FTML Exercices 4 solutions

TABLE DES MATIÈRES

1 CONVEXITY

1.1 C2

1.1.1 *Enoncé*

$x \mapsto \theta^T x$ is convex on $\mathbb{R}^d$ with $\theta \in \mathbb{R}^d$ (linear form)

1.1.2 *Solution*

$$\theta^T(\alpha x + (1-\alpha)y) = \alpha\theta^T x + (1-\alpha)\theta^T y \tag{1}$$

Remark. *Actually, all linear forms are convex.*

1.2 C3

1.2.1 *Enoncé*

if Q is a symmetric definite positive matrix (matrice définie positive) with smallest eigenvalue $\lambda_{min} > 0$, then $x \mapsto x^T Q x$ is $2\lambda_{min}$- strongly convex.

1.2.2 *Solution*

Let $\mu = 2\lambda_{min}$.
We want to show that

$$(\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) \leqslant \alpha x^T Q x + (1-\alpha)y^T Q y - \frac{\mu}{2}\alpha(1-\alpha)||x-y||^2 \tag{2}$$

which means

$$(\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) - \alpha x^T Q x - (1-\alpha)y^T Q y \leqslant -\frac{\mu}{2}\alpha(1-\alpha)||x-y||^2 \tag{3}$$

We compute the left-hand side :

$$
\begin{aligned}
&(\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) - \alpha x^T Qx + (1-\alpha)y^T Qy \\
&= \alpha^2 x^T Qx + (1-\alpha)^2 y^T Qy + \alpha(1-\alpha)(x^T Qy + y^T Qx) - \alpha x^T Qx - (1-\alpha)y^T Qy \\
&= \alpha(\alpha-1)x^T Qx + (1-\alpha)((1-\alpha)-1)y^T Qy + \alpha(1-\alpha)(x^T Qy + y^T Qx) \\
&= \alpha(\alpha-1)x^T Qx + \alpha(\alpha-1)y^T Qy + \alpha(1-\alpha)(x^T Qy + y^T Qx) \\
&= \alpha(1-\alpha)\Big(-x^T Qx - y^T Qy + x^T Qy + y^T Qx\Big) \\
&= -\alpha(1-\alpha)\Big((x-y)^T Q(x-y)\Big) \\
&\leqslant \lambda_{min}\alpha(1-\alpha)\|x-y\|^2
\end{aligned}
\tag{4}
$$

which is exactly what we wanted.

2 LOGISTIC REGRESSION

1]

$$
\begin{aligned}
\forall z \in \mathbb{R}, \sigma'(z) &= \Big(-\frac{1}{(1+e^{-z})^2}\Big)\big(-e^{-z}\big) \\
&= \frac{1}{1+e^{-z}}\frac{e^{-z}}{1+e^{-z}} \\
&= \frac{1}{1+e^{-z}}\frac{e^{-z}e^{z}}{(1+e^{-z})e^{z}} \\
&= \frac{1}{1+e^{-z}}\frac{1}{1+e^{z}} \\
&= \sigma(z)\sigma(-z)
\end{aligned}
\tag{5}
$$

2] We compute the second order derivative.

$$
\begin{aligned}
\frac{\partial l}{\partial \hat{y}}(\hat{y}, y) &= \frac{-ye^{-\hat{y}y}}{1+e^{-\hat{y}y}} \\
&= \frac{-ye^{-\hat{y}y}}{1+e^{-\hat{y}y}}\frac{e^{\hat{y}y}}{e^{\hat{y}y}} \\
&= \frac{-y}{e^{\hat{y}y}+1} \\
&= -y\sigma(-\hat{y}y)
\end{aligned}
\tag{6}
$$

Hence,

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \hat{y}^2}(\hat{y}, y) &= -y\sigma(-\hat{y}y)\sigma(\hat{y}y) \times -y \\
&= y^2\sigma(-\hat{y}y)\sigma(\hat{y}y) > 0
\end{aligned}
\tag{7}
$$

Hence, the second-order derivative is strictly positive, and $l(\hat{y}, y)$ is stricly convex in its first argument.

3] We introduce the following functions :

$$
g_i = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ \theta \mapsto l(x_i^T\theta, y_i) \end{cases}
$$

$$
u_i = \begin{cases} \mathbb{R} \to \mathbb{R} \\ \hat{y} \mapsto l(\hat{y}, y_i) \end{cases}
$$

$$v_i = \begin{cases} \mathbb{R}^d \to \mathbb{R} \\ \theta \mapsto x_i^T \theta \end{cases}$$

Then, $\forall i$

$$l(x_i^T\theta, y_i) = g_i(\theta) = (u_i \circ v_i)(\theta) \tag{8}$$

a] (convexity) It is sufficient to show that each $g_i : \theta \to l(x_i^T\theta, y_i)$ is convex, because the sum of convex functions is convex. By definition, (equation 8), g_i is a convex function u_i applied to a linear mapping v_i, which prooves that g_i is convex.

b] (gradient) By composition of the jacobian matrices,

$$L_\theta^{g_i} = L_{v_i(\theta)}^{u_i} L_\theta^{v_i} = u_i'(v_i(\theta)) L_\theta^{v_i} \tag{9}$$

Or equivalently :

$$\nabla_\theta g_i(\theta) = u_i'(v_i(\theta)) \nabla_\theta v_i(\theta) \tag{10}$$

We already know that $\nabla_\theta v_i(\theta) = x_i$.

In question 2, we have seen that $\forall y, \hat{y}$,

$$\frac{\partial l}{\partial \hat{y}}(\hat{y}, y) = -y\sigma(-\hat{y}y) \tag{11}$$

Hence,

$$\begin{aligned} u_i'(v_i(\theta)) &= -y_i\sigma(-v_i(\theta)y_i) \\ &= -y_i\sigma(-x_i^T\theta y_i) \end{aligned} \tag{12}$$

Finally,

$$\nabla_\theta g_i(\theta) = -y_i\sigma(-x_i^T\theta y_i)x_i \tag{13}$$

and

$$\begin{aligned} \nabla_\theta R_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g_i(\theta) \\ &= \frac{1}{n}\sum_{i=1}^{n} -y_i\sigma(-x_i^T\theta y_i)x_i \end{aligned} \tag{14}$$