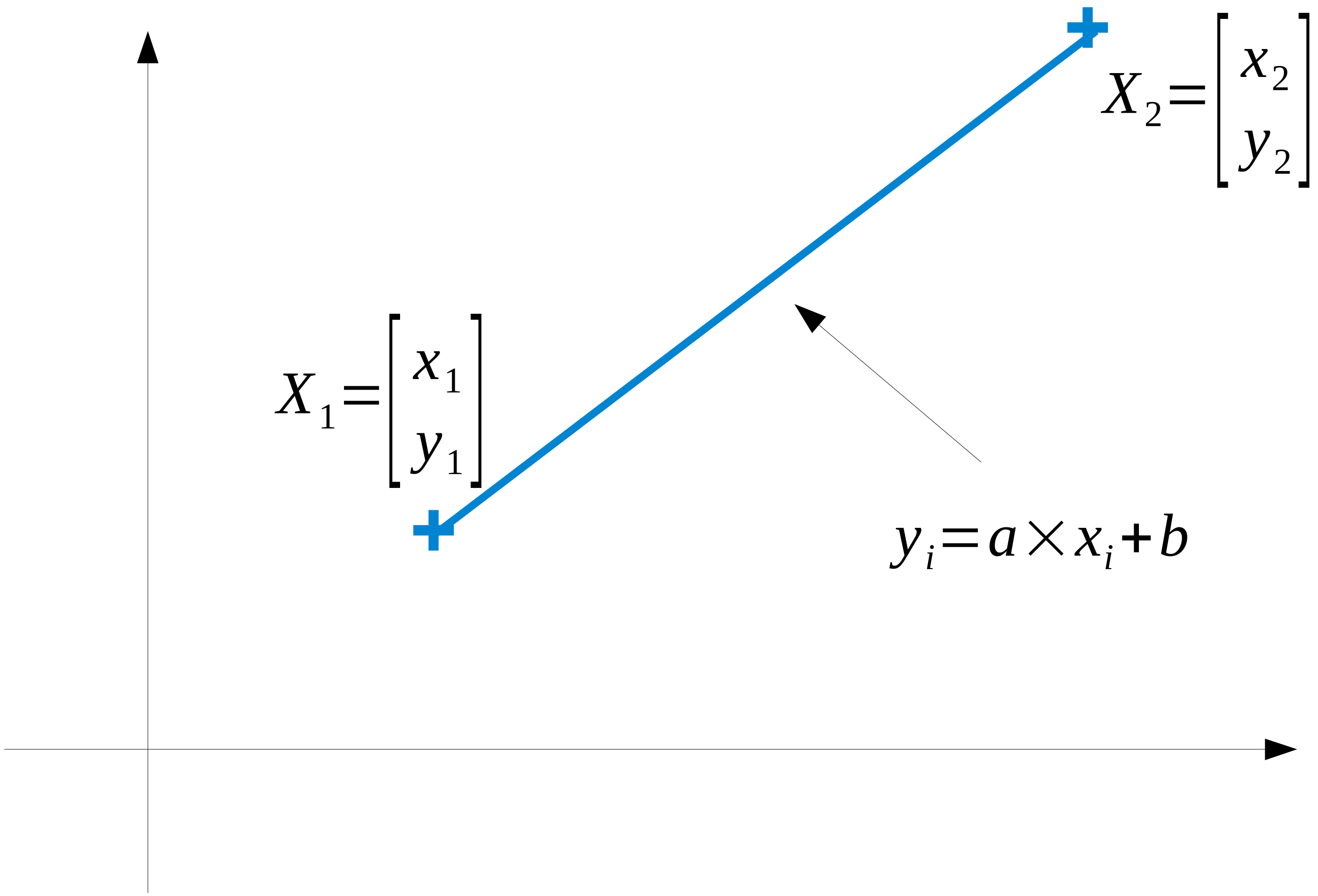
Représentation par Modèles

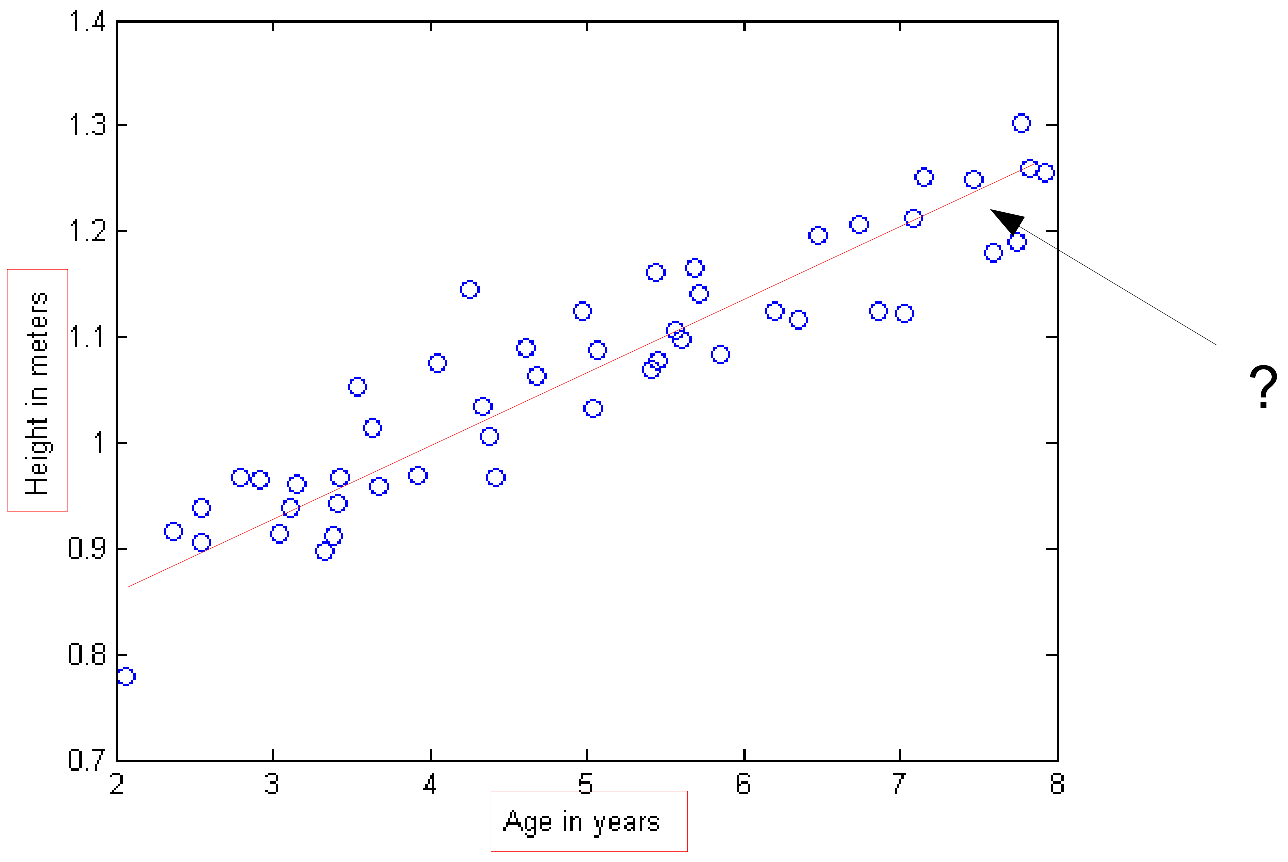
Régression Univariée

Exemple



$X_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$
$X_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$
$y_i = a \times x_i + b$

