3-4-2025

μπορεί να υπολογιστεί κ' αναλυτικά

Αλγόριθμοι για εκτίμηση $b(\lambda,n) = E\left(\frac{1}{\bar{X}_n} - \lambda\right)$

όπου $X_1, \dots, X_n$ iid $\sim Exp(\lambda)$.

(1) Monte Carlo : παραμετρική ως προς $(\lambda, n)$.

Δημιουργούμε $N$ ανεξάρτητα δείγματα μεγέθους $n$ από $Exp(\lambda)$

$\forall$ δείγμα $\Rightarrow$ $\bar{X}_n$ , $\left(\frac{1}{\bar{X}_n} - \lambda\right)$

$\hat{b}(n,\lambda)$ = μέση τιμή του

(2) Έστω ότι έχουμε δείγμα μεγέθους $n$ από $Exp(\lambda)$, (αλλά δεν γνωρίζουμε το $\lambda$)

Υπολογίζω το $\hat{\lambda}_n = \frac{1}{\bar{X}_n}$ (πόσο είναι το bias της $\hat{\lambda}_n$?)

Bootstrap

Παίρνουμε $N$ δείγματα μεγέθους $n$ με επανάθεση από το αρχικό δείγμα

Για κάθε δείγμα υπολογ. $\frac{1}{\bar{X}_n}$ και $\hat{\lambda}_n - \frac{1}{\bar{X}_n}$

Εκτίμηση bias mean των $N$ διαφορών

③ R function boot (library(boot))

boot (dataset, boot.fn, R)

dataset : αρχικά δεδομένα

R = αρ. bootstrap samples

boot.fn : συνάρτηση που περιγράφει πώς υπολογίζεται η ποσότητα που μας ενδιαφέρει από κάθε bootstrap sample

boot.fn = function(dataset, index)

Εδώ boot.fn = function(initial_sample, index)

$$\text{return}\left(\underbrace{\frac{1}{\text{mean(initialsample)}}}_{\hat{\lambda}_n} - \underbrace{\frac{1}{\text{mean(initial\_sample[index])}}}_{\frac{1}{\bar{X}_n}\ \text{(bootstrap)}}\right)$$

# Shrinkage methods

Μοντέλο παλινδρόμησης

$$y = b_0 + b_1 x_1 + \dots + b_p x_p + \varepsilon$$

Ερώτημα: Να βρεθούν ποιες σημαντικές μεταβλητές για την πρόβλεψη της $y$.

Μέθοδοι επιλογής μεταβλητών (π.χ. stepwise) $\|b\|_0$ ?

$$\min_{S \in \{1,\dots,p\}} \left\{ \min_{b_0,\dots,b_p} \left\{ \sum_{j=1}^{n} (y_j - b_0 - b_1 x_{j1} - \dots - b_p x_{jp})^2 : \sum_{i=1}^{n} 1(b_i \neq 0) \leq S \right\} \right\}$$

"Ποιες είναι οι $S$ καλύτερες μεταβλητές"

Αρ. μοντέλων $\displaystyle\sum_{S=1}^{p} \binom{p}{S} = \underline{2^p}$

# Ridge Regression

$$\min_{b_0, b_1, \dots, b_p} \left\{ \sum_{j=1}^{n} (y_j - b_0 \dots)^2 : \sum_{i=1}^{n} b_i^2 \leq s \right\}$$

$$\|b\|_2^2 \leq s^2$$

$$\Updownarrow$$

$$\min_{b} \left\{ \sum_{j=1}^{n} (y_j - b_0 \dots)^2 + \lambda \underbrace{\sum_{i=1}^{n} b_i^2}_{\text{penalty term}} \right\}$$

