13-10-2021

Regularized Regression

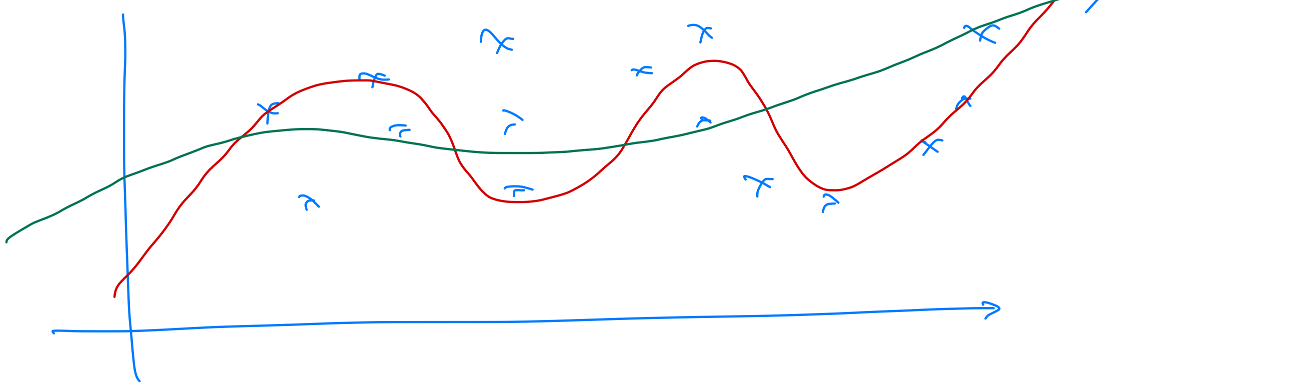
Ridge Regression

$$Y = b_0 + b_1 X_1 + \cdots + b_p X_p \qquad b = \begin{pmatrix} b_0 \\ \vdots \\ b_p \end{pmatrix}$$

$$\min_{b} \; RSS(b) = \sum_{i=1}^{N} (y_i - b_0 - b_1 X_{i1} - \cdots - b_p X_{ip})^2 = (y - Xb)^T (y - Xb)$$

s.t. $\sum_{j=1}^{N} b_j^2 \leq t$

π.χ. $Y = b_0 + b_1 X + b_2 X^2 + \cdots + b_p X^p$



① Scaling $\quad b_1, \ldots, b_p \Leftrightarrow X_1, \ldots, X_p$

π.χ. X_1 (σε km)

Αν εκφραστεί σε m $\Rightarrow X_1' = 1000 \cdot X_1$

αρχικό $\quad Y = \underbrace{b_0}_{} + \underbrace{b_1}_{} X_1 \qquad \gamma_1 = \frac{b_1}{1000}$

νέο $\quad Y = \gamma_0 + \overline{\gamma_1 \cdot 1000}\, X_1$

$$Y = b_0 + b_1 X_1 + \cdots + b_p X_p$$

ⓐ Κανονικοποίηση $\quad x_i' = \dfrac{x_i - x_{min}}{x_{max} - x_{min}} \quad (x_i' \in [0,1])$

$$x_{min} = \min\{x_i : i=1,\dots,N\}$$
$$x_{max} = \max\{x_i : i=1,\dots,N\}$$

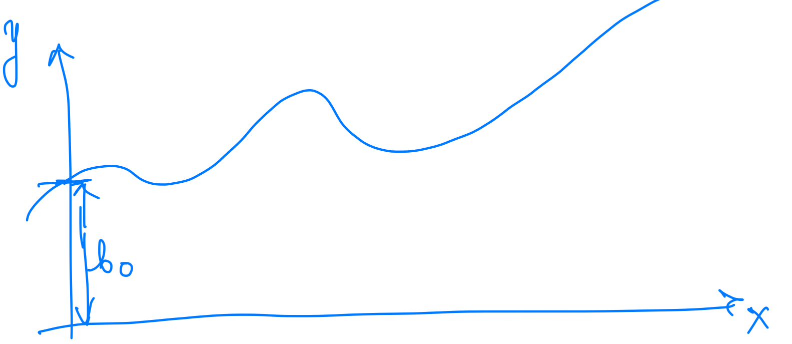
ⓑ Τυποποίηση $\quad \tilde{x}_i = \dfrac{x_i - \bar{x}}{\sigma(x)} \qquad \bar{x} = \dfrac{x_1 + \dots + x_N}{N}$

