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

1. The ridge regression problem is

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

   (a) Show that this problem is equivalent to the minimization of

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

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

   (c) Characterize the solution of this modified criterion.

   (d) Propose a procedure to solve (1) by using

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

   Show that similar result holds for the Lasso task in terms of the Lasso.

2. Le normalized weighted Lasso

   (a) Consider the weighted Lasso optimization problem (on centered data with no intercept) :

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

   with given positive weight $w_j > 0$ $(j = 1, \ldots, p)$. Show that this problem is equivalent to a standard Lasso task

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

   (b) A typical way of using the Lasso is to apply it on normalized data as follows :

   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 $i = 1, \ldots, n$ et $j = 1, \ldots, p$.

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

   Show that this problem is equivalent to a weighted Lasso with a constant term :

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

   Explain how to retreive $\boldsymbol{\beta}$ et β_0 from $\boldsymbol{\beta}^r$.

Exercise 2 — Questions courtes — 5 points

1. What is the difference between test error and generalization ?
2. What is the relationship between the Lasso and quadratic programming ?
3. Under what conditions do you think it appropriate to use Sklearn to solve a two-class discrimination problem ?
4. What are the different components of an autoML type method ?
5. How to measure the proximity between example to reduce dimensionality ?

Exercise 3 — Calculs non reinaux mais matriciels — 7 points

Let W be a diaganoal matrix of size $n$and with general term 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. Explicit f_i the general term of the vector $\mathbf{f} = W\mathbf{e}$ as a function of w_i and e_i, where $\mathbf{e}$ is a $\mathbb{R}^n$ vector of general term e_i.
2. Explicit $\mathbf{e}^\top W\mathbf{e}$ as a function of w_i and e_i.
3. For given vectors $\mathbf{x} = (x_1, \dots, x_n)^\top$, $\mathbf{y} = (y_1, \dots, y_n)^\top$ et $\boldsymbol{w} = (w_1, \dots, w_n)^\top$, compute the solution of the following problem :

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

4. With matrix :
   a) rewrite $J(\alpha)$ avec $\alpha = (a, b)^\top$ with matrices, vectors $\mathbf{x}, \mathbf{y}, \mathbb{1}$ and matrix W (where $\mathbb{1} = (1, \dots, 1)^\top$ is a vector of 1 of size 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.
