$z_i \in \{0, 1\}$ Bernoulli.

$\Leftrightarrow y_i \in \{-1, 1\}$ Rademacher.

Remarque 1 : $y_i = 2z_i - 1$ $\quad z_i = 0 \; y_i = -1$

$z_i = 1 \; y_i = 1$.

$$\mathbb{P}(y_i = 1 / X_i) = \frac{e^{X_i\beta^*}}{1 + e^{X_i\beta^*}}$$

$$= \frac{1}{1 + e^{-X_i\beta^*}}$$

$$\mathbb{P}(y_i = -1 / X_i) = \frac{1}{1 + e^{X_i\beta^*}}$$

$$\frac{1}{1 + \exp(-y_i X_i \beta^*)}$$

$\rightarrow$ vraissemblance pour n obs iid (y_i, X_i)

$$\prod_{i=1}^{n} \frac{1}{1 + \exp(-y_i X_i \beta)} \qquad \ell_n(\beta) = -\frac{1}{n} \log \mathcal{V}(\beta)$$

$$\ell_n(\beta) = -\frac{1}{n} \sum_{i=1}^{n} \log\left(\frac{1}{1 + \exp(-y_i X_i \beta)}\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + \exp(-y_i X_i \beta)\right)$$

But : $\underset{\beta}{\operatorname{argmin}} \; \ell_n(\beta) + \lambda \|\beta\|_1$

Sigmoid.

$$\sigma(u) = \frac{e^u}{1+e^u} = \frac{1}{1+e^{-u}}$$

si $u > 0$ $\sigma(u) = \frac{1}{1+e^{-u}}$

si $u < 0$ $\sigma(u) = \frac{e^u}{1+e^u}$

Calcul du gradient de

$$l_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \log(1+\exp(-y_i X_i \beta))$$

$$\nabla_\beta l_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \nabla_\beta\left[\log(1+\exp(-y_i X_i\beta))\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{\nabla_\beta(\exp(-y_i X_i \beta))}{1+\exp(-y_i X_i \beta)}$$

$$\sigma(u) = \frac{1}{1+e^{-u}} = \frac{e^u}{1+e^u}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{-y_i X_i^T \exp(-y_i X_i \beta))}{1+\exp(-y_i X_i\beta)}$$

$$= 1 - \boxed{\frac{1}{1+\exp(-y_i X_i\beta)}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} -y_i X_i^T(1 - \sigma(y_i X_i \beta))$$

$X_i \beta$ $\quad X_i \in \mathbb{R}^p$ $\quad \langle X_i, \beta\rangle = X_i^T \beta$

$$\longrightarrow \begin{pmatrix} X_1 \\ X_2 \\ \vdots \end{pmatrix} = X \text{ de taille } n \times p.$$

$$X^T = \begin{pmatrix} X_1^T & X_2^T & \dots & X_n^T \end{pmatrix}$$

$$X^T \cdot \begin{pmatrix} y_1 (\sigma(y_1 X_1 \beta) - 1) \\ \vdots \\ y_n (\sigma(y_n X_n \beta) - 1) \end{pmatrix} = \sum_{i=1}^{n} y_i X_i^T \underline{(\sigma(y_i X_i \beta) - 1)}$$

Calcul de la Hessienne

$$\nabla^2_\beta (l_n(\beta)) = \nabla_\beta (\nabla_\beta (l_n(\beta)))$$

$$\sigma(u) = \frac{e^u}{1+e^u} \qquad \sigma'(u) = \frac{e^u(1+e^{+u}) - e^u e^u}{(1+e^u)^2}$$

$$= \frac{e^u}{(1+e^u)^2} = \frac{e^u}{1+e^u} \frac{1}{1+e^u} = \sigma(u)(1-\sigma(u))$$

$$\nabla^2_\beta l_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} y_i X_i^T \nabla_\beta \cdot (\sigma(y_i X_i \beta) - 1)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{y_i^2}_{=1} X_i^T \sigma(y_i X_i \beta)\left[1 - \sigma(y_i X_i \beta)\right] X_i$$

$$= \frac{1}{n} X^T W X. \quad \text{avec } W = \text{diag}(\;)$$

$$\sigma(y_i X_i \beta)(1 - \sigma(y_i X_i \beta))$$

* Calcul de la constante de Lipschitz de $\nabla_\beta l_n(\beta)$

on remarque que $0 \leq \sigma(u)(1-\sigma(u)) \leq \frac{1}{4}$

$$\frac{1}{4n} \sum_{i=1}^{n} \|X_i\|^2$$

$$S_\lambda(v) = \text{sign}(v)\left[|v| - \lambda\right]_+$$

$$x^{(k+1)} = \text{prox}_{\lambda_k g}\left(x^{(k)} - \lambda_k \nabla f(x^{(k)})\right)$$

pas = step = $\frac{1}{L}$

$g = \lambda \|\cdot\|_1$

$$\underbrace{\text{prox}_{\text{step}\cdot\lambda\|\cdot\|_1}}_{\text{sign}(v)\left(|v| - \text{step}\cdot\lambda\right)_+}\left(x^{(k)} - \text{step}\,\nabla f(x^{(k)})\right)$$