La régression linéaire

Cours de machine learning avancé

Régression linéaire

1. Définition

- La régression est un ensemble de méthodes statistiques largement utilisées pour analyser la relation d'une variable continue avec une ou plusieurs autres variables. L'objectif est de mettre en place une fonction utilisant les données d'apprentissage de sorte que la combinaison des variables explicatives permette de prédire avec précision la variable expliquée.

- Par exemple :
  - Prédire le nombre de visite d'un site web basé sur le jour de la semaine ;
  - Prédire le chiffre d'affaires d'un magasin basé sur son historique de vente ;
  - Prédire le prix de location d'une maison en se basant sur ses caractéristiques ;
  - Prédire l'âge d'un patient en fonction d'une image de son cerveau.

2. Data

	Taille (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

Jeux d'entraînement pour la prédiction de prix d'une maison

$x^{(1)}$ = (70, 3, 2010)

$y^{(4)}$ = 178

$x_2^{(3)}$ = 4

$(x^{(2)}, y^{(2)})$ = (40, 3, 2015, 232)

3. Modèle

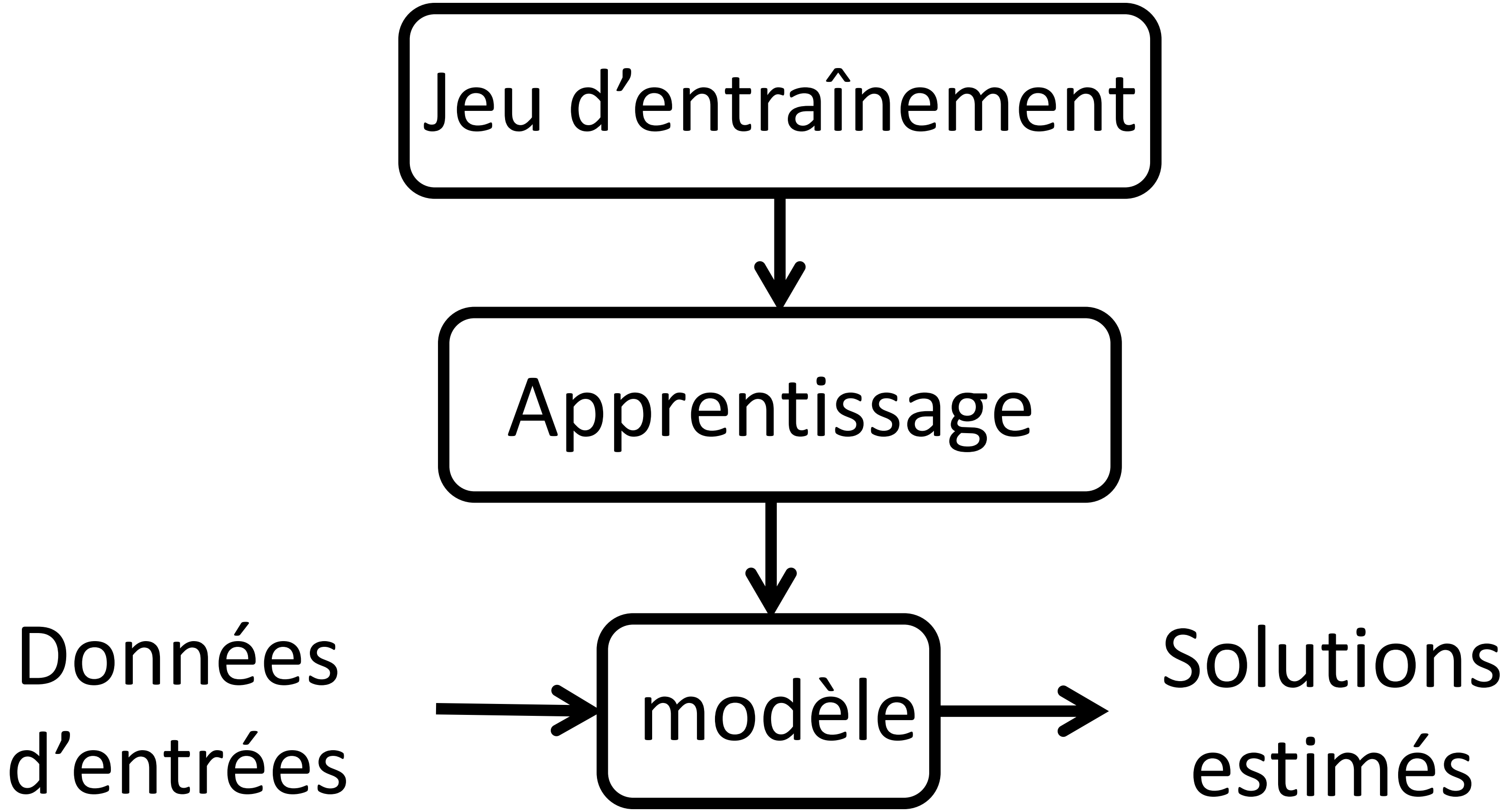


Schéma d’un algorithme de régression

Régression linéaire

1. Problème de régression

- Imaginez que vous avez une société immobilière, votre but est de louer des appartements aux bons prix. Pas trop bas pour ne pas avoir de perte de revenus, pas trop haut pour ne pas le laisser inoccupé.

- **Quel est le juste prix pour chaque appartement ?**

- Le juste prix n'est pas facile à trouver, les humains même les experts ont un parti pris basé sur leurs expériences. Vous voulez une solution neutre et précise pour votre entreprise. Vous voulez donc utiliser un algorithme de machine learning qui trouvera automatiquement le bon prix pour vos appartements.

2. Jeu de données

- Un modèle de régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, appelée variable cible, et une ou plusieurs variables, appelées variables explicatives.

	Taille (x)	Prix (y)
1	70	460
2	40	232
3	25	315
4	12	178
…	…	…
m	25	240

- **m** est le nombre de lignes (observations).
- **x** est appelé variable explicative ou variable d'entrée
- **y** est appelé variable cible ou variable de sortie

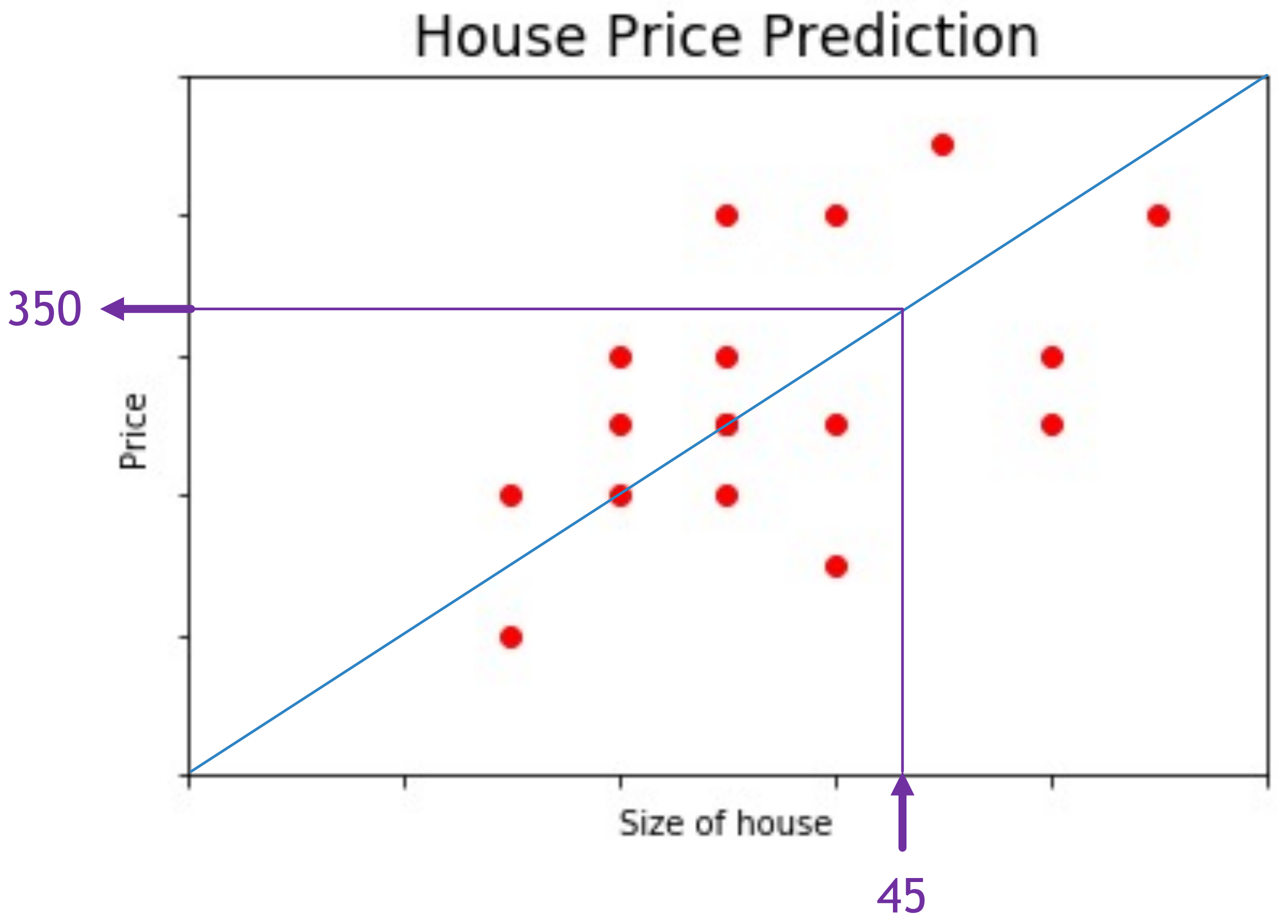
Exemple de lecture du jeu de données :

$x^{(1)}$ = 70

$y^{(4)}$ = 178

$(x^{(3)}, y^{(3)})$ = (40, 232)

3. Une fonction de régression



Régression linéaire

1. Modèle

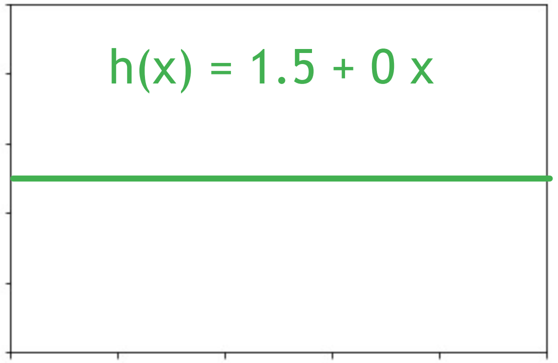
- Le modèle est une fonction qui prend les données d'entrée et estime notre valeur cible.

