Exercise 1 — La Ridge, le lasso et la constante — 8 points

1. Le problème de la ridg régression s'écrit

$$J(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right]^2 + \lambda \sum_{j=1}^{p} |\beta_j|^q. \tag{1}$$

   (a) Montrer que ce problème est équivalent à la minimisation de

$$J^c(\beta_0^c, \boldsymbol{\beta}^c) = \sum_{i=1}^{N} \left[y_i - \beta_0^c - \sum_{j=1}^{p} (x_{ij} - \bar{x}_j)\beta_j^c \right]^2 + \lambda \sum_{j=1}^{p} \left|\beta_j^c\right|^q.$$

   (b) Donner la correspondance entre $[\beta_0^c; \boldsymbol{\beta}^c]$ et $[\beta_0; \boldsymbol{\beta}]$.

   (c) Caractériser la solution de ce critère modifié.

   (d) Proposer une procédure à résoudre (1) en utilisant :

$$\widehat{\boldsymbol{\beta}}^{\mathrm{R}} = (X^\top X + \lambda I)^{-1} X^\top \boldsymbol{y}.$$

   (e) Montrer qu'un résultat analogue estpeut être établi pour le lasso Lasso.

2. Le Lasso pondérée et normalisé

   (a) Considérons le problème d'optimisation du Lasso pondéré (sur des données centrées sans terme contrant) :

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|,$$

   avec les poids positifs donnés $w_j > 0$ $(j = 1, \dots, p)$. Montrer que ce problème est équivalent à une tâche Lasso standard

$$\min_{\boldsymbol{\beta}'} \|\boldsymbol{y} - X'\boldsymbol{\beta}'\|_2^2 + \lambda \sum_{j=1}^{p} |\beta_j'|.$$

   (b) Une manière typique d'utiliser le Lasso est de l'appliquer sur des données normalisées comme suit :

      i. Normalise $\boldsymbol{y}$ et X : $\boldsymbol{y}^r = \frac{\boldsymbol{y} - \bar{y}}{\sigma_y}$ et $x_{ij}^r = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$ pour $i = 1, \dots, n$ et $j = 1, \dots, p$.

      ii. Minimise $\|\boldsymbol{y}^r - X^r\boldsymbol{\beta}^r\|_2^2 + \lambda \|\boldsymbol{\beta}^r\|_1$ par rapport à $\boldsymbol{\beta}^r$.

   Montrer que ce problème est équivalent à un Lasso pondéré avec un terme constant :

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \beta_0 - X\boldsymbol{\beta}\|_2^2 + \lambda \sum_{j=1}^{p} w_j |\beta_j|.$$

   Expliquer comment retrouver $\boldsymbol{\beta}$ et β_0 à partir de $\boldsymbol{\beta}^r$.

Exercise 2 — Questions courtes — 5 points

1. Quelle st la différence entre erreur de test et généralisation ?
2. Quel est le rapport entre le Lasso et la programmation quadratique ?
3. A quelles conditions vous semble-t-il indiqué d'utiliser Sklearn pour résoudre un problème de discrimination à deux classes ?
4. Quelles sont les différents composants d'une méthode de type autoML ?
5. Comment mesurer la proximité entre exemple pour faire de la réduction de dimensionnalité ?

Exercise 3 — Calculs non reinaux mais matriciels — 7 points

Soit W une matrice diagonale de taille n et de terme général w_i.

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 & \dots & \dots & 0 \\ 0 & w_2 & \dots & 0 & \dots & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & w_i & 0 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & \dots & 0 & \dots & w_{n-1} & 0 \\ 0 & \dots & \dots & 0 & \dots & 0 & w_n \end{pmatrix}$$

1. Donner la forme du terme général f_i du vecteur $\mathbf{f} = W\mathbf{e}$ en fonction des w_i et des e_i, où $\mathbf{e}$ est un vecteur de $\mathbb{R}^n$ de terme général e_i.
2. Donner également la forme générale de $\mathbf{e}^\top W\mathbf{e}$ en fonction des w_i et des e_i.
3. Pour des vecteurs $\mathbf{x} = (x_1, \dots, x_n)^\top$, $\mathbf{y} = (y_1, \dots, y_n)^\top$ et $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ donnés, calculer la solution du problème suivant :

$$\min_{a,b} \ J(a,b) \qquad \text{avec } J(a,b) = \frac{1}{2}\sum_{i=1}^{n} w_i\big(y_i - (a + bx_i)\big)^2$$

4. Calcul matriciel :
   a) Écrire matriciellement $J(\alpha)$ avec $\alpha = (a,b)^\top$ et en fonction des vecteurs $\mathbf{x}, \mathbf{y}, \mathbb{1}$ et de la matrice W (où $\mathbb{1} = (1, \dots, 1)^\top$ est un vecteur de 1 de taille n).
   b) En déduire $\nabla_\alpha J$, le gradient de J par rapport à α.
   c) Donner l'expression de α optimal (solution du problème de minimisation) en fonction de $\mathbf{x}$, $\mathbf{y}, \mathbb{1}$ et W.
   d) Donner le code python permettant de résoudre ce problème, les vecteurs $\mathbf{x} = (x_1, \dots, x_n)^\top$, $\mathbf{y} = (y_1, \dots, y_n)^\top$ et $\boldsymbol{w} = (w_1, \dots, w_n)^\top$ étant donnés.
