



Μηχανική Μάθηση

ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ



Ταξινόμηση προτύπων

2

Προβλήματα στην ταξινόμηση με νευρωνικά δίκτυα

- Η ταξινόμηση με perceptrons δουλεύει μόνο με γραμμικά διαχωρίσιμες κλάσεις
- Η ταξινόμηση με δίκτυα MLP υποφέρει από βραδεία εκπαίδευση (μην ξεχνάμε ότι στην περίπτωση αυτή λύνουμε γενικότερο πρόβλημα από αυτό της ταξινόμησης)

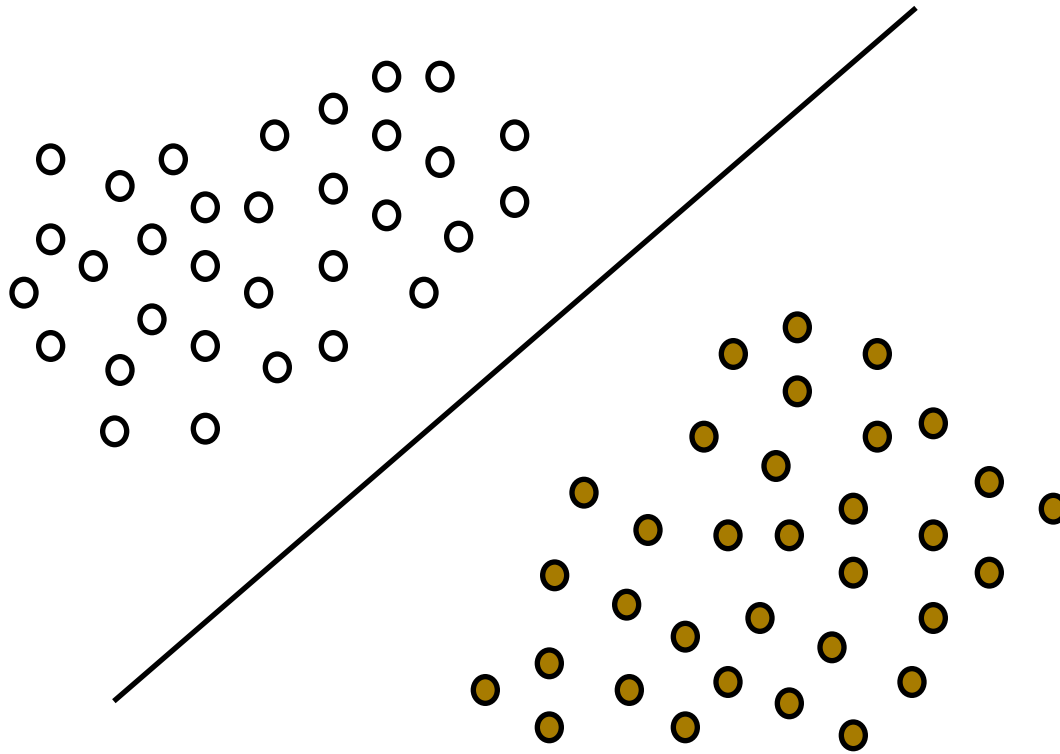
Ιδέα

- Αν επικεντρωθούμε στο πρόβλημα της ταξινόμησης μπορούμε να πετύχουμε καλύτερους χρόνους εκπαίδευσης και καλύτερες ιδιότητες γενίκευσης



