FTML Exercices 2

Pour le 9 mars 2023

TABLE DES MATIÈRES

1 ORDINARY LEAST SQUARES

1.0.1 *Enoncé*

Les question 1 et 2 peuvent être traitées indépendamment.
Soit n et $d \in \mathbb{N}^*$.

1) Soit $X \in \mathbb{R}^{n,d}$, et $y \in \mathbb{R}^n$. Calculer le gradient de

$$g = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ \theta \mapsto ||X\theta - y||^2 \end{cases}$$

C'est la fonction objectif du problème OLS.

2) On veut montrer que la fonction $g : \mathbb{R}^d \to \mathbb{R}$ définie plus haut est convexe. Il y a de nombreuses méthodes pour cela mais ici utiliser les étapes suivantes :

a) montrer que si $s : \mathbb{R}^d \to \mathbb{R}^n$ est linéaire et $f : \mathbb{R}^n \to \mathbb{R}$ est convexe, alors $f \circ s : \mathbb{R}^d \to \mathbb{R}$ est convexe.
b) montrer que toute norme sur $\mathbb{R}^n$ est convexe.
c) montrer que si $w : \mathbb{R} \to \mathbb{R}$ est convexe croissante et $a : \mathbb{R}^n \to \mathbb{R}$ est convexe, alors $f = w \circ a : \mathbb{R}^n \to \mathbb{R}$ est convexe.
d) montrer que si $u : \mathbb{R}^n \to \mathbb{R}$ est convexe, alors si $\beta \in \mathbb{R}^n$, l'application $f : x \mapsto u(x + \beta)$ est convexe.
e) conclure.

1.0.2 *Solution*

1) On connaît déjà le gradient de l'application $f : x \mapsto ||x||^2$, qui vaut $2x$.
Si on considère l'application r :

$$r = \begin{cases} \mathbb{R}^d \to \mathbb{R}^n \\ \theta \mapsto X\theta - y \end{cases}$$

alors $g = f \circ r$. Comme tout est différentiable, on en déduit qu'en notant L les jacobiennes :

$$L_\theta g = L_{X\theta - y} f L_\theta r \tag{1}$$

ou bien on considérant le gradient (qui est la transposée de la jacobienne quand l'application est à valeurs dans $\mathbb{R}$) :

$$\nabla_\theta g = (L_\theta r)^T \nabla_{X\theta - y} f \tag{2}$$

Or $L_\theta r = X$. Donc

$$\nabla_\theta g = 2X^T(X\theta - y) \tag{3}$$

a) Soit $\alpha \in [0, 1]$, et $x, y \in \mathbb{R}^d$.

$$\begin{aligned}(f \circ s)(\alpha x + (1-\alpha)y) &= f(s(\alpha x + (1-\alpha)y)) \\ &= f(\alpha s(x) + (1-\alpha)s(y)) \\ &\leqslant \alpha f(s(x)) + (1-\alpha)f(s(y))\end{aligned} \tag{4}$$

b) Soit $\alpha \in [0, 1]$, et $x, y \in \mathbb{R}^n$.

$$\begin{aligned}||\alpha x + (1-\alpha)y|| &\leqslant ||\alpha x|| + ||(1-\alpha)y|| \\ &= \alpha||x|| + (1-\alpha)||y||\end{aligned} \tag{5}$$

c) Soit $\alpha \in [0, 1]$, et $x, y \in \mathbb{R}$.
Since a is convex,

$$a(\alpha x + (1-\alpha)y) \leqslant \alpha a(x) + (1-\alpha)a(y) \tag{6}$$

Since w is increasing,

$$w(a(\alpha x + (1-\alpha)y)) \leqslant w(\alpha a(x) + (1-\alpha)a(y)) \tag{7}$$

Since w is convex,

$$w(\alpha a(x) + (1-\alpha)a(y)) \leqslant \alpha(w(a(x)) + (1-\alpha)w(a(y)) \tag{8}$$

Finally,

$$(w \circ a)(\alpha x + (1-\alpha)y)) \leqslant \alpha(w \circ a)(x)) + (1-\alpha)(w \circ a)(y) \tag{9}$$

d) Soit $\alpha \in [0, 1]$, et $x, y \in \mathbb{R}^n$.

$$\begin{aligned}f(\alpha x + (1-\alpha)y) &= u(\alpha x + (1-\alpha)y + \beta) \\ &= u(\alpha x + (1-\alpha)y + (\alpha + 1 - \alpha)\beta) \\ &= u(\alpha x + (1-\alpha)y + \alpha\beta + (1-\alpha)\beta) \\ &= u(\alpha(x+\beta) + (1-\alpha)(y+\beta)) \\ &\leqslant \alpha u(x+\beta) + (1-\alpha)u(y+\beta)) \\ &= \alpha f(x) + (1-\alpha)f(y)\end{aligned} \tag{10}$$

e) On utilise :

- le point **c)** avec $w : t \mapsto t^2$ et a l'application norme sur $\mathbb{R}^n$ pour montrer que $u : x \mapsto ||x||^2$ est convexe.
- le point **d)** avec $\beta = -y$ pour montrer que $f : x \mapsto ||x - y||^2$ est convexe de $\mathbb{R}^n$ dans $\mathbb{R}$.
- le point **a)** appliqué à $g = f \circ s$ avec $s : \theta \mapsto X\theta$ linéaire de $\mathbb{R}^d$ dans $\mathbb{R}$.

2 EXPECTED VALUE AS A MINIMIZATION

2.0.1 *Enoncé*

Soit X une variable aléatoire réelle ayant un moment d'ordre 2. Montrer que son espérance $E(X)$ est la quantité minimisant la fonction de variable réelle $t \mapsto E\left((X-t)^2\right)$

2.0.2 *Solution*

All expected values are over X. We remark that

$$\begin{aligned} E\left[(X-t)^2\right] &= E\left[\left(X-E(X)+E(X)-t\right)^2\right] \\ &= E\left[\left(X-E(X)\right)^2 + 2\left(X-E(X)\right)\left(E(X)-t\right) + \left(E(X)-t\right)^2\right] \end{aligned}$$

By linearity, the expected value is separated in 3 terms.

— $E\left[\left(X-E(X)\right)^2\right]$

— $E\left[2\left(X-E(X)\right)\left(E(X)-t\right)\right]$

— $E\left[\left(E(X)-t\right)^2\right]$

We note that the first term $E\left[(X-E(X))^2\right]$ does not depend on t.
Also, $(E(X)-t)^2$, is a fixed scalar, and not a random variable, hence :

$$E\left[\left(E(X)-t\right)^2\right] = \left(E(X)-t\right)^2$$

We also have that

$$E\left[2\left(X-E(X)\right)\left(E(X)-t\right)\right] = 2\left(E(X)-t\right)E\left[\left(X-E(X)\right)\right] = 0$$

As a consequence, the value that minimizes $E\left[(X-t)^2\right]$ is $t = E(X)$.