Définition du problème

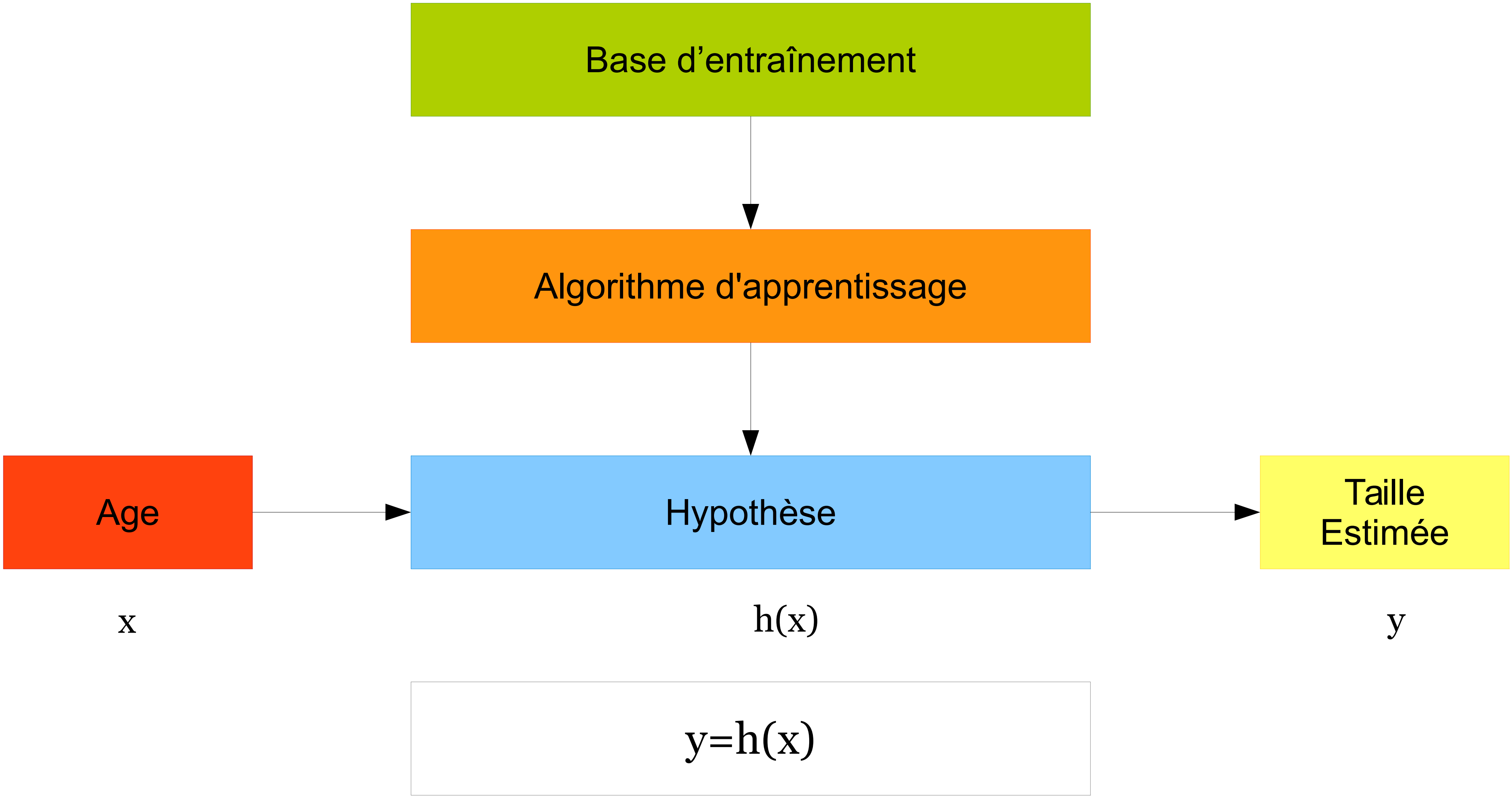


Base d'entraînement

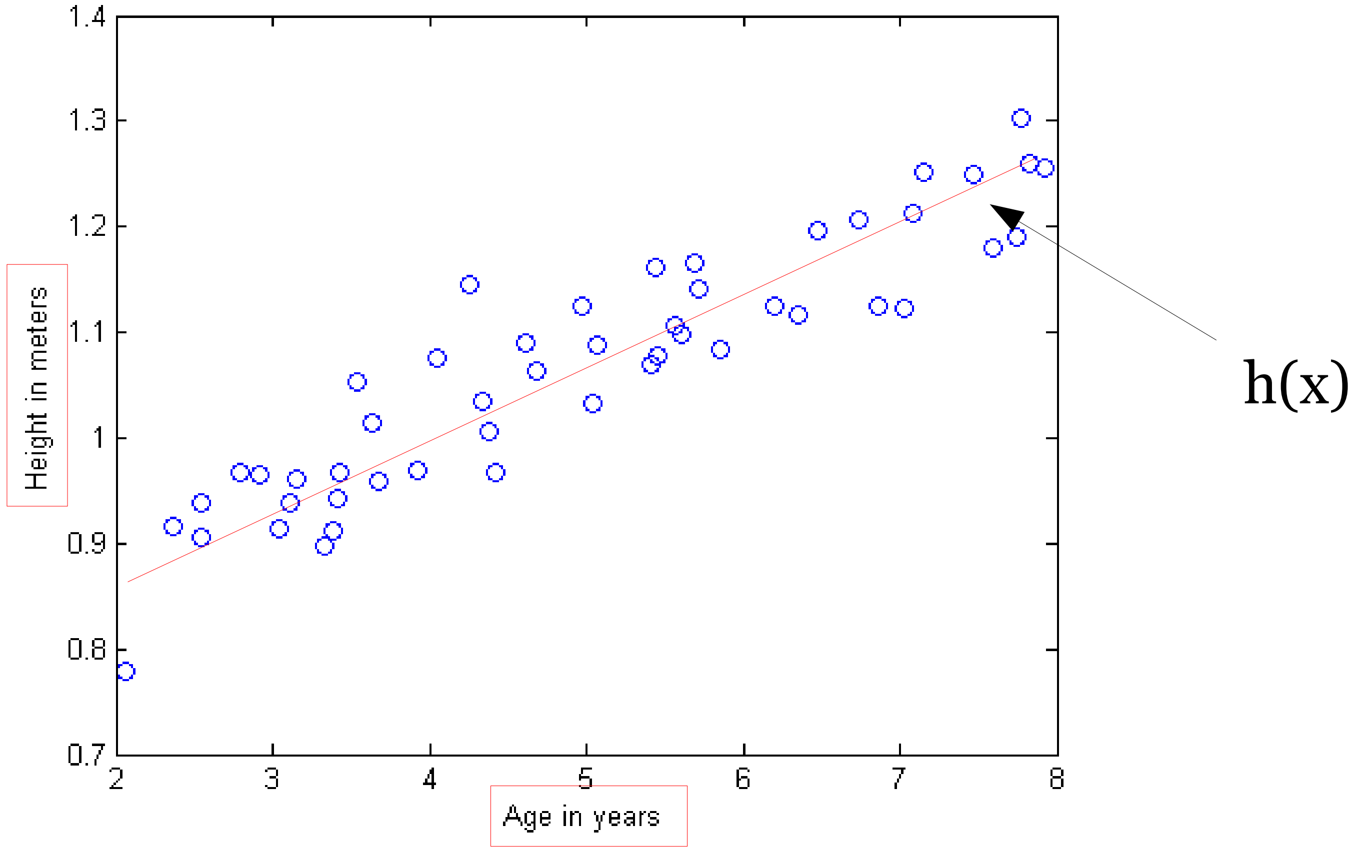
Age (X)	Taille en m (Y)
2.0658746e+00	7.7918926e-01
2.3684087e+00	9.1596757e-01
2.5399929e+00	9.0538354e-01
2.5420804e+00	9.0566138e-01
2.5490790e+00	9.3898890e-01
2.7866882e+00	9.6684740e-01
2.9116825e+00	9.6436824e-01
3.0356270e+00	9.1445939e-01
3.1146696e+00	9.3933944e-01
3.1582389e+00	9.6074971e-01
3.3275944e+00	8.9837094e-01