Το πρόβλημα της ταξινόμησης

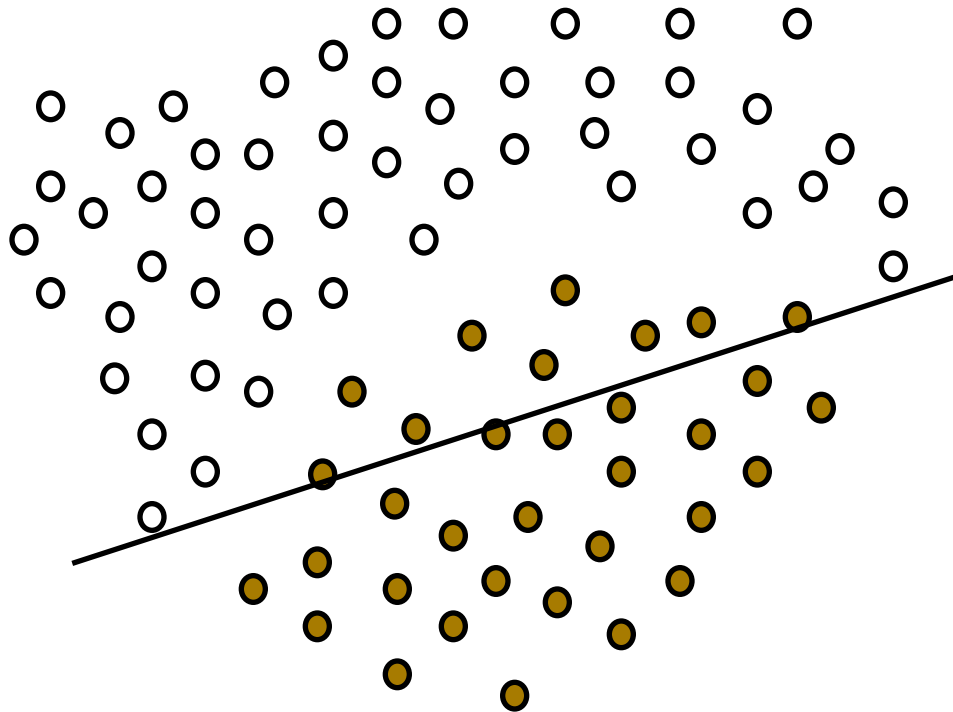
3





Μη γραμμικά διαχωρίσιμες κλάσεις

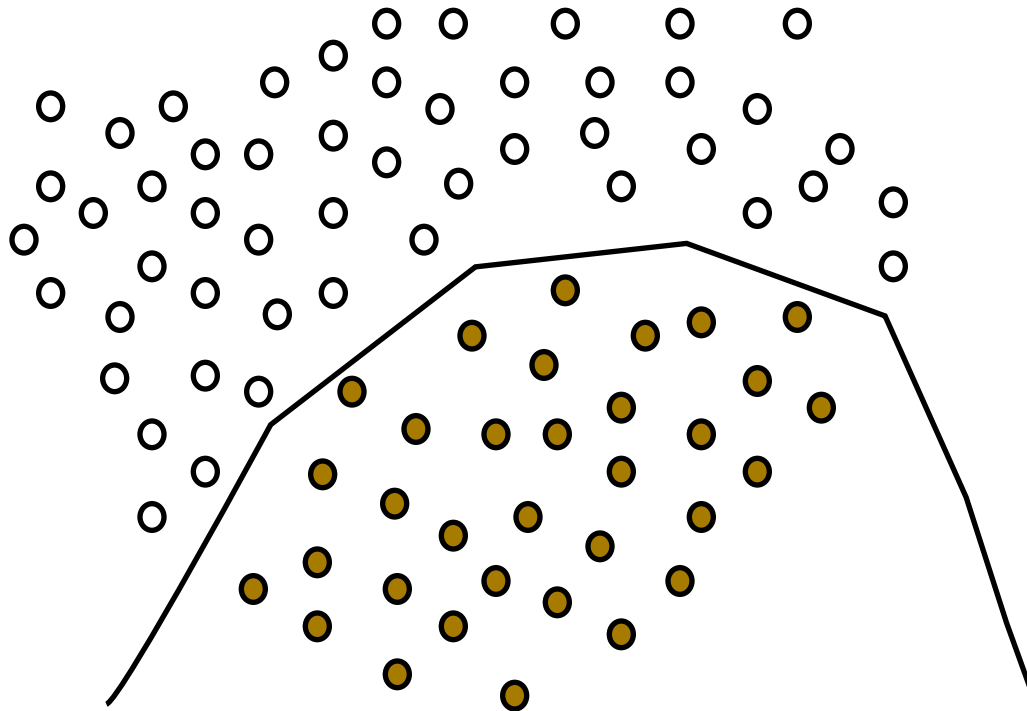
4





Μη γραμμικά διαχωρίσιμες κλάσεις

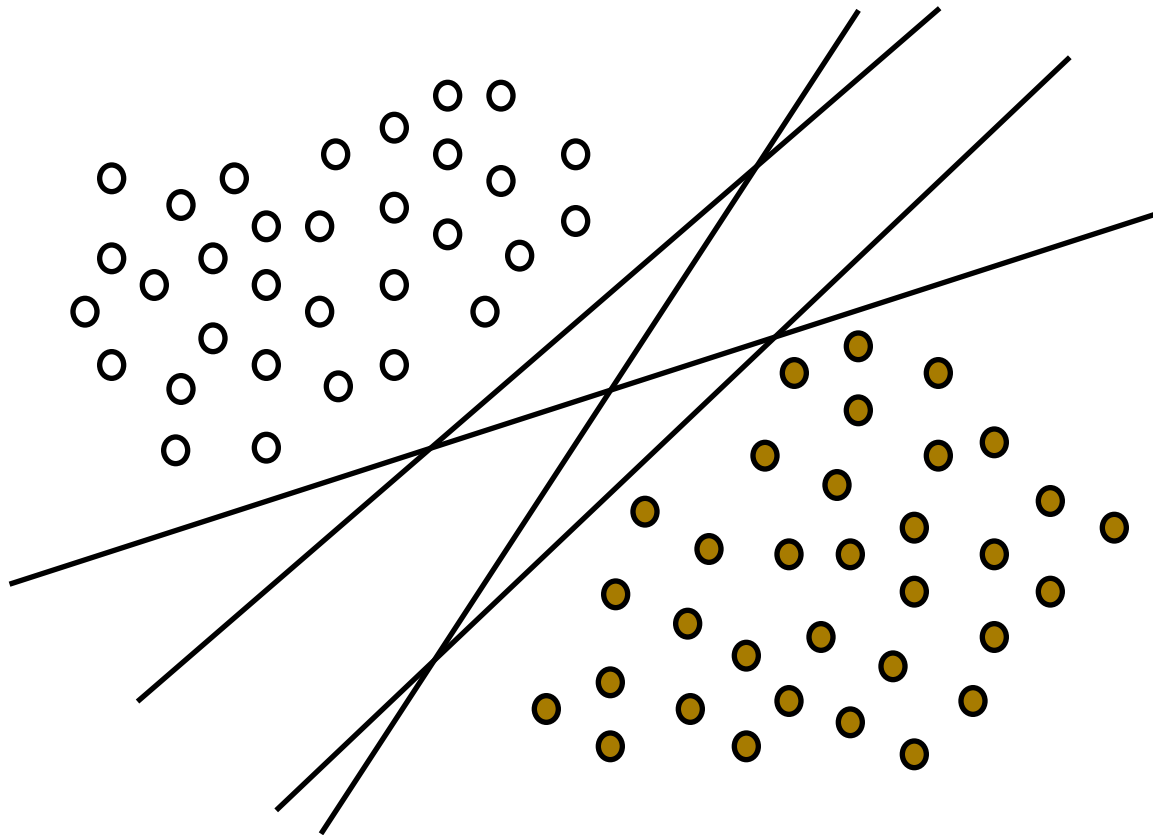
5





Ευθείες διαχωρισμού

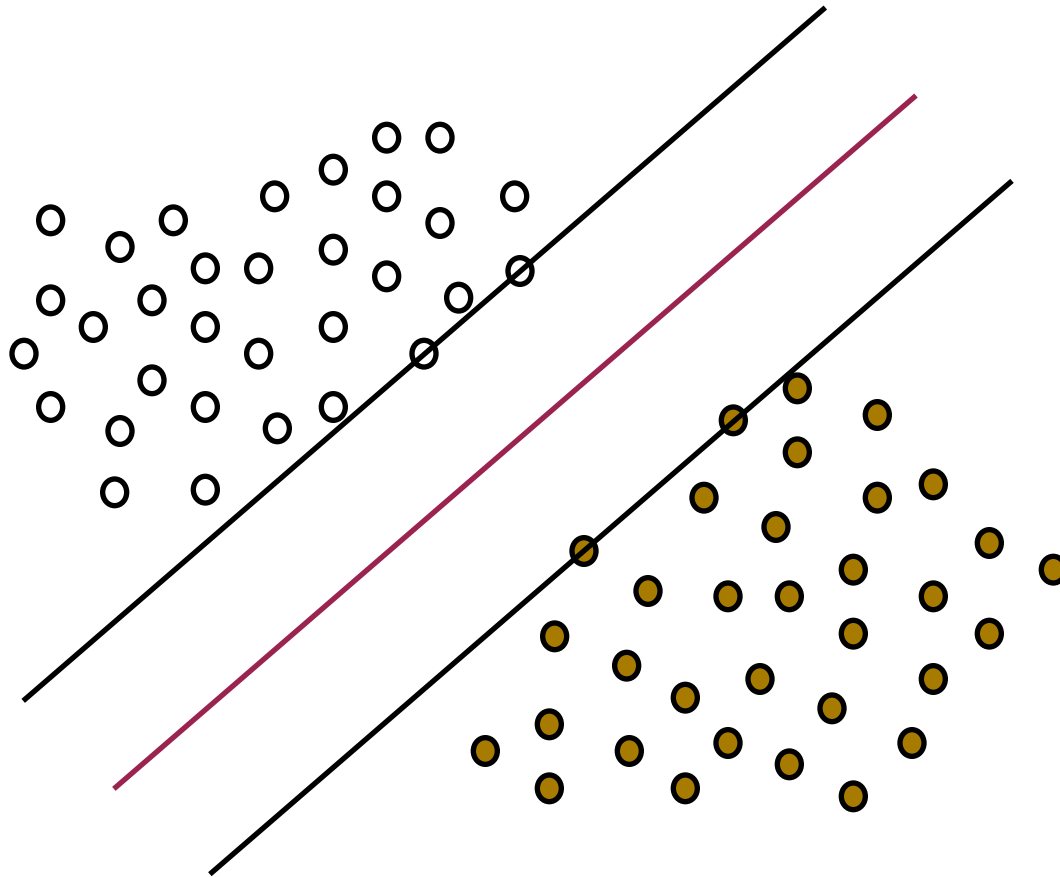
6





Βέλτιστη ευθεία διαχωρισμού

7





Τυπικός ορισμός προβλήματος

8

Διατύπωση προβλήματος

Δίνεται ένα σύνολο ζευγών $(\mathbf{x}_1, d_1), \dots, (\mathbf{x}_p, d_p)$

με $d_i = -1$ αν $\mathbf{x}_i \in \mathcal{C}_0$ και $d_i = 1$ αν $\mathbf{x}_i \in \mathcal{C}_1$

