Tout savoir sur la régression logistique

Partie 2

Présenté par **Morgan Gautherot**

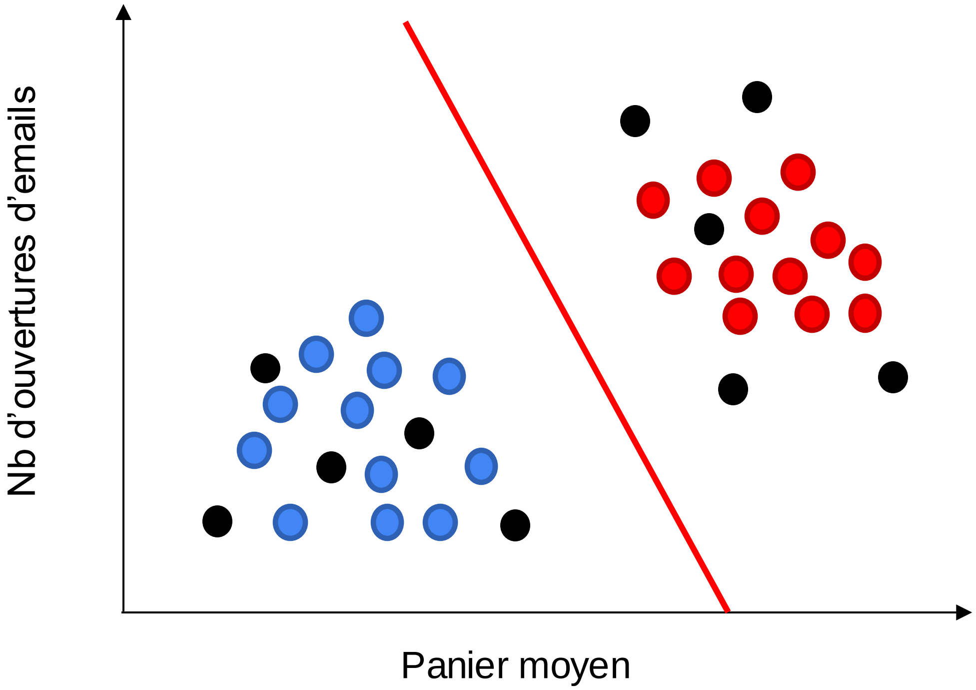
Problème de classification

X $\longrightarrow$ Y

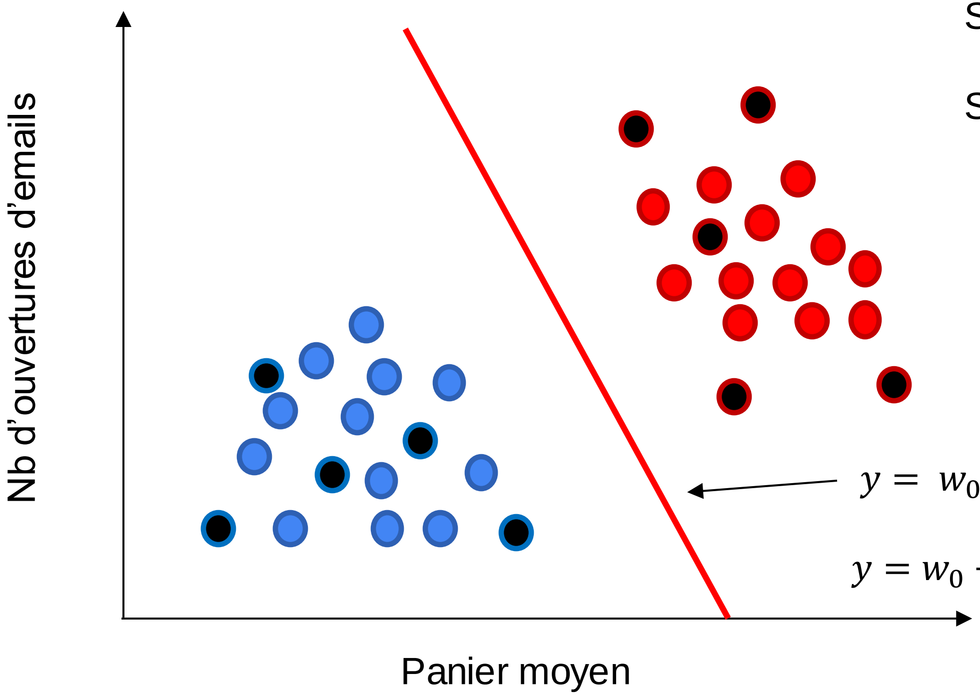
	Nb d'e-mails ouverts (x_1)	Nb de produits achetés (x_2)	Panier moyen (x_3)	Ouverture de l'e-mail (y)
1	12	3	120	1
2	0	1	40	0
3	30	10	1800	1
4	14	5	799	1
…	…	…	…	…
m	25	2	260	0