$\sigma(X)$ = τυπ. απόκλιση της X στο δείγμα.

$$\bar{\tilde{x}} = 0$$
$$\sigma(\tilde{x}) = 1$$

---

② Σταθερός όρος : b_0



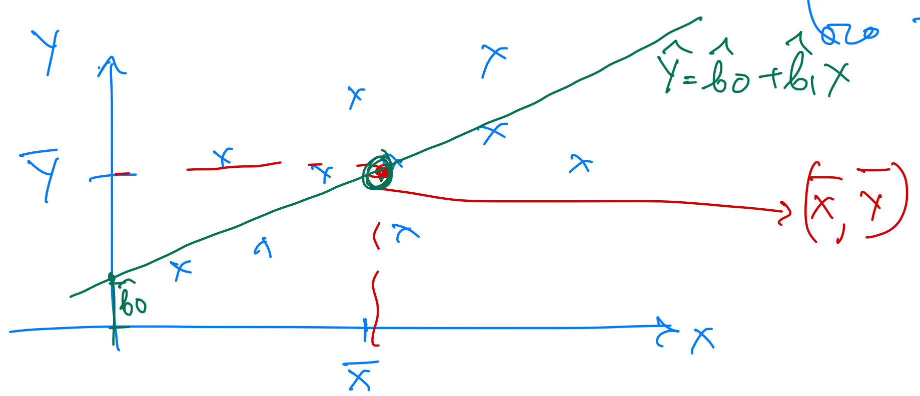
Κεντρικοποίηση $\quad X_1, \dots, X_p$

Στο μοντέλο $\quad Y = b_0 + b_1 X_1 + \dots + b_p X_p$

οι εκτιμήτριες LS έχουν την ιδιότητα

$$\bar{Y} = \hat{b}_0 + \hat{b}_1 \bar{X}_1 + \dots + \hat{b}_p \bar{X}_p$$

$\bar{Y}, \bar{X}_1, \dots, \bar{X}_p$ (δειγματικοί μέσοι στο training set).



κέντρο βάρους δείγματος

$$\hat{b}_0 = \overline{Y} - \hat{b}_1 \overline{X}_1 - \cdots - \hat{b}_p \overline{X}_p$$

$$X_1^c = X_1 - \overline{x}_1 \;, \; \cdots \;, \; X_p^c = X_p - \overline{x}_p$$

$$x_{ij}^c = x_{ij} - \overline{x}_j$$

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \cdots + \hat{b}_p X_p$$

$$= \hat{b}_0 + \hat{b}_1 (X_1^c + \overline{x}_1) + \cdots + \hat{b}_p (X_p^c + \overline{x}_p)$$

$$= \underbrace{(\hat{b}_0 + \hat{b}_1 \overline{x}_1 + \cdots + \hat{b}_p \overline{x}_p)}_{= \overline{y}} + \hat{b}_1 X_1^c + \cdots + \hat{b}_p \cdot X_p^c$$

$$\Rightarrow \hat{Y} = \overline{y} + \sum_{j=1}^{p} \hat{b}_j X_j^c$$

Αν $\overline{X}_j = 0 \Rightarrow$ απαιτούνται μόνο $b_1, \ldots, b_p$.

$$\min \; RSS(b) = (y - Xb)^T (y - Xb) \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_p \end{pmatrix}$$

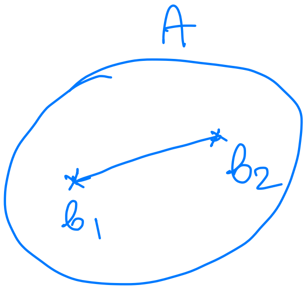
$$\text{s.t.} \quad b^T b \le t$$

Πρόβλημα κυρτού προγραμματισμού

$(y - Xb)^T (y - Xb) = RSS(b) \leftarrow$ κυρτή (b).

$\{ b : b^T b \le t \}$: κυρτό σύνολο.

