Tout savoir sur la régression linéaire

Partie 1 : La théorie

Présenté par **Morgan Gautherot**

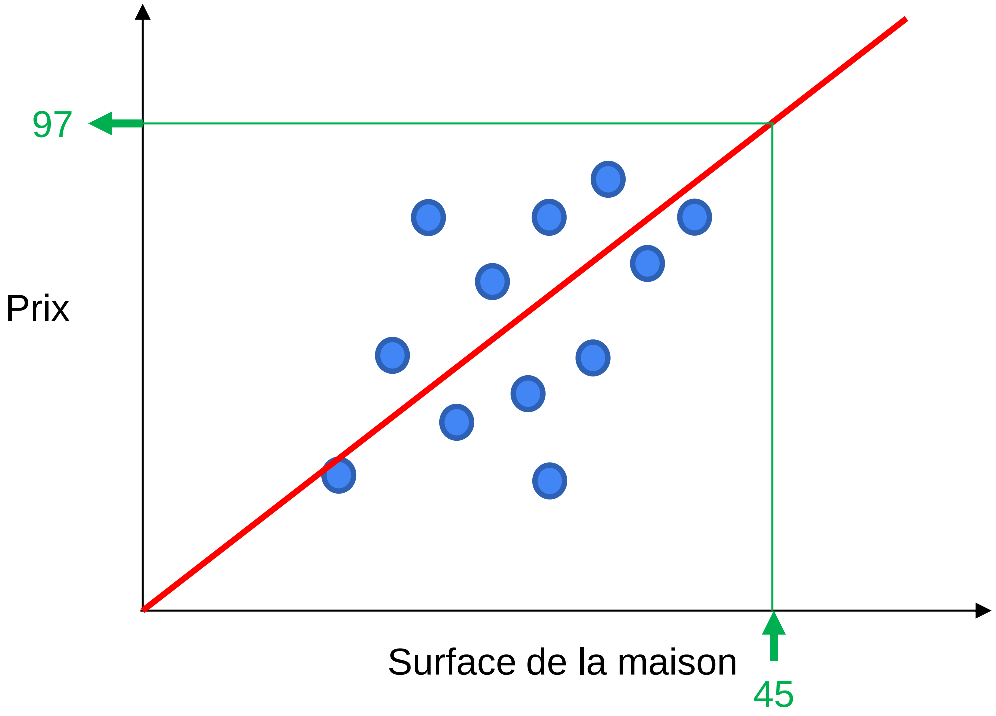
Problème de régression

X → Y

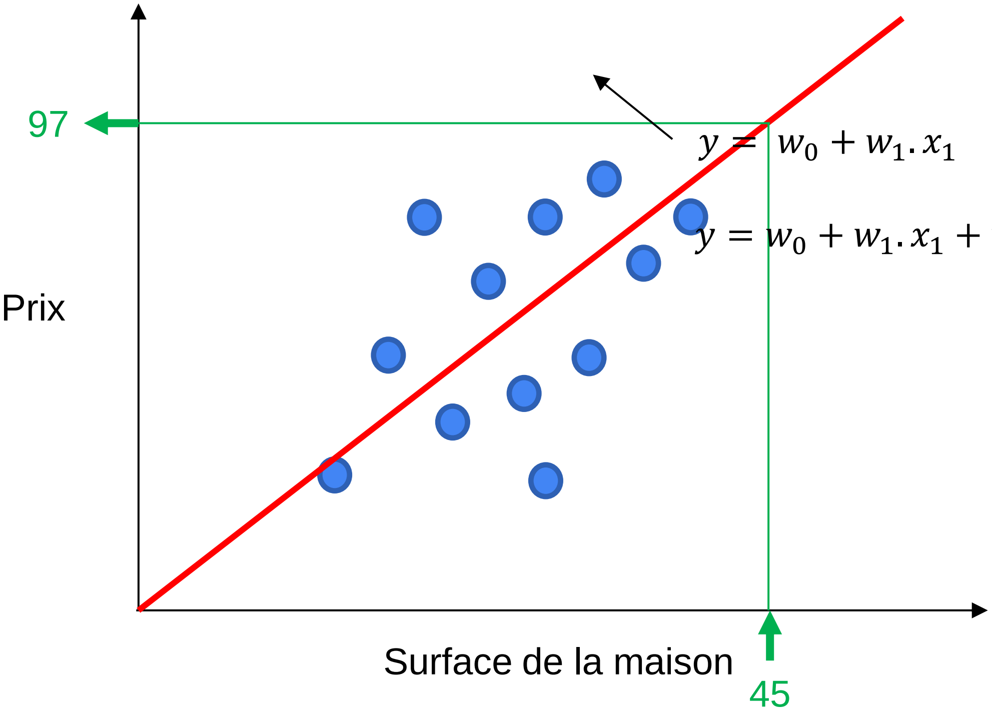
	Surface (x_1)	Nb de pièces(x_2)	Année (x_3)	Prix (y)
1	70	3	2010	460
2	40	3	2015	232
3	45	4	1990	315
4	12	2	2017	178
…	…	…	…	…
m	25	1	2005	240

