

2021-12-13

Συμπληρωματικότητα (Complementarity)

$$\alpha_i \left(y_i (x_i^T b + b_0) - (1 - \xi_i) \right) = 0 \quad i=1, \dots, N$$
$$\mu_i \xi_i = 0 \quad i=1, \dots, N$$

Έχουμε

$$b = \sum_{i=1}^N \alpha_i x_i y_i$$

$$y_i (x_i^T b + b_0) - (1 - \xi_i) \geq 0 \quad i=1, \dots, N$$
$$\xi_i \geq 0 \quad i=1, \dots, N$$

$$\mu_i = \alpha_i - C \geq 0 \Rightarrow 0 \leq \alpha_i \leq C \quad i=1, \dots, N$$

$$\sum \alpha_i y_i = 0$$

Λύνεται με αλγόριθμο quadratic programming

Παρατήρηση

$$b = \sum_{i=1}^N \alpha_i y_i x_i$$

Ουμπίδα ποσεία
regression/classification

$$\hat{b} = G y$$
$$G = (X^T X)^{-1} X^T$$

κρί complementarity:

$$\alpha_i > 0 \text{ μόνο αν } \underline{y_i (x_i^T b + b_0) = 1 - \xi_i}$$

Επίσης $\mu_i \xi_i = 0 \Rightarrow (C - \alpha_i) \xi_i = 0 \quad \forall i$

$$\Rightarrow \underline{\text{αν } \xi_i > 0 \Rightarrow \alpha_i = C}$$

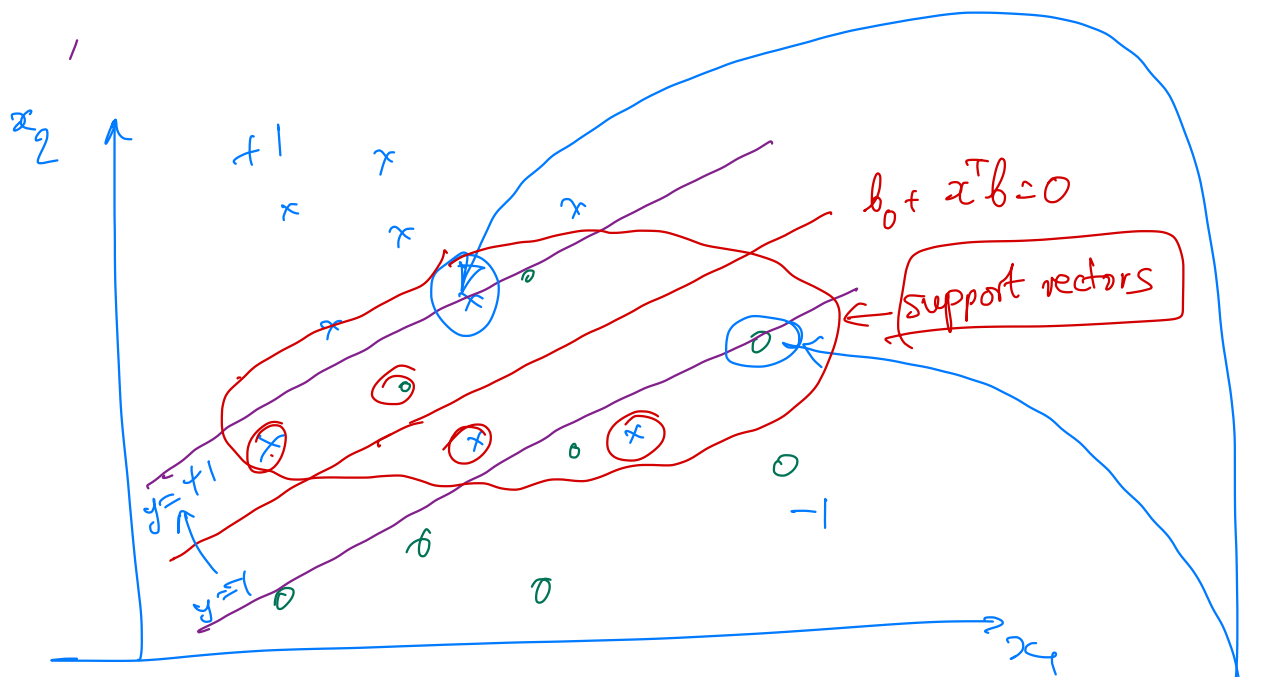
$$(\text{αν } \xi_i = 0 \Rightarrow 0 < \alpha_i < C)$$