Jeu d'entraînement pour la prédiction de prix de maison

Utiliser une droite



Utiliser une droite



Si y=1 :

$$w_0 . x_0 + \dots + w_n . x_n \geq 0$$

Si y=0 :

$$w_0 . x_0 + \dots + w_n . x_n < 0$$

$$y = w_0 + w_1 . x_1 + w_2 . x_2$$

$$y = w_0 + w_1 . x_1 + w_2 . x_2 + \cdots + w_n . x_n$$

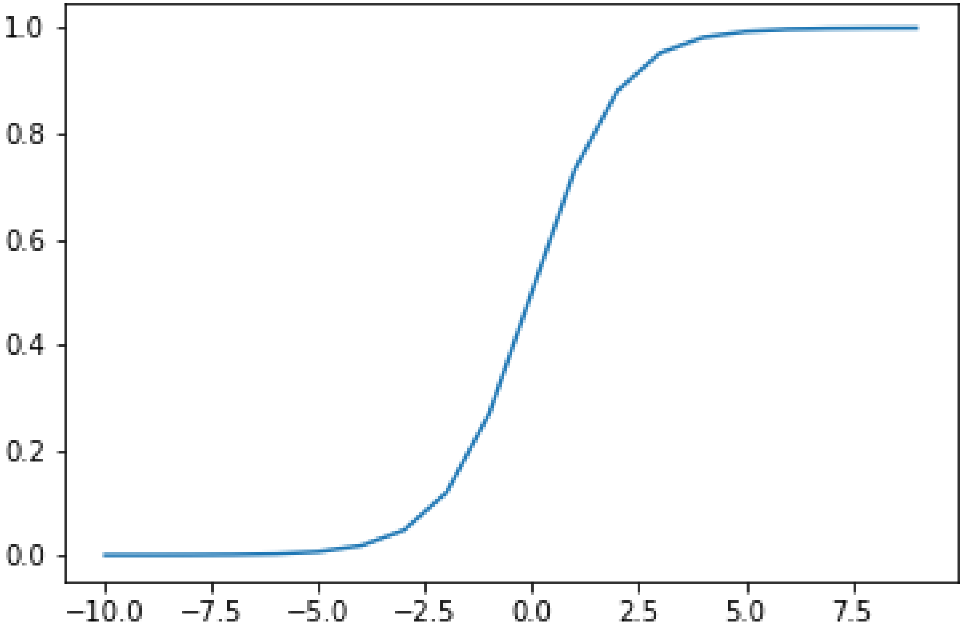
Similitude avec la régression linéaire

- Nous voulons une ligne séparant nos deux classes.
- Pour cela, nous utiliserons donc l'équation de la régression linéaire :

$$\hat{y} = w_0 . x_0 + w_1 . x_1 + w_2 . x_2 + \ldots + w_n . x_n$$

- Pour la régression linéaire $\hat{y} \in R$
- Pour la régression logistique $0 \leq \hat{y} \leq 1$

La fonction sigmoïde



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

$$0 \leq g(z) \leq 1$$

Expression du modèle

- Cette fonction prend en entrée les variables et retourne une valeur entre 0 et 1.

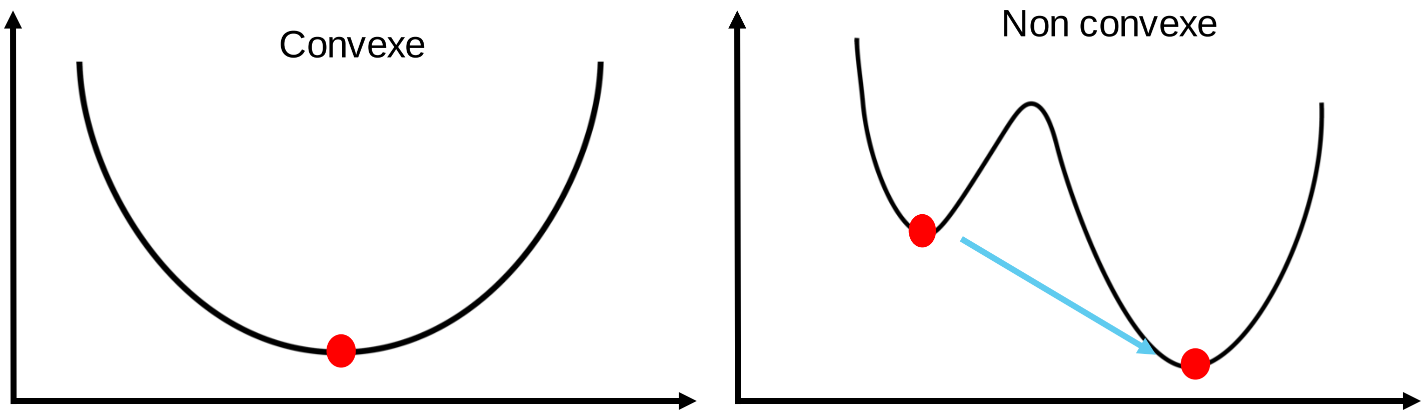
$$\hat{y} = g(W^T X) = \frac{1}{1+e^{-W^T X}}$$

- Plus $\hat{y}$ est proche de 1, plus l’observation a de chance d’appartenir à la classe 1.
- Plus $\hat{y}$ est proche de 0, plus l’observation a de chance d’appartenir à la classe 0.