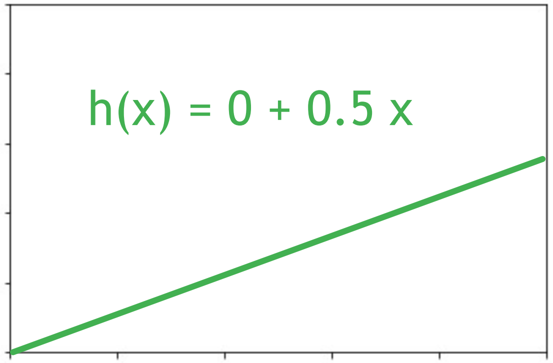
$$h(x) = w_0 + w_1 x$$

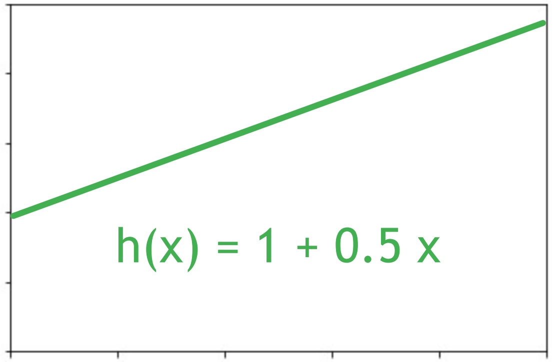
- Cette fonction est l'expression de la régression linéaire univariée.

2. Exemple de fonctions

h(x) = $w_0 + w_1 x$







w_0 = 1.5

w_1 = 0

w_0 = 0

w_1 = 0.5

w_0 = 1

w_1 = 0.5

Régression linéaire

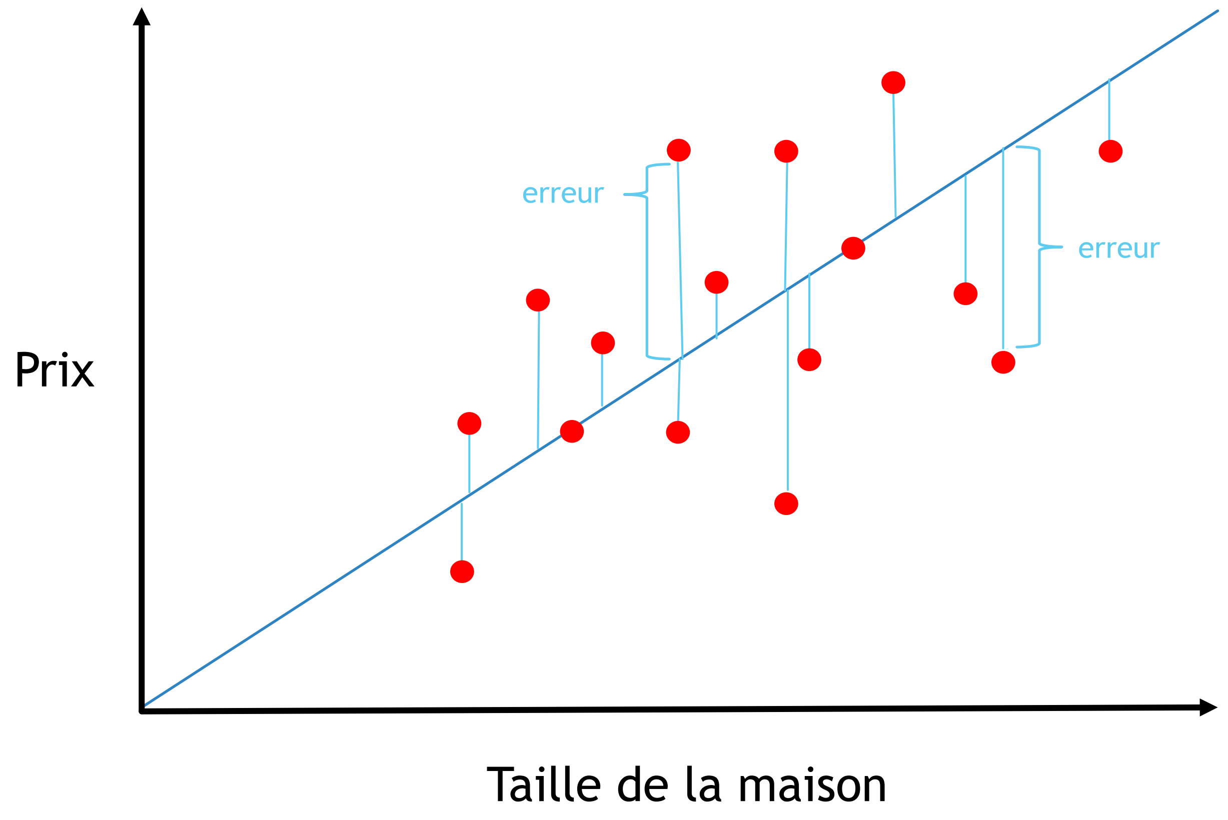
1. Choisissez vos paramètres avec soin

- Les paramètres de la fonction lui permettent de s'adapter facilement au problème et donc de prévoir le phénomène étudié.

Comment choisir nos paramètres $w_0,\ w_1$ pour avoir la meilleure prédiction ?

- Le but est de choisir les paramètres afin de minimiser l'erreur de notre prédiction, minimisant ainsi h(x) – y.