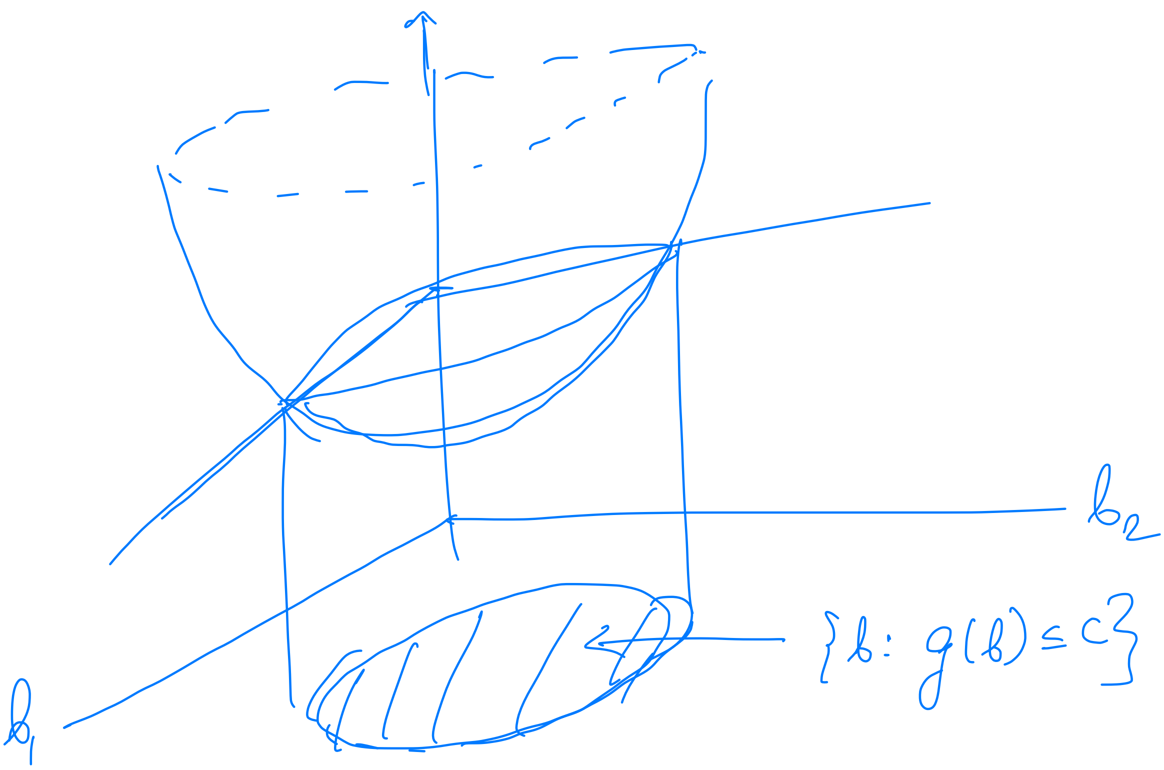
$A \subset \mathbb{R}^p$ κυρτό



$g(b) = b^T b$: κυρτή συνάρτηση του b

<u>Θεώρημα</u> Αν $g(b)$ κυρτή $(b \in \mathbb{R}^p)$

τότε $\forall \, c \in \mathbb{R}$: $\{ b : g(b) \le c \}$ κυρτό σύνολο



Lagrangean

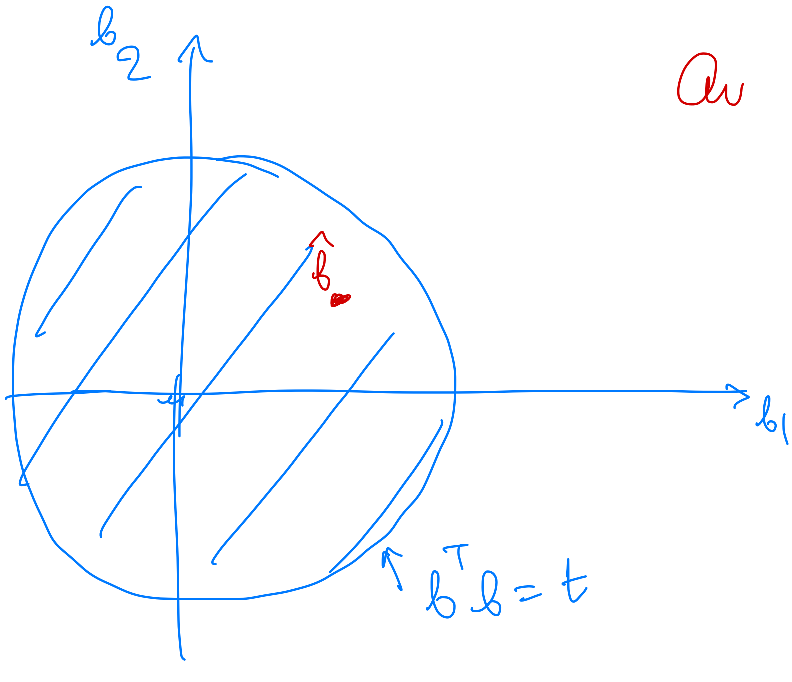
$$L(b,\lambda) = RSS(b) + \lambda\,(b^T b - t \le 0)$$