Une fonction de coût de la régression linéaire

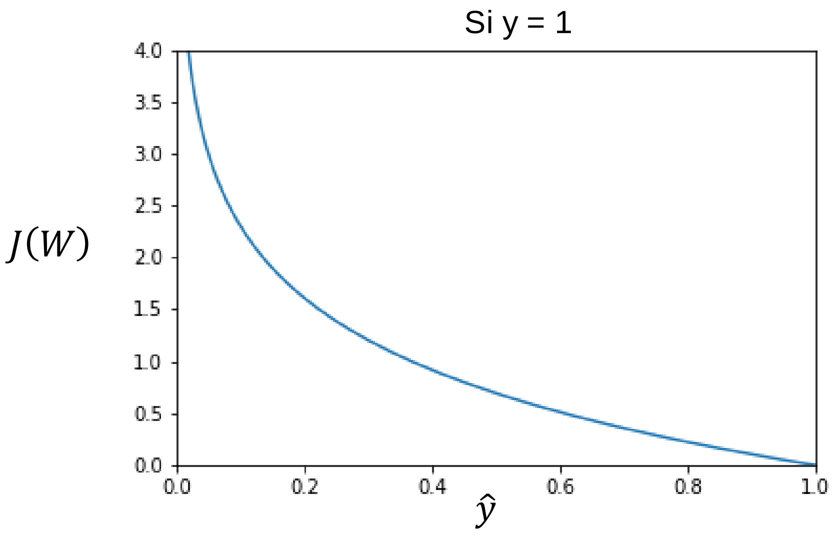
$$\min J(W) = \frac{1}{2m}\sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2$$

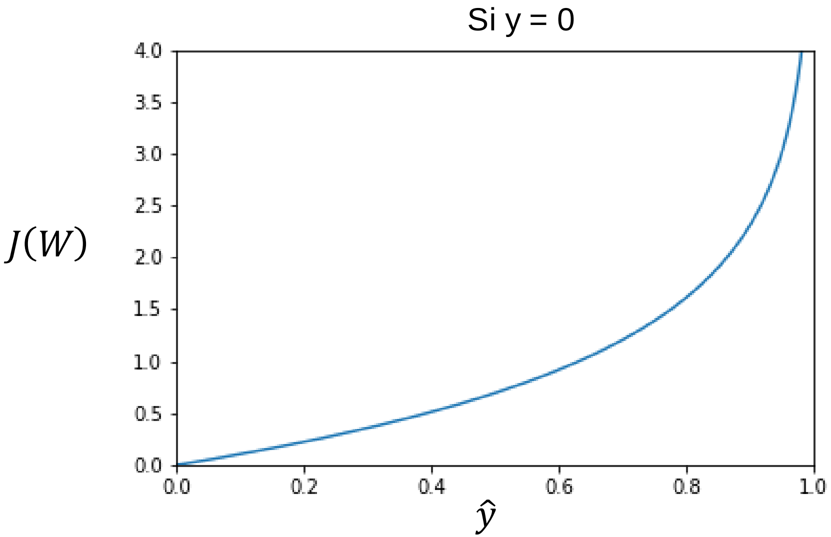
$$\min J(W) = \frac{1}{2m}\sum_{i=1}^{m} (\frac{1}{1+e^{-w^T x^{(i)}}} - y^{(i)})^2$$



Fonction convexe pour la classification

$$Coût(\hat{y}, y) = \begin{cases} -\log(\hat{y}) & si\ y = 1 \\ -\log(1-\hat{y}) & si\ y = 0 \end{cases}$$





Fonction de coût pour la régression logistique

$$Coût(\hat{y}, y) = \begin{cases} -\log(\hat{y}) & si\ y = 1 \\ -\log(1-\hat{y}) & si\ y = 0 \end{cases}$$

$$Coût(\hat{y}, y) = \boxed{-y\log(\hat{y})} - \boxed{(1-y)\log(1-\hat{y})}$$

Si y = 1 $$Coût(\hat{y}, y) = -1\log(\hat{y})$$

Si y = 0 $$Coût(\hat{y}, y) = -1\log(1-\hat{y})$$

$$min\, J(W) = \frac{1}{m}\sum_{i=1}^{m} -y^{(i)}\log(\hat{y}^{(i)}) - (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

Le gradient descent

Répéter jusqu'à la convergence {

$w_j \coloneqq w_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)}) \,.\, x_j^{(i)}$(simultanément j = (0, …, n)

}