
TD 2

Régression logistique

Exercice 1

On cherche à construire un classifieur qui prend en entrée $x \in [-1, 1]$ et qui le classifie en la classe 0 ou la classe 1. Ce classifieur va être paramétré par $\alpha, \beta \in \mathbb{R}$. La classe associée à l'entrée x sera 1 si $C_{\alpha, \beta}(x) \geq 1/2$ et 0 sinon, en définissant

$$C_{\alpha, \beta}(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}.$$

1. Tracer approximativement la courbe de la fonction $t \mapsto C_{0,1}(t) = e^t / (1 + e^t)$ ($t \in \mathbb{R}$).
2. Montrer que $e^t / (1 + e^t) \geq 1/2 \iff t \geq 0$.
3. Pour $\alpha = 0$ et $\beta = 1$, quels x sont classés en 0 et quels x sont classés en 1.
4. Même question pour $\alpha = 1$ et $\beta = -1$.
5. Pour un $x \neq 0$ et $\alpha \in \mathbb{R}$ fixés, lorsque β devient assez grand ($\beta \rightarrow +\infty$), comment est classé x .
6. Même question lorsque $\beta \rightarrow -\infty$.

Exercice 2

Maintenant, on cherche à construire un classifieur qui prend en entrée $x = (x_1, x_2) \in [-1, 1]^2$ et qui le classifie en la classe 0 ou la classe 1. Ce classifieur va être paramétré par $\alpha, \beta_1, \beta_2 \in \mathbb{R}$. La classe associée à l'entrée x sera 1 si $C_{\alpha, \beta_1, \beta_2}(x) \geq 1/2$ et 0 sinon, en définissant

$$C_{\alpha, \beta_1, \beta_2}(x) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2}}.$$

1. On prend $\alpha = 0, \beta_1 = 1, \beta_2 = 0$. Représenter graphiquement quels x sont classés 0 et quels x sont classés 1.
2. Même question avec $\alpha = 1, \beta_1 = 2, \beta_2 = -1$.
3. En général, pour α, β_1, β_2 quelconques tels que $(\beta_1, \beta_2) \neq (0, 0)$, déterminer quels x sont classés 0 et quels x sont classés 1. La réponse à cette question permet de dire que l'on étudie un *classifieur linéaire*.

Exercice 3

On propose maintenant un modèle probabiliste qui correspond à ce classifieur. On fixe une dimension $d \in \mathbb{N}$, $d \geq 2$. On considère $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}^d$. On considère un couple de variables aléatoires $(X, Y) \in [-1, 1]^d \times \{0, 1\}$ où X suit la loi uniforme sur $[-1, 1]^d$ et, pour tout $x \in [-1, 1]^d$,

$$\mathbb{P}(Y = 1 | X = x) = 1 - \mathbb{P}(Y = 0 | X = x) = C_\beta(x) = \frac{e^{\beta^\top x}}{1 + e^{\beta^\top x}}.$$

1. On considère un premier classifieur $f : [-1, 1]^d \rightarrow \mathbb{R}$ qui attribue la classe 1 à x si $C_\beta(x) \geq 1/2$ et la classe 0 sinon. Prouver que le risque de classification $\mathbb{P}(f(X) \neq Y)$ s'écrit

$$\mathbb{P}(f(X) \neq Y) = \frac{1}{2^d} \int_{[-1, 1]^d} \min(C_\beta(x), 1 - C_\beta(x)) dx.$$

On pourra d'abord calculer $\mathbb{P}(f(X) \neq Y | X = x)$ pour tout $x \in [-1, 1]^d$ et ensuite utiliser la formule de l'espérance totale

$$\mathbb{P}(f(X) \neq Y) = \mathbb{E}(\mathbb{P}(f(X) \neq Y | X)).$$

1. Matériel de base créé par François Bachoc.

On pourra aussi utiliser (sans démonstration)

$$\mathbb{P}(f(X) \neq Y | X = x) = \mathbb{P}(f(x) \neq Y | X = x)$$

et, que si $0 \leq t \leq 1/2$ alors $t = \min(t, 1 - t)$.

2. *Question bonus plus difficile.* Prouver que lorsque $\|\beta\| \rightarrow +\infty$,

$$\mathbb{P}(f(X) \neq Y) \rightarrow 0.$$

3. On considère le risque de classification $\mathbb{P}(g(X) \neq Y)$ où $g : [-1, 1]^d \rightarrow \mathbb{R}$ est un autre classifieur qui attribue la classe 1 à x si $C_\gamma(x) \geq 1/2$ et la classe 0 sinon. Ici $\gamma \in [-1, 1]^d$ est un vecteur différent de β . Prouver que

$$\mathbb{P}(g(X) \neq Y) = \frac{1}{2^d} \int_{[-1, 1]^d} [\mathbf{1}_{g(x)=f(x)} \min(C_\beta(x), 1 - C_\beta(x)) + \mathbf{1}_{g(x) \neq f(x)} \max(C_\beta(x), 1 - C_\beta(x))] dx.$$

4. Prouver que

$$\mathbb{P}(f(X) \neq Y) \leq \mathbb{P}(g(X) \neq Y).$$

Interpréter ce résultat.

Exercice 4

On considère n couples $(X_1, Y_1), \dots, (X_n, Y_n)$ iid et de même loi que (X, Y) dans l'exercice précédent. On fixe $x_1, \dots, x_n \in [-1, 1]^d$.

1. On fixe $y_1, \dots, y_n \in \{0, 1\}$. Calculer

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n | X_1 = x_1, \dots, X_n = x_n).$$

en fonction de β . On admettra que

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n | X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(Y_1 = y_1 | X_1 = x_1) \times \dots \times \mathbb{P}(Y_n = y_n | X_n = x_n).$$

Indication : on pourra montrer que

$$\mathbb{P}(Y_i = y_i | X_i = x_i) = C_\beta(x_i)^{y_i} (1 - C_\beta(x_i))^{1-y_i}.$$

2. On note $\mathcal{L}(\beta) = -\log(P(\beta))$ où $P(\beta)$ est la probabilité à calculer dans la question précédente. On note $M(\beta)$ sa matrice Hessienne calculée en β . La matrice M est donc de taille $d \times d$ et son élément i, j est égal à $\partial^2 \mathcal{L}(\beta) / \partial \beta_i \partial \beta_j$. Montrer que

$$M(\beta) = \sum_{i=1}^n \frac{e^{x_i^\top \beta}}{(1 + e^{x_i^\top \beta})^2} x_i x_i^\top.$$

3. On pose

$$c(\beta) = \left(\min_{i=1, \dots, n} \frac{e^{x_i^\top \beta}}{(1 + e^{x_i^\top \beta})^2} \right) \lambda_{\min}(X^\top X)$$

où $\lambda_{\min}(X^\top X)$ est la plus petite valeur propre de la matrice $X^\top X$. Montrer que pour tout vecteur v on a

$$v^\top M v \geq c(\beta) \|v\|^2.$$

4. *Question bonus plus difficile.* On suppose que

$$\lim_{\|\beta\| \rightarrow +\infty} \mathcal{L}(\beta) = +\infty.$$

Montrer que le problème d'optimisation

$$\min_{\beta \in \mathbb{R}^d} \mathcal{L}(\beta)$$

admet un unique minimiseur. *On dira alors que l'estimateur du maximum de vraisemblance est unique dans le modèle de régression logistique.*