Βέλτιστη λύση : $\dfrac{\partial L}{\partial b_j} = 0$, $j = 1, \dots, p$

p εξισώσεις $\Rightarrow$ $b_1^*(\lambda)$, $b_2^*(\lambda)$, ..., $b_p^*(\lambda)$

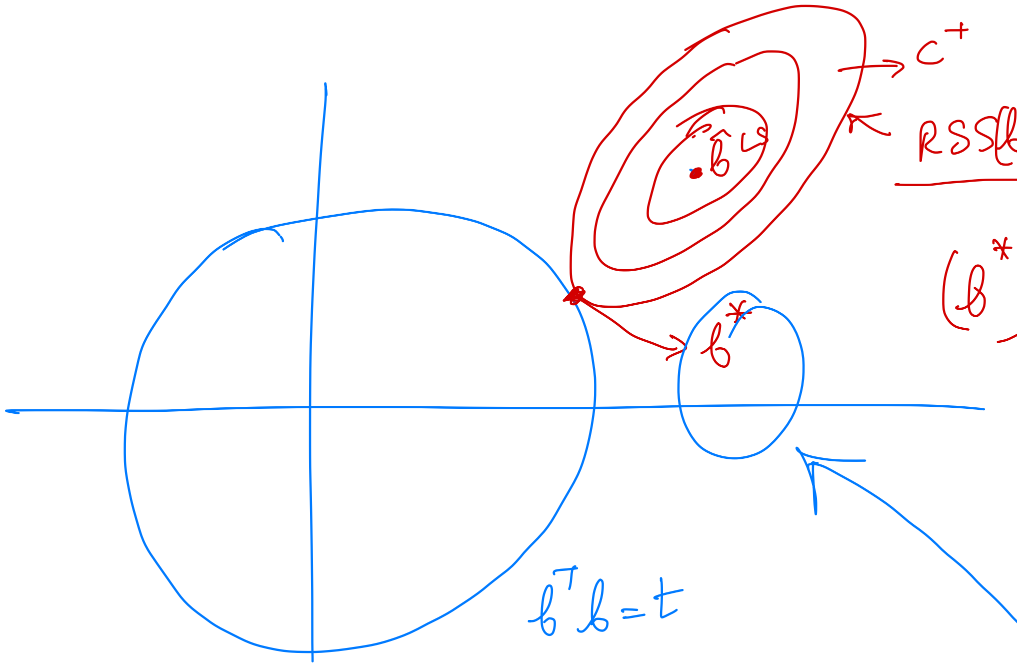
Για το λ :

Έστω $\hat{b}^{LS}$: οι εκτ. ελαχ. τετραγ.: $\min_b RSS(b)$



Αν $(\hat{b}^{LS})^T(\hat{b}^{LS}) \le t$

$\Rightarrow$ $\hat{b}^* = \hat{b}^{LS}$, $\lambda = 0$.



$RSS(b) = c$

$(b^*)^T (b^*) = t$

Επομένως επιπλέον εξίσωση

$$\sum_{j=1}^{p} b_j^*(\lambda) = t \Rightarrow \lambda^*(t) \Rightarrow \left\{ b_1^*, \ldots, b_p^* \right\}$$

$\exists$ ~~μια~~ 1–1 αντιστοιχία μεταξύ t κ λ.

και ισοδύναμα

$$\min RSS(b) \quad \sum b_j^2 \le t \iff \min_{b} \left\{ RSS(b) + \lambda \sum_{j=1}^{p} b_j^2 \right\}$$

$$\underbrace{\min_{b} (y - Xb)^T (y - Xb) + \lambda \underbrace{b^T b}} = RSS(\lambda)$$

λ : παράμετρος

$$\nabla_b RSS = -2X^T y + \underline{2(X^T X) \cdot b} + \underline{2\lambda b}$$

