Régression logistique :

Jeu de données : $(x_1, y_1), \text{------}, (x_n, y_n)$

où $x \in \mathbb{R}^d$ et $y \in \{0, 1\}$.

Modèle probabiliste :

$$\mathbb{P}\left[Y = 1 \mid X = x\right] = g\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p\right)$$

où $g(z) = \dfrac{1}{1 + e^{-z}} = \dfrac{e^{z}}{1 + e^{z}}$ fonction logistique

On peut voir $x = (1, x_1, \dots, x_p)^T$.

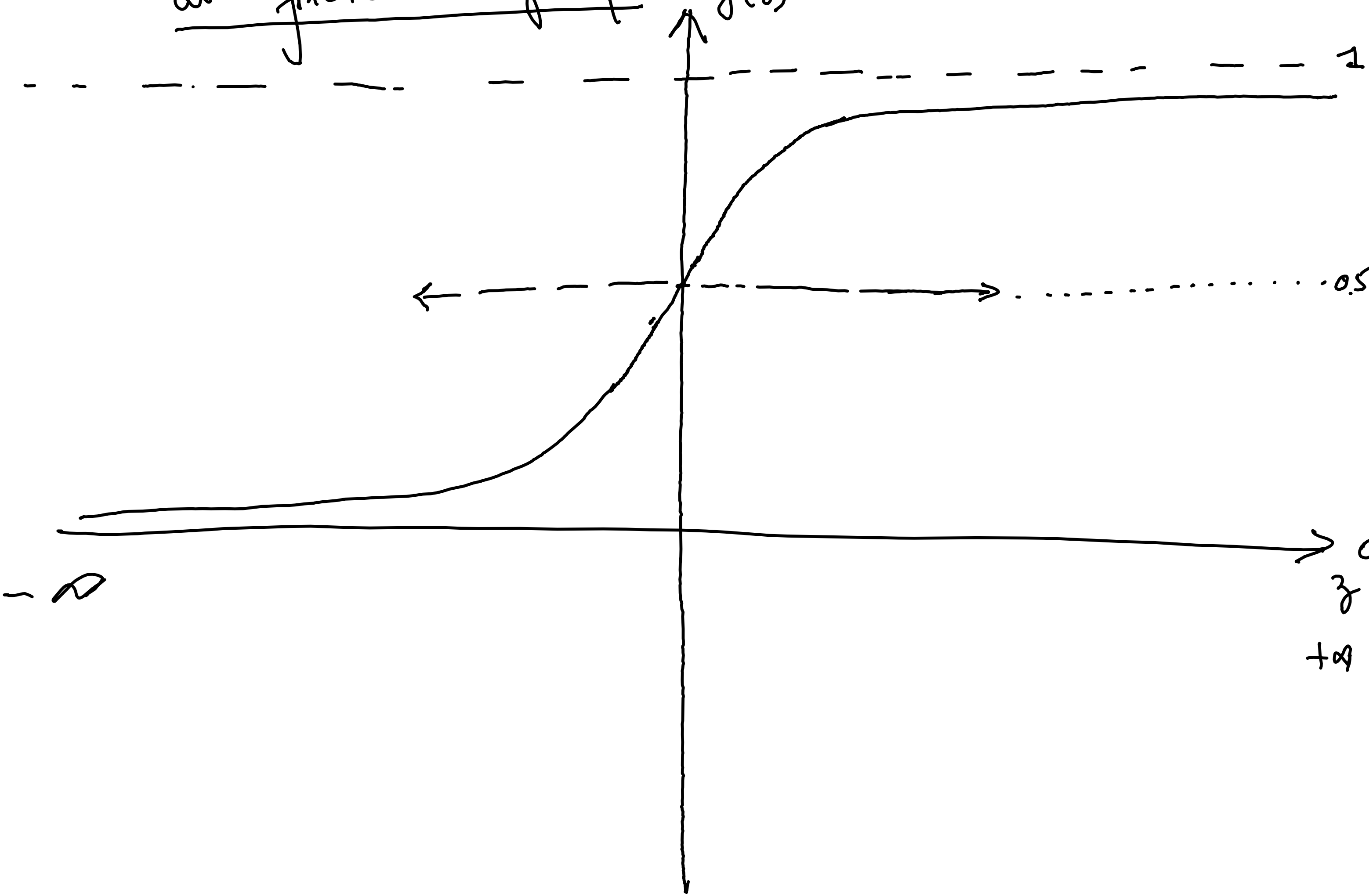
et noter $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$

$= \beta^T x$ où $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \beta_p \end{bmatrix}$.