2. Prédiction de l'erreur

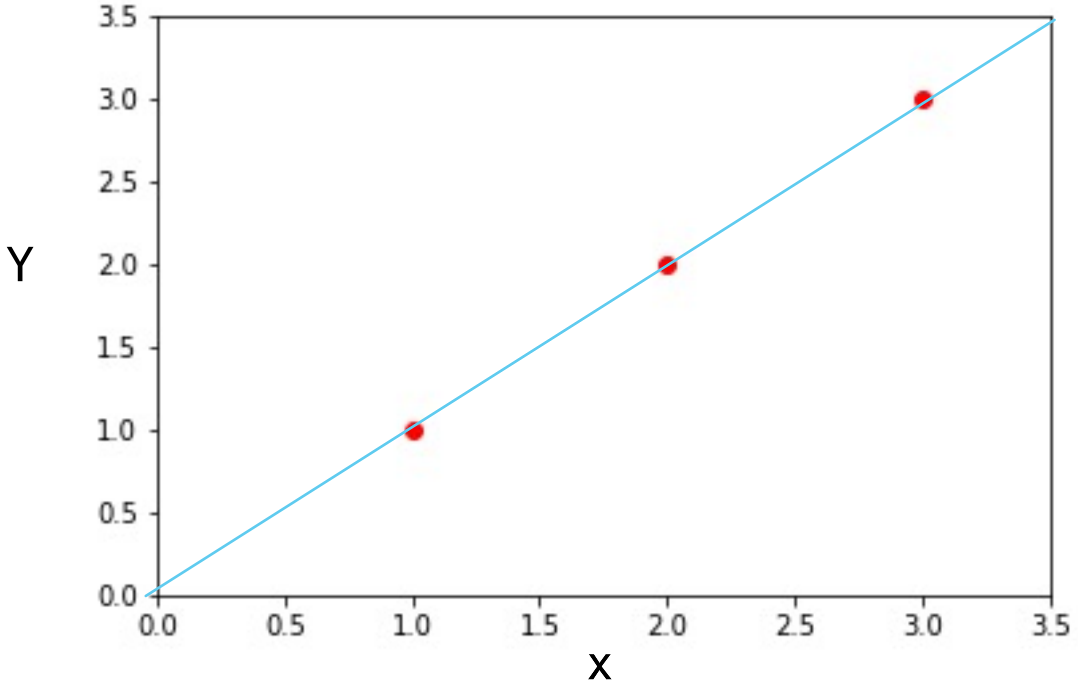


3. Fonction de coût (Erreur moyenne au carré)

$$Minimiser\ J(W) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

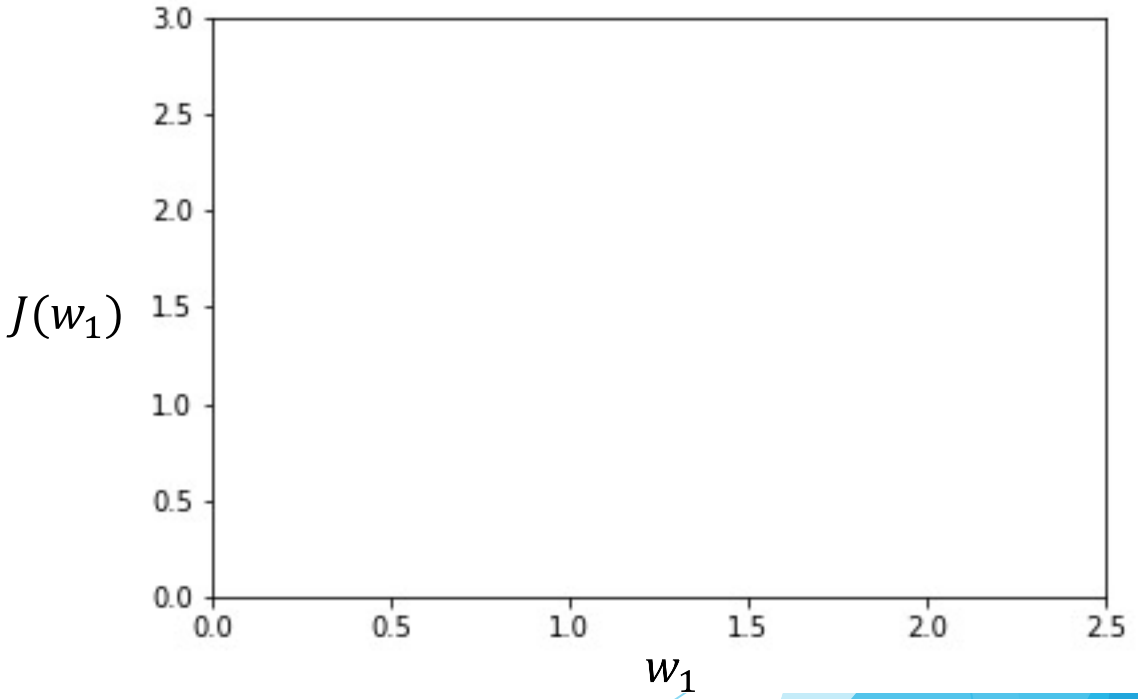
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x) = 0 + 1\ x$

Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre

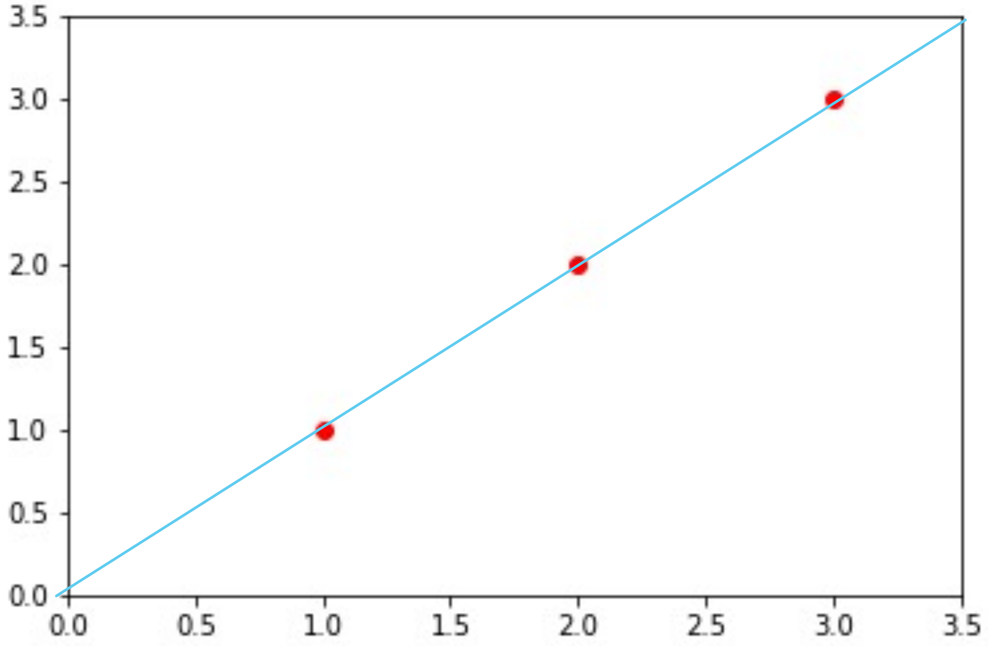


$J(1) = \frac{1}{2m}[(1-1)^2 + (2-2)^2 + (3-3)^2] = 0$

4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre

Y

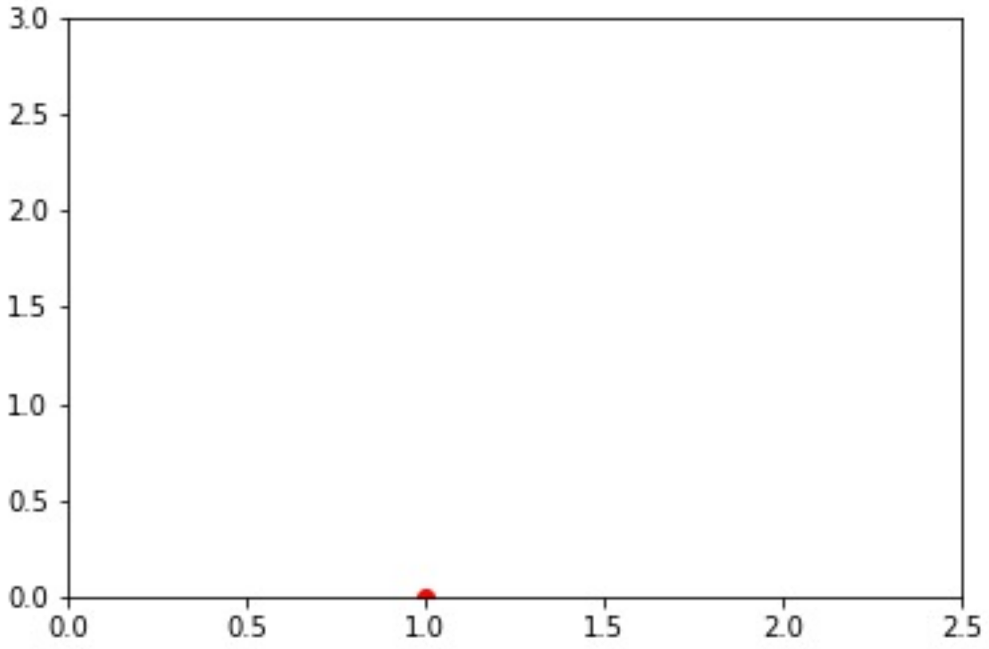


x

$h(x)$ = 0 + 1 x

Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre

$J(w_1)$

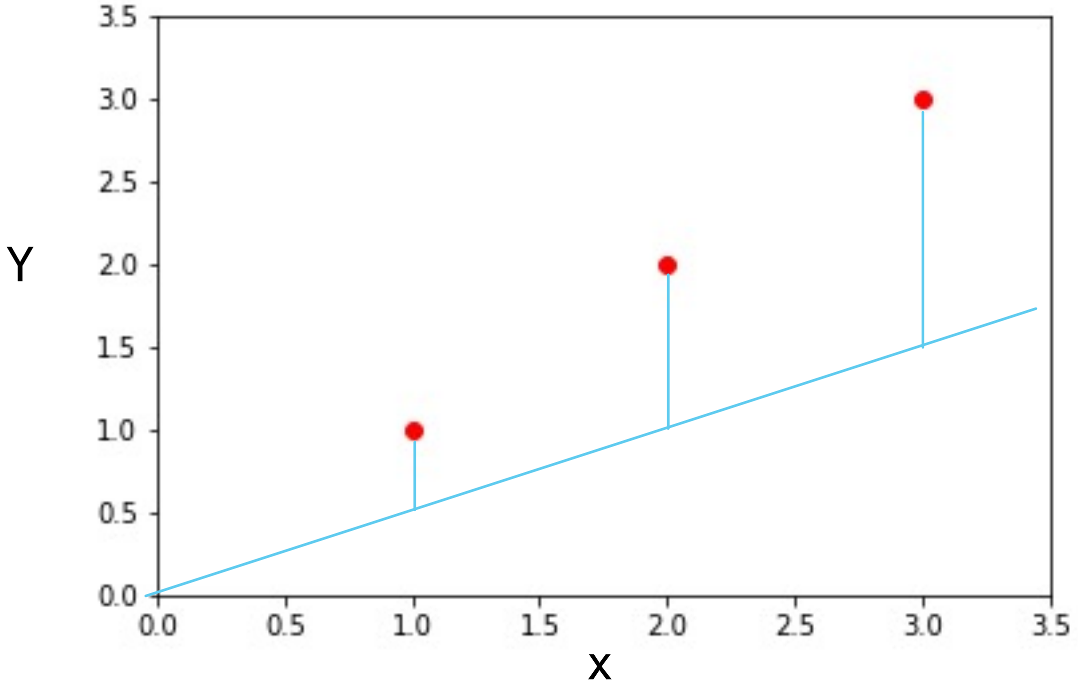


w_1

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

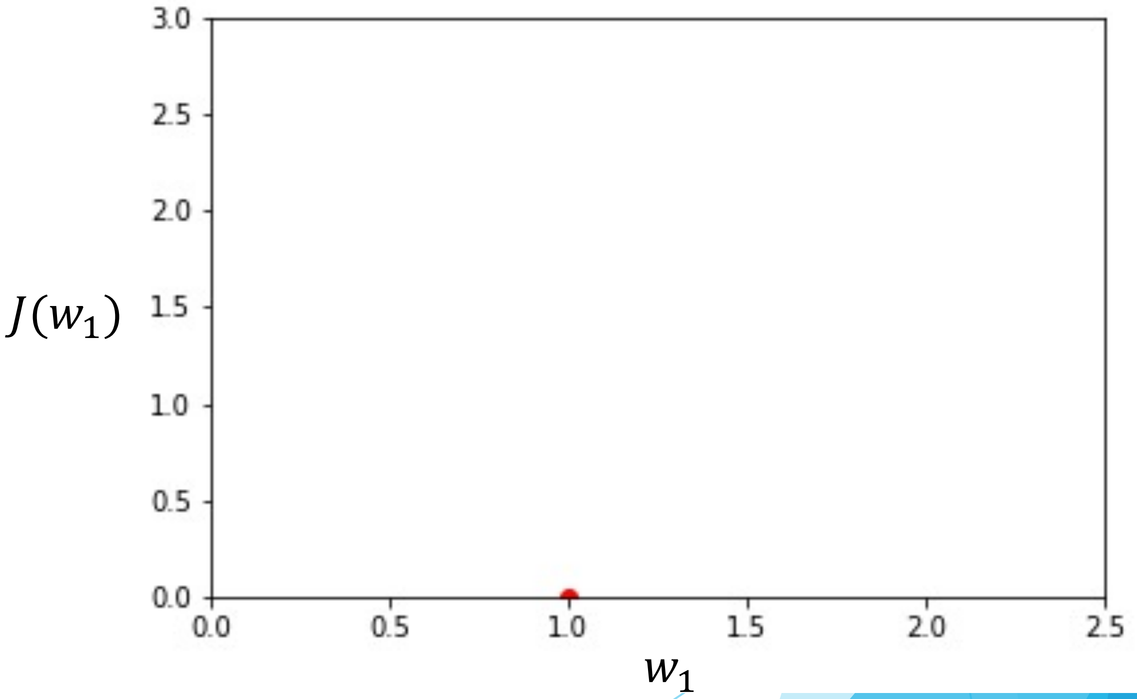
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x)$ = 0 + 0.5 x

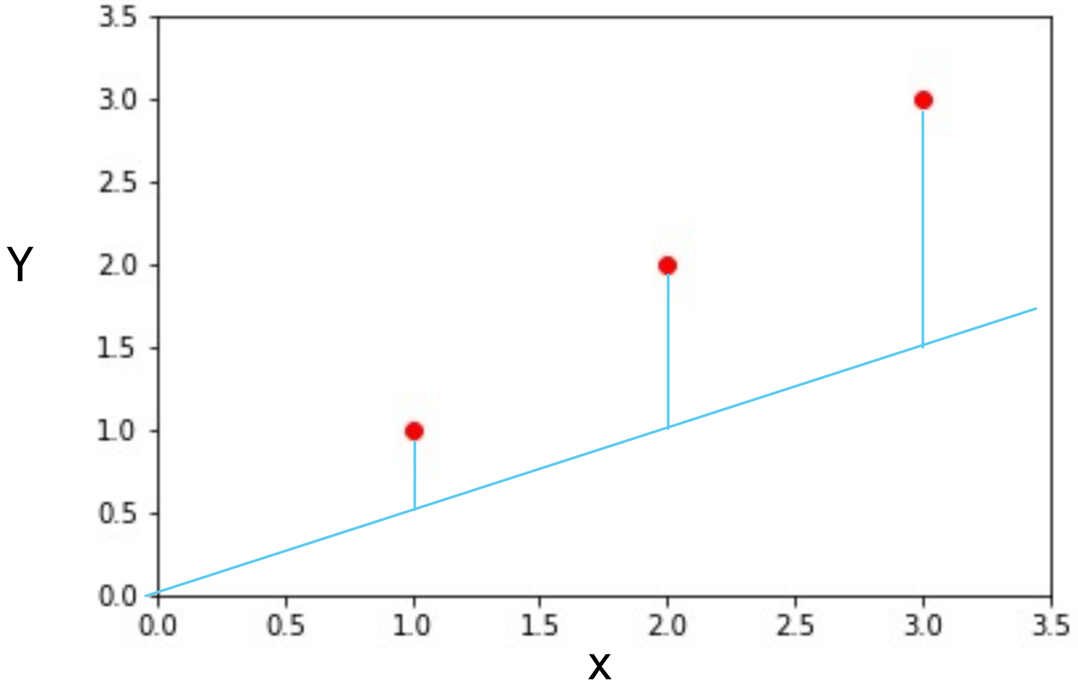
Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



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

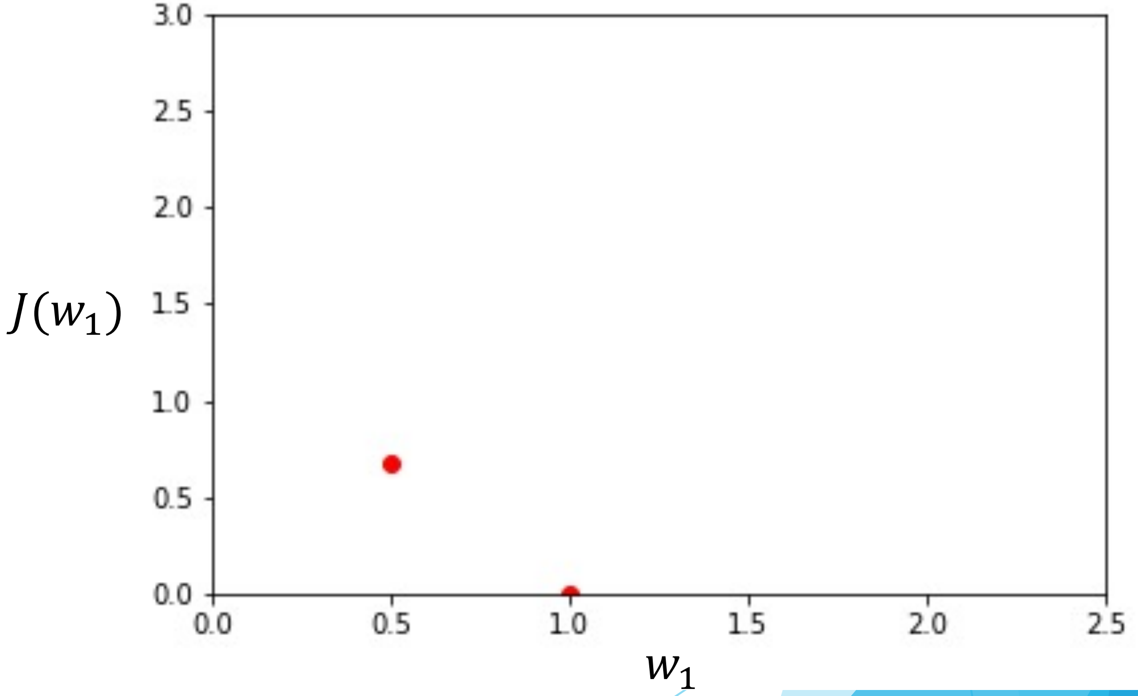
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x)$ = 0 + 0.5 x

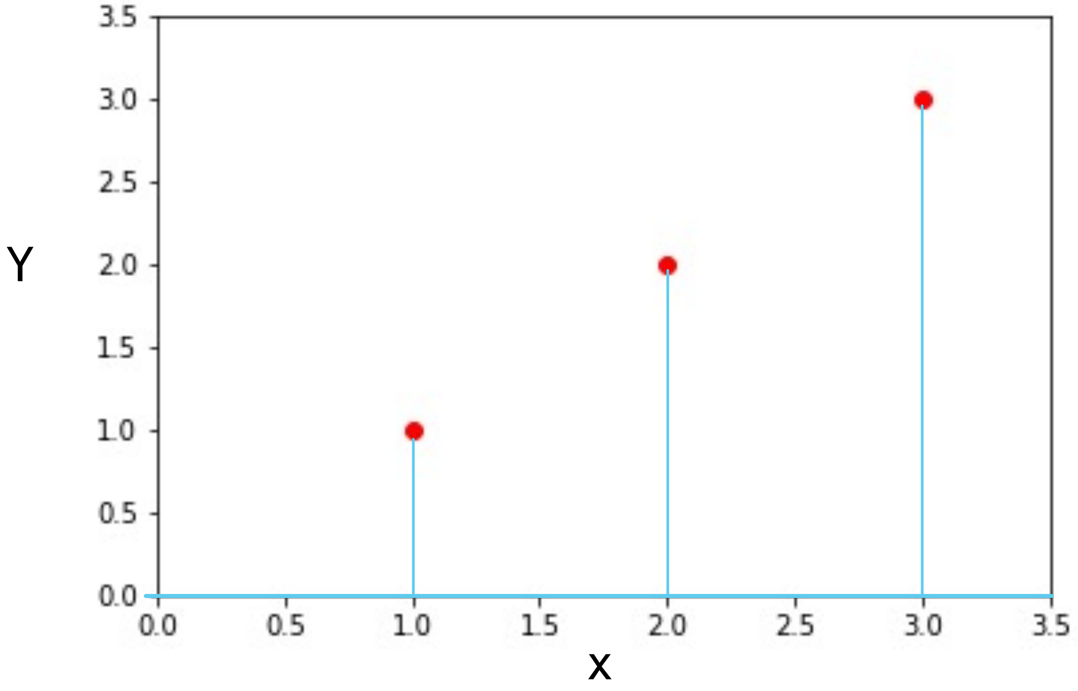
Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



