FTML Exercices 4 solutions

Pour le 7 avril 2023

TABLE DES MATIÈRES

1 LOGISTIC REGRESSION

1]

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

2] We compute the second order derivative.

$$\begin{aligned}
\frac{\partial l}{\partial \hat{y}}(\hat{y}, y) &= \frac{-ye^{-\hat{y}y}}{1+e^{-\hat{y}y}} \\
&= \frac{-ye^{-\hat{y}y}}{1+e^{-\hat{y}y}}\frac{e^{\hat{y}y}}{e^{\hat{y}y}} \\
&= \frac{-y}{e^{\hat{y}y}+1} \\
&= -y\sigma(-\hat{y}y)
\end{aligned} \tag{2}$$

Hence,

$$\begin{aligned}
\frac{\partial^2 l}{\partial \hat{y}^2}(\hat{y}, y) &= -y\sigma(-\hat{y}y)\sigma(\hat{y}y) \times -y \\
&= y^2\sigma(-\hat{y}y)\sigma(\hat{y}y) > 0
\end{aligned} \tag{3}$$

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 :

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

$$u = \left\{ \begin{array}{l} \mathbb{R} \to \mathbb{R} \\ \hat{y} \mapsto l(\hat{y}, y_i) \end{array} \right.$$

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

Then, $\forall i$

$$l(x_i^T \theta, y_i) = g_i(\theta) = (u \circ v_i)(\theta) \tag{4}$$

Hence, by composition of the jacobian matrices,

$$L_\theta^{g_i} = L_{v_i(\theta)}^{u} L_\theta^{v_i} = u'(v_i(\theta)) L_\theta^{v_i} \tag{5}$$

Or equivalently :

$$\nabla_\theta g_i(\theta) = u'(v_i(\theta)) \nabla_\theta (v_i(\theta)) \tag{6}$$

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

We have seen that $\forall y, \hat{y}$,

$$\frac{\partial l}{\partial \hat{y}}(\hat{y}, y) = -y\sigma(-\hat{y}y) \tag{7}$$

Hence,

$$\begin{aligned} u'(v_i(\theta)) &= -y_i \sigma(-v_i(\theta) y_i) \\ &= -y_i \sigma(-x_i^T \theta y_i) \end{aligned} \tag{8}$$

Finally,

$$\nabla_\theta g_i(\theta) = -y_i \sigma(-x_i^T \theta y_i) x_i \tag{9}$$

and

$$\begin{aligned} \nabla_\theta R_n(\theta) &= \frac{1}{n} \sum_{i=1}^{n} \nabla_\theta g_i(\theta) \\ &= \frac{1}{n} \sum_{i=1}^{n} -y_i \sigma(-x_i^T \theta y_i) x_i \end{aligned} \tag{10}$$