Η πιθανότητα διαχωρισμού των ω_1, ω_2 με πρότυπα διαστάσεων $\underline{x} \in R^l$ αυξάνει με την αύξηση της διάσταση των προτύπων από l σε k (k>l).
  - Έστω ο μετασχηματισμός $\underline{x} \in R^l \rightarrow \underline{z} \in R^k,\ k > l$

  Μπορούμε να εφαρμόσουμε, τότε, τη μεθοδολογία SVM στο χώρο R^k

- Το δυαδικό πρόβλημα βελτιστοποίησης γίνεται:

$$\underline{\lambda}^* = \arg\max_{\underline{\lambda}} \left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \underline{z}_i^T \underline{z}_j \right)$$

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

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

☑ Εισαγωγή
☑ Γραμμικά διαχωρίσιμες κλάσεις
☑ Μη γραμμικά διαχωρίσιμες κλάσεις
★ Kernel functions

SVM με συναρτήσεις πυρήνα (II)

- Η συνάρτηση διαχωρισμού θα είναι:

$$g(\underline{z}) = \underline{w}^T \underline{z} + w_0 = \sum_{i=1}^{N_s} \lambda_i y_i \underline{z}_i^T \underline{z}, \qquad \text{όπου } \underline{x} \to \underline{z} \in R^k$$

- Το πρόβλημα με την πιο πάνω σχέση είναι ότι περιλαμβάνει N_s εσωτερικά γινόμενα ($\underline{z}_i^T \underline{z}$) σε ένα χώρο μεγάλης διάστασης (R^k) και επομένως είναι υπολογιστικά πολύπλοκη.
- Κάτω από ορισμένες προϋποθέσεις μπορούμε να υπολογίσουμε τα ανωτέρω εσωτερικά ως συναρτήσεις των εσωτερικών γινομένων ($\underline{x}_i^T \underline{x}$) όπως φαίνεται στο επόμενο παράδειγμα:

$$\text{Έστω } \underline{x} = [x_1, x_2]^T \in R^2$$

$$\text{Μετασχηματίζοντας το } \underline{x} \to \underline{z} = \begin{bmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{bmatrix} \in R^3$$

- Είναι εύκολο να δείξουμε ότι:

$$\underline{y}_i^T \underline{y}_j = (\underline{x}_i^T \underline{x}_j)^2$$

☑ Εισαγωγή
☑ Γραμμικά διαχωρίσιμες κλάσεις
☑ Μη γραμμικά διαχωρίσιμες κλάσεις
★ Kernel functions

Θεώρημα Mercer

$$\text{Έστω } \underline{x} \to \underline{\Phi}(\underline{x}) \in H$$

- Το εσωτερικό γινόμενο στο χώρο Η δίνεται από τη σχέση:

$$\sum_r \Phi_r(\underline{x}) \Phi_r(\underline{y}) = K(\underline{x}, \underline{y})$$

όπου για τη συνάρτηση K(x,y) ισχύει: $\int K(\underline{x}, \underline{y}) g(\underline{x}) g(\underline{y}) d\underline{x} d\underline{y} \geq 0$

για κάθε συνάρτηση g($\underline{x}$) τέτοια ώστε: $\int g^2(\underline{x}) d\underline{x} < +\infty$

Η συμμετρική συνάρτηση K($\underline{x}$,$\underline{y}$) είναι μια συνάρτηση πυρήνα (kernel function)

Χρησιμότητα συναρτήσεων πυρήνα

- Το ενδιαφέρον με τις συναρτήσεις πυρήνα είναι ότι αντιστοιχούν πάντοτε σε εσωτερικό γινόμενο σε ΚΑΠΟΙΟ χώρο μικρότερης διάστασης:

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

- Παραδείγματα συναρτήσεων πυρήνα:
  - Πολυωνυμικές: $K(\underline{x}, \underline{z}) = (\underline{x}^T \underline{z} + 1)^q,\ q > 0$
  - Ακτινικές συναρτήσεις βάσης: $K(\underline{x}, \underline{z}) = \exp\left(-\frac{\|\underline{x} - \underline{z}\|^2}{\sigma^2}\right)$
  - Υπερβολική εφαπτομένη: $K(\underline{x}, \underline{z}) = \tanh(\beta \underline{x}^T \underline{z} + \gamma)$

  για κάποιες κατάλληλες τιμές των παραμέτρων β, γ

SVM με συναρτήσεις πυρήνα

- Η χρήση των συναρτήσεων πυρήνα μας βοηθά να μεταβούμε σε κάποιο χώρο υψηλότερης διάστασης στον οποίο είναι πιθανό το πρόβλημα μας να είναι γραμμικό.
  - Ο χώρος αυτός δεν μας είναι γνωστός (και στην πραγματικότητα δεν θα θέλαμε να είναι γνωστός)
- Η μεθοδολογία συνοψίζεται στα πιο κάτω βήματα:
  - Βήμα 1^0: Επιλογή κάποιων συναρτήσεων πυρήνα. Οι πιο συχνά χρησιμοποιούμενες είναι οι ακτινικές συναρτήσεις (Radial Basis Functions)
  - Βήμα 2^0: Επίλυση του δυαδικού προβλήματος

$$\underline{\lambda}^* = \arg\max_{\underline{\lambda}} \left(\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j K(\underline{x}_i^T \underline{x}_j) \right)$$

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

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

SVM με συναρτήσεις πυρήνα (II)

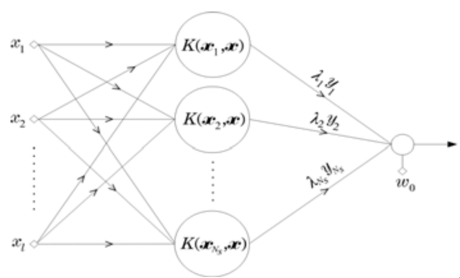
- Το βήμα 2 στην ουσία οδηγεί σε έμμεσο συνδυασμό της μορφής:

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

  - Βήμα 3^0: Το διάνυσμα $\underline{x}$ ταξινομείται σύμφωνα με τον πιο κάτω κανόνα:

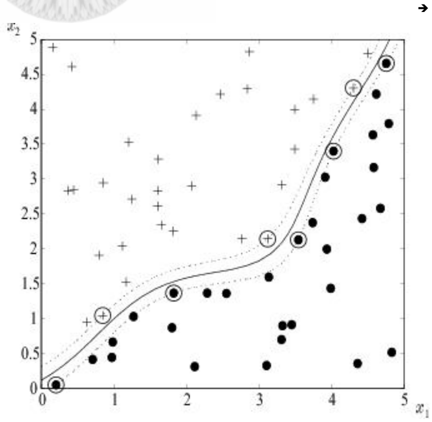
$$\omega_1(\omega_2) \text{ if } g(\underline{x}) = \sum_{i=1}^{N_s} \lambda_i y_i \, K(\underline{x}_i, \underline{x}) + w_0 > (<) 0$$

- Αρχιτεκτονική SVM:



ΚΕΣ 03: Αναγνώριση Προτύπων και Ανάλυση Εικόνας

Παράδειγμα



- Το παράδειγμα του σχήματος δείχνει πως δυο γραμμικά μη διαχωρίσιμες κλάσεις διαχωρίζονται με τη χρήση συναρτήσεων πυρήνα:
  - 8 συναρτήσεις πυρήνα (8 support vectors) – 5 αντιστοιχούν στην κλάση **o** και τρεις στην κλάση +