$$b = \sum_{i: \alpha_i > 0} \alpha_i x_i y_i$$

Οι $\alpha_i > 0$ αντιστοιχούν με

$$y_i (x_i^T b + b_0) = 1 - \xi_i$$

Support vectors



Support vector

① αν $\xi_i = 0 \Rightarrow y_i (x_i^T b + b_0) = 1$

η παρατηρούμε ότι ορισμένα σημεία είναι
στο όριο όριο της ζώνης

② αν $\xi_i > 0 \quad y_i (x_i^T b + b_0) = 1 - \xi_i$

SVC Method

↳ support vector classifiers.

παράγει boundaries

Συνδυασμός με basis expansion methods

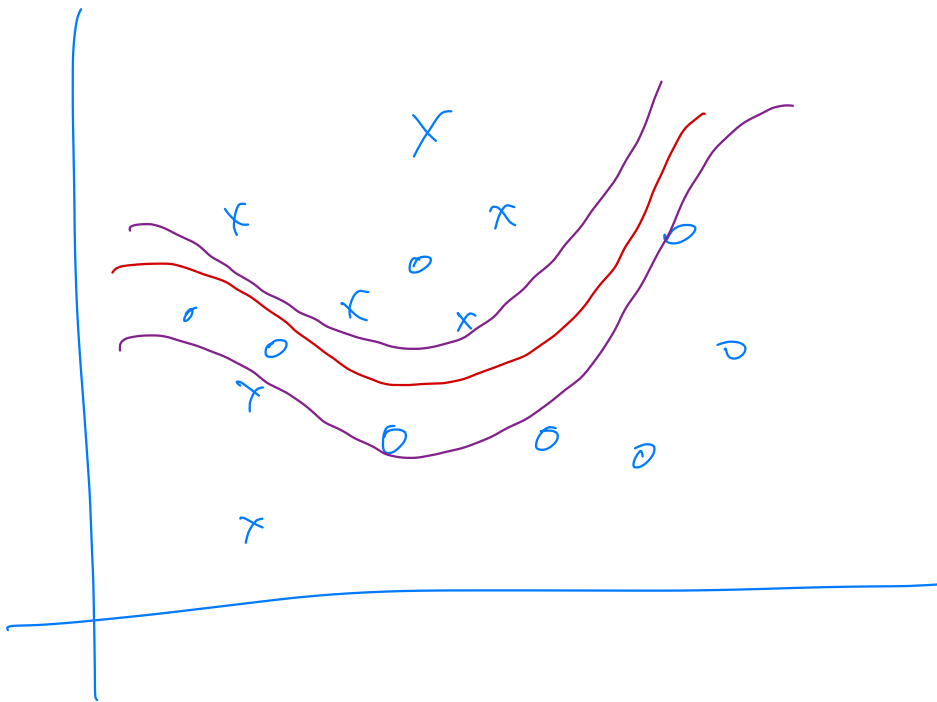
$$x_1, x_2, \dots, x_p$$

↓

$$h_1(x), h_2(x), \dots, h_M(x)$$

↓

Support Vector μεθοδολογία με μη γραμμικά boundaries



Support
vector
Machines

Unsupervised Learning

επιβλ.
παιχνίδι

X : inputs/features

Y : outcome

$$P(X, Y)$$



$$E(Y|X=x)$$

αν $Y = G$: κατηγορία

$$P(G=g_k | X=x)$$

Unsupervised Learning