Jeu d'entraînement pour la prédiction de prix de maison

Utiliser une droite



Equation d’une droite

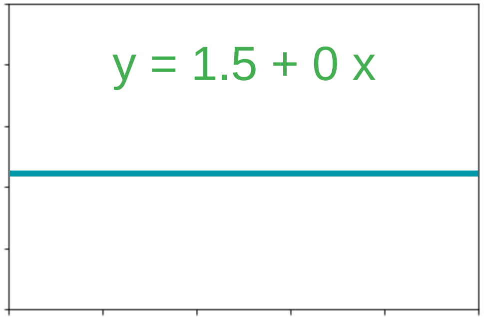


$$y = w_0 + w_1.x_1 \qquad f(x) = a.x + b$$

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

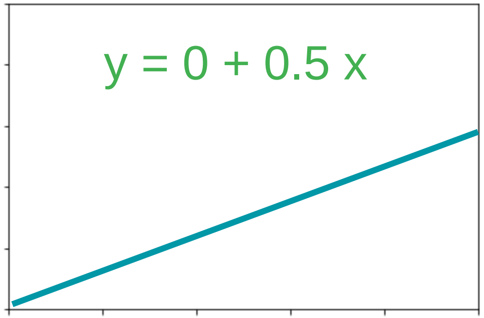
Les paramètres de la régression linéaire

