18-10-2021

Γραμμικές Μέθοδοι Ταξινόμησης

μεταβλητές $\underline{X} = (X_1, \dots, X_p)$: predictors / ανεξ. μεταβλητές
features

G : εξαρ. μεταβλητή (outcome)

$$G \in \underbrace{\{1, 2, \dots, K\}}_{\text{κατηγορίες (classes)}} \equiv \{g_1, \dots, g_K\}$$

Παρατήρηση i στο training set: $(x_i^T) = (x_1, \dots, x_p)$, $i = 1, \dots, N$

$g_i \in \{1, \dots, K\}$

$\Rightarrow \dots \Rightarrow$ συνάρτηση πρόβλεψης

$\hat{G}(\underline{x})$ = πρόβλεψη για την κατηγορία μιας νέας παρατήρησης με τιμή $\underline{x}$

Συνάρτηση απώλειας $L(G, \hat{G}) =$

κόστος/απώλεια αν σε μια παρατήρηση κατηγορίας G προβλέψουμε $\hat{G}$

Επιλέγουμε $L(G, \hat{G}) = 1(\hat{G} \neq G)$

Μέσο σφάλμα πρόβλεψης [Θεωρώντας (X, G) από κοινού κατανομή]

$$EPE = E_{(X,G)}\left[L(G, \hat{G})\right] = E\left[1(\hat{G} \neq G)\right] =$$

$$= P(\hat{G} \neq G)$$

Δεδομένου $X = x$

$$EPE(x) = P\left(\hat{G}(x) \neq G \mid X = x\right) =$$

$$= \sum_{k=1}^{K} P(G = k \mid X = x) \cdot P\big(\underbrace{\hat{G}(x)}_{\text{σταθερά}} \neq k\big)$$

$$P(\hat{G}(x) \neq k) = \begin{cases} 1 & \text{αν } k \neq \hat{G}(x) \\ 0 & \text{αν } k = \hat{G}(x) \end{cases}$$

$$= P\left(G \neq G(x) \mid X = x\right) = \underline{1 - P\left(G = \hat{G}(x) \mid X = x\right)}$$

$$\min_{\hat{G}} EPE(x) \Rightarrow \boxed{\max_{\hat{G} = 1, \dots, K} P\left(G = \hat{G}(x) \mid X = x\right)}$$

Επομένως αν $G|X=x$ είναι γνωστή κατανομή

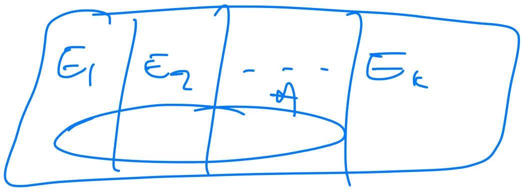
$\Rightarrow$ **Bayes classifier:**

$$\boxed{\hat{G}(x) = \arg\max_{k} P(G=k|X=x)}$$

$$\arg\max_{x \in A} f(x) = \{x^* : f(x^*) \geq f(x) \ \forall x \in A\}$$

Συμβατικά θεωρούμε argmax οποιαδήποτε τιμή που μεγιστοποιεί

Γιατί Bayes?



Γνωρίζω $\boxed{P(E_1), \dots, P(E_k)}$ → prior

$P(A|E_1) \dots P(A|E_k)$

Με ενδιαφέρει $P(E_1|A), \dots, P(E_k|A)$ → posterior

$$P(E_j|A) = \frac{P(A|E_j)\,P(E_j)}{\sum_j P(A|E_j)\,P(E_j)} \sim P(E_j)\,P(A|E_j)$$

Έστω prior κατανομή του G :

$$(\pi_1, \dots, \pi_K) \qquad \pi_j = P(G=j)$$

Έστω $f_k(x)$: σππ του $(X \mid G=k)$

γνωστά

και ζητάμε $\underline{P(G=k} \mid X=x) = \dfrac{\pi_k f_k(x)}{\sum_{k=1}^{K} \pi_k f_k(x)} =$

$$= \underline{C(x) \cdot \pi_k f_k(x)}$$

1 Classification via Linear Regression

$$\forall k=1,\dots,K \qquad Y_k = 1(G=k) = \begin{cases} 1 & G=k \\ 0 & G\neq k. \end{cases}$$

$$\underline{G} \iff \underline{Y} = (Y_1,\dots,Y_K) = (0,\dots,\underset{\substack{\uparrow\\ k=G}}{1},\dots 0)$$