- Notations :
  - m : nombre d'observations,
  - X : variable d'entrée ou attribut,
  - Y : variable de sortie ou objectif,
  - (X,Y) : une observation,
  - (X_i, Y_i) : i^ième^ observation.
- Exemple
  - X_2 =2,3684087,
  - Y_{10} =0,96074971.
  - (X_3, Y_3)=(2.5399929,0.90538354)

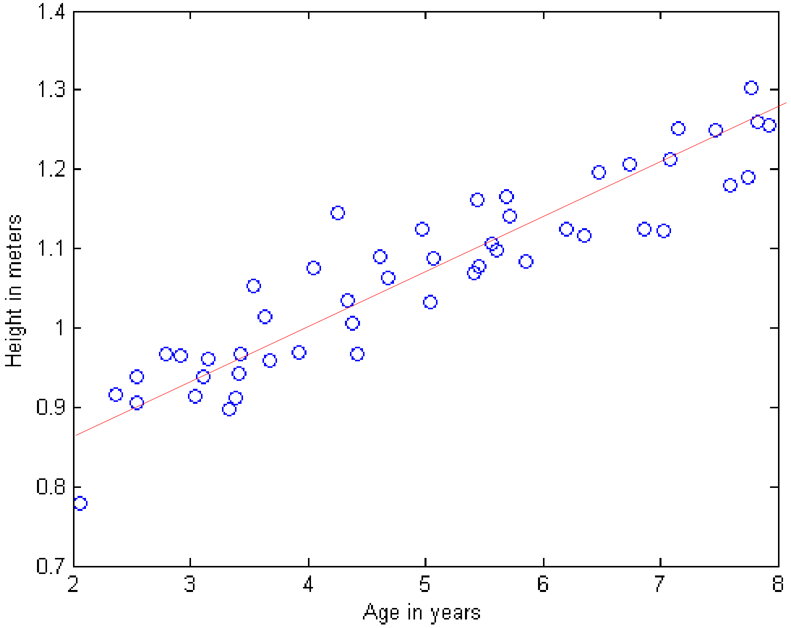
Apprentissage



Définition du problème



Définition du problème



$$h_\theta(X) = \theta_0 + \theta_1 \times x$$

$$h_\theta(X) = X^T\theta$$

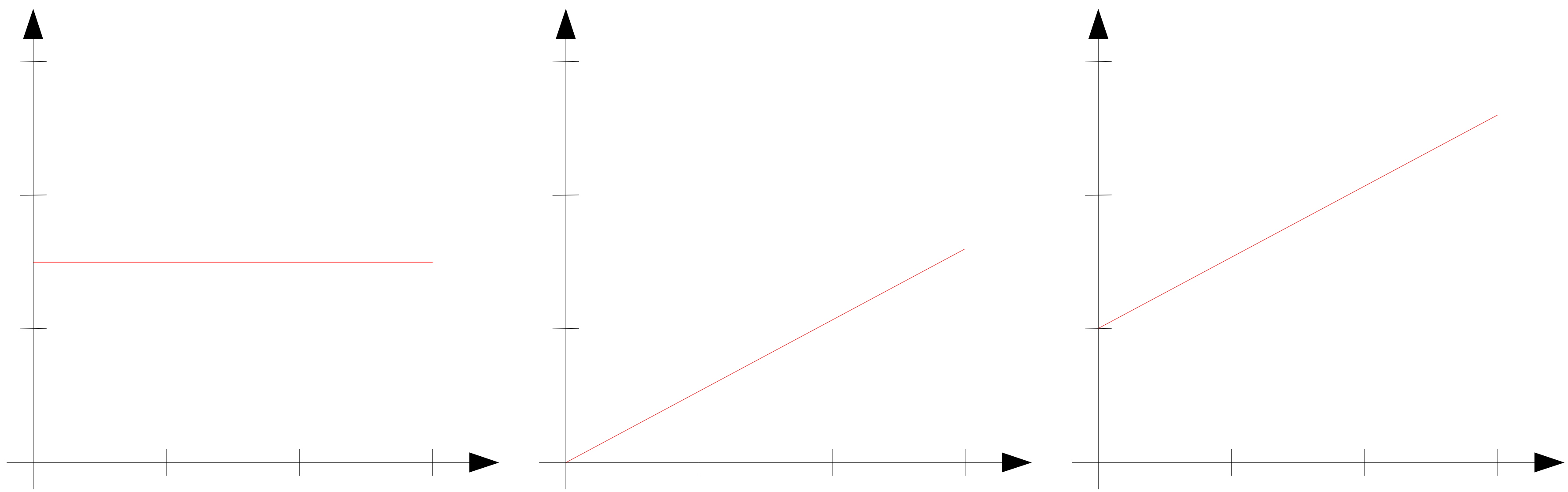
$$\theta = [\theta_0 \ \theta_1]^T$$

$$X = [1 \ x]^T$$

$$\theta \ ?$$

Définition du problème

$$h(X) = \theta_0 + \theta_1 \times x$$

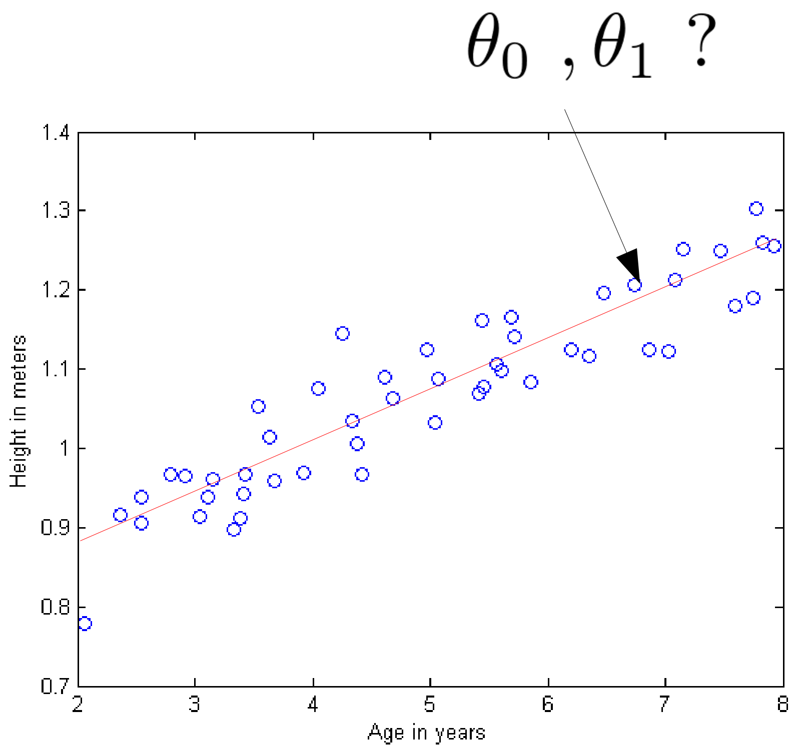


$\theta_0 = 1,5$
$\theta_1 = 0$

$\theta_0 = 0$
$\theta_1 = 0,5$

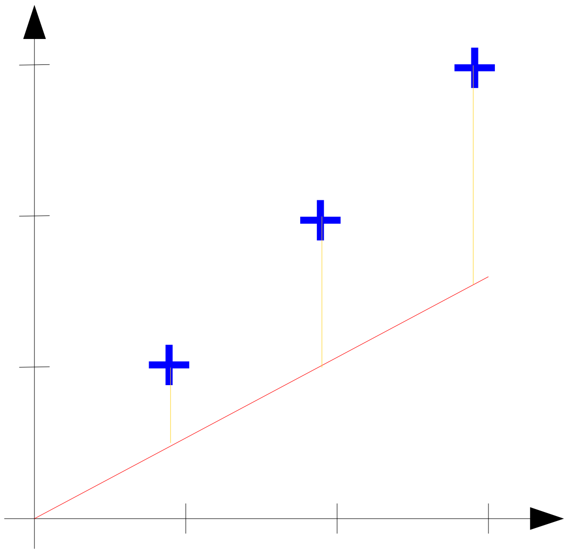
$\theta_0 = 1$
$\theta_1 = 0,5$

Intuition



- Trouver θ de façon à ce que *h(X)* soit proche de *Y* pour nos exemples d'entraînement.

