

5-6-2025

Διαγώνισμα 2

① Training set $\begin{cases} \rightarrow n \text{ παρατηρήσεις} \\ \rightarrow p \text{ μεταβλητές } (X_1, \dots, X_p) \end{cases}$

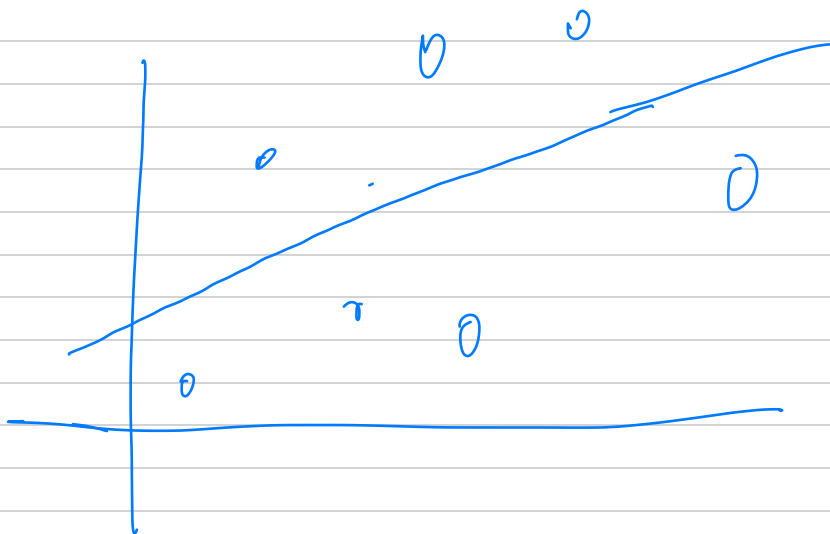
α) 2k : μεγαλύτερη ευελιξία

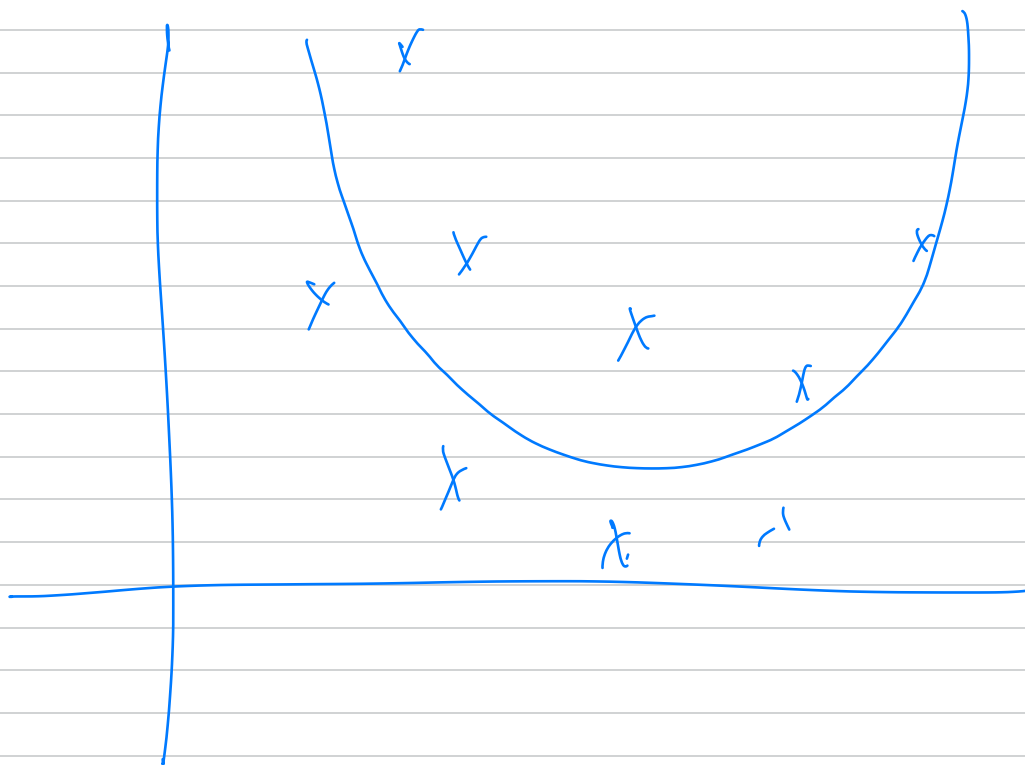
β) ① $\begin{cases} n \text{ μεγάλο} \\ p \text{ μικρό} \end{cases} \rightarrow \text{αποτελεσματική μεγαλύτερη ευελιξία}$

② $\begin{cases} n \text{ μικρό} \\ p \text{ μεγάλο} \end{cases} \rightarrow \text{μικρότερη ευελιξία}$

③ $\begin{cases} \text{ισχυρά μη} \\ \text{στατιστικά όρια} \end{cases} \rightarrow \begin{cases} \text{μικρή ευελιξία} \rightarrow \text{poor fit} \\ \rightarrow \text{bias} \end{cases}$

④ $\sigma^2 = \text{Var}(\varepsilon)$ από μεγάλο





Problem 2 (b)

$\text{glmnet} \rightarrow \alpha \in [0, 1]$
 $\alpha = 0 \Rightarrow \text{ridge}$
 $\alpha = 1 \Rightarrow \text{lasso}$

$\text{glmnet}(\underline{x, y, \alpha = ., \lambda = (\text{range})})$

$cv_{\text{aet}} = cv.glmnet(x^{\text{train}}, y^{\text{train}}, \alpha = 0)$



$cv_{\text{out}} \& \lambda_{\text{min}}$ ←

$x_{\text{train}} = \text{model.matrix}(\text{lpse} \sim -id - \text{train},$
 $\text{data} = \text{prostate}[, \text{prostate} \& \text{train}],$

$y_{\text{train}} = \text{prostate}[\text{train},] \& \text{lpse}$