

Introduction au Machine Learning

Régression logistique pour la classification

Emmanuel Dellandréa

emmanuel.dellandrea@ec-lyon.fr

Version du 10/01/2025



Classification

- Régression : la variable cible à prédire est continue (variable quantitative)
- Classification : la variable cible à prédire est discrète (variable qualitative)
 - ➔ objectif : prédire un nombre réduit de valeurs discrètes
 - Problèmes multi-classes :
 - Chien / chat / cheval / vache /
 - Problèmes binaires :
 - Tumeur maligne / tumeur sein
 - Bon client / mauvais client

➔ Focus ici sur la classification binaire avec la régression logistique

Classification binaire

- Une solution envisageable : considérer la fonction de prédiction linéaire f_θ vue dans le cours précédent sur la régression et une valeur de seuil (par exemple 0.5)
 - Si $f_\theta(x) \geq 0.5$ alors prédire $y=1$
 - et si $f_\theta(x) < 0.5$ alors prédire $y=0$
 - Problème :
 - $f_\theta(x) \geq$ peut prendre des valeurs très supérieures à 1 et très inférieures à 0 alors que y appartient à $\{0,1\}$
- ➔ Régression logistique : $0 \leq f_\theta(x) \leq 1$

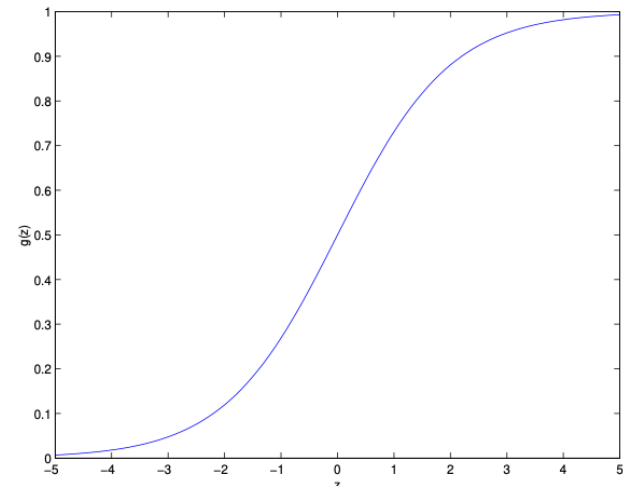
Régression logistique

→ Exprimer f_θ à l'aide de la fonction sigmoïde (fonction logistique) :

$$f_\theta(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

▪ Où g est la fonction sigmoïde :

$$g(z) = \frac{1}{1 + e^{-z}}$$



Régression logistique

- Prédire 1 si $f_{\theta}(x) \geq 0.5$ (soit $\theta^T x \geq 0$)
- Prédire 0 si $f_{\theta}(x) < 0.5$ (soit $\theta^T x < 0$)

- On considère donc :

$$P(y = 1/x; \theta) = f_{\theta}(x)$$

$$P(y = 0/x; \theta) = 1 - f_{\theta}(x)$$

- Soit : $p(y/x; \theta) = (f_{\theta}(x))^y (1 - f_{\theta}(x))^{1-y}$

Vraisemblance

- Evaluation de la réussite par la vraisemblance :

$$L(\theta) = p(Y/X; \theta)$$

→ La vraisemblance est la probabilité prédite du résultat observé étant donné les caractéristiques X et les paramètres du modèle θ

Vraisemblance

- Elle peut encore s'exprimer par (indépendance des exemples d'apprentissage) :

$$L(\theta) = \prod_{i=1}^N p(y^{(i)} / x^{(i)}; \theta)$$

- Soit :

$$L(\theta) = \prod_{i=1}^N (f_{\theta}(x^{(i)}))^{y^{(i)}} (1 - f_{\theta}(x^{(i)}))^{1-y^{(i)}}$$

Log vraisemblance

- Utilisation du log pour simplifier :

$$l(\theta) = \log L(\theta)$$

- D'où :

$$l(\theta) = \sum_{i=1}^N y^{(i)} \log(f_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - f_{\theta}(x^{(i)}))$$

Calcul du gradient

- Rappel : $f_{\theta}(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$

- Avec : $g(z) = \frac{1}{1+e^{-z}}$

- Dérivée de g :
$$\begin{aligned}g'(z) &= \frac{d}{dz} \frac{1}{1+e^{-z}} \\&= \frac{1}{(1+e^{-z})^2} (e^{-z}) \\&= \frac{1}{(1+e^{-z})} \cdot \left(1 - \frac{1}{(1+e^{-z})}\right) \\&= g(z)(1 - g(z)).\end{aligned}$$