Fonction de coût



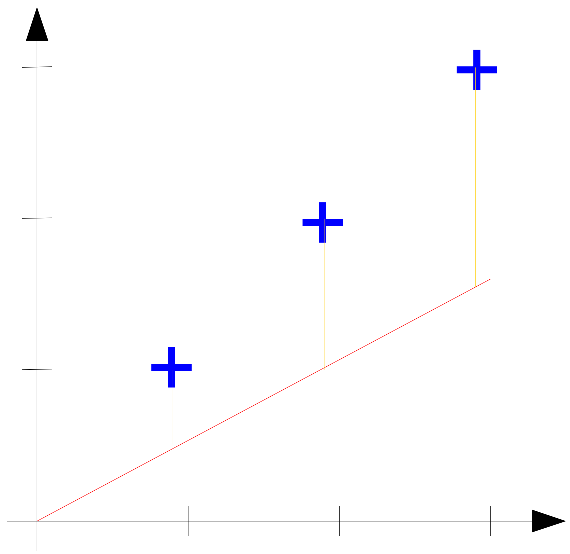
$$h_\theta(X) - Y$$

$$\frac{1}{m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)$$

$$\frac{1}{m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)^2$$

$$\min_{\theta_0,\theta_1} \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)^2$$

Fonction de coût



$$\min_{\theta_0,\theta_1} \; \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)^2$$

$$h_\theta(X_i) = \theta_0 + \theta_1 \times x_i^1$$

$$J(\theta_0, \theta_1) = \; \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)^2$$

$$\min_{\theta_0,\theta_1} \; J(\theta_0, \theta_1)$$

Fonction de Coût

Contexte

- Hypothèse :

$$h_\theta(X) = \theta_0 + \theta_1 \times x$$

- Paramètres :

$$\theta_0 \ \ , \theta_1$$

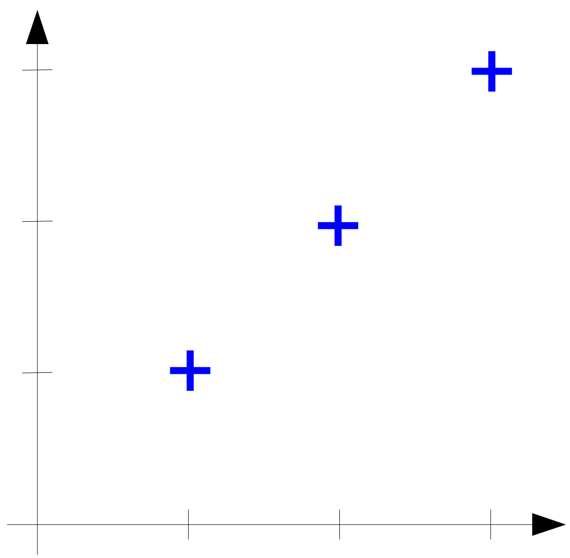
- Fonction de coût (fonctionnelle) :

$$J(\theta_0, \theta_1) = \ \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

- Objectif :