$\lambda$: (Lagrange multiplier)

penalty parameter

$\lambda \to 0 \quad \Rightarrow \quad \hat{b}$ : LSE estimates

$\lambda \to \infty \quad \Rightarrow \quad \hat{b} \to 0$

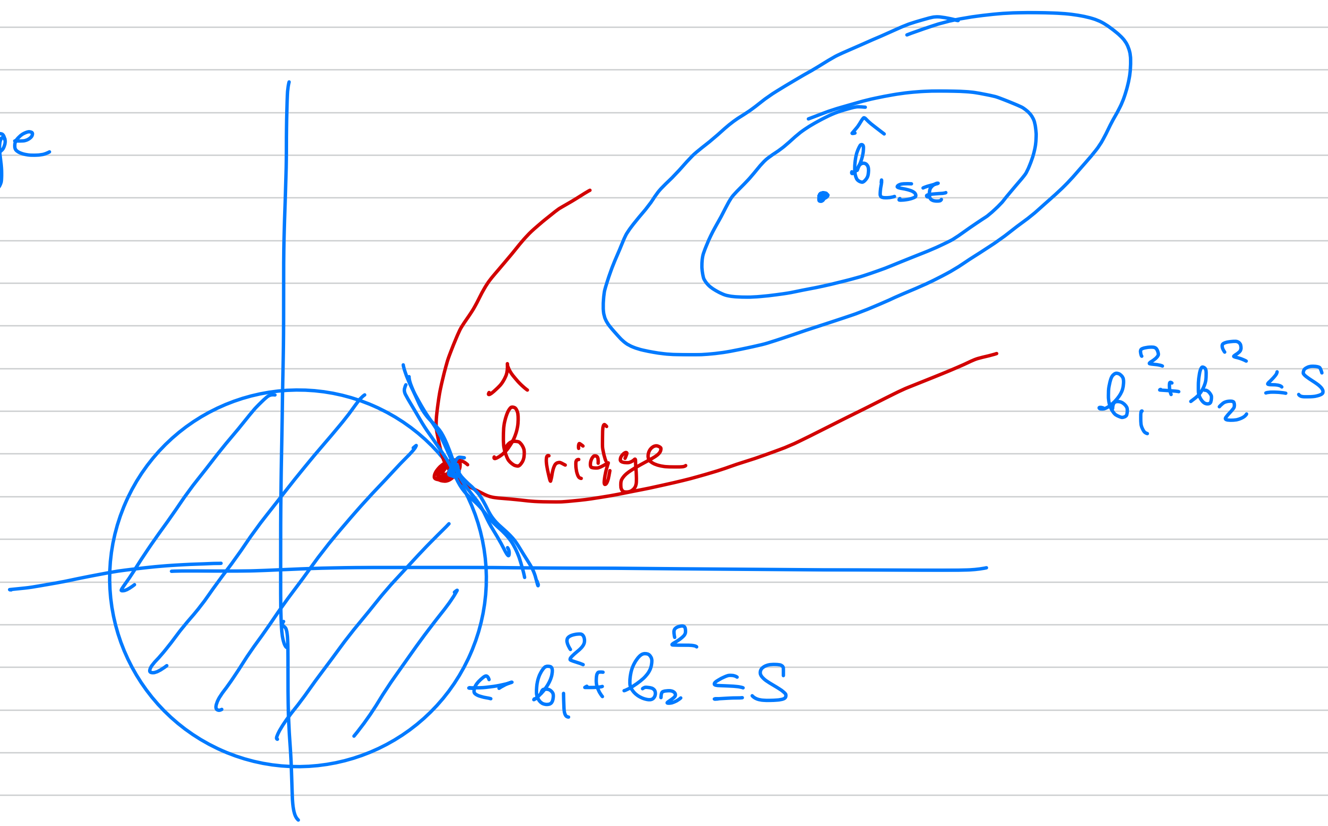
## Lasso Regression

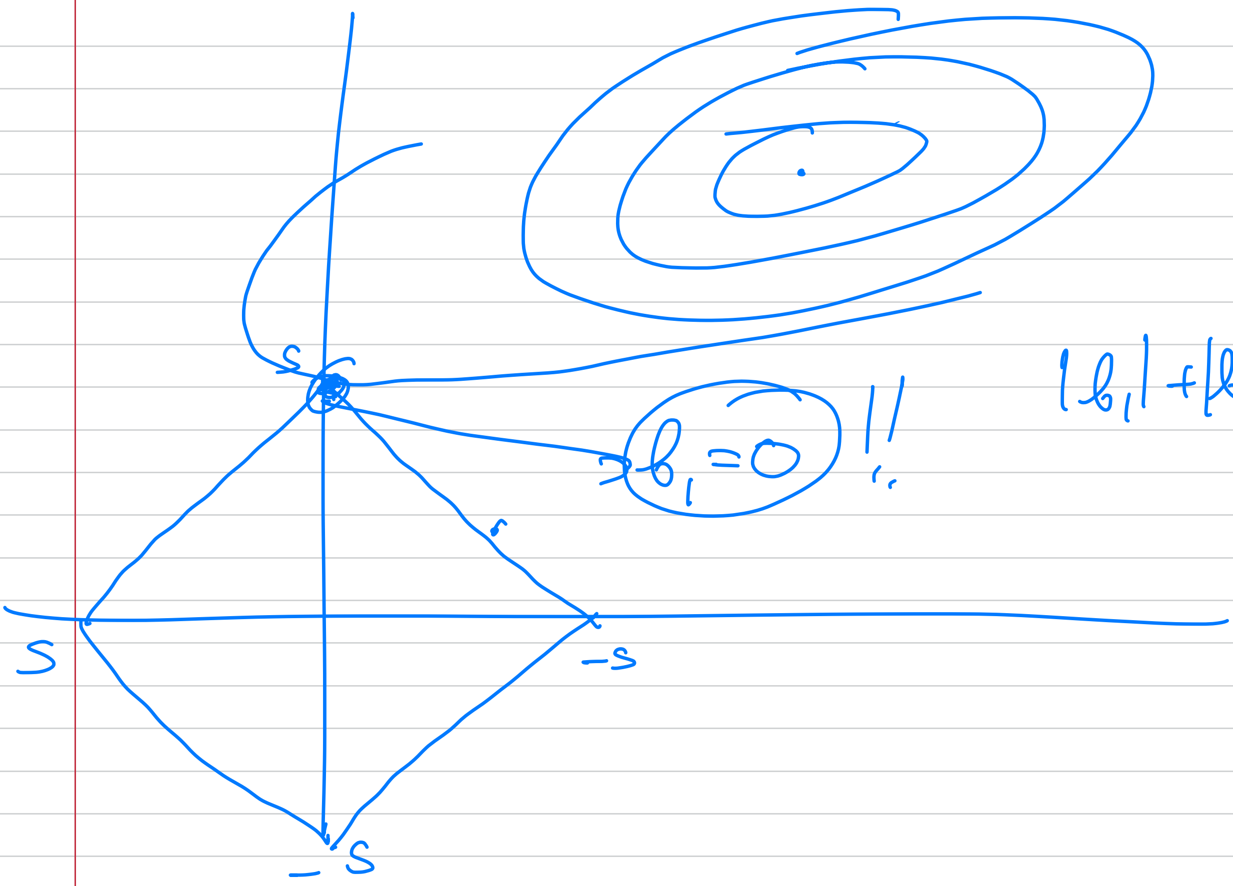
$$\min \left\{ \sum (y_j - b_0 \cdots)^2 : \underbrace{\sum_{i=1}^{n} |b_i|}_{\|b\|_1} \le S \right\}$$

$$\Rightarrow \quad \min \left\{ \sum (y_j - b_0 \cdots)^2 + \lambda \sum_{i=1}^{n} |b_i| \right\}$$

$p=2$

Ridge






$|\beta_1| + |\beta_2| \leq s$