Practical 5, Group A
March, 26th

1. Theoretical part

We want to prove that (1) $\Leftrightarrow$ (2), where

(1) $\min_{j\in P\setminus S} ||y - X^*\hat{\beta}^* - x_j\beta_j||_2^2$

(2) $\max_{j\in P\setminus S} |x_j^T(y - X^*\hat{\beta}^*)|$

Proof

Define the residual as $e^* = y - X^*\hat{\beta}^*$.
Let us compute the First Order Condition for

$$f(\beta_j) = ||y - X^*\hat{\beta}^* - x_j\beta_j||_2^2.$$

Expanding the L2 norm, we get

$$\begin{aligned} f(\beta_j) =&||y - X^*\hat{\beta}^* - x_j\beta_j||_2^2 \\ &= ||e^* - x_j\beta_j||_2^2 \\ &= (e^* - x_j\beta_j)^T(e^* - x_j\beta_j) \\ &= e^{*T}e^* - 2\beta_j^T x_j^T e^* + \beta_j^T x_j^T x_j \beta_j \\ &= e^{*T}e^* - 2\beta_j^T x_j^T e^* + \beta_j^T \beta_j, \end{aligned}$$

since we assume all predictors unitary. Differentiating according to β_j we get

$$\frac{\partial f(\beta_j)}{\partial \beta_j} = -2x_j^T e^* + 2\beta_j = 0 \Leftrightarrow \hat{\beta}_j = x_j^T e^*.$$

Therefore minimising $f(\beta_j), \forall j$ is equivalent to

$$\begin{aligned} &\min_{j\in P\setminus S} e^{*T}e^* - 2\beta_j^T\beta_j + \beta_j^T\beta_j \\ &\Leftrightarrow \min_{j\in P\setminus S} e^{*T}e^* - 2||\beta_j||_2^2 + ||\beta_j||_2^2 \\ &\Leftrightarrow \min_{j\in P\setminus S} e^{*T}e^* - ||\beta_j||_2^2 \\ &\Leftrightarrow \min_{j\in P\setminus S} -||\beta_j||_2^2, \end{aligned}$$

since e^* is independent of j. Using some optimisation result, we obtain

$$\min_{j\in P\setminus S} -||\beta_j||_2^2 \Leftrightarrow \max_{j\in P\setminus S} ||\beta_j||_2^2.$$

Since optimising over the L2 norm is equivalent to optimising over the L1 norm, we get the following optimisation problem

$$\max_{j \in P \setminus S} |\beta_j|.$$

Using the FOC, First Optimality Condition, this results in

$$\max_{j \in P \setminus S} |x_j^T e^*|.$$

Using the definition of e^*, we find the optimisation problem (2):

$$\max_{j \in P \setminus S} |x_j^T (y - X^* \hat{\beta}^*)|,$$

which completes the proof. $\square$