4-10-2021

X, Y , $X \in \mathbb{R}^p$, $Y \in \mathbb{R} \leftarrow$ εξαρτημένη

$p(X,Y)$

Πρόβλεψη: Given $X=x$, ποια είναι η "καλύτερη" πρόβλεψη για το Y?

(διαισθητικά $\hat{y}(x) = E(Y|X=x)$)

Συνάρτηση απώλειας

$\underbrace{L(f(X),Y)}_{\text{τ.μ.}}$ = ποινή/κόστος αν η πραγματική τιμή είναι Y κι' εμείς προβλέψουμε $f(X)$.

$(L(Y,Y)=0.)$

$$\min_f \overbrace{\boxed{E_{(X,Y)} L(f(X),Y)}}^{EPE(f)} = \min_f EPE(f)$$

Περίπτωση 1

Y συνεχής, $L(f(X),Y) = (Y-f(X))^2$

$$EPE(f) = E(Y-f(X))^2$$

$$= \underbrace{\int_x \overbrace{E\big((Y-f(x))^2 \,\big|\, X=x\big)}^{\min_x \ \forall x} \cdot f_X(x)\, dx}_{\min} .$$

$$\min_{(x_1,\dots,x_n)} \left[\sum_j f_j(x_j)\right] \iff \min_{x_j} f_j(x_j) \Rightarrow x_j^*$$

$x^* = (x_1^*, \dots, x_n^*)$ βέλτιστο

$\forall x \in \mathbb{R}^p$: $\min_c E\left[(Y - \underset{f(x)}{c})^2 \,|\, X = x\right] \Rightarrow \dots$ $\boxed{c^* = E(Y|X=x)}$

Αν χρησιμοποιήσουμε $L(f(X), Y) = |Y - f(X)|$

$\Rightarrow \dots \Rightarrow$ $f(X) = \text{median}(Y|X=x)$ [δείξτε το!]

$\boxed{E(Y|X=x) : \text{συνάρτηση παλινδρόμησης της } Y \text{ ως προς } X}$

Linear models $E(Y|X=x)$ άγνωστη

υποθέτουμε $E(Y|X=x) = x^T \cdot b$ (b: άγνωστες παράμετροι)

$\Rightarrow \hat{b}$ LSE $\Rightarrow$ $\boxed{\hat{y} = x^T \cdot \hat{b}}$ ← εξ. παλινδρόμησης

Περίπτωση 2 : Classification

(X, G) $\quad X \in \mathbb{R}^p$

G: εξαρτ. μεταβλητή $\in \{G_1, G_2, \dots, G_m\}$
(κατηγορική)

$E(G)$ δεν έχει νόημα

Συνάρτηση απώλειας $\quad L(\hat{G}(X), G)$

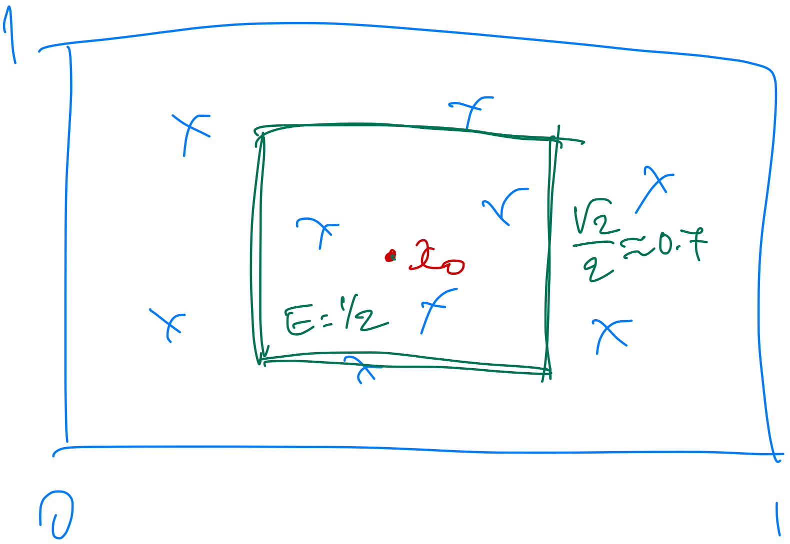
$$\min_{\hat{G}} EPE(\hat{G}) = E_X\left[E_{G|X}\left(L(\hat{G}(X), G)\right)\right]$$