Ζητάμε την εύρεση των βαρών \mathbf{w} και του κατωφλίου w_o , έτσι ώστε:

$\mathbf{w}^\top \mathbf{x}_i + w_o \geq 0$ αν $d_i = 1$ ($\mathbf{x}_i \in \mathcal{C}_0$)

$\mathbf{w}^\top \mathbf{x}_i + w_o < 0$ αν $d_i = -1$ ($\mathbf{x}_i \in \mathcal{C}_1$)

Υπόθεση

Υπάρχει τέτοια ευθεία (οι κλάσεις είναι γραμμικά διαχωρίσιμες)

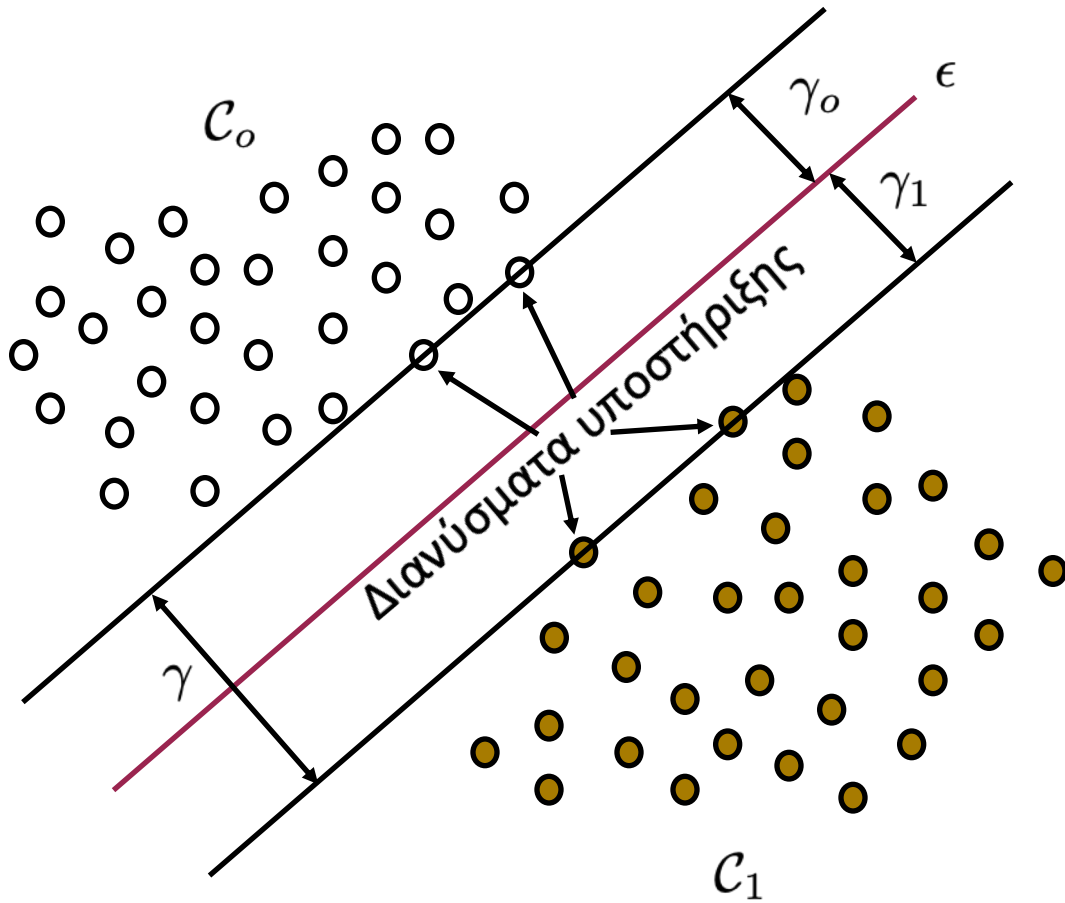
Απαίτηση

Η ευθεία που θα κατασκευαστεί πρέπει να έχει όσο το δυνατόν μεγαλύτερο περιθώριο ταξινόμησης



Βέλτιστο υπερεπίπεδο διαχωρισμού

9



$$\gamma_0 = \min_{\mathbf{x} \in C_0} d(\mathbf{x}, \epsilon)$$

$$\gamma_1 = \min_{\mathbf{x} \in C_1} d(\mathbf{x}, \epsilon)$$

$$\gamma = \gamma_0 + \gamma_1$$

Κανονικό υπερεπίπεδο

$$\gamma_0 = \gamma_1$$

$$\mathbf{w}^T \mathbf{x}_i + w_0 \geq 1 \text{ αν } \mathbf{x}_i \in C_0$$

$$\mathbf{w}^T \mathbf{x}_i + w_0 \leq -1 \text{ αν } \mathbf{x}_i \in C_1$$



Υπολογισμός περιθωρίου ταξινόμησης

10

Η συνάρτηση $g(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_o$ ένα μέτρο της απόστασης του \mathbf{x} από το βέλτιστο υπερεπίπεδο (όπου \mathbf{w} και w_o τα βέλτιστα βάρη).

Υπολογίζουμε το \mathbf{x} ως $\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|}$, όπου r η απόσταση του \mathbf{x} από το βέλτιστο υπερεπίπεδο

$$\text{Συνεπώς } g(\mathbf{x}) = \mathbf{w}^\top \left(\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + w_o$$

$$\Rightarrow g(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}_p + w_o + r \frac{\mathbf{w}^\top \mathbf{w}}{\|\mathbf{w}\|}$$

$$\Rightarrow g(\mathbf{x}) = r \frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} \Rightarrow r = \frac{g(\mathbf{x})}{\|\mathbf{w}\|}$$

Άρα αφού για τα διανύσματα υποστήριξης έχουμε $g(x) = 1$ ($\mathbf{x}_i \in \mathcal{C}_o$) και $g(x) = -1$ ($\mathbf{x}_i \in \mathcal{C}_1$)

Τελικά:

$$\gamma = \frac{2}{\|\mathbf{w}\|}$$



Βέλτιστο διαχωριστικό υπερεπίπεδο

11

Ορισμός προβλήματος βελτιστοποίησης

Υπολόγισε το ελάχιστο της συνάρτησης:

$$\mathcal{J}(\mathbf{w}, w_o) = \frac{1}{2} \|\mathbf{w}\|^2$$

υπό τους περιορισμούς των P ανισοτήτων:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) \geq 1, \quad i = 1, \dots, P$$