(A) $\forall k$

$$Y_k = b_{k0} + b_{k1}X_1 + \dots + b_{kp}X_p \implies \hat{b}_k = (\hat{b}_{k0}, \hat{b}_{k1}, \dots, \hat{b}_{kp})^T \text{ LSE}$$

$$k=1,\dots,K \quad , \quad Y_k = (1, X^T)\, b_k$$

(B) Πολυμεταβλητή Παλινδρόμηση (Multivariate Regression)
(Πολλές εξαρτημένες μεταβλητές)

$$\underline{Y} = (\underline{Y_1}, \dots, Y_K) =$$

$$= \left((1,\underline{X}^T)\,\underline{b}_1, \dots, (1,\underline{X}^T)\,\underline{b}_K\right)$$

$$= (1, X^T)\, B \quad , \quad B = (\underset{\in \mathbb{R}^{p+1}}{\underline{b}_1}, \dots, b_K)_{(p+1)\times K}$$

$$= \begin{pmatrix} b_{10} & b_{20} & & b_{K0} \\ b_{11} & b_{21} & \dots & \vdots \\ \vdots & \vdots & & \vdots \\ b_{1p} & b_{2p} & & b_{Kp} \end{pmatrix} \in B$$

Training set: $\underline{Y} = \begin{pmatrix} Y_1 \dots Y_K \\ \\ \\ \end{pmatrix}_{N\times K}$ $\quad G = \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix}$ $\quad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix}$

$\forall \; i=1,\dots,N \Rightarrow \hat{Y}_i \in \mathbb{R}^K$

$$= \underline{\hat{Y}_i} = (\hat{Y}_{i1}, \dots, \hat{Y}_{iK}) = \underline{(1 \; x_i^T)\hat{B}}$$

Loss function $\displaystyle\sum_{i=1}^{N} \sum_{k=1}^{K} (\hat{Y}_{ik} - Y_{ik})^2 \leftarrow \min_{B}$

$$= \sum_{k=1}^{K} \underbrace{\left[\sum_{i=1}^{N} (\hat{Y}_{ik} - Y_{ik})^2 \right]}_{RSS(\underline{b_k})}$$

(μονοδιάστατο regression της Y_k)

Ⓐ, Ⓑ ισοδύναμες.

$\hat{B}$: σε κάθε $k=1,\dots,K$

υπολογίζω $\hat{b}_k = \underbrace{(X^TX)^{-1}X^T}_{G} \underline{y_k}$, $\hat{y}_k = X(X^TX)^{-1}X^T y_k$

$= H y_k$

$$\hat{B} = (\hat{b}_1 \; \dots \; \hat{b}_K)$$

$$= (G\underline{y_1}, \dots, G\underline{y_K}) = GY$$

$\hat{B} = GY$, $\boxed{\hat{Y} = HY}$

$$\hat{Y} = HY \qquad Y, \hat{Y} \in \mathbb{R}^{N \times (p+1)}, \quad H \in \mathbb{R}^{(p+1)\times(p+1)}$$

$\hat{y}_1$ εκτίμηση του $y_1 = \begin{cases} 0 \\ 1 \end{cases}$

$$\forall x \Rightarrow \hat{y}_1(x), \underline{\hat{y}_2(x)}, \dots, \hat{y}_K(x)$$

$$\hat{g}(x) = k \quad \text{αν} \quad \hat{y}_k(x) = \max_j \hat{y}_j(x), \quad j = 1, \dots K$$

