Exercices 2 solutions

TABLE DES MATIÈRES

1 CONVEXITY

1.1 C1

1.1.1 *Enoncé*

Show that all norms are convex.

1.1.2 *Solution*

$\alpha \in [0, 1]$

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

1.2 C2

1.2.1 Enoncé

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

1.2.2 Solution

$$\theta^T(\alpha x + (1-\alpha)y) = \alpha\theta^T x + (1-\alpha)\theta^T y \tag{2}$$

1.3 C3

1.3.1 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.

1.3.2 Solution

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

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

which means

$$(\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) - \alpha x^T Q x - (1-\alpha)y^T Q y \leqslant -\frac{\mu}{2}\alpha(1-\alpha)||x-y||^2 \tag{4}$$

We compute the left-hand side :

$$\begin{aligned}
&(\alpha x + (1-\alpha)y)^T Q(\alpha x + (1-\alpha)y) - \alpha x^T Q x + (1-\alpha)y^T Q y \\
&= \alpha^2 x^T Q x + (1-\alpha)^2 y^T Q y + \alpha(1-\alpha)(x^T Q y + y^T Q x) - \alpha x^T Q x - (1-\alpha)y^T Q y \\
&= \alpha(\alpha-1)x^T Q x + (1-\alpha)((1-\alpha)-1)y^T Q y + \alpha(1-\alpha)(x^T Q y + y^T Q x) \\
&= \alpha(\alpha-1)x^T Q x + \alpha(\alpha-1)y^T Q y + \alpha(1-\alpha)(x^T Q y + y^T Q x) \\
&= \alpha(1-\alpha)\Big(-x^T Q x - y^T Q y + x^T Q y + y^T Q x\Big) \\
&= -\alpha(1-\alpha)\Big((x-y)^T Q(x-y)\Big) \\
&\leqslant \lambda_{min}\alpha(1-\alpha)||x-y||^2
\end{aligned} \tag{5}$$

1.4 C4

1.4.1 Enoncé

If f is increasing and convex and g is convex, then $f \circ g$ is convex.

1.4.2 Solution

Since g is convex,

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

Since f is increasing,

$$f(g(\alpha x + (1-\alpha)y)) \leqslant f(\alpha g(x) + (1-\alpha)g(y)) \tag{7}$$

Since f is convex,

$$f(\alpha g(x) + (1-\alpha)g(y)) \leqslant \alpha(f(g(x)) + (1-\alpha)f(g(y)) \tag{8}$$

Finally,

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

1.5 C5

1.5.1 *Enoncé*

Is f in convex and g is linear, then $f \circ g$ is convex.

1.5.2 *Solution*

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

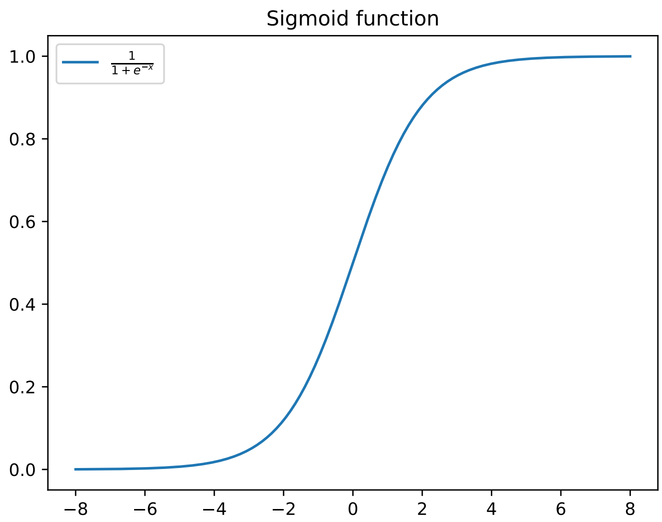
2 LOGISTIC REGRESSION

Definition 1. Cross entropy loss
$l : \mathbb{R}^2 \to \mathbb{R}$

$$l(\hat{y}, y) = y \log(1 + e^{-\hat{y}}) + (1-y)\log(1 + e^{\hat{y}}) \tag{11}$$

Definition 2. Sigmoid function
$\sigma : \mathbb{R} \to \mathbb{R}$.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{12}$$



2.1 L1

2.1.1 *Enoncé*

Show that σ is differentiable and that

$$\forall z, \sigma'(z) = \sigma(z)\sigma(-z) \tag{13}$$

2.1.2 *Solution*

$$\begin{aligned}
\sigma'(z) &= \Big(-\frac{1}{(1+e^{-z})^2}\Big)\big(-e^{-z}\big) \\
&= \frac{1}{1+e^{-z}}\frac{e^{-z}}{1+e^{-z}} \\
&= \frac{1}{1+e^{-z}}\frac{e^{-z}e^{z}}{(1+e^{-z})e^{z}} \\
&= \frac{1}{1+e^{-z}}\frac{1}{1+e^{z}} \\
&= \sigma(z)\sigma(-z)
\end{aligned} \tag{14}$$

2.2 L2

2.2.1 *Enoncé*

Show that $l(\hat{y}, y)$ is convex in its first argument, which means for fixed y, $\hat{y} \mapsto l(\hat{y}, y)$ is convex.

2.2.2 *Solution*

We compute the second order derivative.

$$\begin{aligned}
\frac{\partial l}{\partial \hat{y}}(\hat{y}, y) &= \frac{-ye^{-\hat{y}}}{1+e^{-\hat{y}}} + \frac{(1-y)e^{\hat{y}}}{(1+e^{\hat{y}})} \\
&= \frac{-ye^{-\hat{y}}}{1+e^{-\hat{y}}} + \frac{(1-y)e^{\hat{y}}e^{-\hat{y}}}{(1+e^{\hat{y}})e^{-\hat{y}}} \\
&= \frac{-ye^{-\hat{y}}}{1+e^{-\hat{y}}} + \frac{1-y}{1+e^{-\hat{y}}} \\
&= \frac{1}{1+e^{-\hat{y}}} - y\frac{1+e^{-\hat{y}}}{1+e^{-\hat{y}}} \\
&= \sigma(\hat{y}) - y
\end{aligned} \tag{15}$$

$$\begin{aligned}
\frac{\partial^2 l}{\partial \hat{y}^2}(\hat{y}, y) &= \sigma'(\hat{y}) \\
&= \sigma(\hat{y})\sigma(-\hat{y})
\end{aligned} \tag{16}$$

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