Παρατηρήσεις

- Η συνάρτηση κόστους είναι κυρτή
- Οι περιορισμοί είναι γραμμικοί

Καλούμαστε να επιλύσουμε ένα πρόβλημα τετραγωνικού προγραμματισμού



Μέθοδος πολλαπλασιαστών Lagrange

12

Ορίζουμε τη συνάρτηση κόστους:

$$\mathcal{L}(\mathbf{w}, w_o, \lambda_1, \dots, \lambda_p) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^P \lambda_i [d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) - 1]$$

$$\text{με } \lambda_i \geq 0, i = 1, \dots, P$$

Η συνάρτηση αυτή πρέπει να ελαχιστοποιηθεί ως προς τα \mathbf{w} , w_o και να μεγιστοποιηθεί ως προς τα λ_i

Συνθήκες Karush-Kuhn-Tucker (για το βέλτιστο σημείο)

$$\frac{\partial L}{\partial w_o} = 0 \quad \frac{\partial L}{\partial \mathbf{w}} = 0 \quad \lambda_i [d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) - 1] \geq 0, i = 1, \dots, P$$



Βέλτιστη διαχωριστική επιφάνεια

13

Από τις συνθήκες KKT έχουμε:

$$\frac{\partial L}{\partial w_o} = 0 \quad \longrightarrow \quad \sum_{i=1}^P \lambda_i d_i = 0$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \quad \longrightarrow \quad \mathbf{w} = \sum_{i=1}^P \lambda_i d_i \mathbf{x}_i$$

Συνεπώς η βέλτιστη διαχωριστική επιφάνεια δίνεται από τη σχέση:

$$g^*(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + w_o = \sum_{i=1}^P \lambda_i d_i \mathbf{x}_i^\top \mathbf{x} + w_o$$



Βέλτιστη πόλωση

14

Για τα διανύσματα υποστήριξης ισχύει ότι:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) = 1 \longrightarrow w_o = \frac{1}{d_i} - \mathbf{w}^\top \mathbf{x}_i$$

Για λόγους αριθμητικής ευστάθειας, χρησιμοποιούμε τη σχέση:

$$w_o = \frac{1}{|I_{sv}|} \sum_{i \in I_{sv}} \left(\frac{1}{d_i} - \mathbf{w}^\top \mathbf{x}_i \right)$$

όπου:

$$I_{sv} = \{i : \mathbf{x}_i \text{ διάνυσμα υποστήριξης}\}$$

Παρατήρηση

Οι μόνοι πολλαπλασιαστές λ_i που μπορούν να είναι θετικοί είναι αυτοί που αντιστοιχούν σε κάποιο διάνυσμα υποστήριξης \mathbf{x}_i .

Για τους υπόλοιπους ισχύει $\lambda_i = 0$.



Δυσικό πρόβλημα (1)

15

Από τα παραπάνω έχουμε:

$$\frac{1}{2}\|w\|^2 = \frac{1}{2}\mathbf{w}^\top \mathbf{w} = \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j$$

$$\begin{aligned} \sum_{i=1}^P \lambda_i [d_i (\mathbf{w}^\top \mathbf{x}_i + w_o) - 1] &= \sum_{i=1}^P \lambda_i d_i \sum_{j=1}^P \lambda_j d_j \mathbf{x}_j^\top \mathbf{x}_i + w_o \sum_{i=1}^P \lambda_i d_i - \sum_{i=1}^P \lambda_i \\ &= \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^P \lambda_i \end{aligned}$$

Επομένως:

$$\mathcal{L}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j$$



Δυσικό πρόβλημα (2)

16

Ορισμός δυσικού προβλήματος βελτιστοποίησης

Υπολόγισε το ελάχιστο της συνάρτησης:

$$\mathcal{L}^d(\lambda_1, \dots, \lambda_P) = \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_{i=1}^P \lambda_i$$

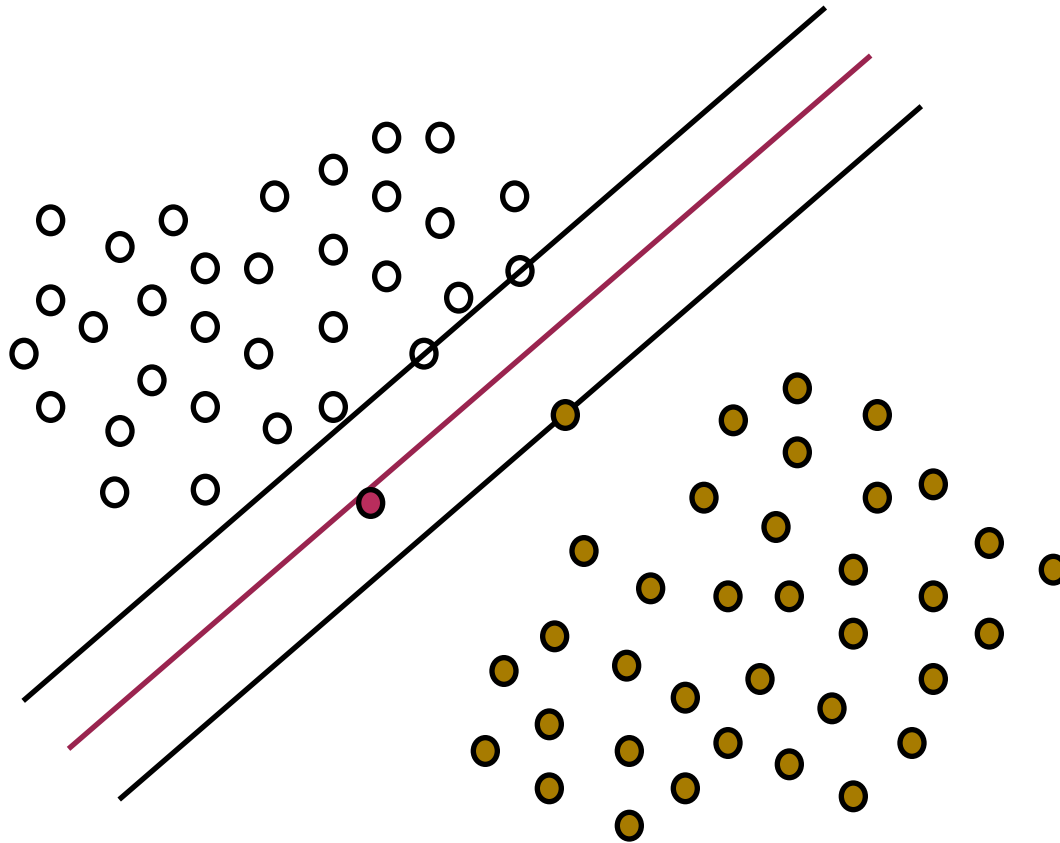
ως προς τα $\lambda_1, \dots, \lambda_P$, υπό τους περιορισμούς