Σύνολο μεταβλητών

$$X = (X_1, \dots, X_p)$$

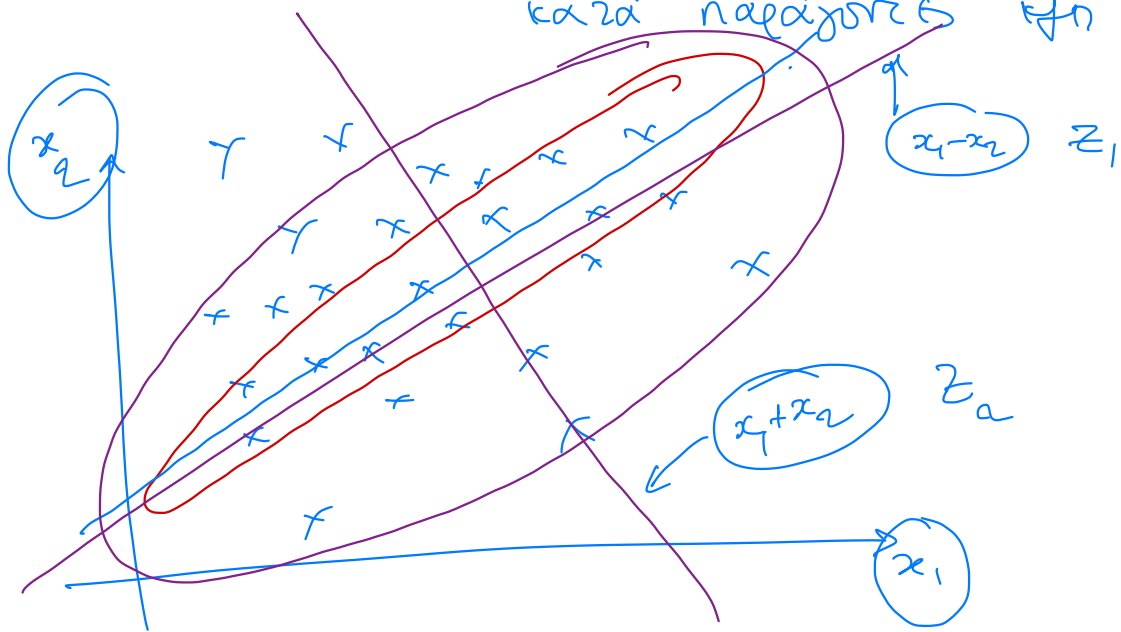
Data set (x_1, x_2, \dots, x_N)

Πρόβλημα / Στόχος

Να βρεθούν / παρακρυφθούν πληροφορίες των $x \in \mathbb{R}^p$

που έχουν μεγάλη πιθανότητα

Μέθοδοι όπως ανάλυση σε κύριες συνιστώσες
κατά παράγοντες κ.λπ.

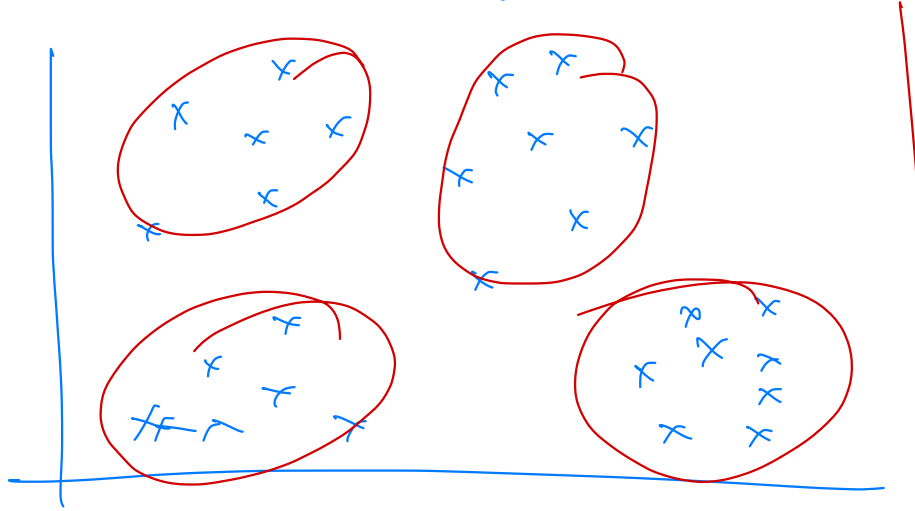


(x_1, \dots, x_{36}) : βαθμοί στα 36 μαθ. του πραγματικού

$$z_1 = \frac{x_1 + \dots + x_{36}}{36}$$

$$z_2 = \text{MO}(\text{Θωπρ}) - \text{MO}(\text{Εφαρμ}).$$

Μέθοδοι τώνου clustering



$G \in \{1, 2, 3, 4\}$

Association Rules

Παράδειγμα

Market basket analysis

K προϊόντα

Αγορά	P1	P2	Pk
1	x	x	x			
2	x		x			
3	x				x	x
4			x	x		

2^k υποσύνολα του $\{1, \dots, k\}$

n.x. $N = 10^9$

R: `a = read.csv(...)`

↓

variable a : dataframe