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

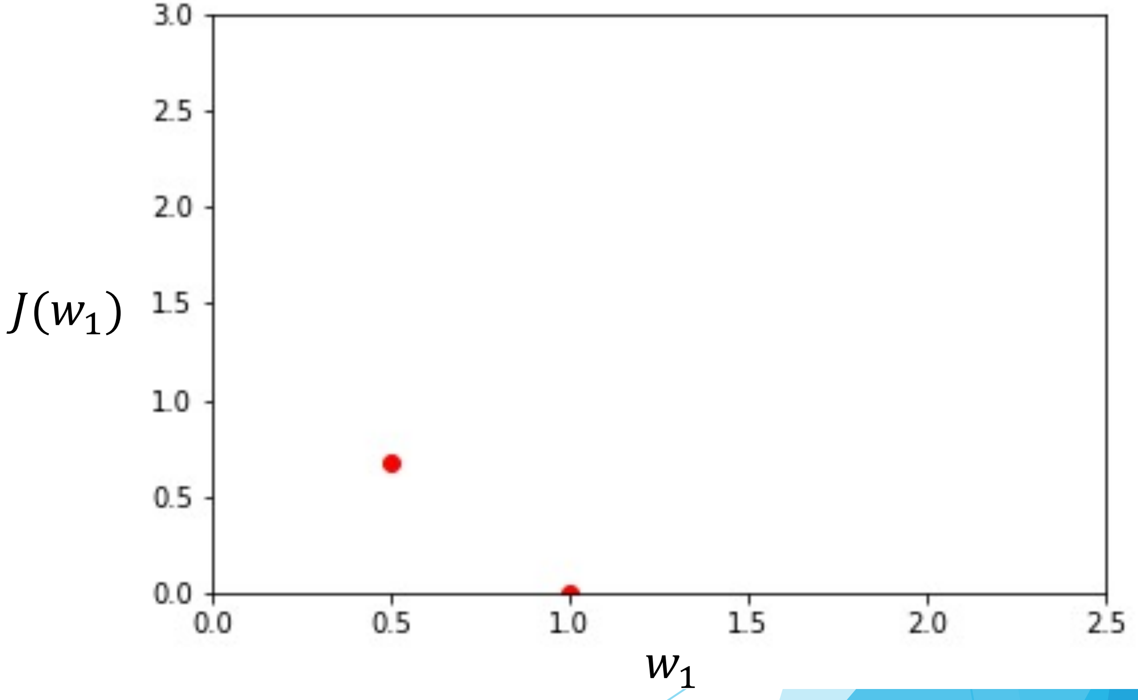
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x) = 0 + 0$ x

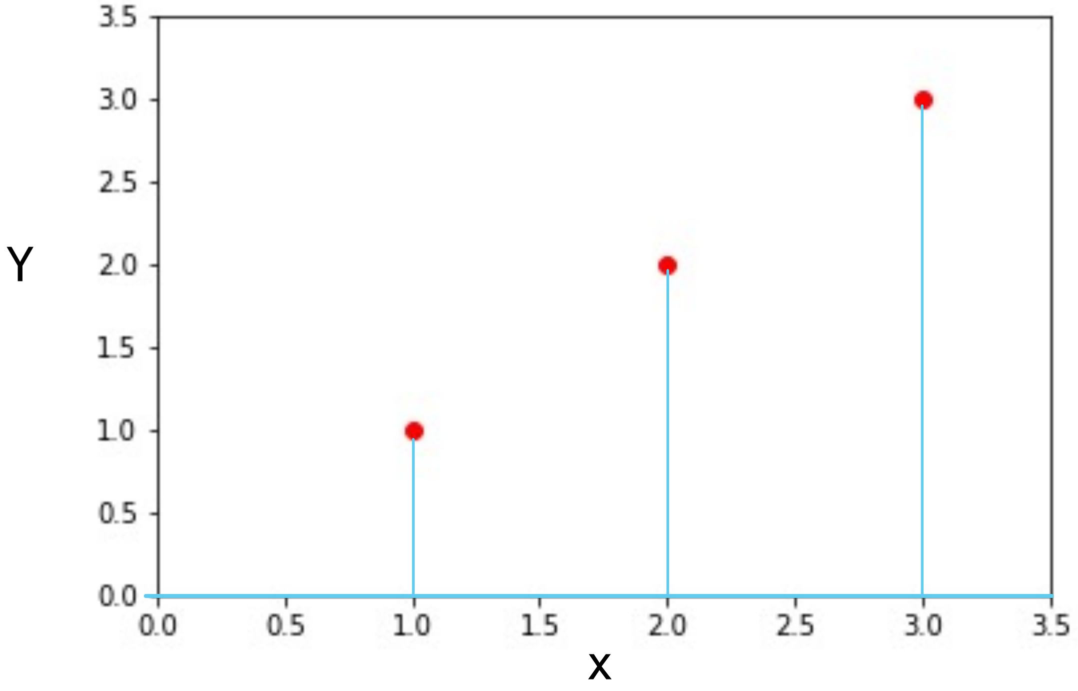
Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



$J(0.5) = \frac{1}{2m}[(1)^2 + (2)^2 + (3)^2] = 2.3$

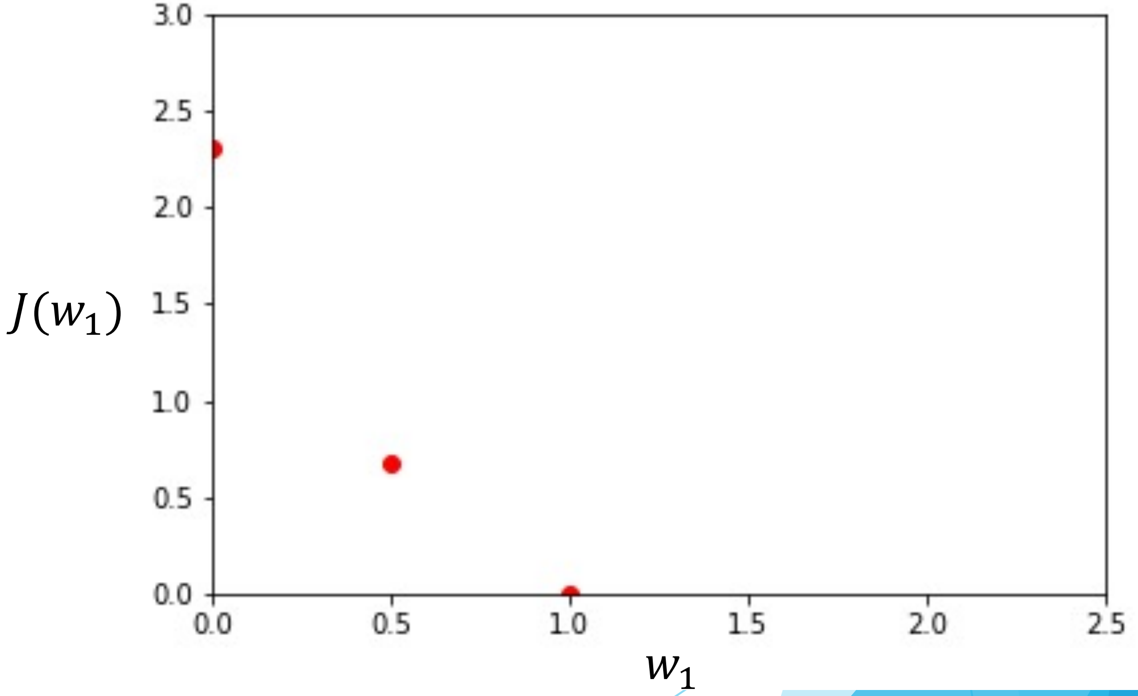
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x) = 0 + 0$ x

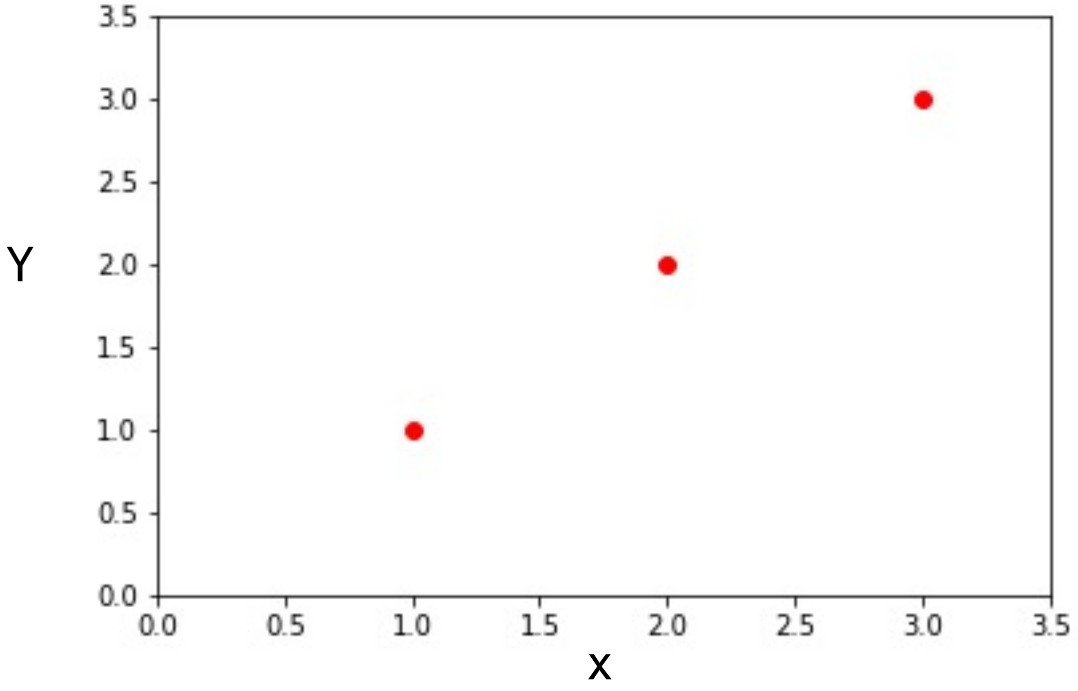
Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