$$= E_X\left[\sum_{k=1}^{m} L(\hat{G}(X, G_k)) \cdot P(G = G_k \mid X)\right]$$

Ισοδύναμα

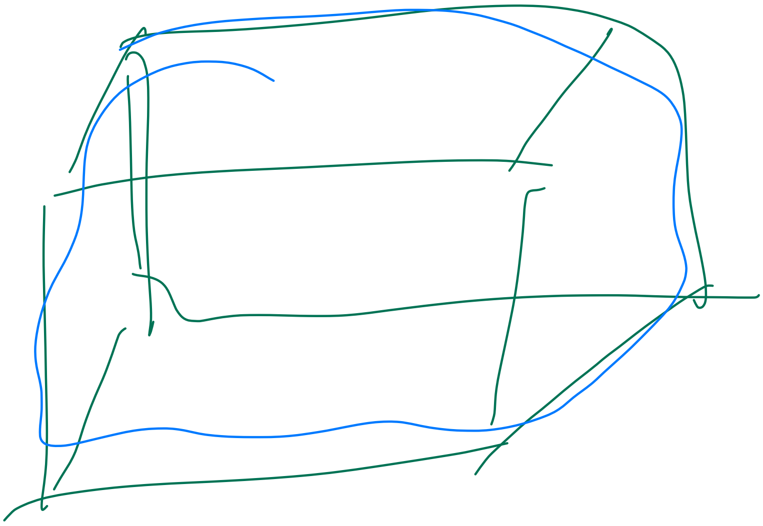
$$\forall x : \quad E \sum_{k=1}^{m} L(\hat{G}(x), k) \cdot P(G = G_k \mid X = x)$$

Μεγάλες διαστάσεις $(p >>)$



k: επιλέγω $r = 50\%$ των σημείων.

$p = 100$



"κατάρα των διαστάσεων"

Curse of dimensionality

0-1 απώλεια

$$L(\hat{G}(x), k) = \begin{cases} 0, & \hat{G} = G_k \\ 1, & \hat{G} \neq k \end{cases}$$

$$\sum_{k=1}^{m} L(\overset{g}{\overset{\parallel}{\hat{G}(x)}}, \underline{G_k}) P(G = G_k \mid X = x) =$$

$$= \sum_{\substack{k=1 \\ G_k \neq g}}^{m} 1 \cdot P(G = G_k \mid X = x) + 0 \cdot P(G = g \mid X = x)$$

$$= P(G \neq g \mid X = x) = \underbrace{1 - P(G = g \mid X = x)}_{\min}$$

$$g^*(x) : \boxed{\max_g P(G = g \mid \underline{X = x})}$$

Bayes classifier

Στατιστικά Μοντέλα

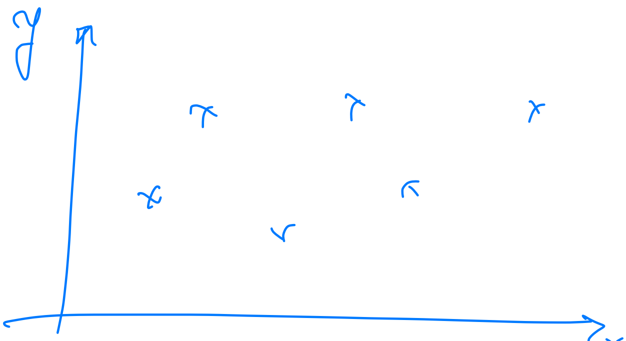
$$f(x) = E(Y|X=x)$$

Αν δεν γνωρίζουμε την $P(X,Y)$ η $f(x)$ πρέπει να χρησιμοποιήσουμε πληροφορία από δείγμα

