TP STATISTIQUES

Question 6 : Régression Ridge

Anaëlle CROISANT, Amr STITE, Antoine HUSSER

11 décembre 2025

Problème d'optimisation

On cherche à minimiser :

$$\min_{(\beta_0,\boldsymbol{\beta})\in\mathbb{R}^{p+1}} \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \boldsymbol{\beta}^T\mathbf{x}_i)^2 + \lambda\sum_{j=0}^{p}\beta_j^2 \tag{1}$$

Notation matricielle

Soit $\tilde{\mathbf{X}}$ la matrice augmentée avec une colonne de 1 :

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ 1 & x_{2,1} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}_{n\times(p+1)} \tag{2}$$

Et $\tilde{\boldsymbol{\beta}} = (\beta_0, \beta_1, \ldots, \beta_p)^T \in \mathbb{R}^{p+1}$
Le problème devient :

$$J(\tilde{\boldsymbol{\beta}}) = \frac{1}{n}\|\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}\|^2 + \lambda\|\tilde{\boldsymbol{\beta}}\|^2 \tag{3}$$

En développant nous avons :

$$\begin{aligned} J(\tilde{\boldsymbol{\beta}}) &= \frac{1}{n}(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}})^T(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}) + \lambda\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} && (4) \\ &= \frac{1}{n}\left[\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}\right] + \lambda\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} && (5) \end{aligned}$$

Calcul du gradient

Pour trouver le minimum, on calcule le gradient par rapport à $\tilde{\boldsymbol{\beta}}$:

$$\begin{aligned} \nabla J(\tilde{\boldsymbol{\beta}}) &= \frac{\partial}{\partial\tilde{\boldsymbol{\beta}}}\left[\frac{1}{n}\left(\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}\right) + \lambda\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}}\right] && (6) \\ &= \frac{1}{n}\left[0 - 2\tilde{\mathbf{X}}^T\mathbf{y} + 2\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}\right] + 2\lambda\tilde{\boldsymbol{\beta}} && (7) \\ &= \frac{2}{n}\left[\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} - \tilde{\mathbf{X}}^T\mathbf{y}\right] + 2\lambda\tilde{\boldsymbol{\beta}} && (8) \end{aligned}$$

En posant $\nabla J(\tilde{\boldsymbol{\beta}}) = \mathbf{0}$:

$$\frac{1}{n}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} - \frac{1}{n}\tilde{\mathbf{X}}^T\mathbf{y} + \lambda\tilde{\boldsymbol{\beta}} = \mathbf{0} \tag{9}$$

$$\frac{1}{n}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \lambda\tilde{\boldsymbol{\beta}} = \frac{1}{n}\tilde{\mathbf{X}}^T\mathbf{y} \tag{10}$$

$$\left(\frac{1}{n}\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} + \lambda\mathbf{I}_{p+1}\right)\tilde{\boldsymbol{\beta}} = \frac{1}{n}\tilde{\mathbf{X}}^T\mathbf{y} \tag{11}$$

En multipliant par n :

$$\left(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} + n\lambda\mathbf{I}_{p+1}\right)\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{X}}^T\mathbf{y} \tag{12}$$

Nous obtenons finalement :

$$\boxed{\tilde{\boldsymbol{\beta}}^{\text{Ridge}} = \left(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} + n\lambda\mathbf{I}_{p+1}\right)^{-1}\tilde{\mathbf{X}}^T\mathbf{y}} \tag{13}$$

Remarque sur l'inversibilité

La matrice $(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} + n\lambda\mathbf{I}_{p+1})$ est toujours inversible si $\lambda > 0$ car :

- Les valeurs propres de $\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$ sont $\mu_1, \ldots, \mu_{p+1}$ avec $\mu_i \geq 0$
- Les valeurs propres de $(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}} + n\lambda\mathbf{I})$ sont $\mu_i + n\lambda > 0$ pour tout i
- Donc la matrice est définie positive et inversible