$J(0.5) = \frac{1}{2m}[(1)^2 + (2)^2 + (3)^2]$ = 2.3

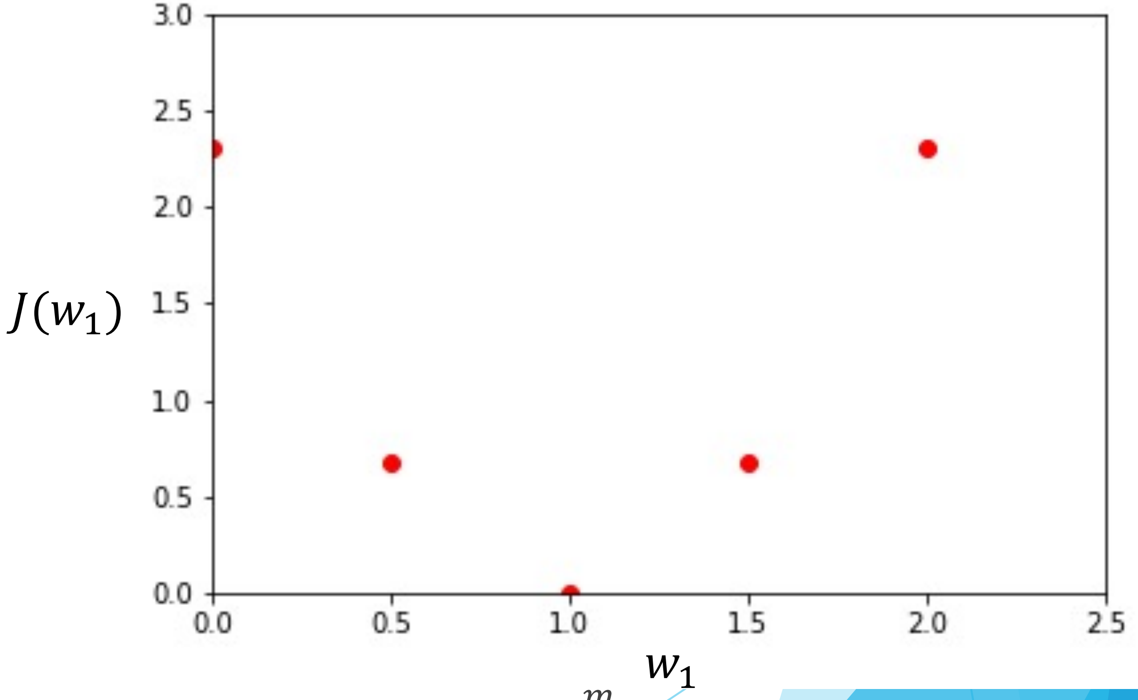
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x) = 0 + w_1$ x

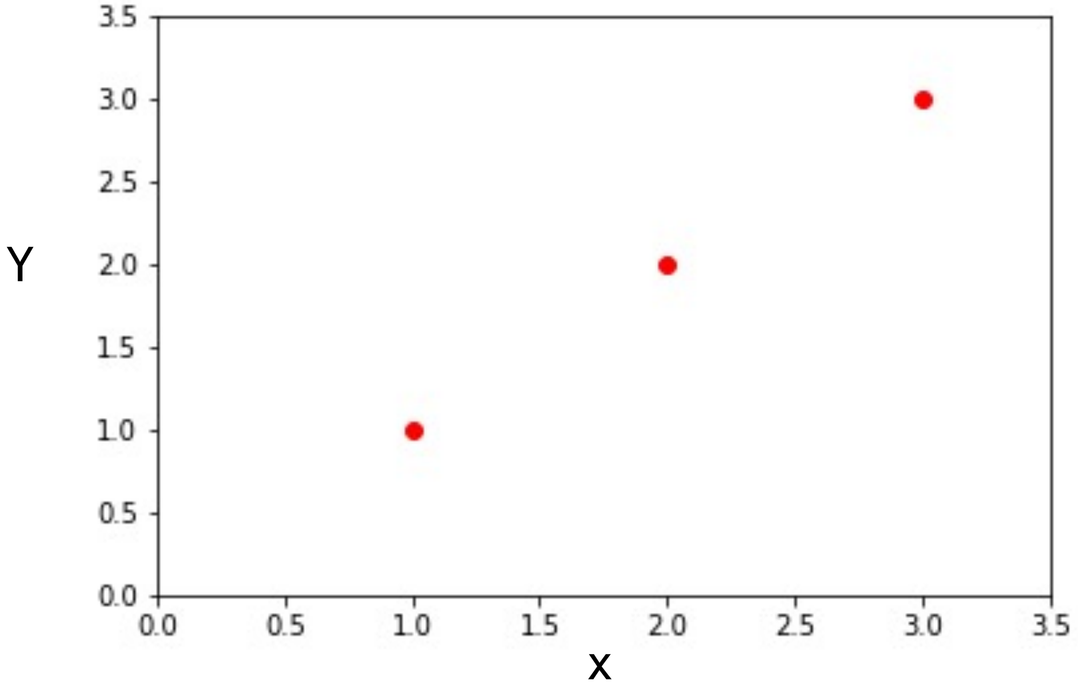
Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