On peut écrire donc :

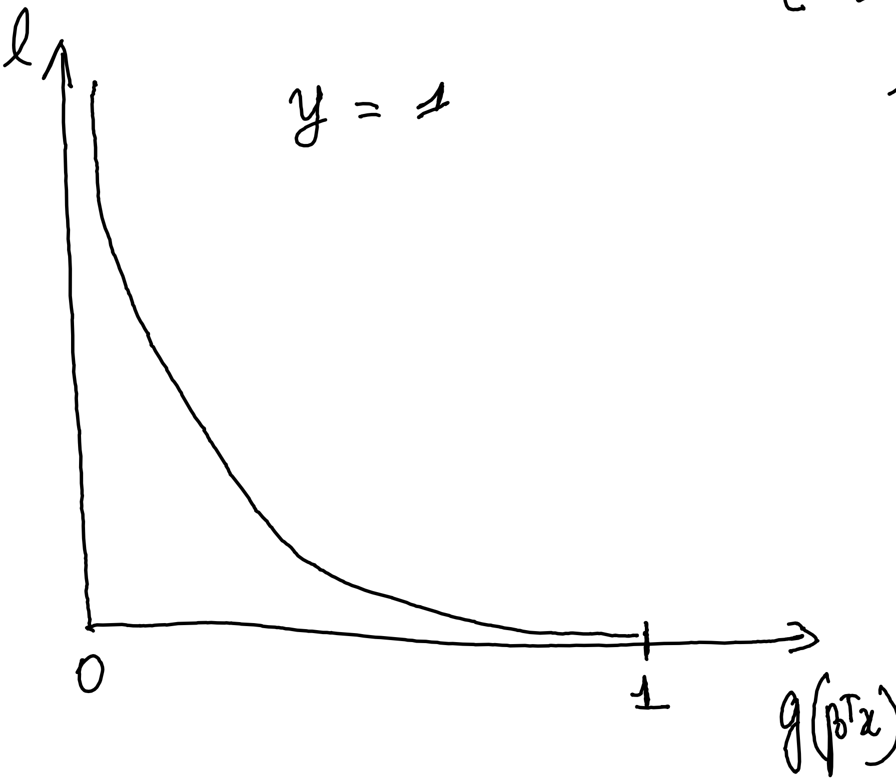
$$\mathbb{P}(Y = y \mid X = x) = g(\beta^T x)^y \times (1 - g(\beta^T x))^{1-y}$$

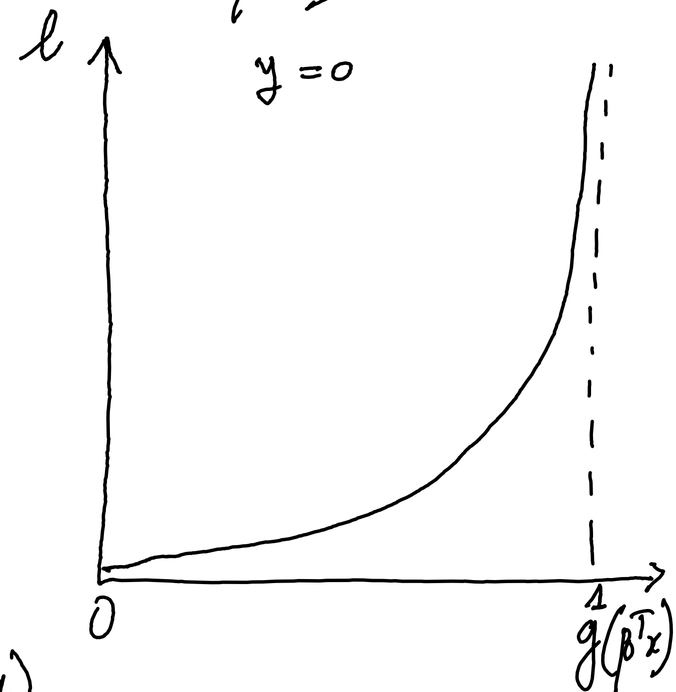
La fonction logistique



La fonction de perte :

