

TD 1: LASSO ET RÉGRESSION RIDGE

Nous allons utiliser pour ce travail les packages R appelés *glmnet* et *MASS* (pour générer des Gaussiennes multivariées).

- (i) Dans un premier temps, générer des observations provenant du modèle de régression linéaire

$$Y_i = \mathbf{X}_i' \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n,$$

où $\boldsymbol{\beta} \in \mathbb{R}^p$ et où les covariables sont indépendantes et Gaussiennes standard. On génère également des erreurs $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ i.i.d. de loi normale standard également.

- (ii) Calculer les estimateurs Ridge et Lasso de $\boldsymbol{\beta}$ pour plusieurs valeurs du paramètre de régularisation λ après avoir généré des observations comme ci-dessus.
- (iii) Décomposer l'échantillon en deux parties (training/test) et sélectionner λ par cross-validation.
- (iv) Répéter l'opération plusieurs fois (Monte-Carlo) et calculer l'erreur de prédiction moyenne sur ces répétitions.