$$\sum_{i=1}^P \lambda_i d_i = 0 \quad \lambda_i \geq 0, \quad i = 1, \dots, P$$



Μη γραμμικά διαχωρίσιμες κλάσεις

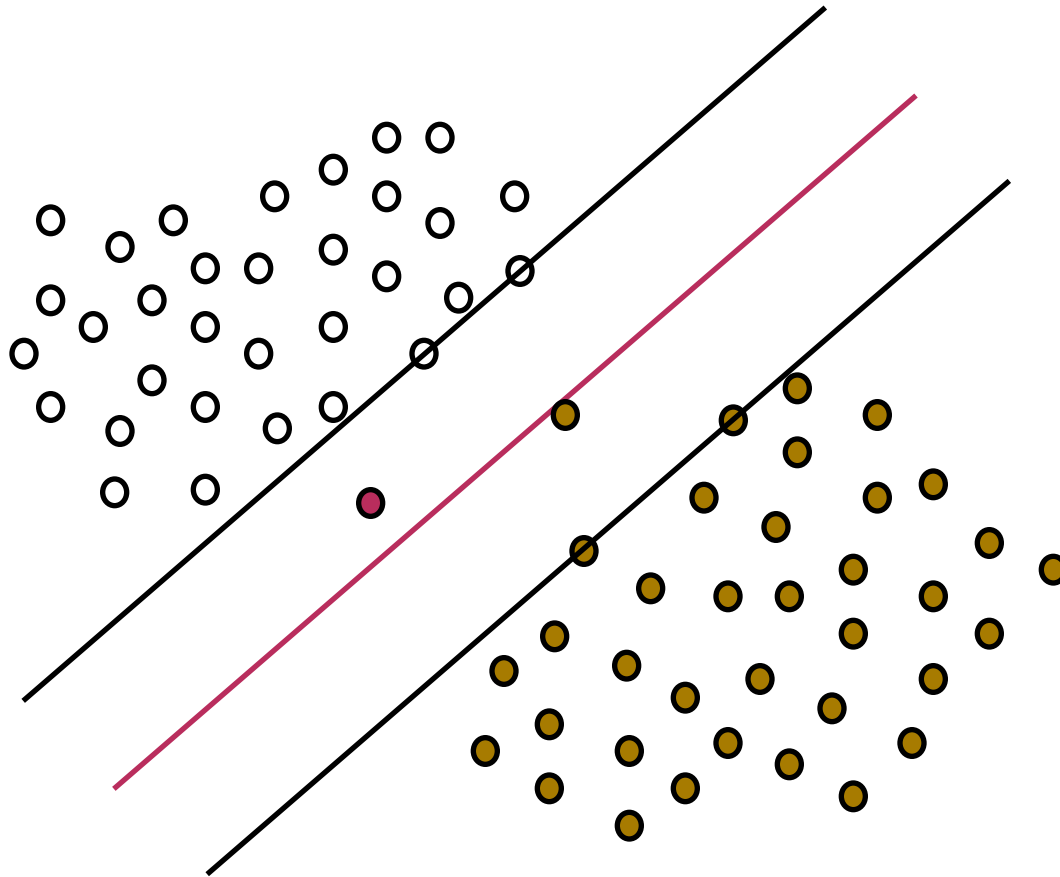
17





Μη γραμμικά διαχωρίσιμες κλάσεις

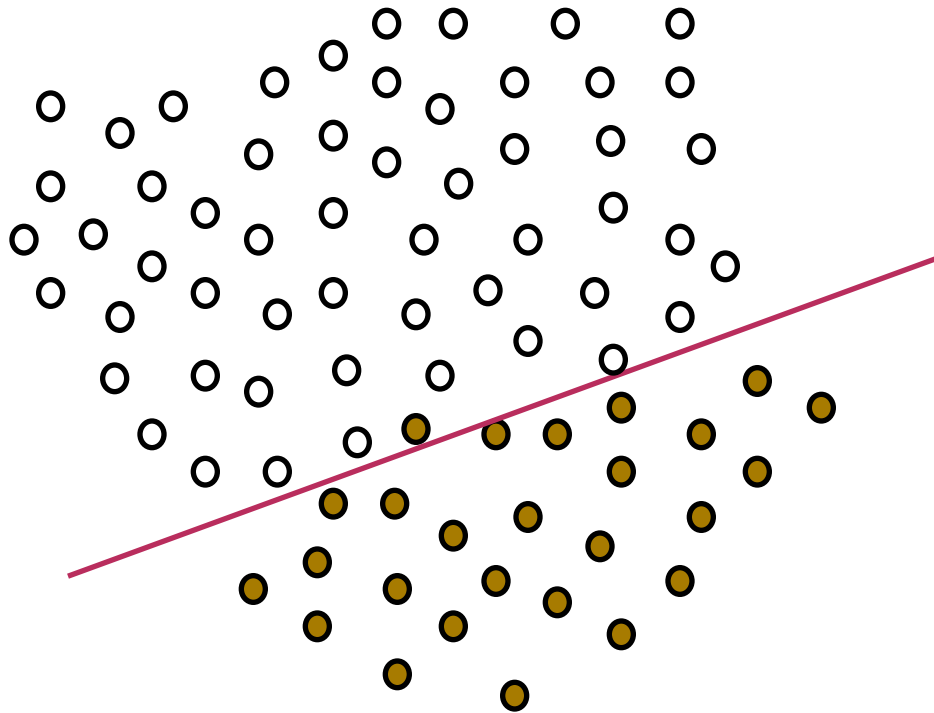
18





Μη γραμμικά διαχωρίσιμες κλάσεις

19





Βέλτιστο υπερεπίπεδο

20

Μεταβλητές χαλαρότητας

Ορίζουμε ένα σύνολο $\{\xi_i\}_{i=1}^N$ από θετικές τιμές και τις εισάγουμε στην εξίσωση της βέλτιστης ευθείας διαχωρισμού ως εξής:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) \geq 1 - \xi_i, i = 1, \dots, P$$

με $\xi_i \geq 0, i = 1, \dots, P$

Παρατηρούμε ότι:

Αν $\xi_i \leq 1$ δεν υπάρχει λάθος ταξινόμηση

Αν $\xi_i > 1$ υπάρχει λάθος ταξινόμηση
και το πρότυπο \mathbf{x}_i ταξινομείται σε λάθος κλάση



Βέλτιστο υπερεπίπεδο

21

Ορισμός προβλήματος βελτιστοποίησης

Υπολόγισε το ελάχιστο της συνάρτησης:

$$\mathcal{J}(\mathbf{w}, w_o) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^P \xi_i$$

υπό τους περιορισμούς των P ανισοτήτων:

$$d_i(\mathbf{w}^\top \mathbf{x}_i + w_o) \geq 1 - \xi_i, i = 1, \dots, P$$

όπου η παράμετρος C επιλέγεται από το χρήστη και είναι το βάρος του κόστους των λανθασμένων ταξινομήσεων

Αν $C = 0$ τότε αγνοούμε τελείως τις παραμέτρους χαλαρότητας, επομένως δεν μας ενδιαφέρει αν έχουμε λανθασμένες ταξινομήσεις

Αν $C \rightarrow \infty$ τότε δίνουμε έμφαση στη σωστή ταξινόμηση των προτύπων



Δυσικό πρόβλημα

22

Ορισμός δυσικού προβλήματος βελτιστοποίησης

Υπολόγισε το ελάχιστο της συνάρτησης:

$$\mathcal{L}_{ns}^d(\lambda_1, \dots, \lambda_P) = \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \mathbf{x}_i^\top \mathbf{x}_j - \sum_{i=1}^P \lambda_i$$

ως προς τα $\lambda_1, \dots, \lambda_P$, υπό τους περιορισμούς

$$\sum_{i=1}^P \lambda_i d_i = 0 \quad 0 \leq \lambda_i \leq C, \quad i = 1, \dots, P$$