Calcul du gradient

- La vraisemblance étant :

$$l(\theta) = \sum_{i=1}^N y^{(i)} \log(f_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - f_{\theta}(x^{(i)}))$$

- Pour un exemple (x, y) on a :

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \ell(\theta) &= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) \frac{\partial}{\partial \theta_j} g(\theta^T x) \\ &= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) g(\theta^T x)(1 - g(\theta^T x)) \frac{\partial}{\partial \theta_j} \theta^T x \\ &= (y(1 - g(\theta^T x)) - (1 - y)g(\theta^T x)) x_j \\ &= (y - g(\theta^T x)) x_j \end{aligned}$$

Calcul du gradient

- On obtient donc le gradient de la vraisemblance :

$$\rightarrow \frac{\partial}{\partial \theta_j} l(\theta) = \sum_{i=1}^N (y^{(i)} - f_{\theta}(x^{(i)})) x_j^{(i)}$$

Descente du gradient

- Fonction de coût à minimiser :

$$J(\theta) = -l(\theta)$$

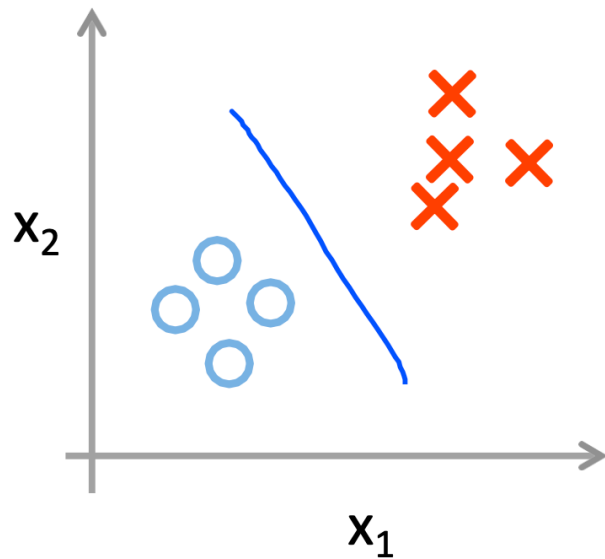
- D'où l'algorithme de la descente du gradient :
Répéter itérativement jusqu'à convergence (simultanément pour chaque j) :

$$\theta_j = \theta_j - \alpha \frac{1}{N} \sum_{i=1}^N (f_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

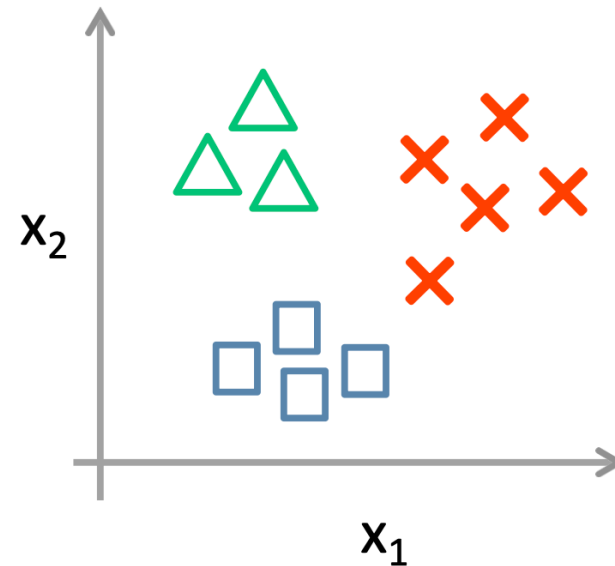
→ expression identique à celle de la régression linéaire mais avec f_{θ} différente (sigmoïde)

Classification à plusieurs classes

Classification binaire



Classification multiple



Classification à plusieurs classes

- Principe de **UN CONTRE TOUS**

→ Entraîner un classifieur de type régression logistique $f_{\theta}^{(k)}(x)$ pour chaque classe k afin de prédire la probabilité $y=k$

→ Pour une nouvelle donnée x , choisir la classe k qui maximise $f_{\theta}^{(k)}(x)$



ÉCOLE
CENTRALE LYON

36, avenue Guy de Collongue 69130 Écully - France
+33 (0)4 72 18 60 00

www.ec-lyon.fr