$$\min_{\theta_0, \theta_1} \ J(\theta_0, \theta_1)$$

Simplification



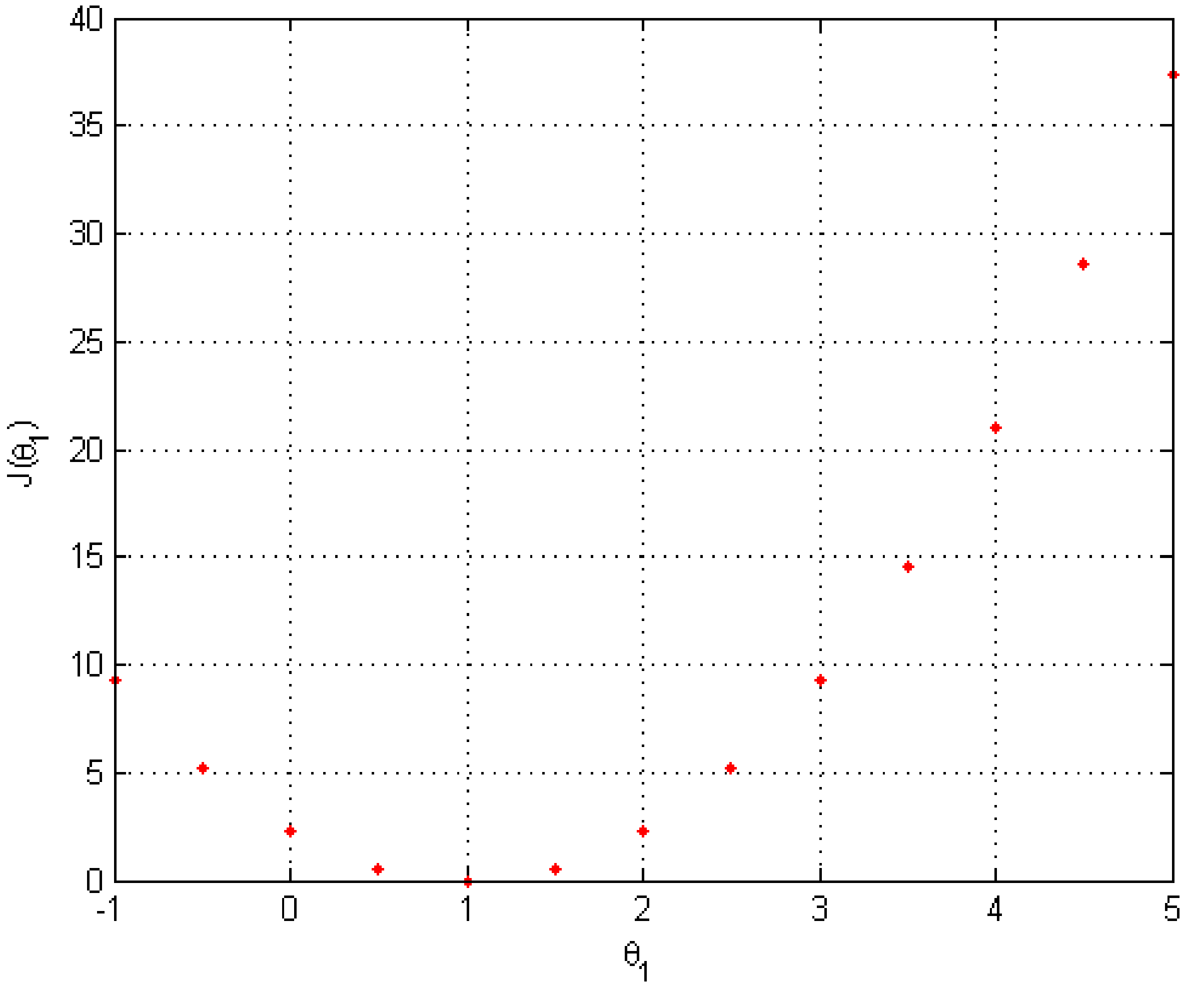
$$\theta_0 = 0$$

$$h_\theta(X) = \theta_1 \times x$$

$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(X_i) - Y_i)^2$$

$$\min_{\theta_1} \; J(\theta_1)$$

$J(\theta_1)$



Contexte

- Hypothèse :

$$h_\theta(X) = \theta_0 + \theta_1 \times x$$

- Paramètres :

$$\boxed{\theta_0 \ \ , \theta_1}$$

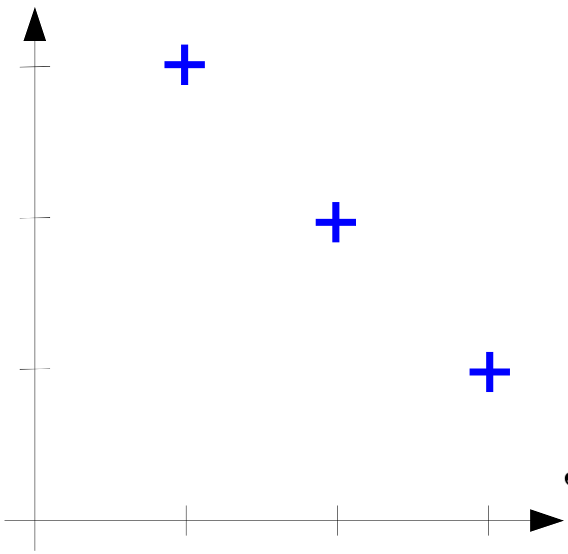
- Fonction de coût (fonctionnelle) :

$$J(\theta_0, \theta_1) = \ \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

- Objectif :