$$y = w_0 + w_1 x$$



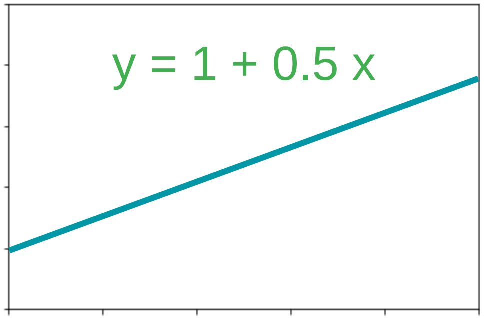
$w_0 = 1.5$

$w_1 = 0$



$w_0 = 0$

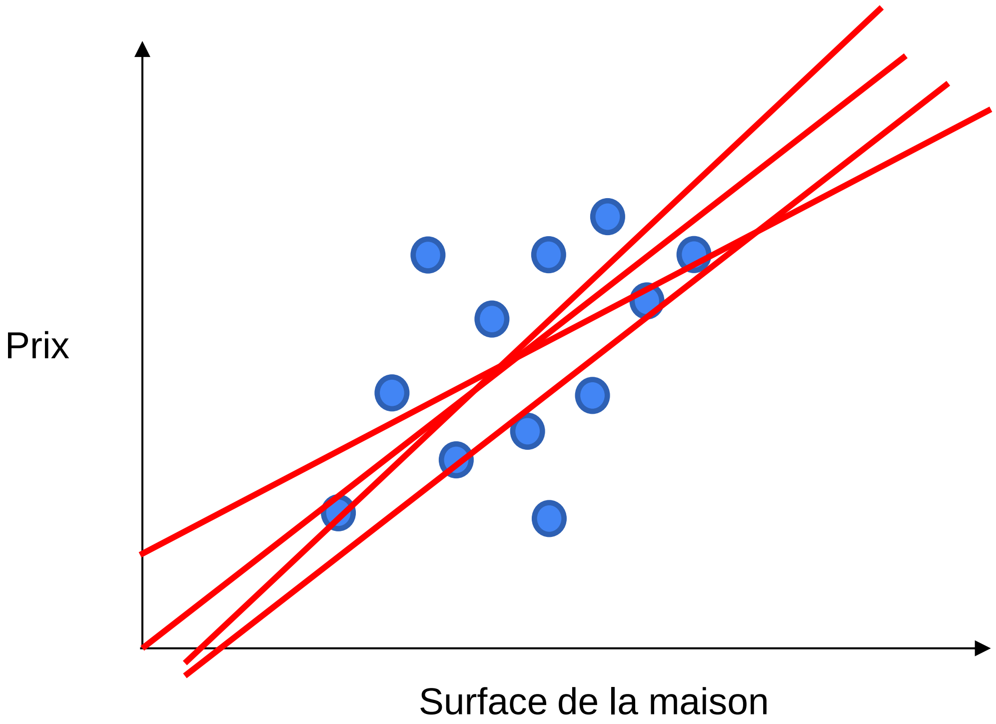
$w_1 = 0.5$



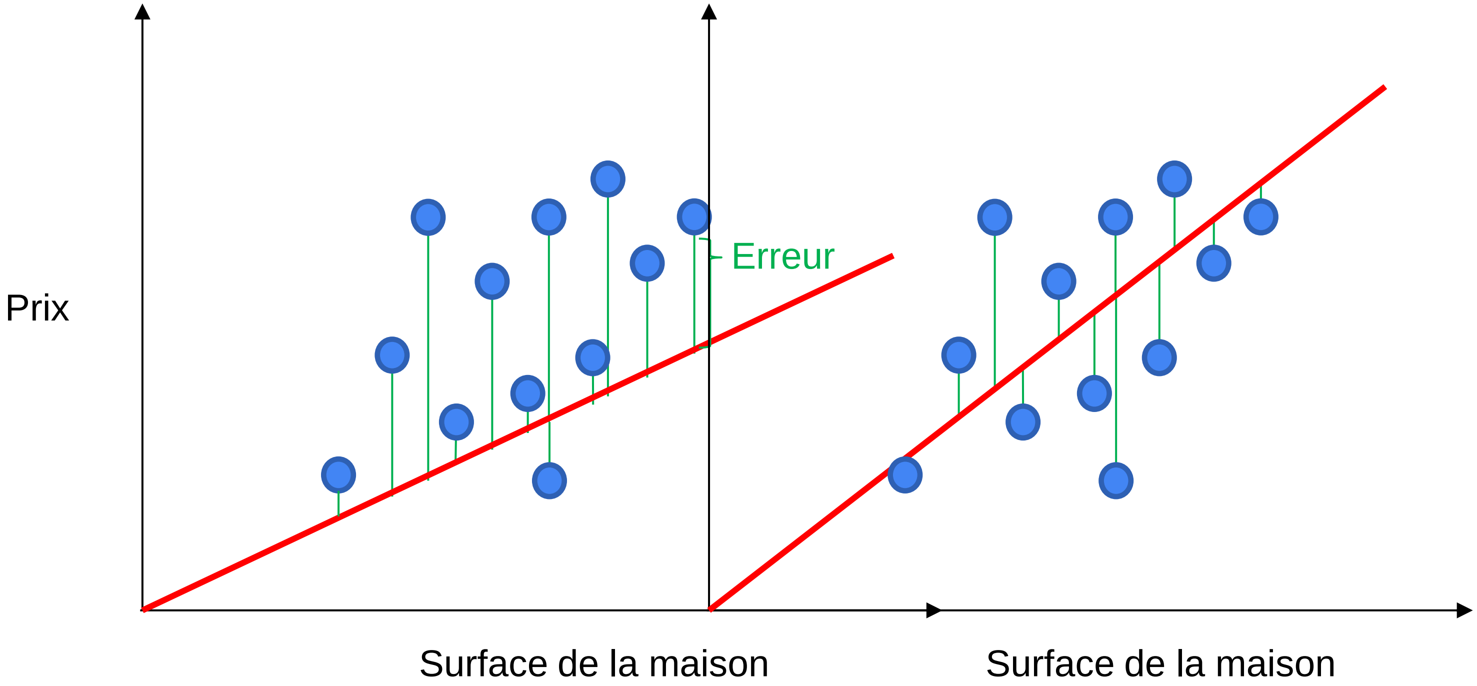
$w_0 = 1$

$w_1 = 0.5$

Quels paramètres choisir ?