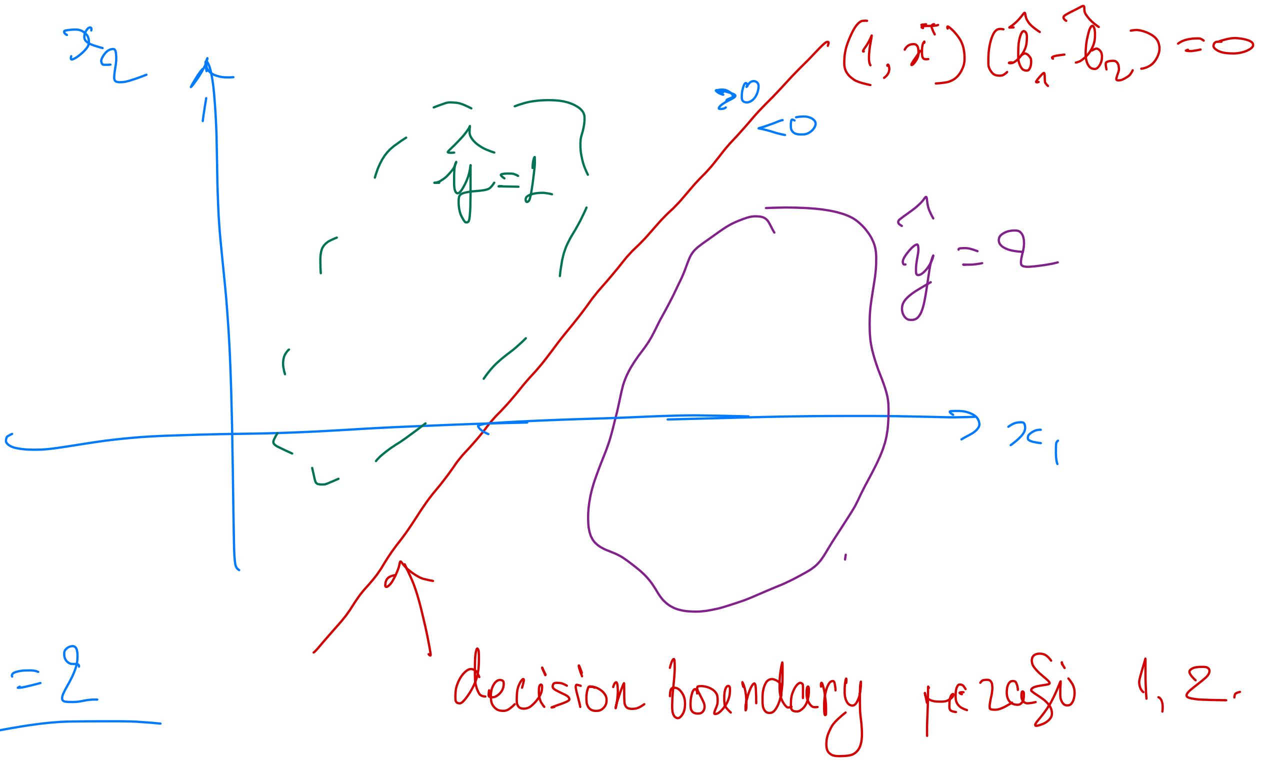
Έστω $k, l \in \{1, \dots, K\}$

Μεταξύ των k, l μια παρατήρηση x θα ταξινομηθεί ως k αν $\hat{y}_k(x) \geq \hat{y}_l(x)$

$$\hat{b}_{k0} + \hat{b}_{k1}x_1 + \dots + \hat{b}_{kp}x_p = (1, x^T)\hat{b}_k \geq (1, x^T)\hat{b}_l.$$

$$\Rightarrow \boxed{(1, x^T)(\hat{b}_k - \hat{b}_l) \geq 0} \quad \begin{cases} > & k \\ = 0 & \text{αδιάφορο} \\ < 0 & l. \end{cases}$$