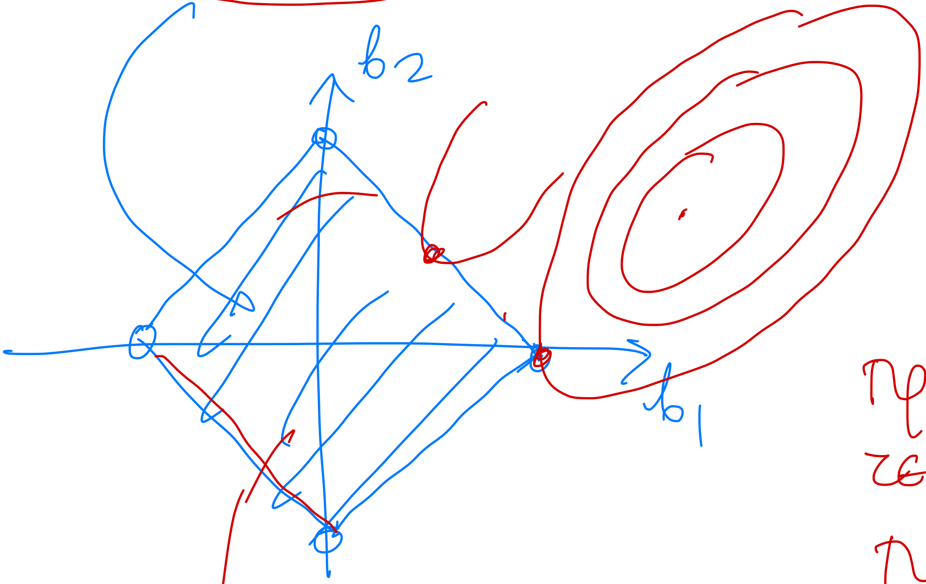
$$= 2\left[(X^T X + \lambda I) b - X^T y \right] = 0$$

$$\Rightarrow \boxed{\hat{b}^{ridge} = (X^T X + \lambda I)^{-1} \cdot X^T y}$$

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

Lasso Regression

$$\min \; RSS(b) \quad \text{s.t.} \; \sum_{j=1}^{p} |b_j| \le t \iff \min_{b_1,\dots,b_p} RSS(b) + \lambda \sum_{j=1}^{p} |b_j|$$



πρόβλημα τετραγωνικού προγραμματισμού

$$b_1 + b_2 \le t$$
$$-b_1 - b_2 \le t$$
$$b_1 - b_2 \le t$$
$$-b_1 + b_2 \le t$$

Ridge Regression

X

singular value decomposition

$$X = UDV^T$$

$$X^T = VDU^T$$

D : διαγώνιος

U, V ορθογώνιοι

$U^TU = I \quad , \quad V^TV = I$

$U, V \in \mathbb{R}^{N \times p}$

$$X^TX = VD\underbrace{U^TU}DV^T = VD^2V^T$$

$$D^{-2} = (D^2)^{-1}$$

$$(X^TX)^{-1} = VD^{-2}V^T$$

$$(VD^{-2}\underline{V^T)(V}D^2V^T) = I$$

LSE

$$\hat{b}^{LSE} = (X^TX)^{-1} \cdot X^Ty =$$

$$= VD^{-2}V^TVDU^Ty = VD^{-1}U^Ty$$

$$\hat{y}^{LSE} = X\hat{b}^{LSE} = \dots = UU^T y = \sum_j u_j u_j^T \cdot y$$

$$\hat{y}^{ridge} = X\hat{b}^{ridge} = \dots U D (D^2 + \lambda I)^{-1} D U^T y$$

$$= \sum_{j=1}^{p} u_j \frac{d_j^2}{d_j^2 + \lambda} u_j^T \cdot y \qquad U = [u_1, \dots, u_p]$$

d_j = singular values του πίνακα X.

↳ σχετίζονται με κύριες συνιστώσες του X.

Άσκ. 21

G κατηγορική

κωδικοποίηση

$$\left\{\begin{array}{l} g_1 \\ \vdots \\ g_k \end{array}\right. \qquad \left.\begin{array}{ll} (1, 0, \ldots 0) & X_1 \\ (0, 1, \ldots 0) & \vdots \\ (0, 0, \ldots 1) & X_k \end{array}\right\}$$

$$Y = b_0 + b_1 X_1 + \cdots + b_k X_k$$

$\hat{y}$: prediction

$$\hat{G}_i = \qquad \hat{y}_i = \begin{bmatrix} \hat{y}_{i1} \\ \vdots \\ \hat{y}_{ik} \end{bmatrix}$$