La notion d’erreur



L’erreur moyenne au carré

$$\hat{y} - y$$

$$\frac{1}{m}\sum \hat{y} - y$$

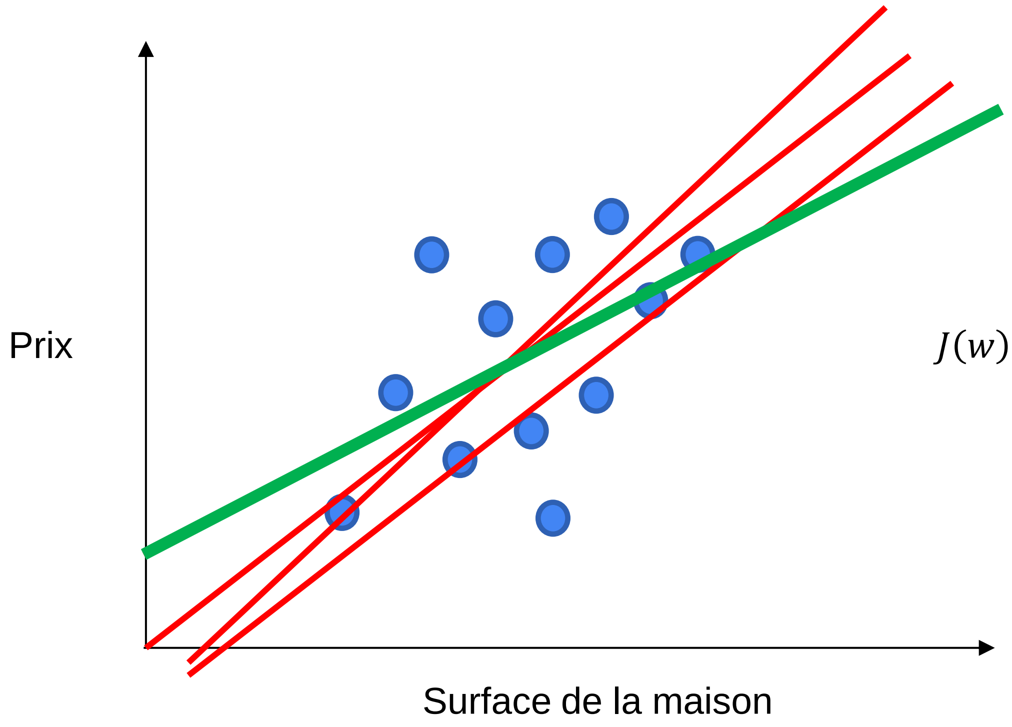
$$\hat{y}_1 = 500 \qquad y_1 = 520 \qquad \hat{y}_1 - y_1 = -20$$

$$\hat{y}_2 = 350 \qquad y_2 = 320 \qquad \hat{y}_1 - y_1 = 30$$

$$\frac{1}{2}\sum_{i=1}^{2} \hat{y}_{(i)} - y_{(i)} = \frac{10}{2} = 5$$

$$J(w) = \frac{1}{m}\sum_{i=1}^{m} (\hat{y}_{(i)} - y_{(i)})^2$$

Quels paramètres choisir ?



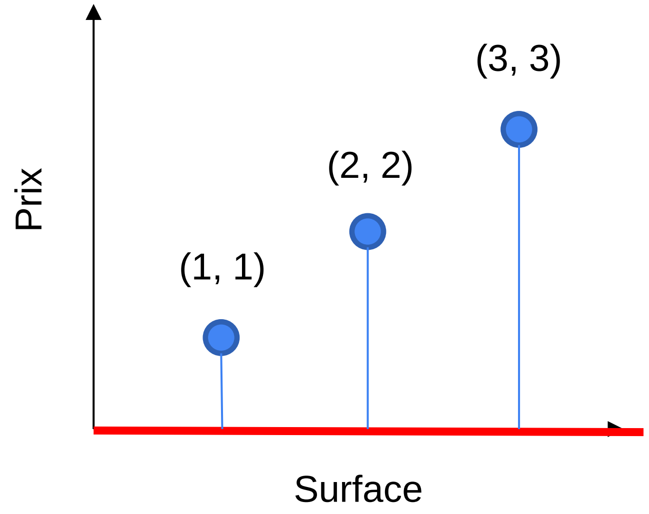
$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_{(i)} - y_{(i)})^2$$

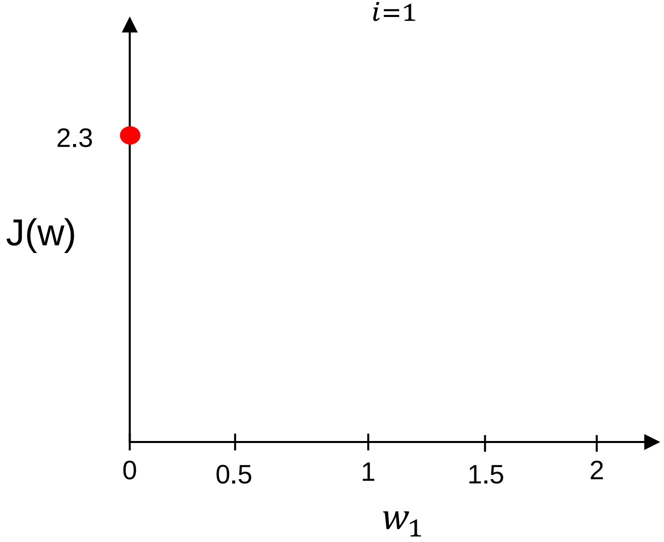
Tracer la fonction de coût

$$\hat{y} = w_0 + w_1 . x_1$$

w_0 est fixé à 0
w_1 va varier

$$J(w) = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_{(i)} - y_{(i)})^2$$