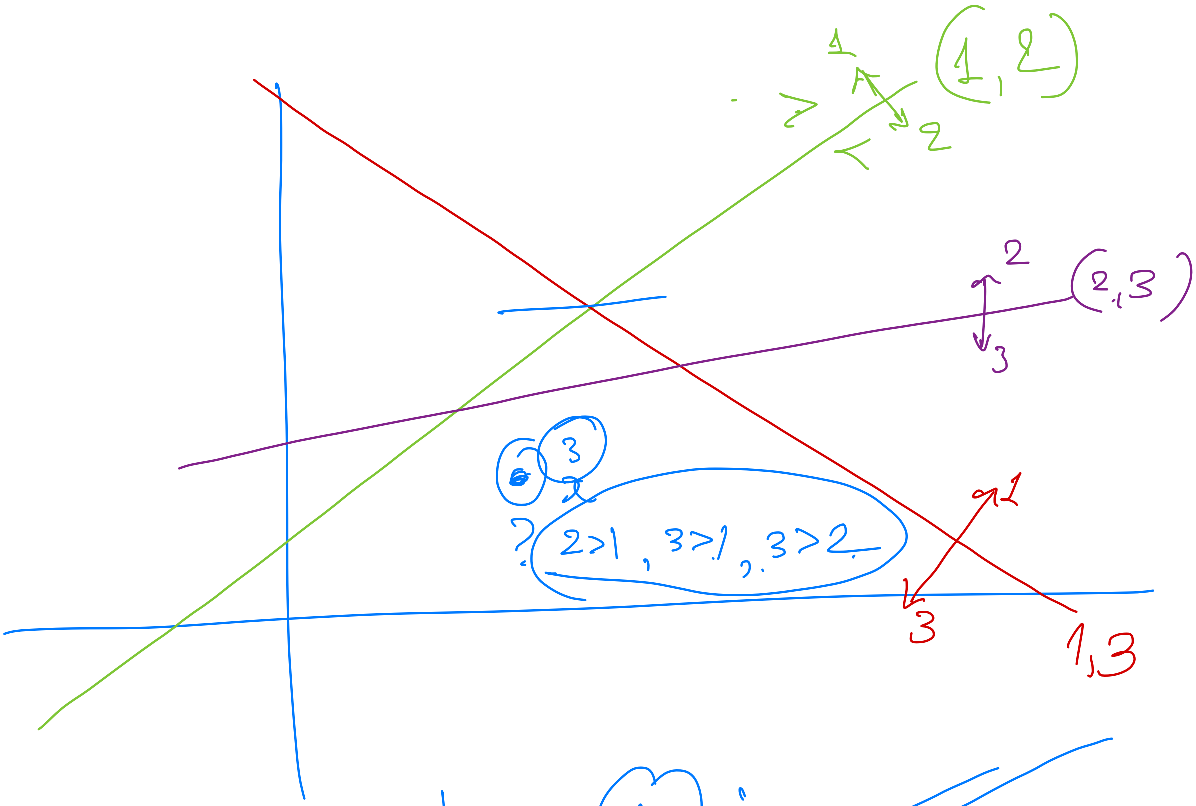
Κανόνας για διάκριση μεταξύ k, l.

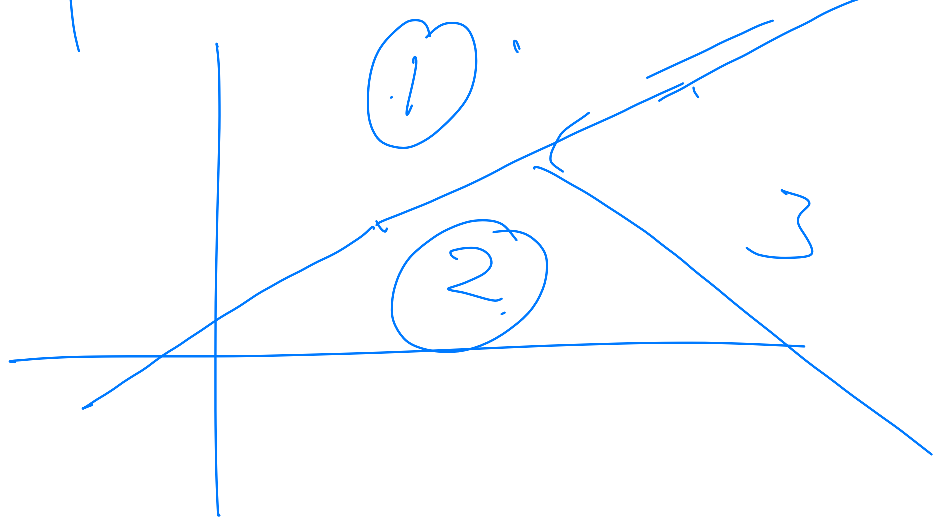


K=2

decision boundary μεταξύ 1, 2.

K=3





Παράδειγμα $x \in \mathbb{R}$, $K=3$