
Feuille de TP 4 – Validation croisée

Régression Ridge

Dans cette partie, on cherche à calibrer le paramètre λ du problème de régression avec pénalisation Ridge

$$J_n(\vartheta) = R_n(\vartheta) + \lambda \text{pen}(\vartheta) = \frac{1}{2n} \sum_{i=1}^n (y_i - \langle x_i, \vartheta \rangle)^2 + \lambda \frac{1}{2} \sum_{j=1}^d |\vartheta_j|^2.$$

Pour $k = 1, \dots, n$, on note $\widehat{\vartheta}_{\lambda, n}^k$ le minimum de

$$J_{\lambda, n}^k(\vartheta) = \frac{1}{2n} \sum_{i=1, i \neq k}^n (y_i - \langle x_i, \vartheta \rangle)^2 + \lambda \frac{1}{2} \sum_{j=1}^d |\vartheta_j|^2.$$

Dans notre cas, on a une expression explicite de $\widehat{\vartheta}_{\lambda, n}^k$. On cherche ensuite à minimiser l'erreur de validation croisée

$$T(\lambda) = \sum_{k=1}^n \left(y_k - \langle x_k, \widehat{\vartheta}_{\lambda, n}^k \rangle \right)^2.$$

1. Pour le jeu de données `trees`, tracer la fonction $T(\lambda)$ en fonction de λ . On prendra des valeurs de $\lambda \in [0.00005, 0.0004]$.
2. Trouver (numériquement) le minimum de cette fonction avec la fonction `optimize`.