$$\hat{y} = 0 + 0 . x_1$$

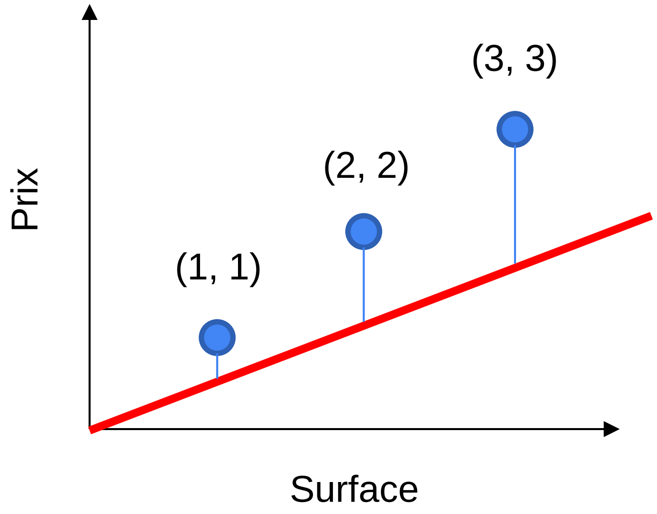
$$J(0) = \frac{1}{2m} [(1)^2 + (2)^2 + (3)^2] = 2.3$$

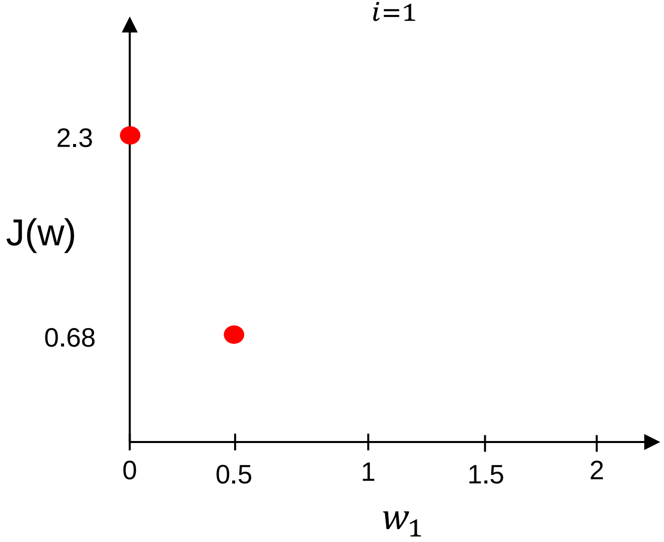
Tracer la fonction de coût

$\hat{y} = w_0 + w_1 . x_1$

w_0 est fixé à 0
w_1 va varier

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_{(i)} - y_{(i)})^2$$





$\hat{y} = 0 + 0.5 . x_1$

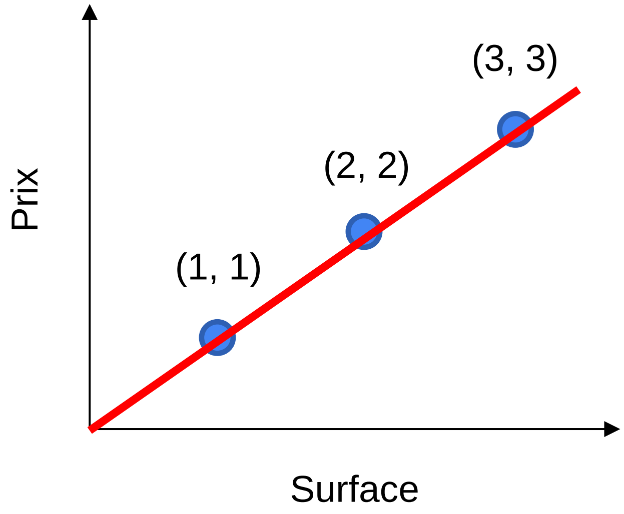
$$J(0.5) = \frac{1}{2m}[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2] = 0.68$$

