

# MINI-PROJET

Le but de ce mini-projet est de mettre en évidence les différences de performance prédictive entre deux méthodes de régression régularisée : la régression *Ridge* (pénalisation L2) et la régression *Lasso* (pénalisation L1), à l'aide d'une figure de synthèse. Pour ce faire,  $n = 100$  données synthétiques ont été simulées selon un modèle linéaire, où la variable réponse  $y \in \mathbb{R}^n$  est liée linéairement à une matrice de design  $X \in \mathbb{R}^{n \times p}$  contenant  $p = 20$  variables explicatives simulées. Les entrées de  $X$  sont générées indépendamment selon une loi normale standard  $\mathcal{N}(0, 1)$ . Le vecteur de coefficients  $\beta \in \mathbb{R}^p$  est défini de manière parcimonieuse, avec seulement  $k = 8$  coefficients non nuls tirés selon une loi  $\mathcal{N}(0, 1)$ , les autres étant fixés à zéro. Un bruit additif  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ , avec  $\sigma = 1.0$ , est ajouté à la réponse. Pour chaque valeur du paramètre de régularisation  $\lambda$ , les modèles Ridge et Lasso sont ajustés sur un échantillon d'apprentissage, puis évalués sur un échantillon test à l'aide de la racine de l'erreur quadratique moyenne (RMSE). Cette opération est répétée 50 fois pour chaque valeur de  $\lambda$ , et la figure produite représente la RMSE moyenne observée en fonction de  $\lambda$ , permettant de comparer les comportements respectifs des deux approches. Le code source de cette analyse est disponible sur GitHub à l'adresse suivante : <https://github.com/Saadem/Mini-Projet-IFT-7021.git>

La figure ci-dessous révèle que, pour de très faibles valeurs de  $\lambda$ , les deux modèles affichent des RMSE faibles et proches, correspondant à une régularisation quasi inexistante. À mesure que  $\lambda$  augmente, la performance reste stable un certain temps, puis se dégrade rapidement sous l'effet d'une pénalisation excessive. Dans une zone intermédiaire, Lasso surpasse Ridge en raison de sa capacité à sélectionner automatiquement les variables pertinentes. En revanche, lorsque  $\lambda$  devient très élevé, les deux modèles présentent une hausse marquée de la RMSE et leurs performances tendent à converger, témoignant d'un fort sous-ajustement.