$$J(w_1) = \frac{1}{2m}\sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$$

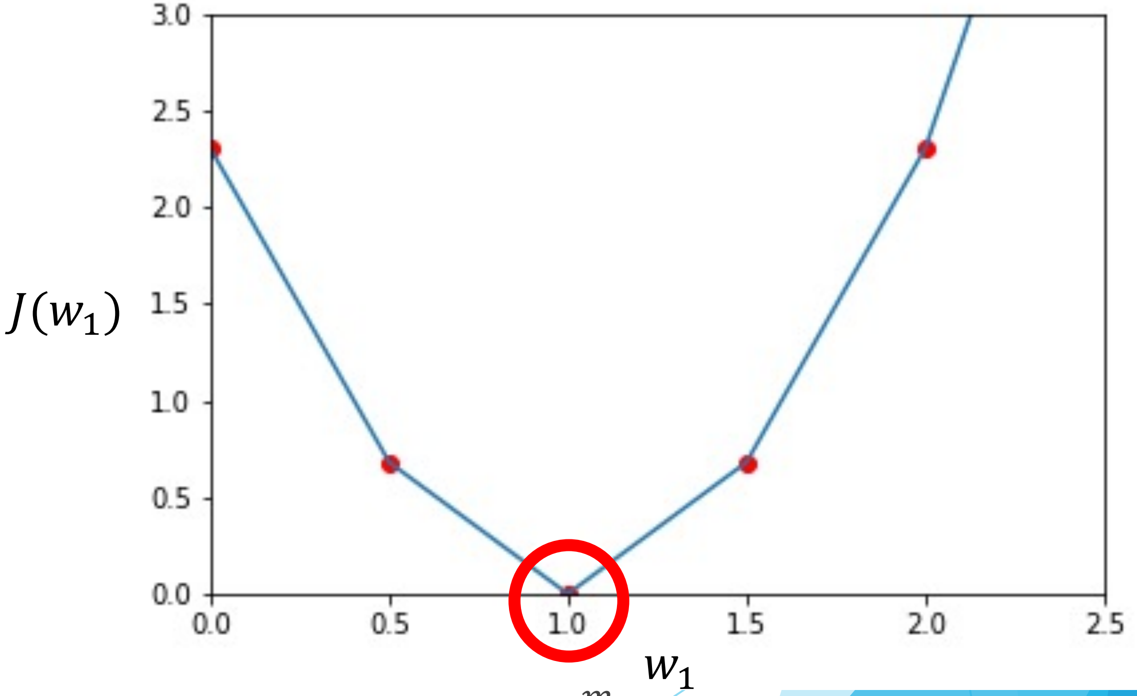
4. Tracé une fonction de coût

Fonction H(x)
w_0 = 0, w_1 fixe, x est un paramètre



$h(x) = 0 + w_1$ x

Fonction J(w_1)
w_0 = 0, x fixe, w_1 est un paramètre



$$J(w_1) = \frac{1}{2m}\sum_{i=1}^{m}(h(x^{(i)}) - y^{(i)})^2$$

Régression linéaire

1. L'idée générale d'optimisation

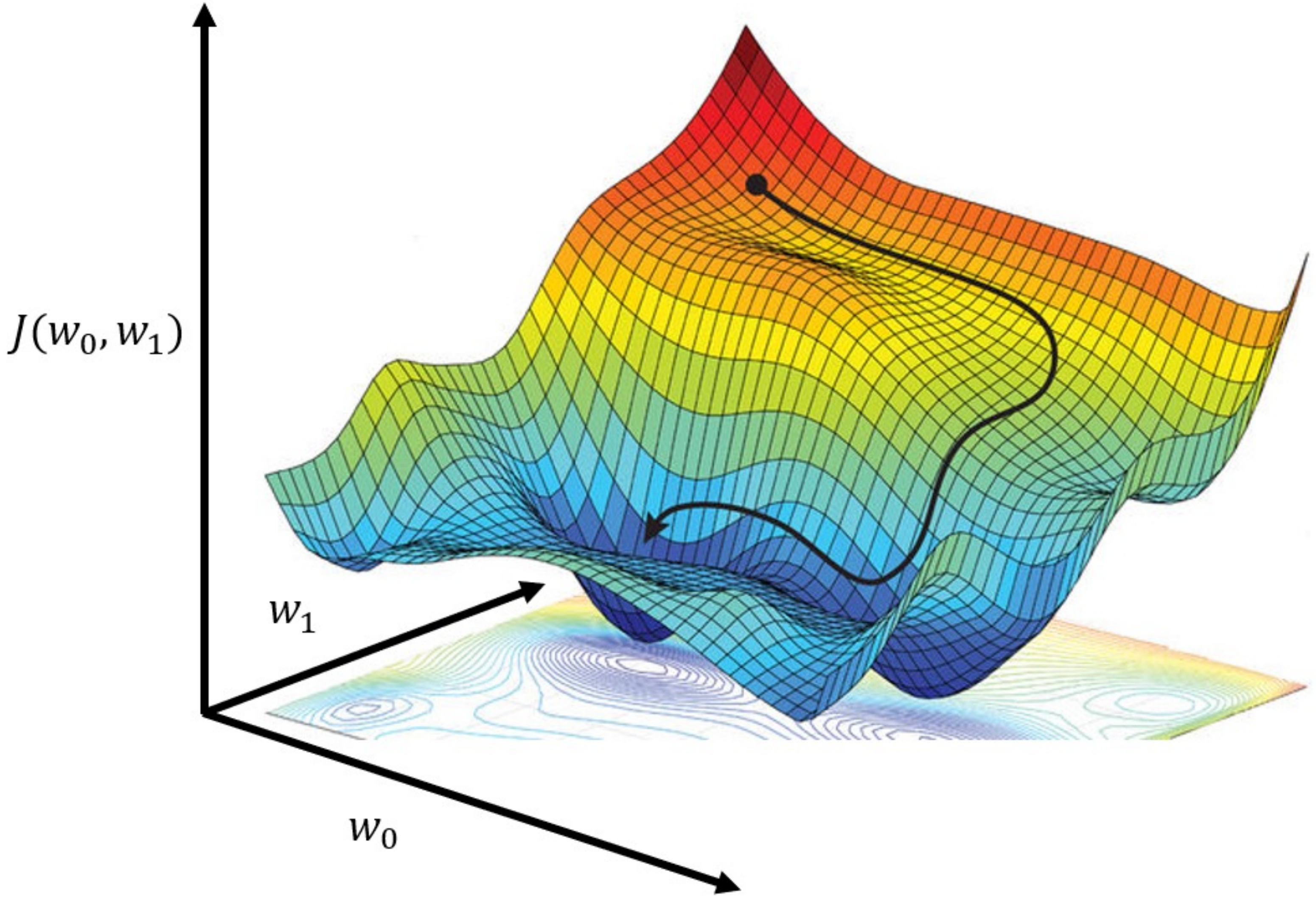
Nous avons une fonction J(W)

Nous voulons optimiser J(W)

Algorithme :

- Commençons avec un vecteur $W \in R^n$
- Nous voulons l'optimiser W pour minimiser J(W) jusqu'à atteindre l'erreur minimum.

2. Fonction de coût avec deux paramètres



3. Algorithme de la descente de gradient

Algorithme :

n correspond au nombre de variables d'entrée du modèle.

Répéter jusqu’à la convergence {

$$w_j := w_j - \underbrace{\boxed{\alpha}}_{\text{Learning rate}} \underbrace{\boxed{\frac{\partial}{\partial w_j} J(W)}}_{\text{Dérivé partiel}} \quad \text{(simultanément j = (0, …, n)}$$