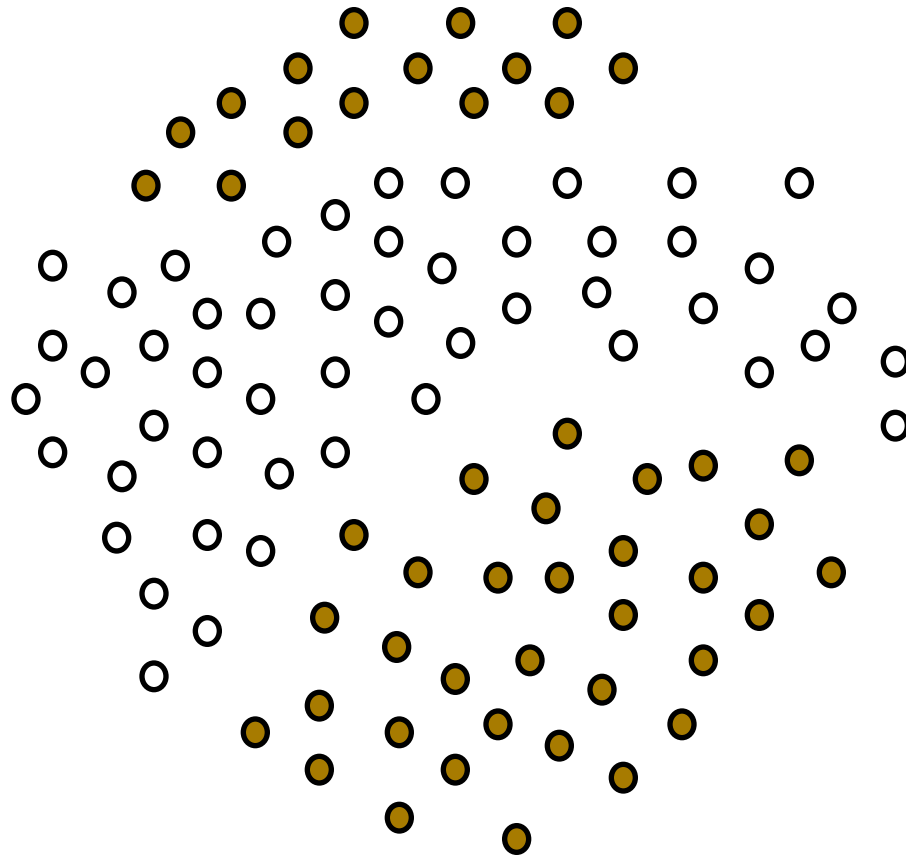
Παρατήρηση

Παρατηρούμε ότι τα ξ_i εμφανίζονται μόνο στο δεύτερο περιορισμό



Μη γραμμικά διαχωρίσιμες κλάσεις

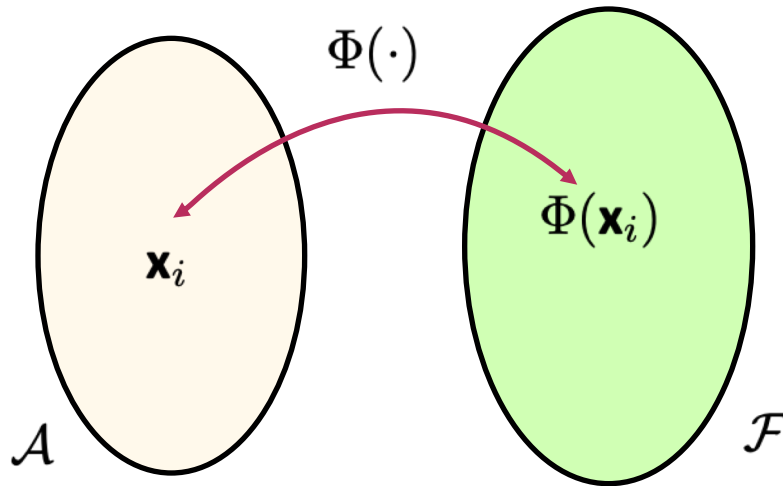
23





Απεικόνιση σε γραμμικά διαχωρίσιμες κλάσεις

24



\mathcal{A} : χώρος εισόδου

\mathcal{F} : χώρος χαρακτηριστικών

$\Phi(\cdot)$: μη-γραμμική συνάρτηση απεικόνισης

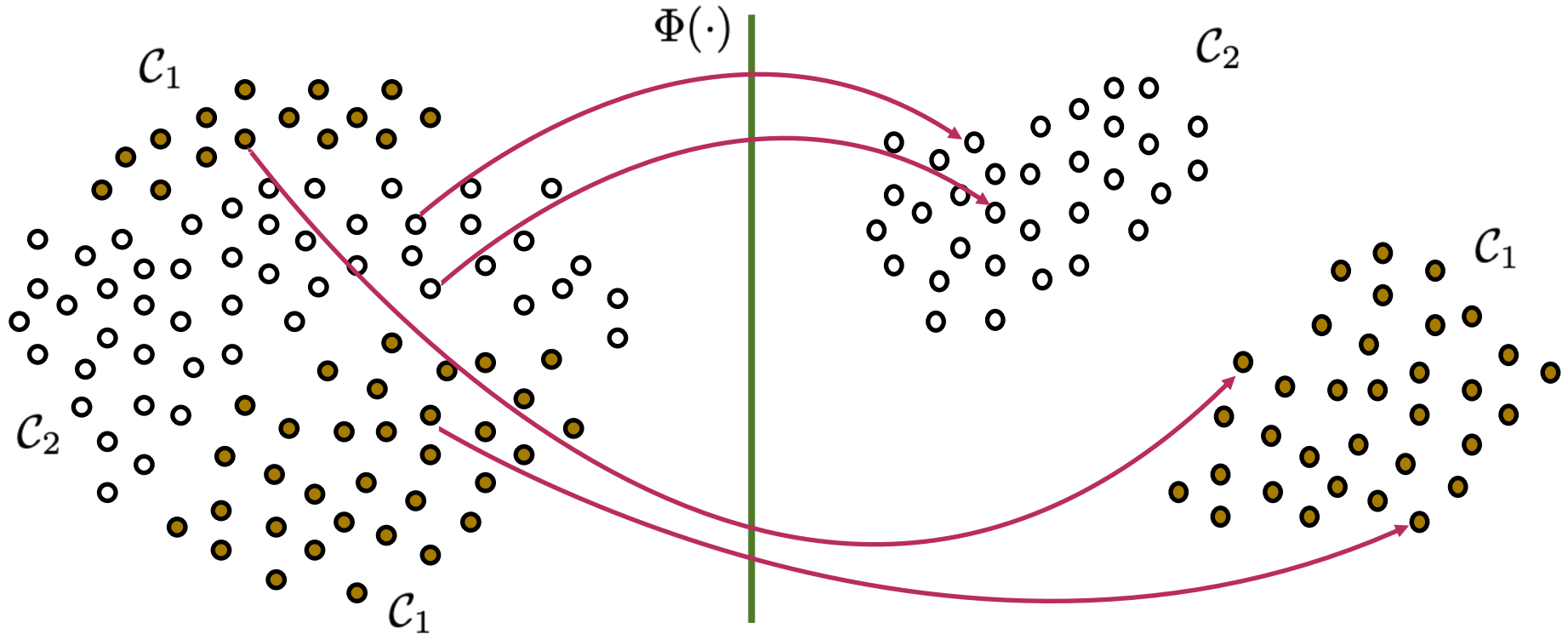
Θεώρημα Cover

Κάθε πολυδιάστατος χώρος με μη γραμμικά διαχωρίσιμα πρότυπα, μπορεί να μετασχηματιστεί σε ένα νέο χώρο στον οποίο τα πρότυπα είναι γραμμικά διαχωρίσιμα με *υψηλή πιθανότητα*, αρκεί ο μετασχηματισμός να είναι μη γραμμικός και ο νέος αυτός χώρος να έχει την απαραίτητη διάσταση