$$
\ell\left(y, g(\beta^T x)\right) = \begin{cases} -\ln\left(g(\beta^T x)\right) & \text{si } y = 1 \\ -\ln\left(1 - g(\beta^T x)\right) & \text{si } y = 0 \end{cases}
$$





donc, si on note $x_i = (1, x_{i1}, \ldots, x_{ip})$

la perte du ième individu du jeu de données est donnée par

$$l(y_i, g(\beta^T x_i)) = -y_i \ln(g(\beta^T x_i)) - (1-y_i) \ln(1 - g(\beta^T x_i))$$

pour un jeu de données de taille n ; la la perte est donnée par $L(\beta) = \frac{1}{n} \sum_{i=1}^{n} l(y_i, g(\beta^T x_i))$.

On souhaite calculer

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} L(\beta)$$

On va calculer

$$\nabla_\beta L(\beta) = \underbrace{\begin{bmatrix} \frac{\partial}{\partial \beta_0} L(\beta) \\ \vdots \\ \vdots \\ \frac{\partial}{\partial \beta_p} L(\beta) \end{bmatrix}}_{\text{le gradient}}$$

$$\frac{\partial}{\partial \beta_j} L(\beta) = \frac{\partial}{\partial \beta_j} \left(\frac{1}{n} \sum_{i=1}^{n} \ell(y_i, g(\beta^T x_i)) \right)$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \left[y_i \ln(g(\beta^T x_i)) + (1 - y_i) \ln(1 - g(\beta^T x_i)) \right]$$

$$\frac{\partial}{\partial \beta_j}\left(y_i \ln\left(g\left(\beta^T x_i\right)\right) + (1-y_i) \ln\left(1-g\left(\beta^T x_i\right)\right)\right)$$

$$= \left(y_i \frac{1}{g(\beta^T x_i)} - \frac{(1-y_i)}{1-g\left(\beta^T x_i\right)}\right) \frac{\partial}{\partial \beta_j} g\left(\beta^T x_i\right)$$

on sait que $g(z) = \dfrac{1}{1+e^{-z}}$ et $1-g(z) = \dfrac{e^{-z}}{1+e^{-z}}$

on peut donc écrire :

$$\frac{\partial}{\partial \beta_j}\left(g\left(\beta^T x_i\right)\right) = \frac{e^{-\beta^T x_i}}{\left(1+e^{-\beta^T x_i}\right)^2} \frac{\partial}{\partial \beta_j}\left(\beta^T x_i\right)$$

$$= \frac{e^{-\beta^T x_i}}{1+e^{-\beta^T x_i}} \times \frac{1}{1+e^{-\beta^T x_i}} \, x_{ij}$$

donc : $$\frac{\partial}{\partial \beta_j}\left(g(\beta^T x_i)\right) = g(\beta^T x_i)\left(1 - g(\beta^T x_i)\right) x_{ij}$$

On obtient donc :

$$\frac{\partial}{\partial \beta_j}\left(y_i \ln\left(g(\beta^T x_i)\right) + (1-y_i)\ln\left(1-g(\beta^T x_i)\right)\right)$$

$$= \left(\frac{y_i}{g(\beta^T x_i)} - \frac{(1-y_i)}{1-g(\beta^T x_i)}\right)\frac{\partial}{\partial \beta_j}\left(g(\beta^T x_i)\right)$$

$$= \frac{y_i\left(1-g(\beta^T x_i)\right) - (1-y_i)\, g(\beta^T x_i)}{\cancel{g(\beta^T x_i)\left(1-g(\beta^T x_i)\right)}} \times \cancel{g(\beta^T x_i)\left(1-g(\beta^T x_i)\right)}\, x_{ij}$$

$$= \left(y_i - g(\beta^T x_i)\right) x_{ij} = -\left(g(\beta^T x_i) - y_i\right) x_{ij}$$

On déduit que :

$$\frac{\partial L(\beta)}{\partial \beta_j} = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \left(y_i \ln\left(g(\beta^T x_i)\right) + (1-y_i) \ln\left(1-g(\beta^T x_i)\right) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(g(\beta^T x_i) - y_i \right) x_{ij} .$$

Descente du gradient :

$$\beta_j \leftarrow \beta_j - \eta \frac{1}{n} \sum_{i=1}^{n} \left(g(\beta^T x_i) - y_i \right) x_{ij}$$