Introduction au Machine Learning

Régression Logistique

Théo Lopès-Quintas

BPCE Payment Services,
Université Paris Dauphine

2023

Motivation

Complétude

Quoique la doctrine des logarithmes soit si solidement établie, que les vérités qu'elle renferme semblent aussi rigoureusement démontrées que celles de la Géométrie ; les Mathématiciens sont pourtant encore fort partagés sur la nature des logarithmes négatifs et imaginaires.

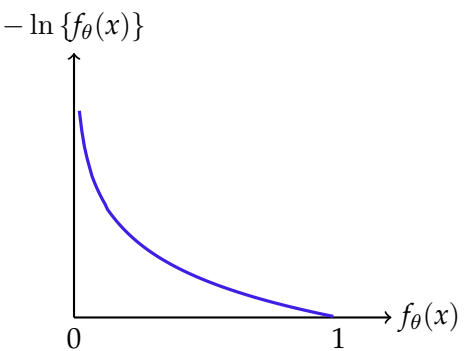
— Leonhard Euler (1749)

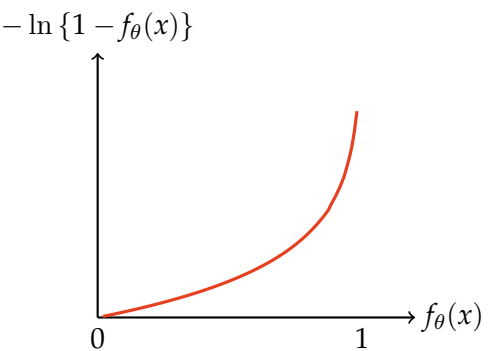
Problème
Modélisation

Observation positive

$$\mathcal{L}(\theta; x, y) = -\left[y \ln \{f_\theta(x)\} + (1-y) \ln \{1 - f_\theta(x)\} \right]$$

Observation négative



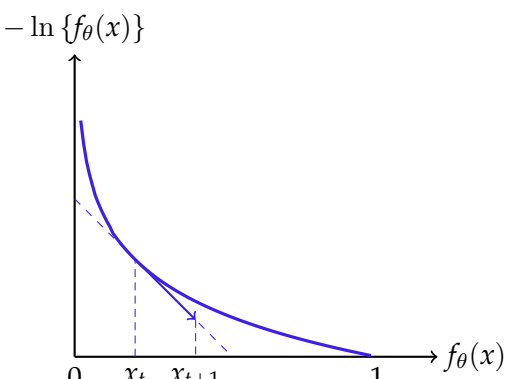


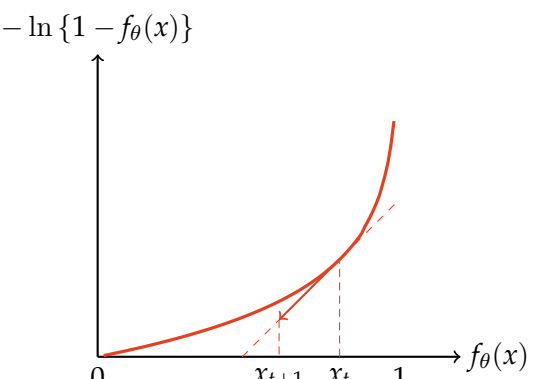
Problème

Descente de gradient

$$x^* = \arg\min_{x \in \mathbb{R}^d} f(x)$$

$$x_{t+1} = x_t - \eta \nabla f(x_t) \quad , \eta > 0$$





Problème

Solution

Exercice 1

On rappelle que $f_\theta(x) = \dfrac{1}{1+e^{-<\theta,x>}}$. *Montrer que :*

1. $f_\theta(x) = \dfrac{e^{<\theta,x>}}{1+e^{<\theta,x>}}$
2. $f_\theta(-x) = 1 - f_\theta(x)$
3. $\dfrac{\partial \ln}{\partial \theta_j}(f_\theta(x)) = x_j\,(1 - f_\theta(x))$
4. $\dfrac{\partial \ln}{\partial \theta_j}(1 - f_\theta(x)) = -x_j f_\theta(x)$
5. $\dfrac{\partial \mathcal{L}}{\partial \theta_j}\left(\theta; x^{(i)}, y_i\right) = x_j^{(i)}\left(f_\theta(x^{(i)}) - y_i\right)$
6. *Conclure que la descente de gradient pour le problème avec la fonction de coût est :*

$$\theta_j^{t+1} = \theta_j^t - \eta \sum_{i=1}^{n} x_j^{(i)}\left(f_\theta(x^{(i)}) - y_i\right)$$

Mesurer la performance d'une régression

Accuracy

		Prédit	
		Classe 0 (baisse)	Classe 1 (hausse)
Réel	Classe 0	TN	FP
	Classe 1	FN	TP

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

Exercice 2

On souhaite prédire une hausse exceptionnelle, et dans le dataset que l'on a à disposition, il y a 1% de classe 1 (hausse exceptionnelle). Construire un algorithme qui permet d'atteindre 99% d'accuracy.

Mesurer la performance d'une régression

Précision, Recall et F1-Score

		Prédit	
		Classe 0 (baisse)	Classe 1 (hausse)
Réel	Classe 0	TN	FP
	Classe 1	FN	TP

$$\text{Précision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = \frac{2}{\frac{1}{\text{Précision}} + \frac{1}{\text{Recall}}}$$

Exercice 3

Vous avez trop de mails, et vous demandez à votre data scientist de concevoir un algorithme qui va prioriser les mails en essayant de prédire les mails qui sont les plus importants. Vous lui donnez un dataset d'entraînement et un dataset de test. Dans le dataset de test, il y a 1000 mails dont 200 sont importants. Il vous présente un premier modèle qui pour un certain seuil (A) présente la matrice de confusion suivante :

		Prédit	
		Classe 0	*Classe 1*
Réel	*Classe 0*	700	100
	Classe 1	50	150

Pour un autre seuil (B), il présente cette matrice de confusion :

		Prédit	
		Classe 0	*Classe 1*
Réel	*Classe 0*	760	40
	Classe 1	80	120

1. *Calculer l'accuracy, la précision, le recall et le F1-score de chacun des seuils.*
2. *Conclure sur le seuil que vous souhaitez conserver.*