$\boxed{\text{προσέγγιση της } f(x)}$

προσέγγιση συναρτήσεων

π.χ. Αρ. Ανάλυση   πολυωνυμική παρεμβολή



$(x_1, y_1) \ldots (x_n, y_n) \Rightarrow$ πολυώνυμο βαθμού m

$$f(x) = b_0 + b_1 x + \ldots + b_m x^m \quad ;$$

$$\sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 \leftarrow \min$$

Κατηγορίες Συναρτήσεων Πρόβλεψης

1 Αναπτύγματα

$$f(x) = \sum_{k=0}^{K} \theta_k h_k(x)$$

$\underbrace{h_k(x),\ k=1,\dots,K}_{\text{γνωστές συναρτήσεις}}$

π.χ. $h_k(x) = x^k$

$$f(x) = \sum \theta_k x^k$$

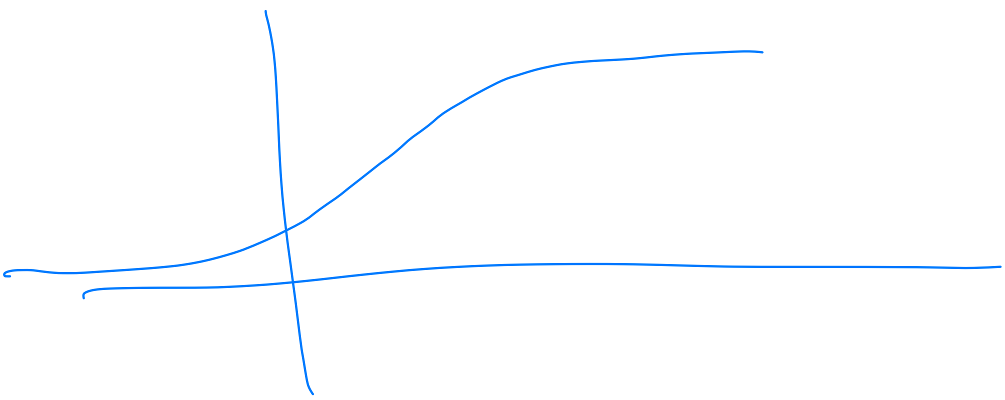
π.χ. $h_k(x) = \sin(kx + \varphi)$

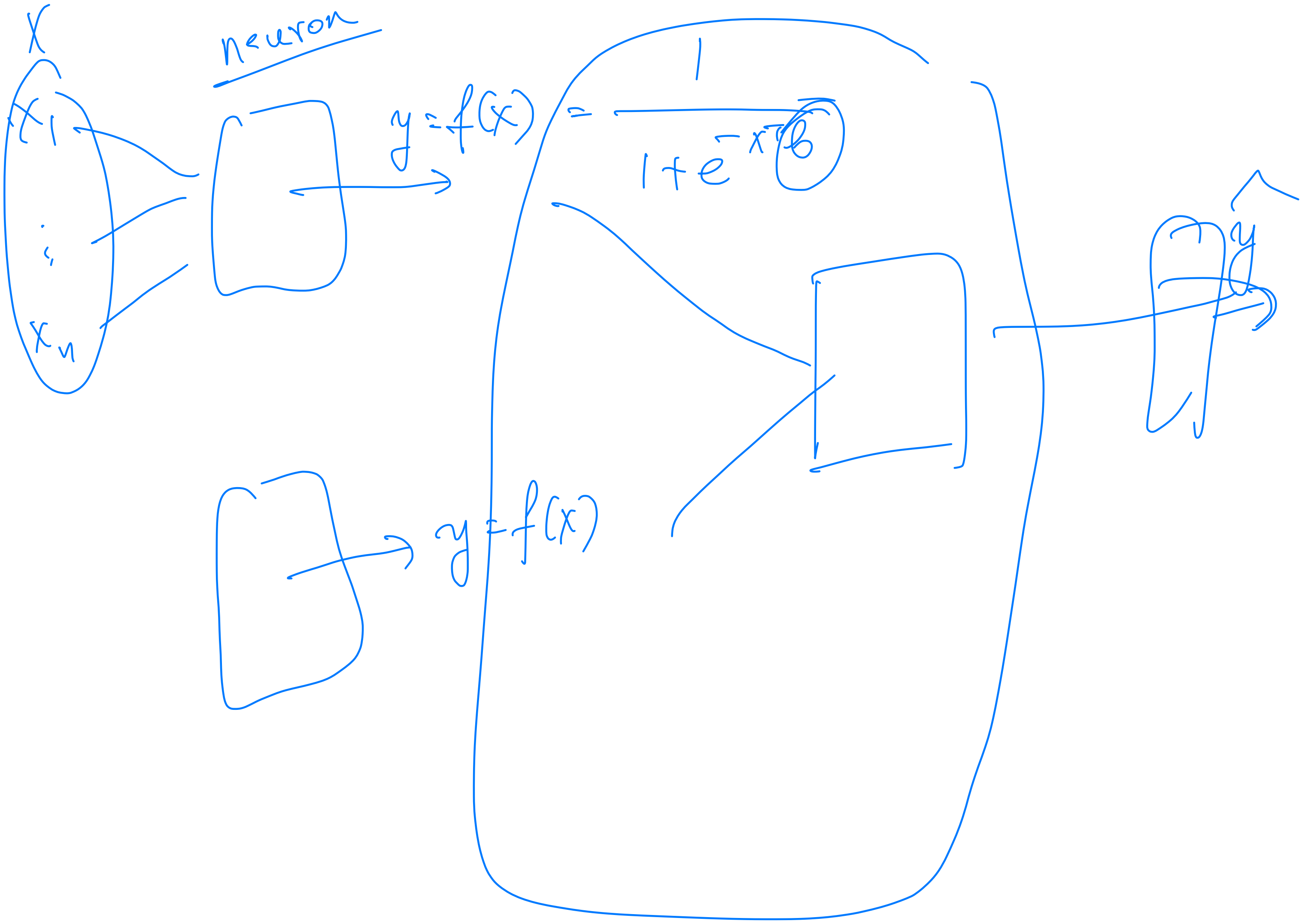
Γραμμικό Μοντέλο

$$f(x) = x^T \cdot b = \sum_j b_j x_j$$

Νευρωνικά δίκτυα

$$h_k(x) = \frac{1}{1 + e^{-x^T b_k}}, \quad b_k \in \mathbb{R}^p$$




X
x_1
x_n
neuron