}

Learning rate

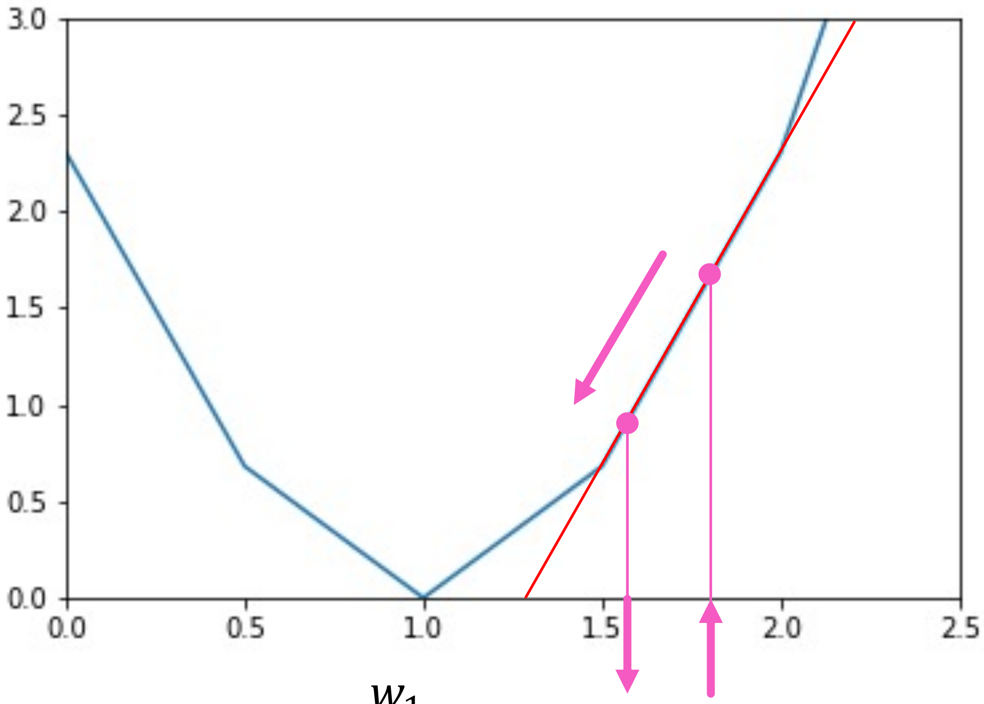
Dérivé partiel

≥ 0

4. Comprendre la descente de gradient

Fonction J(w_1)
w_0 = 0, x fixé, w_1 est un paramètre

$J(w_1)$



w_1

w_1 w_1

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

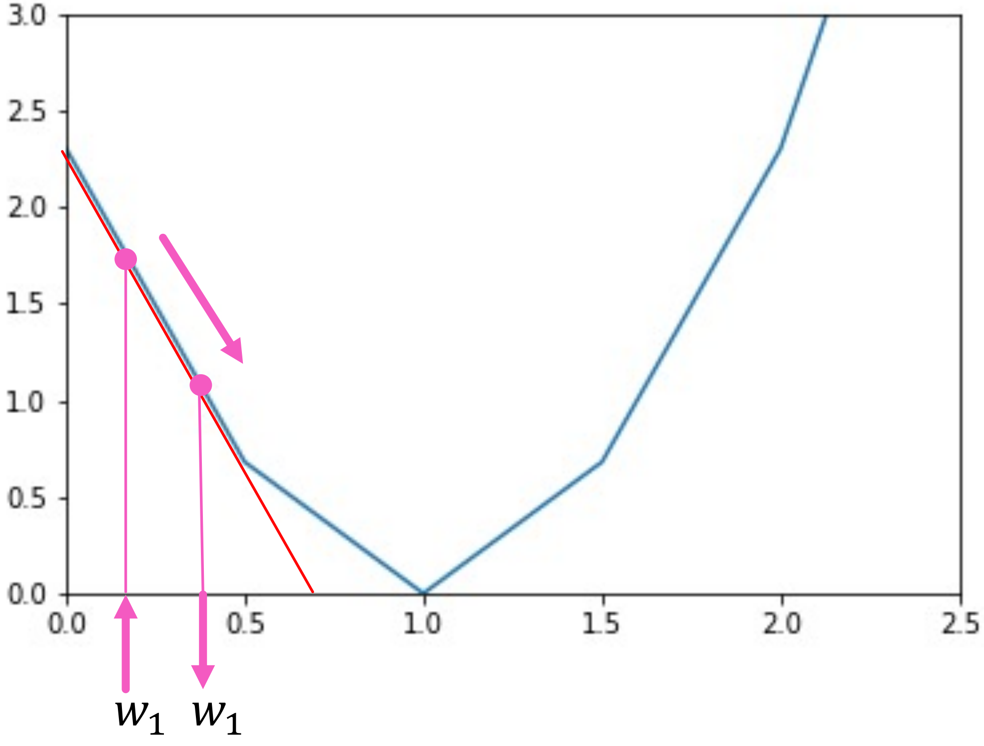
$$\geq 0$$

$$w_1 \coloneqq w_1 - \alpha\,(positif)$$

$$w_1 \leq w_1$$

4. Comprendre la descente de gradient

Fonction J(w_1)
w_0 = 0, x fixé, w_1 est un paramètre



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

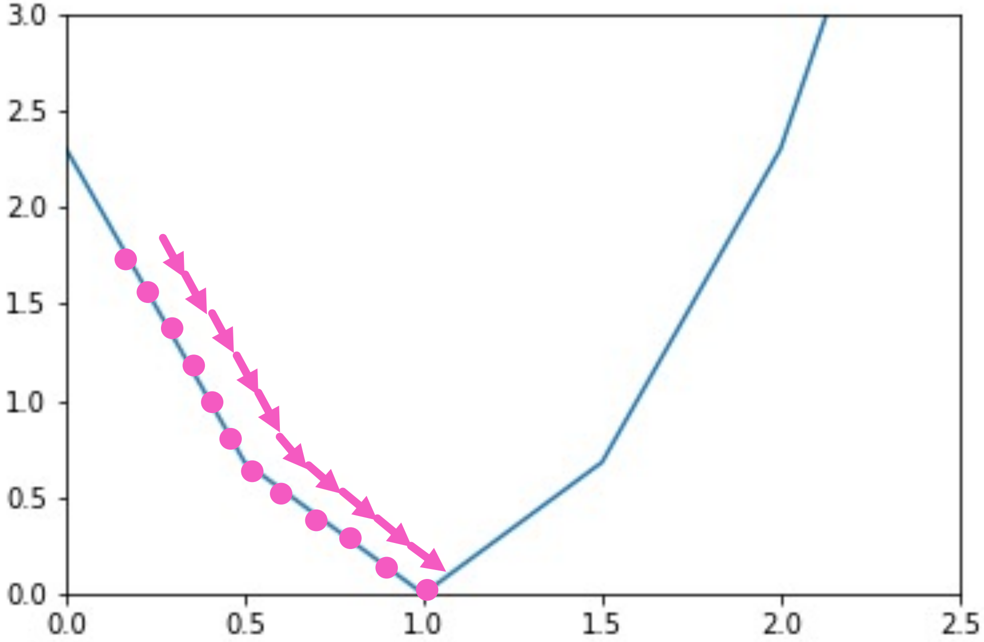
$$\leq 0$$

$$w_1 := w_1 - \alpha \, (negatif)$$

$$w_1 \geq w_1$$

5. L’impact du learning rate

Fonction J(w_1)
w_0 = 0, x fixé, w_1 est un paramètre



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

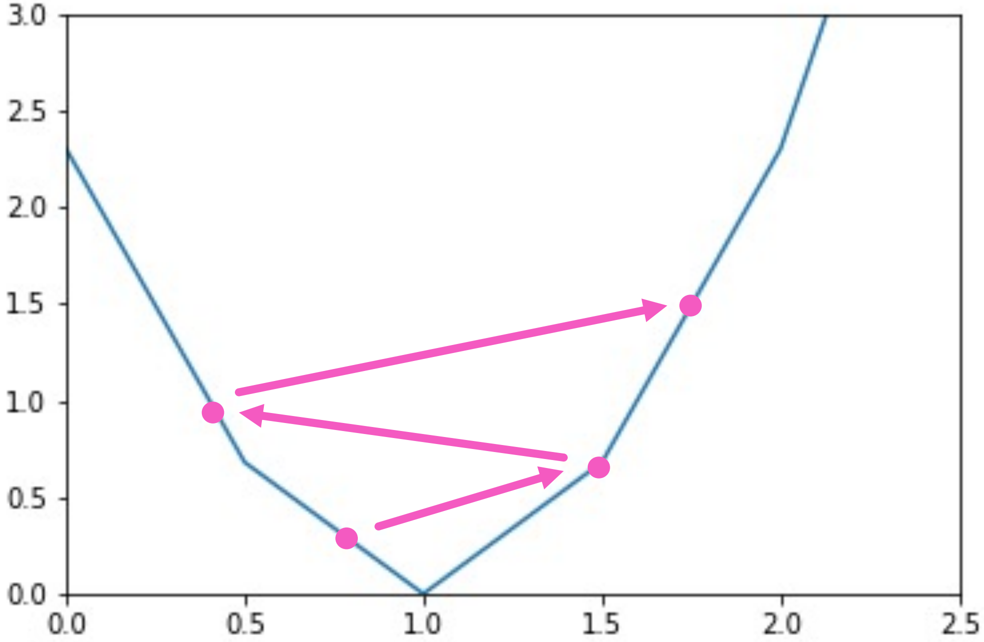
Trop petit

Converge trop lentement

5. L'impact du learning rate

Fonction J(w_1)

w_0 = 0, x fixé, w_1 est un paramètre



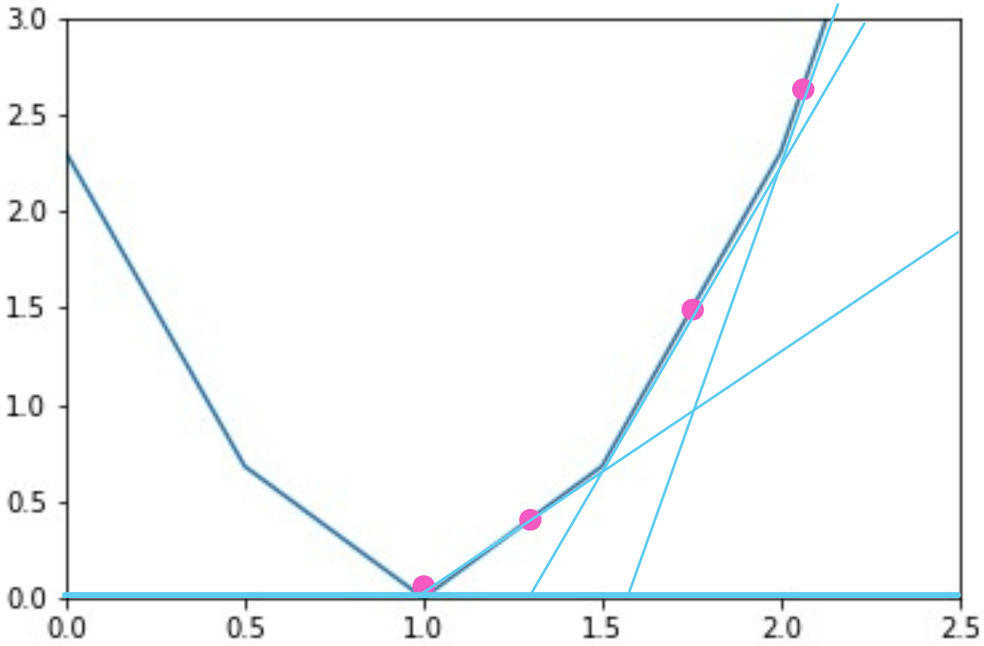
$$w_1 := w_1 - \boxed{\alpha} \frac{\partial}{\partial w_1} J(W)$$

Trop grande

N'arrive pas à converger voir diverge

6. Converger vers le minimum

Fonction J(w_1)
w_0 = 0, x fixé, w_1 est un paramètre

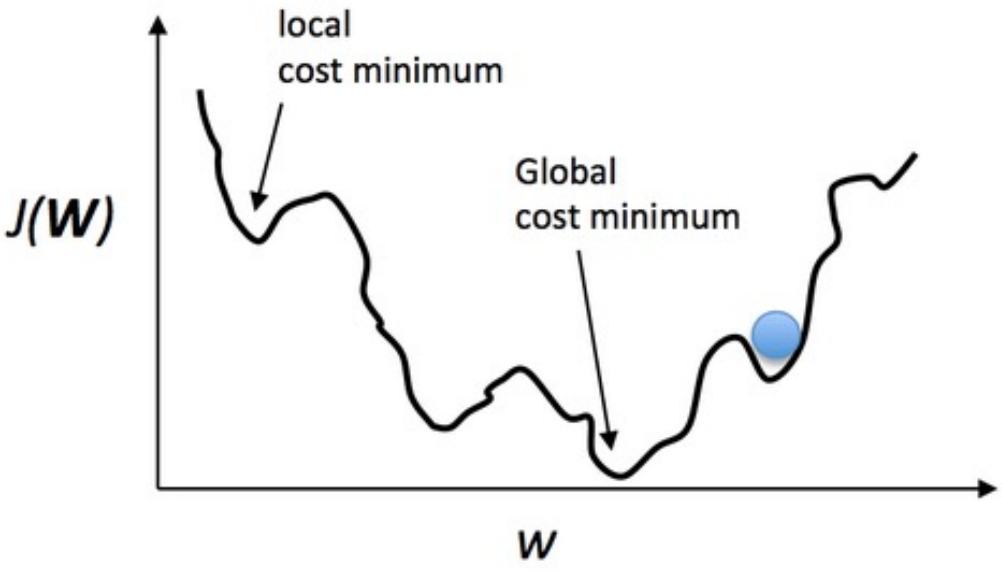


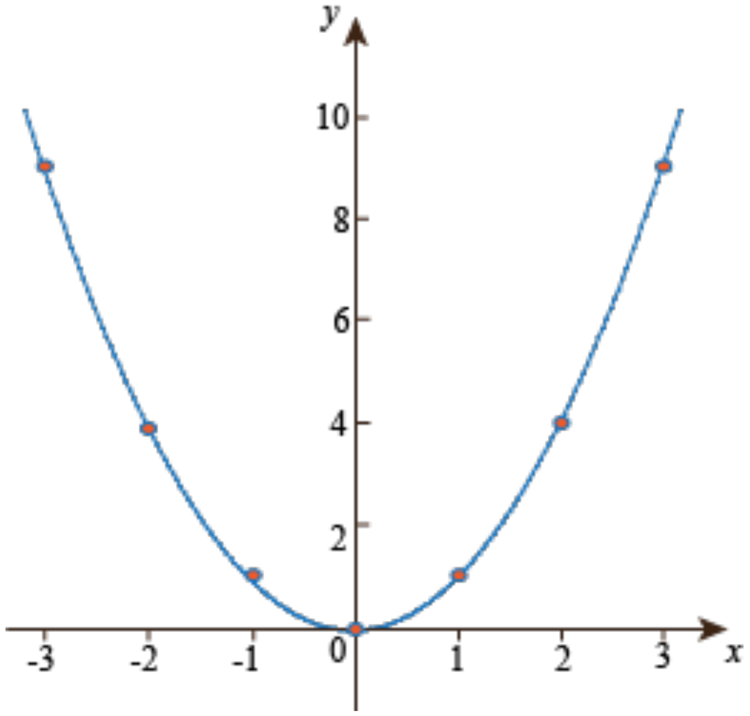
$$w_1 := w_1 - \boxed{\alpha} \boxed{\frac{\partial}{\partial w_1} J(W)}$$

Fixé

Réduit quand il approche du minimum

7. Fonction non convexe





Régression linéaire

1. Hypothèse

$$H(X) = w_0 . x_0 + w_1 . x_1 + w_2 . x_2 + \dots + w_n . x_n$$

$$H(X) = 80 * 1 + 1.5\ x_1 + 10\ x_2 + \dots + -2\ x_n$$

- w_0 est le prix de base d'un appartement sans autres variables.
- w_1 est le coefficient de la taille de l'appartement.
- w_2 est le coefficient du nombre de pièces.
- w_n est le coefficient de la date de construction de la maison.