Université de Sfax
Ecole Nationale d'Electronique
et des Télécommunications de Sfax
Département de Mathématiques
et Informatique Décisionnelle

A.U.: 2022-2023
Section: 2 IDSD
Date: Octobre 2022

Feuille d'exercices $n°1$
Machine Learning

Nous considérons $Y = (Y_1, Y_2, \ldots, Y_n)^t$ et $X = (x_{ij})_{1\leq i\leq n; 1\leq j\leq d}$ non aléatoire vérifiant

$$Y = X\beta^* + \varepsilon,$$

avec $\varepsilon \sim N(0, \sigma^2 I_n)$ et $\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_d^*) \in \mathbb{R}^d$.

Exercice 1. (Moindres carrés et ridge et Lasso en dimension 1)
Nous supposons ici que $d = 1$, le modèle considéré peut-être réecrit sous la forme

$$Y_i = \beta^* x_i + \varepsilon_i,\ i = 1, 2, \ldots, n,$$

avec $\beta \in \mathbb{R}$. L'objectif de cet exercice est de comparer dans ce cadre l'estimateur des moindres carrés

$$\hat{\beta}^{MC} \in \arg\min_{\beta\in\mathbb{R}} \sum_{i=1}^{n} (Y_i - \beta x_i)^2,$$

l'estimateur ridge $\hat{\beta}_\lambda^R$ et l'estimateur Lasso $\hat{\beta}_\lambda^L$.

1. Donner l'écriture de l'estimateur $\hat{\beta}^{MC}$ en fonction de $\{(Y_i, x_i),\ i = 1, 2, \ldots, n\}$.
2. Calculer le biais et la variance de l'estimateur $\hat{\beta}^{MC}$.
3. Écrire le problème de minimisation que doit vérifier l'estimateur ridge dans ce cadre-là et donner son écriture.
4. Calculer son biais, sa variance et son risque quadratique.
5. Nous considérons maintenant l'estimateur Lasso, c'est-à-dire la solution du critère de minimisation
$$\hat{\beta}_\lambda^L \in \arg\min_{\beta\in\mathbb{R}} \sum_{i=1}^{n} (Y_i - \beta x_i)^2 + \lambda \mid \beta \mid .$$
Calculer la solution du problème de minimisation.

Exercice 2. (Propriétés de l'estimateur Ridge)
Nous considérons, pour $\lambda > 0$, l'estimateur Ridge

$$\hat{\beta}_\lambda^R = (X^t X + \lambda I)^{-1} X^t Y.$$

L'objectif de cet exercice est de prouver les propriétés de l'estimateurs Ridge.

1. Montrer que le problème de minimisation
$$\min_{\beta\in\mathbb{R}^d} \left\{ \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \lambda \parallel \beta \parallel^2 \right\} \tag{1}$$
admet $\hat{\beta}_\lambda^R$ comme unique solution.

2. Montrer que toute solution du problème d'optimisation sous contrainte suivant

$$\min_{\beta\in\mathbb{R}^d,\ \|\beta\|\leq M_\lambda} \left\{ \sum_{i=1}^{n} (Y_i - X_i\beta)^2 \right\}, \tag{2}$$

avec $M_\lambda = \| (X^tX + \lambda I)^{-1} X^tY \|$ est aussi solution du problème (1).

3. En déduire que $\hat{\beta}_\lambda^R$ est aussi l'unique solution du problème (2).

4. Exprimer la norme au carré du biais de $\hat{\beta}_\lambda^R$ en fonction des valeurs propres $\lambda_1, \lambda_2, \ldots, \lambda_d$ (comptées avec multiplicité) de X^tX.

$$B_\lambda^R := \| \mathbb{E}[\hat{\beta}_\lambda^R] - \beta^* \|^2 .$$

5. Exprimer le variance

$$\mathbb{V}[\hat{\beta}_\lambda^R] = \mathbb{E}[\| \hat{\beta}_\lambda^R - \mathbb{E}[\hat{\beta}_\lambda^R] \|^2]$$

de $\hat{\beta}_\lambda^R$ en fonction de la variance du bruit σ^2 et des valeurs propres $\lambda_1, \lambda_2, \ldots, \lambda_d$.

Exercice 3. (Propriétés de l'estimateur Lasso)

L'objectif de cet exercice est de montrer que les problèmes d'optimisation

$$\min_{\beta\in\mathbb{R}^d} \left\{ \sum_{i=1}^{n} (Y_i - X_i\beta)^2 + \lambda \| \beta \|_1 \right\} \tag{3}$$

et

$$\min_{\beta\in\mathbb{R}^d,\ \|\beta\|_1\leq M_\lambda} \left\{ \sum_{i=1}^{n} (Y_i - X_i\beta)^2 \right\} \tag{4}$$

sont équivalents lorsque M_λ est bien choisi dans le sens où l'ensemble des solutions des deux problèmes est identique (on rappelle que pour le LASSO on n'a pas unicité de la solution).

1. Montrer que, pour toutes solutions $\hat{\beta}_\lambda^{L,1}$ et $\hat{\beta}_\lambda^{L,2}$ du problème pénalisé (3), on a

$$\| \hat{\beta}_\lambda^{L,1} \|_1 = \| \hat{\beta}_\lambda^{L,2} \|_1$$

Nous noterons par la suite cette valeur commune N_λ.

2. Montrer que toute solution de (3) est aussi solution de (4) lorsque $M_\lambda = N_\lambda$.

3. Montrer que toute solution de (4) est aussi solution de (3) lorsque $M_\lambda = N_\lambda$.

Exercice 4. (Elastic net)

Nous considérons le problème de minimisation suivant

$$\hat{\beta}_{\lambda,\mu}^{(EN)} \in \arg\min_{\beta\in\mathbb{R}^d} \{\mathcal{L}_{EN}(\beta)\},$$

avec

$$\mathcal{L}_{EN}(\beta) = \sum_{i=1}^{n} (Y_i - x_i^t\beta)^2 + \lambda \| \beta \|^2 + \mu \| \beta \|_1 .$$

1. Que se passe-t'il dans le cas $\lambda = 0$? Dans le cas $\mu = 0$?

2. Supposons que la condition ORT est vérifiée. Calculer la valeur minimale de $\mathcal{L}_{EN}$.

3. Pour tout $j = 1, 2, \ldots, d$, calculer la dérivée partielle de $\mathcal{L}_{EN}(\beta)$ par rapport à $\beta_j \neq 0$.

4. En déduire un algorithme de descente de gradient coordonnées par coordonnées pour approcher $\hat{\beta}_{\lambda,\mu}^{(EN)}$ lorsque, pour tout $j = 1, 2, \ldots, d$, $\sum_{i=1}^{n} x_{ij}^2 = \frac{1}{n}$.