$y = f(x) = \frac{1}{1+e^{-x^T\theta}}$
$y = f(x)$
$\hat{y}$

② Structured regression methods

Επιπλέον κριτήρια στο γραμμικό μοντέλο.
(επιλογής μεταβλητών)

① Stepwise Regression

② Penalty methods.

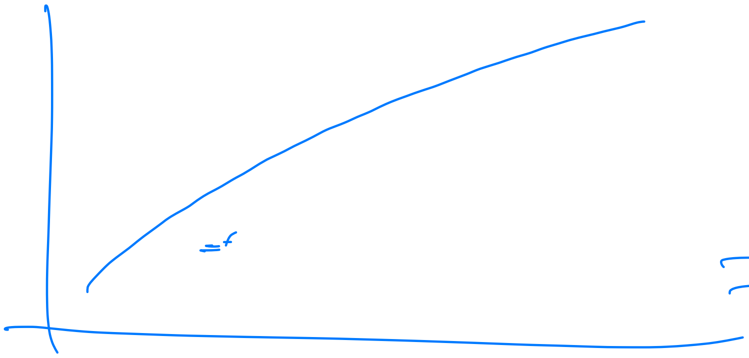
Συνάρτηση απώλειας / σφάλματος

$$\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2 \quad \text{κλασική} \Rightarrow \text{LSE}$$

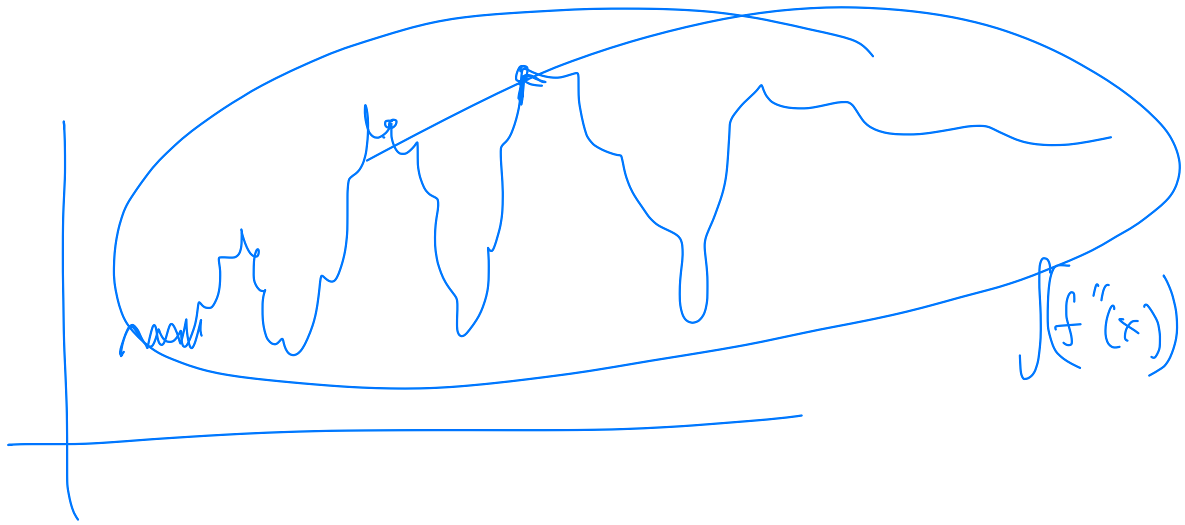
$$f(x_i) = x_i^T \cdot b$$

$$\sum_{i=1}^{N}\left(y_i - f(x_i)\right)^2 + \lambda \int \left(f''(x)\right)^2 dx \quad \leftarrow \min.$$

$f(x)$ γραμμική $\Rightarrow f''(x) = 0$.



$\int (f''(x))^2 \, dx$ μικρό



$\int (f''(x))^2 \, dx$ μεγάλο

Lasso regression

$$\min \sum (y_i - f(x_i))^2$$

s.t. αρ. μεταβλητών

③ Kernel methods

$$\sum_{i=1}^{N} (y_i - f(x_i))^2 \cdot K(x_i, y_i)$$

kernel function