$$\min_{\theta_0, \theta_1} \ J(\theta_0, \theta_1)$$

Résolution



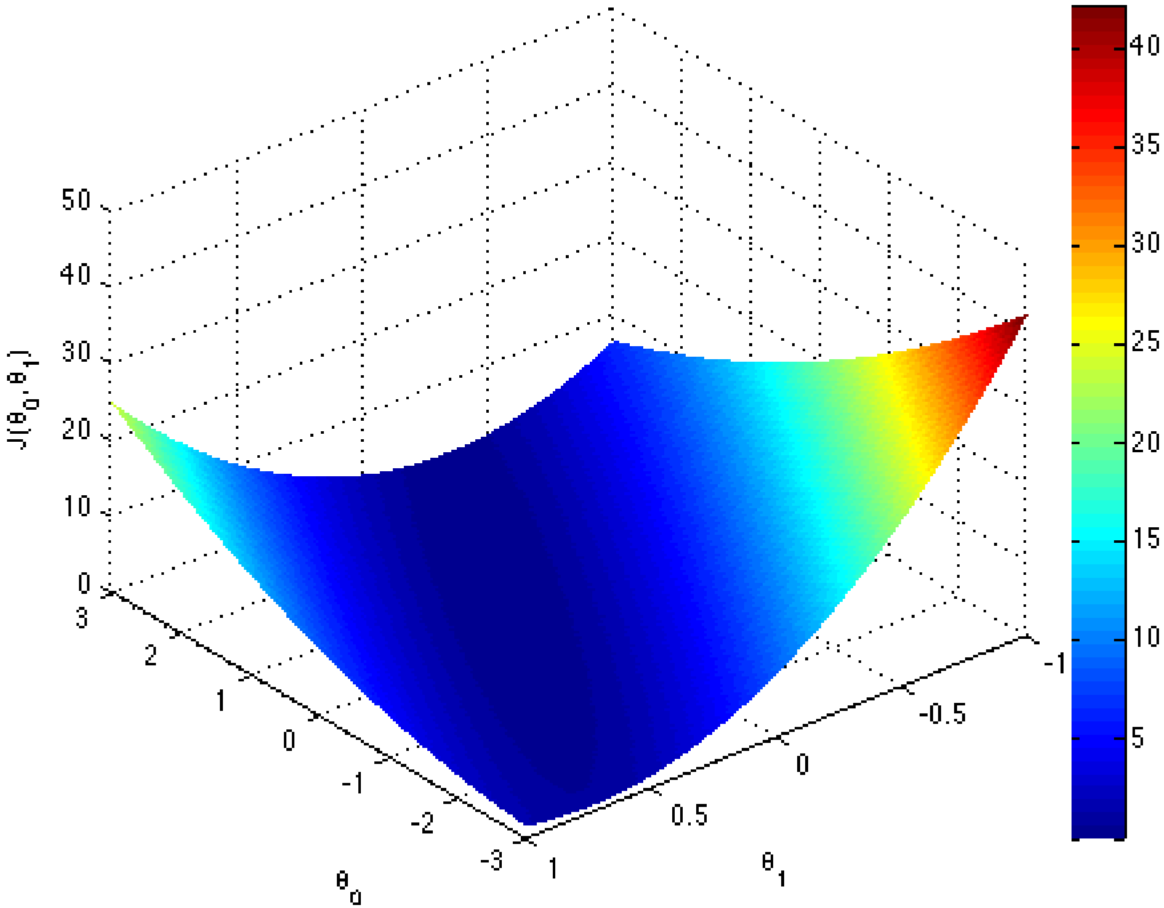
$$\theta_0 \ , \theta_1$$

$$h_\theta(X) = \theta_0 + \theta_1 \times x$$

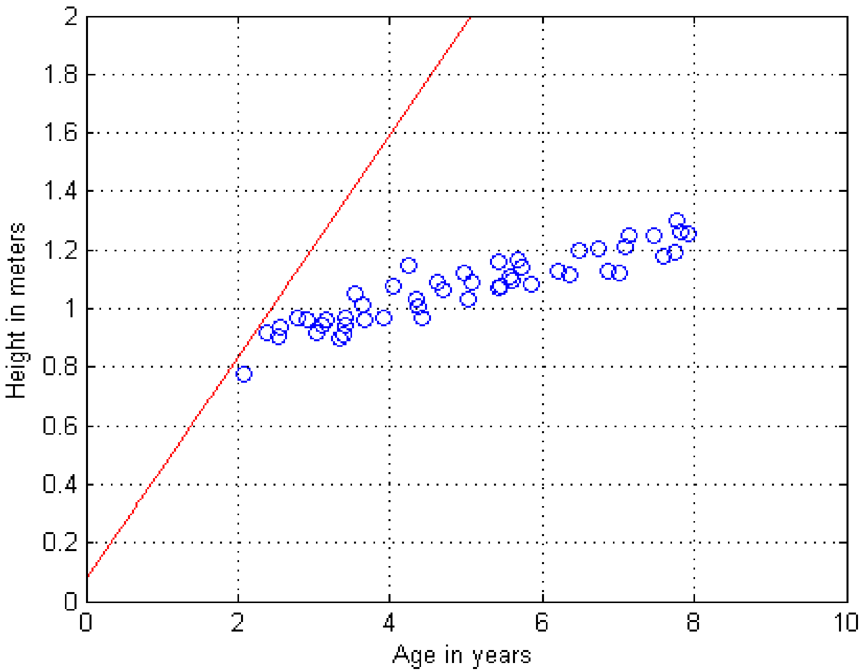
$$J(\theta_0, \theta_1) = \ \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

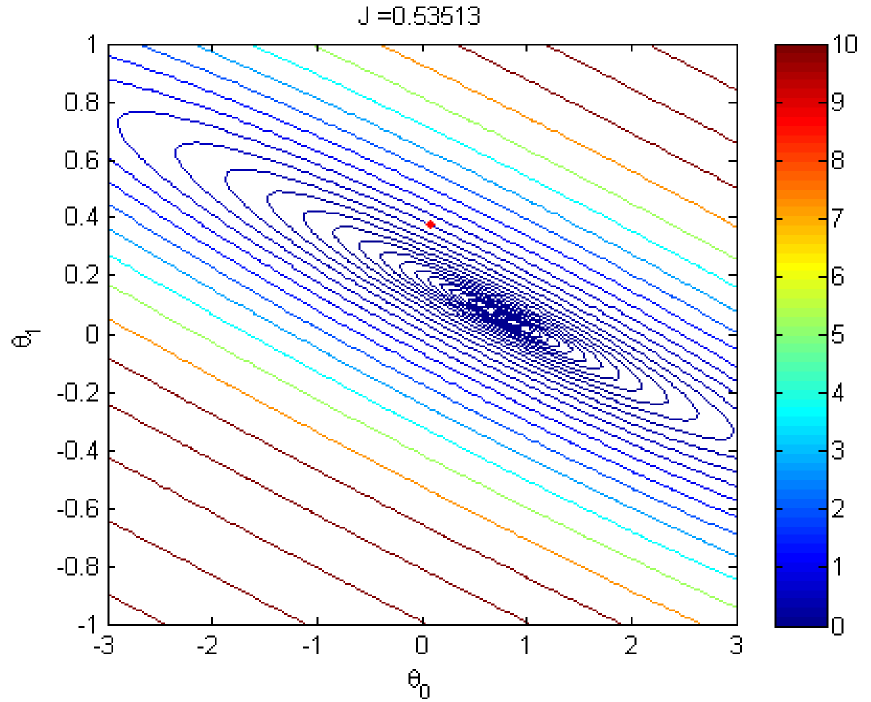
$$\min_{\theta_0, \theta_1} \ J(\theta_0, \theta_1)$$