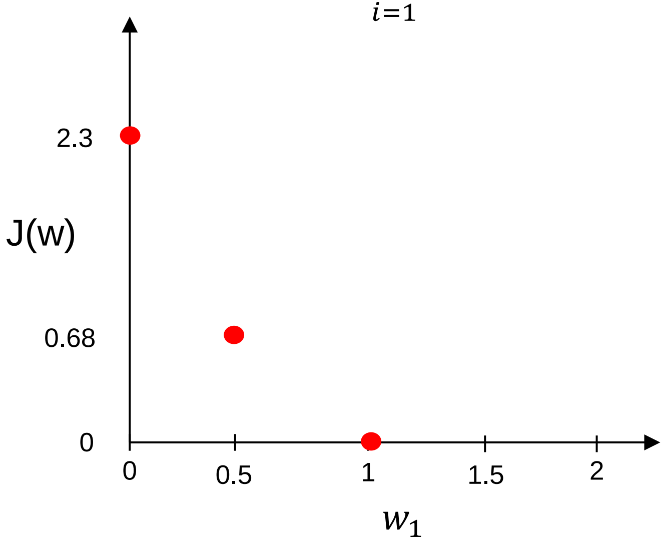
Tracer la fonction de coût

$$\hat{y} = w_0 + w_1 . x_1$$

w_0 est fixé à 0
w_1 va varier

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_{(i)} - y_{(i)})^2$$





$$\hat{y} = 0 + 1 . x_1$$

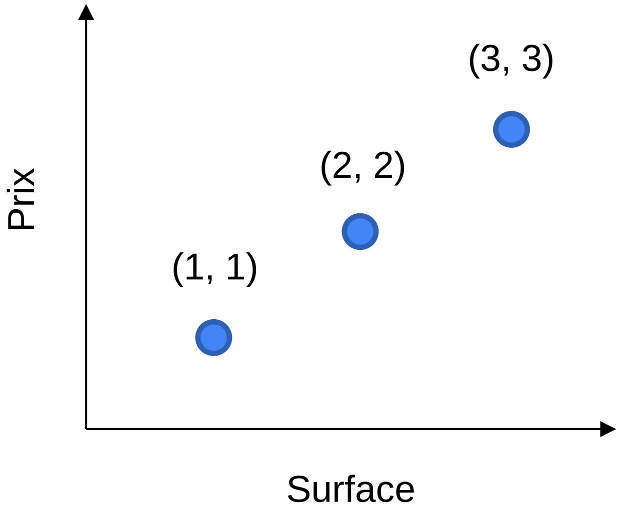
$$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2] = 0$$