Απεικόνιση σε γραμμικά διαχωρίσιμες κλάσεις

25





Λύση δυϊκού προβλήματος

26

Βέλτιστη διαχωριστική επιφάνεια:

$$g^*(\mathbf{x}) = \mathbf{w}^\top \Phi(\mathbf{x}) + w_o = \sum_{i=1}^P \lambda_i d_i \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}) + w_o$$

Κατώφλι:

$$w_o = \frac{1}{|I_{sv}|} \sum_{i \in I_{sv}} \left(\frac{1}{d_i} - \mathbf{w}^\top \Phi(\mathbf{x}_i) \right)$$

Συνάρτηση κόστους του δυϊκού προβλήματος:

$$\mathcal{L}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i \lambda_j d_i d_j \Phi(\mathbf{x}_i)^\top \Phi(\mathbf{x}_j)$$



Χρήση συναρτήσεων πυρήνα

27

Παρατήρηση

Παρατηρούμε ότι σε όλες τις εξισώσεις που χρησιμοποιούμε εμφανίζονται γινόμενα της μορφής $\Phi(\mathbf{x})^\top \Phi(\mathbf{y})$.

Η συνάρτηση $\Phi(\cdot)$ δεν εμφανίζεται ποτέ μόνη της.

Ορισμός

Ορίζουμε τη συνάρτηση $k(x, y) = \Phi(\mathbf{x})^\top \Phi(\mathbf{y})$, την οποία θα ονομάζουμε συνάρτηση πυρήνα.

Χρησιμοποιώντας τη συνάρτηση πυρήνα κάνουμε οικονομία πράξεων ειδικά όταν η διάσταση του \mathbf{x} είναι μεγαλύτερη από τη διάσταση του $\Phi(\mathbf{x})$ (το οποίο συνήθως συμβαίνει)



Παράδειγμα

28

$$\text{Έστω } \mathbf{x} = [x_1 \quad x_2]^\top$$

$$\Phi(\mathbf{x}) = [x_1^2 \quad \sqrt{2}x_1x_2 \quad x_2^2]^\top$$

$$\text{Για } \mathbf{x} = [1 \quad 2]^\top$$

$$\Phi([1 \ 2]^\top) = [1 \quad 2\sqrt{2} \quad 4]^\top$$

$$\begin{aligned} k(x, y) &= \Phi(\mathbf{x})^\top \Phi(\mathbf{y}) = (x_1^2 y_1^2 + 2x_1 y_1 y_2 + x_2^2 y_2^2) \\ &= (x_1 y_1 + x_2 y_2) = (\mathbf{x}^\top \mathbf{y})^2 \end{aligned}$$



Επιλογή συναρτήσεων πυρήνα

29

Θεώρημα Mercer

Έστω $k(\mathbf{x}, \mathbf{y})$ ένας συνεχής συμμετρικός πυρήνας, με $\mathbf{a} \leq \mathbf{x}, \mathbf{y} \leq \mathbf{b}$.

Ο πυρήνας $k(\mathbf{x}, \mathbf{y})$ μπορεί να γραφεί ως:

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\infty} \alpha_i \Phi_i(\mathbf{x}) \Phi_i(\mathbf{y})$$

με $\alpha_i > 0$, $\forall i$, αν και μόνο αν:

$$\int_b^a \int_b^a k(\mathbf{x}, \mathbf{y}) \psi(\mathbf{x}) \psi(\mathbf{y}) dx dy \geq 0$$

για κάθε $\psi(\cdot)$ για την οποία $\int_b^a \psi^2(\mathbf{x}) dx \leq \infty$



Παραδείγματα συναρτήσεων πυρήνα

30

Γκαουσιανή RBF:

$$e^{-\|\mathbf{x}-\mathbf{y}\|^2/(2\sigma^2)}$$

Πολυωνυμική:

$$[\mathbf{x}^\top \mathbf{y} + \theta]^p$$

Σιγμοειδής:

$$\tanh(\alpha \mathbf{x}^\top \mathbf{y} + \theta)$$

Αντίστροφη πολυτετραγωνική:

$$\frac{1}{\sqrt{\|\mathbf{x} - \mathbf{y}\|^2 + c^2}}$$



Πρόβλημα SVM

31

Υπολόγισε το μέγιστο της συνάρτησης:

$$\mathcal{L}_{SVM}(\lambda_1, \dots, \lambda_P) = \sum_{i=1}^P \lambda_i - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \lambda_i q_{ij} \lambda_j$$

υπό τους περιορισμούς:

$$0 \leq \lambda_i \leq C \quad \sum_{i=1}^P \lambda_i d_i = 0$$

όπου: $q_{ij} = d_i d_j k(\mathbf{x}_i, \mathbf{x}_j)$

Παρατήρηση

Το πλήθος των στοιχείων του πίνακα $\mathbf{Q} = [q_{ij}]$ είναι P^2 , συνεπώς είναι αρκετά πολύπλοκη η επίλυση του προβλήματος.



Μέθοδοι υλοποίησης SVM

32

Μέθοδος τεμαχισμού

Η συνάρτηση κόστους δεν αλλάζει αν αφαιρέσουμε τις γραμμές και τις στήλες του \mathbf{Q} που αντιστοιχούν σε μηδενικές τιμές του λ_i

Διαλέγουμε σε κάθε βήμα την επίλυση του προβλήματος για το τμήμα του \mathbf{Q} που αντιστοιχεί στα μη μηδενικά λ_i από το προηγούμενο πρόβλημα και επιπλέον στα K χειρότερα λ_i (που παραβιάζουν περισσότερο τις συνθήκες ΚΚΤ)