$J(\theta_o,\theta_1)$

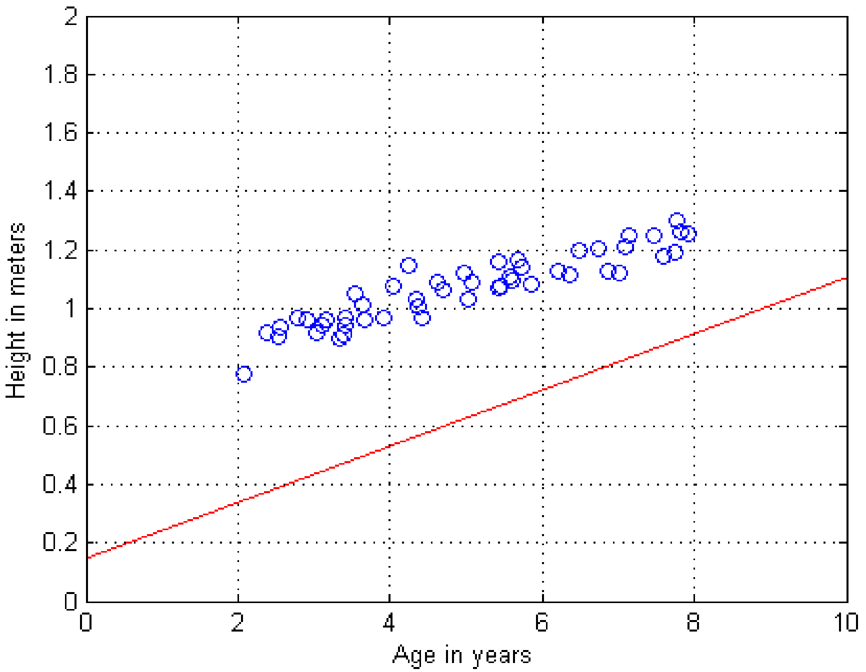


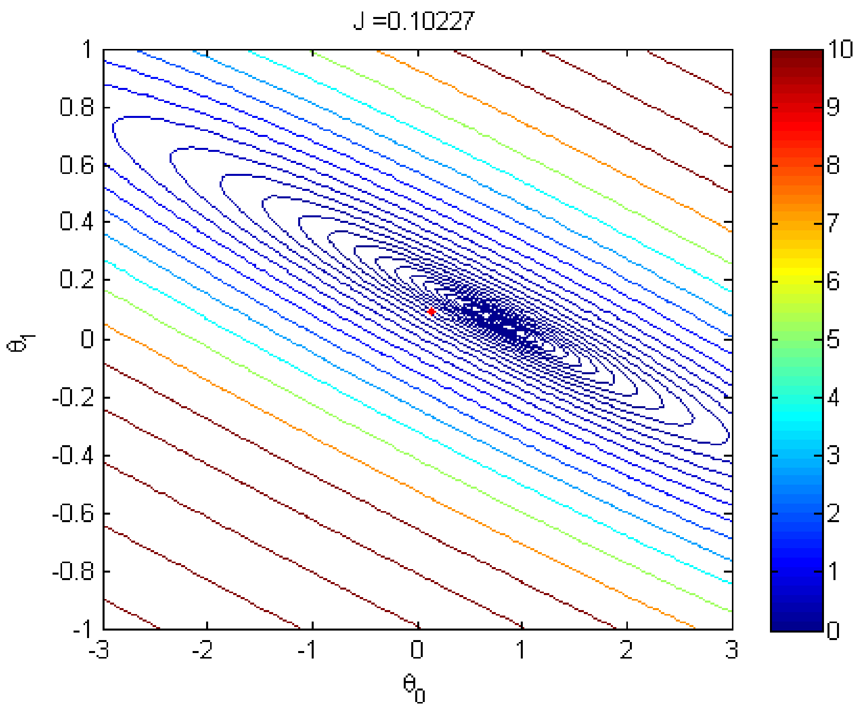
Recherche de solution



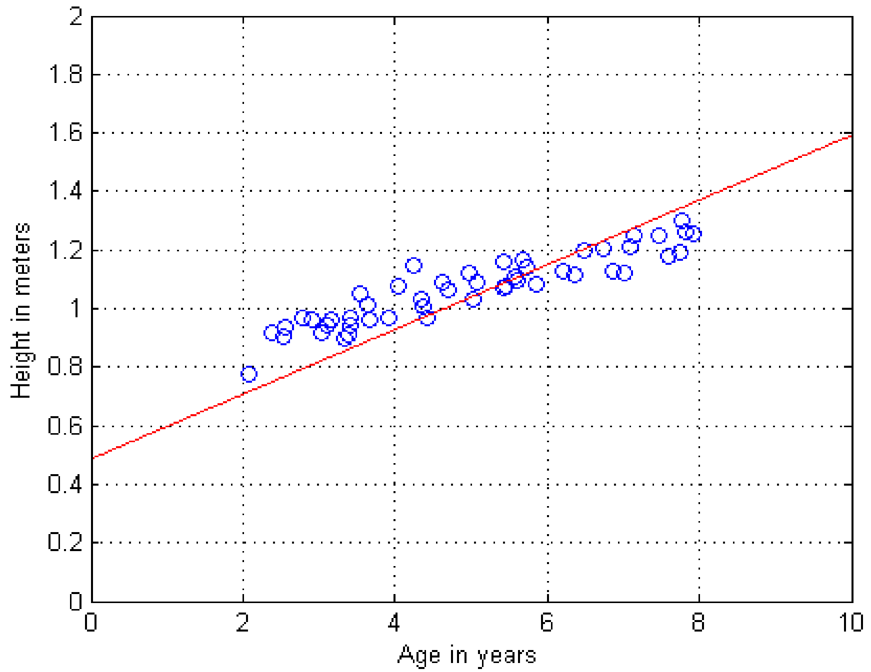


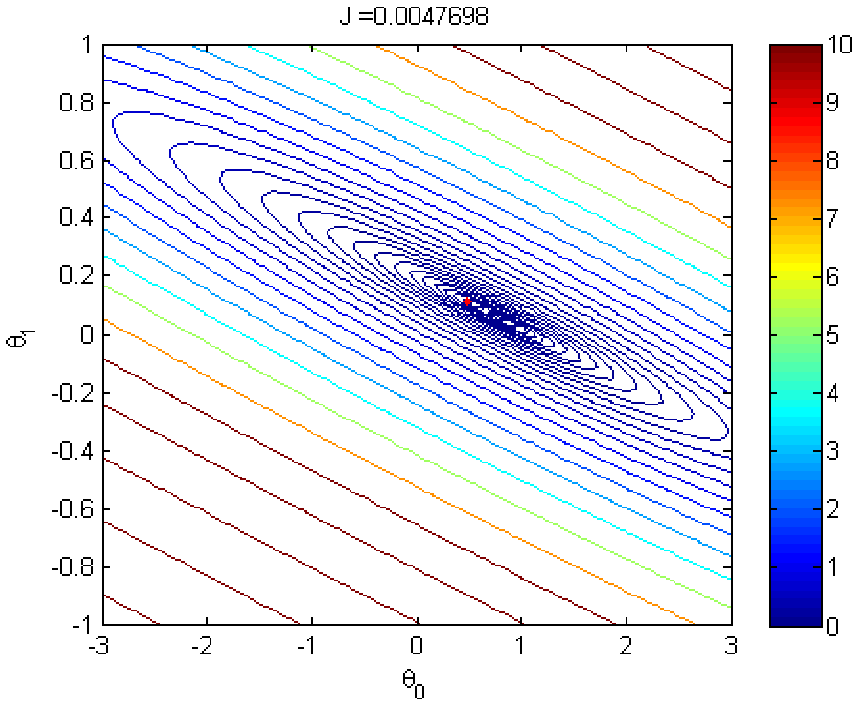
Recherche de solution



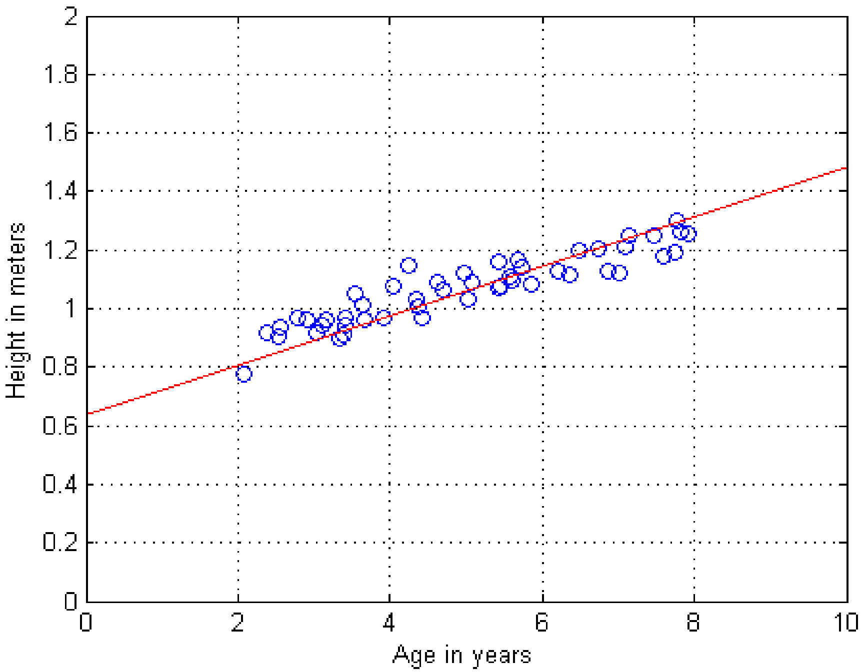


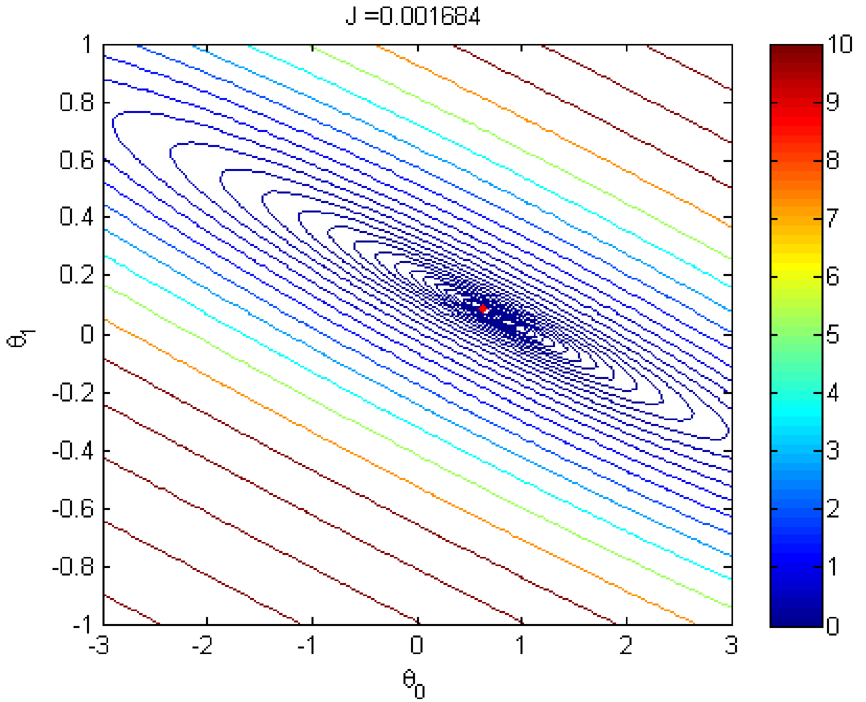
Recherche de solution





Recherche de solution





Méthode des Moindres Carrés

Contexte

- Une fonction :

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

- Un objectif :

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Comment ?

Solution (1)

$$Y = X^T \theta = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$J(\theta_0, \theta_1) = \frac{1}{2m} \boxed{\sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2}$$

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Solution (2)