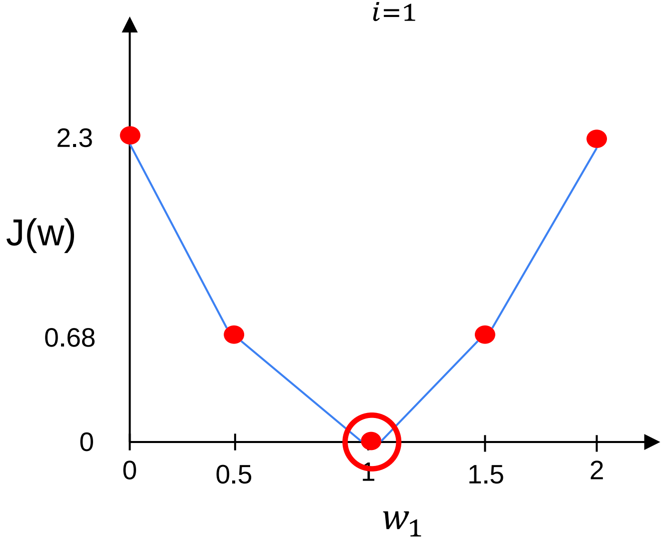
Tracer la fonction de coût

$$\hat{y} = w_0 + w_1 . x_1$$

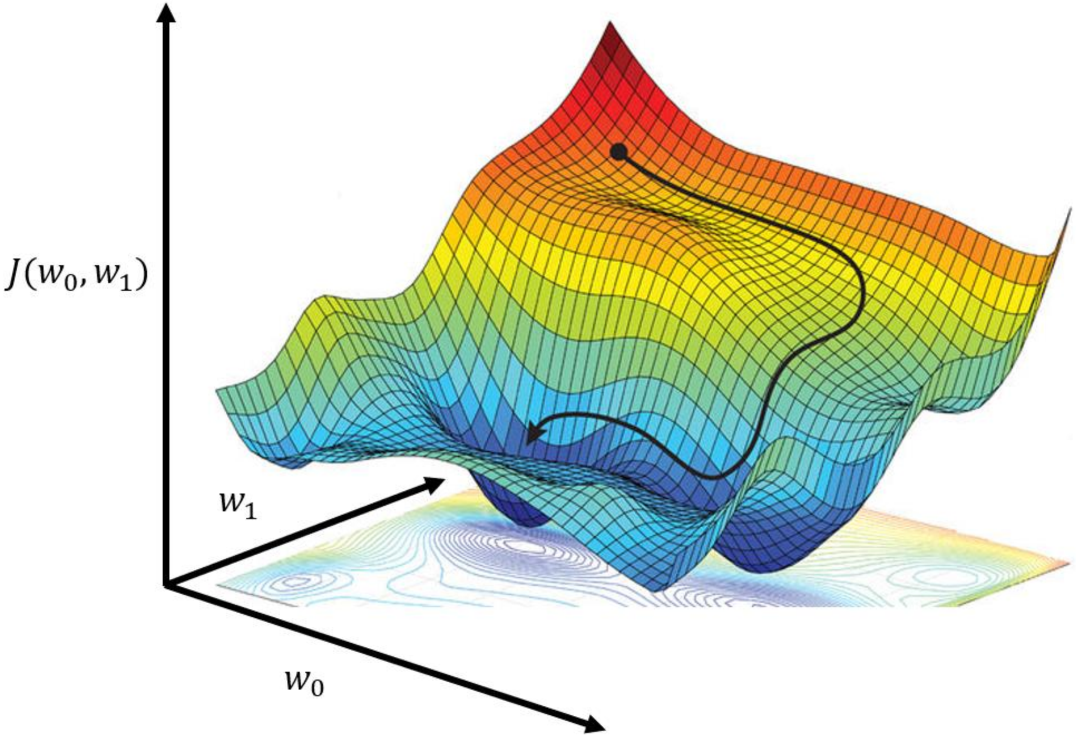
w_0 est fixé à 0
w_1 va varier

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_{(i)} - y_{(i)})^2$$





Fonction de coût avec deux paramètres



Le gradient descent

Soit n le nombre de variables

Répéter ce processus jusqu’à la convergence :

$$\boxed{w_j} \coloneqq \boxed{w_j} - \boxed{\alpha} \boxed{\frac{\partial}{\partial w_j} J(W)} \qquad j \in [1, n]$$

