

FTML Exercices 2 solutions

Pour le 21 mars 2024

TABLE DES MATIÈRES

1	Ordinary least squares	1
1.0.1	Enoncé	1
1.0.2	Solution	1
2	Expected value as a minimization	3
2.0.1	Enoncé	3
2.0.2	Solution	3

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

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

C'est la fonction objectif du problème OLS.

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

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

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

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

$$L_{\theta}g = L_{X\theta - y}fL_{\theta}r \quad (1)$$

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

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

Or $L_{\theta}r = X$. Donc

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

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

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

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

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

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

Since a is convex,

$$a(\alpha x + (1 - \alpha)y) \leq \alpha a(x) + (1 - \alpha)a(y) \quad (6)$$

Since w is increasing,

$$w(a(\alpha x + (1 - \alpha)y)) \leq w(\alpha a(x) + (1 - \alpha)a(y)) \quad (7)$$

Since w is convex,

$$w(\alpha a(x) + (1 - \alpha)a(y)) \leq \alpha(w(a(x)) + (1 - \alpha)w(a(y))) \quad (8)$$

Finally,

$$(w \circ a)(\alpha x + (1 - \alpha)y) \leq \alpha(w \circ a)(x) + (1 - \alpha)(w \circ a)(y) \quad (9)$$

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

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

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 $t \mapsto E((X-t)^2)$

2.0.2 Solution

All expected values are over X . We remark that

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

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

$$\begin{aligned} &— E[(X-E(X))^2] \\ &— E[2(X-E(X))(E(X)-t)] \\ &— E[(E(X)-t)^2] \end{aligned}$$

We note that the first term $E[(X-E(X))^2]$ does not depend on t .

Also, $(E(X)-t)^2$, is a fixed scalar, and not a random variable, hence :

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

We also have that

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

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