Μέθοδος Osuna

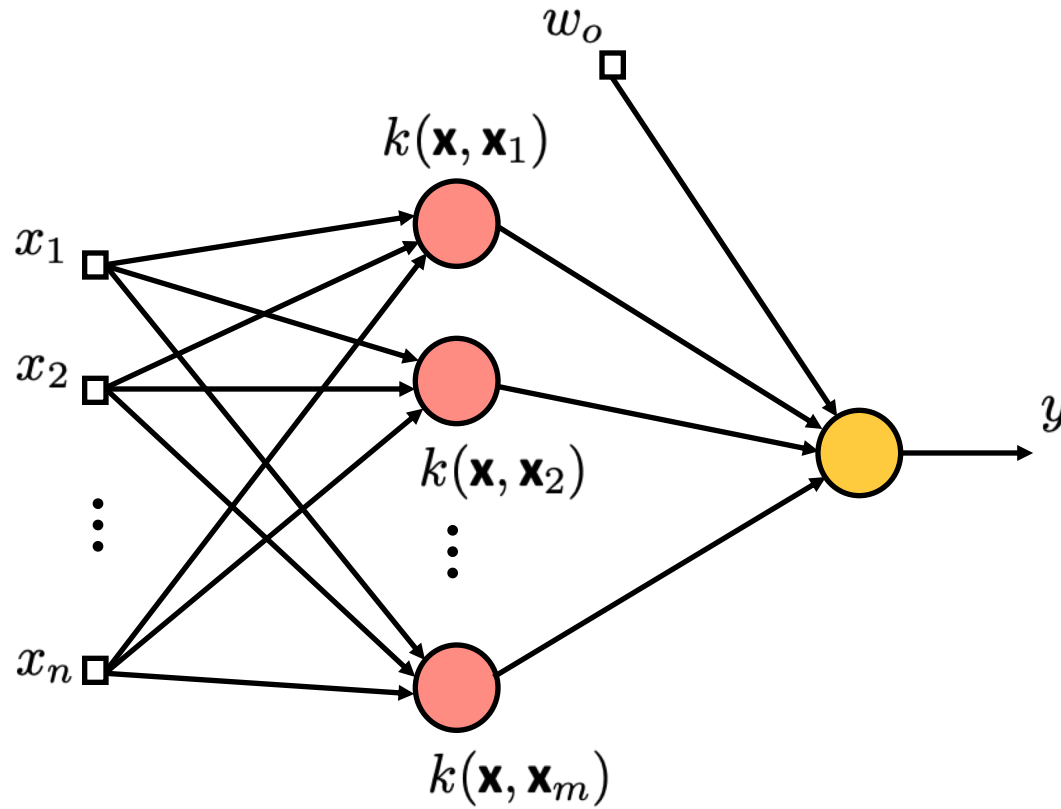
Αν επιλύσουμε ένα μικρότερο πρόβλημα, επιλέγοντας μερικές μόνο γραμμές του \mathbf{Q} έτσι ώστε να περιέχεται τουλάχιστον ένα λ_i που παραβιάζει τις συνθήκες ΚΚΤ τότε η συνάρτηση κόστους μειώνεται και όλοι οι περιορισμοί συνεχίζουν να ικανοποιούνται

Επιλύουμε το πρόβλημα προσθέτοντας μία μεταβλητή λ_i που παραβιάζει τις συνθήκες και αφαιρώντας μία μεταβλητή για την οποία $\lambda_i = 0$ ή $\lambda_j = C$



Δίκτυα SVM

33



Ταξινόμηση σε περισσότερες από δύο κλάσεις



- ❑ Τα μέχρι τώρα αφορούν σε προβλήματα δυαδικής ταξινόμησης (binary classification)
- ❑ Ωστόσο πολλές πρακτικές εφαρμογές περιλαμβάνουν περισσότερες από δύο κλάσεις (multi-class classification)
- ❑ Επομένως, είναι εμφανής η ανάγκη γενίκευσης του προβλήματος δυαδικής ταξινόμησης σε ταξινόμηση με περισσότερες κλάσεις
- ❑ Διάφορες προσεγγίσεις – δύο απλές είναι οι:
 - ❑ One-against-all (ή one-against-the-rest)
 - ❑ One-against-one



One-against-all

- Κατασκευάζει δυαδικά μοντέλα SVM με μια κατηγορία ως θετική και τις υπόλοιπες ως αρνητικές
- Π.χ. για 4 κλάσεις, θα δημιουργηθούν τα ακόλουθα 4 SVM:

$y_i = 1$	$y_i = -1$	Συνάρτηση Απόφασης
Κλάση 1	Κλάσεις 2,3,4	$f^1(x) = (w^1)^T x + b^1$
Κλάση 2	Κλάσεις 1,3,4	$f^2(x) = (w^2)^T x + b^2$
Κλάση 3	Κλάσεις 1,2,4	$f^3(x) = (w^3)^T x + b^3$
Κλάση 4	Κλάσεις 1,2,3	$f^4(x) = (w^4)^T x + b^4$

- Για οποιοδήποτε δεδομένο που ανήκει στην κλάση i , περιμένουμε ότι:

$$f^i(x) \geq 1 \text{ και } f^j(x) \leq -1, i \neq j$$

- Επομένως, ο κανόνας απόφασης είναι:

$$\text{Αναμενόμενη κλάση} = \arg \max_{i=1,\dots,4} f^i(x)$$



One-against-one

- Εδώ κατασκευάζονται συνολικά $\binom{k}{2} = \frac{k(k-1)}{2}$ SVMs, για τη δυαδική ταξινόμηση όλων των κλάσεων ανά δύο
- Κάθε δυαδικός ταξινομητής εκπαιδεύεται με δεδομένα 2 κλάσεων
- Κάθε νέα παρατήρηση x προς ταξινόμηση δοκιμάζεται σε όλους τους ταξινομητές
- Αν το πρόβλημα των κλάσεων i και j δείξει ότι η x παρατήρηση θα πρέπει να είναι στην i , η κλάση i παίρνει μία ψήφο
- Στο τέλος, η παρατήρηση x αντιστοιχίζεται στην κλάση που έχει λάβει τις περισσότερες ψήφους

Π.χ.

Κλάσεις	Νικητής
1 2	1
1 3	1
1 4	1
2 3	2
2 4	4
3 4	3

Κλάση	1	2	3	4
Πλήθος ψήφων	3	1	1	1

Νικήτρια η κλάση 1



Παλινδρόμηση με διανύσματα υποστήριξης (Support Vector Regression)

- Εδώ οι τιμές των στόχων ανήκουν σε συνεχές σύνολο τιμών
- Έστω σύνολο προτύπων x_i και στόχων $t_i, i = 1, \dots, N$.
- Συνάρτηση σφάλματος με ανοχή ε :

$$l_\varepsilon(t, g) = \begin{cases} 0, & \text{αν } |t - g| \leq \varepsilon \\ |t - g| - \varepsilon, & \text{αν } |t - g| > \varepsilon \end{cases}$$

- Η συνάρτηση τιμωρεί τη διαφορά μεταξύ του στόχου και της εκτιμώμενης τιμής g μόνο αν η απόλυτη διαφορά είναι μεγαλύτερη από μια θετική σταθερά ε (ανοχή στο σφάλμα)
- Για γραμμικά επιλύσιμο πρόβλημα, ξεκινάμε θεωρώντας ότι υπάρχει (\mathbf{w}, w_0) , ώστε $L_\varepsilon(\mathbf{w}, w_0) = \sum_{i=1}^N l_\varepsilon(t_i, g(x_i; \mathbf{w}, w_0)) = 0$. Επιλέγεται η λύση με τη μικρότερη τιμή $\|\mathbf{w}\|^2$.
- Πρόβλημα παλινδρόμησης με ανοχή ε :

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$

$$t_i - \mathbf{w}^T x_i - w_0 \leq \varepsilon$$

$$t_i - \mathbf{w}^T x_i - w_0 \geq -\varepsilon$$