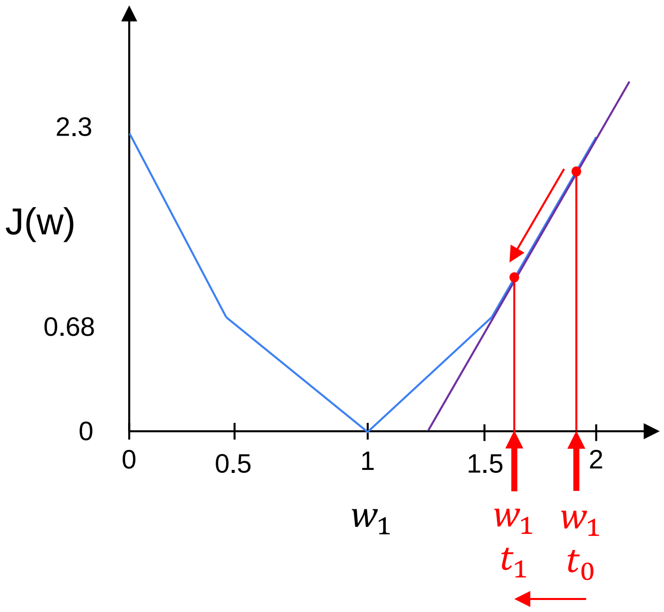
Nouvelle valeur de w_j

Ancienne valeur de w_j

Le learning rate

Dérivée partielle

La dérivée partielle



$$w_1 \coloneqq w_1 - \alpha \boxed{\frac{\partial}{\partial w_1} J(W)}$$

t_1 t_0

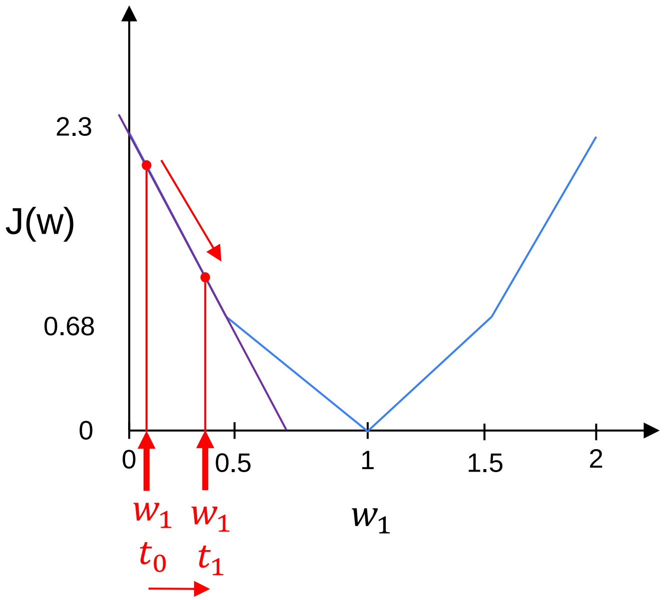
$$w_1 \coloneqq w_1 - positif$$

t_1 t_0

$$w_1 < w_1$$

t_1 t_0

La dérivée partielle

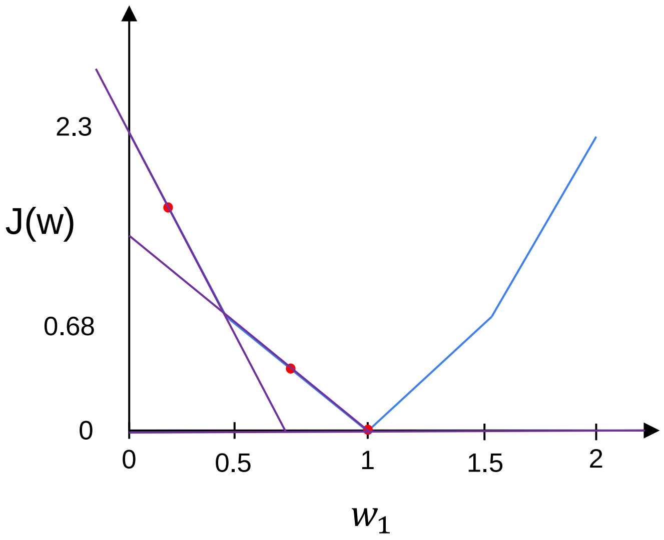


$$\underset{t_1}{w_1} := \underset{t_0}{w_1} - \alpha \boxed{\frac{\partial}{\partial w_1} J(W)}$$

$$\underset{t_1}{w_1} := \underset{t_0}{w_1} - negatif$$

$$\underset{t_1}{w_1} > \underset{t_0}{w_1}$$

Atteindre le minimum

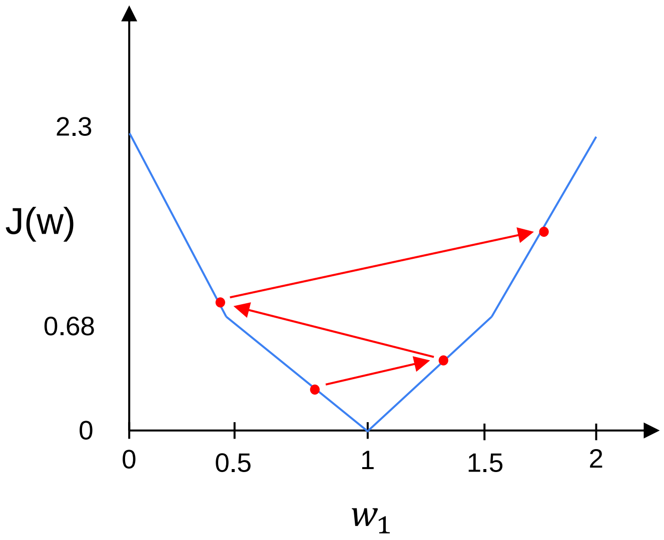


Réduit vers le minimum

$$w_1 := w_1 - \boxed{\alpha}\boxed{\frac{\partial}{\partial w_1} J(W)}$$

Est fixe

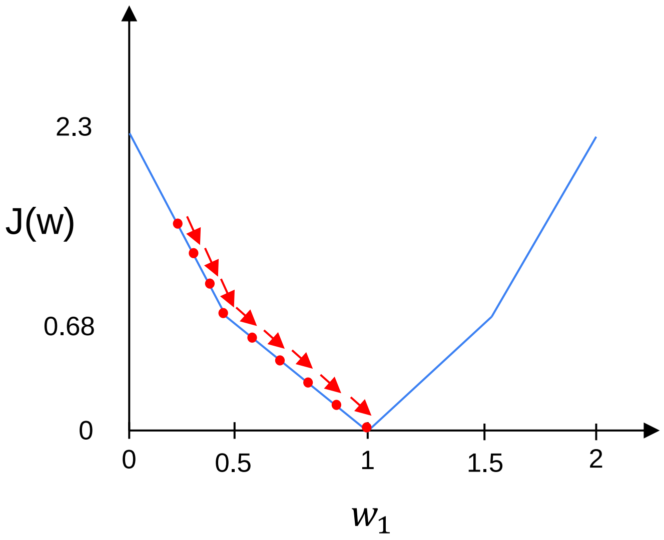
Impact du learning rate



$$w_1 := w_1 - \boxed{\alpha} \frac{\partial}{\partial w_1} J(W)$$

Learning rate trop grand

Impact du learning rate



$$w_1 := w_1 - \boxed{\alpha}\,\frac{\partial}{\partial w_1}J(W)$$

Learning rate trop petit

Impact du learning rate