$$Y = X^T \theta = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$J(\theta_{0,} \theta_1) = \frac{1}{2m} \sum_1^m (h_\theta(X_i) - Y_i)^2 = \frac{1}{2m} (Y - X^T \theta)^T (Y - X^T \theta)$$

$$J(\theta_{0,} \theta_1) \approx (Y - X^T \theta)^T (Y - X^T \theta)$$

$$\min_{\theta_{0,} \theta_1} J(\theta_{0,} \theta_1) \Rightarrow \frac{\partial J(\theta_{0,} \theta_1)}{\partial \theta} = 0$$

Solution (3)

$$J(\theta_{0,}\theta_1) \approx (Y - X^T\theta)^T (Y - X^T\theta)$$

$$\min_{\theta_{0,}\theta_1} J(\theta_{0,}\theta_1) \Rightarrow \frac{\partial J(\theta_{0,}\theta_1)}{\partial\theta} = 0$$

$$\theta = (XX^T)^{-1} XY$$

Descente de Gradient

Contexte

- Une fonction :

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

- Un objectif :

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

Comment ?

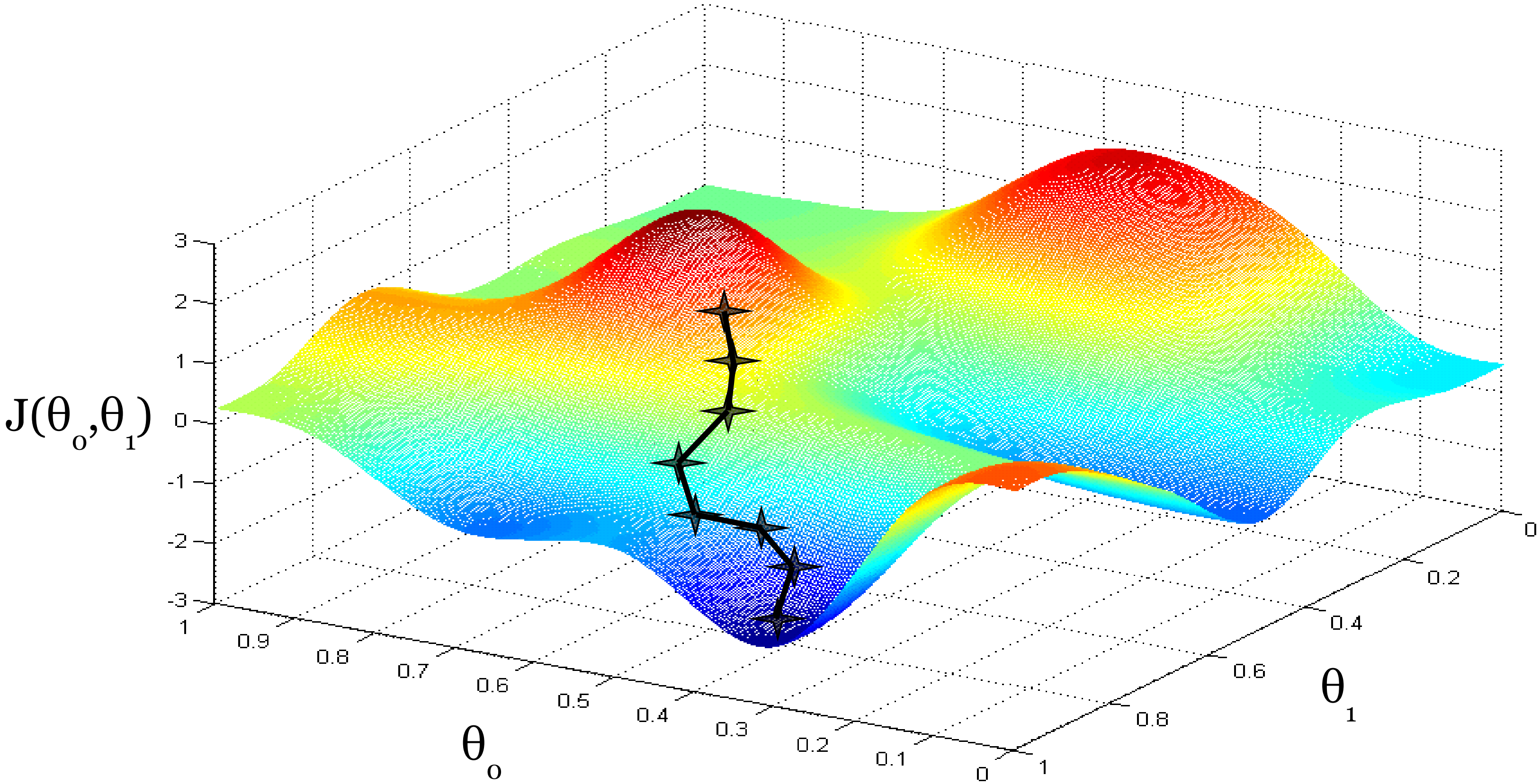
Idée

- Démarrer avec :

$$\theta = [\theta_0 \ \theta_1]^T$$

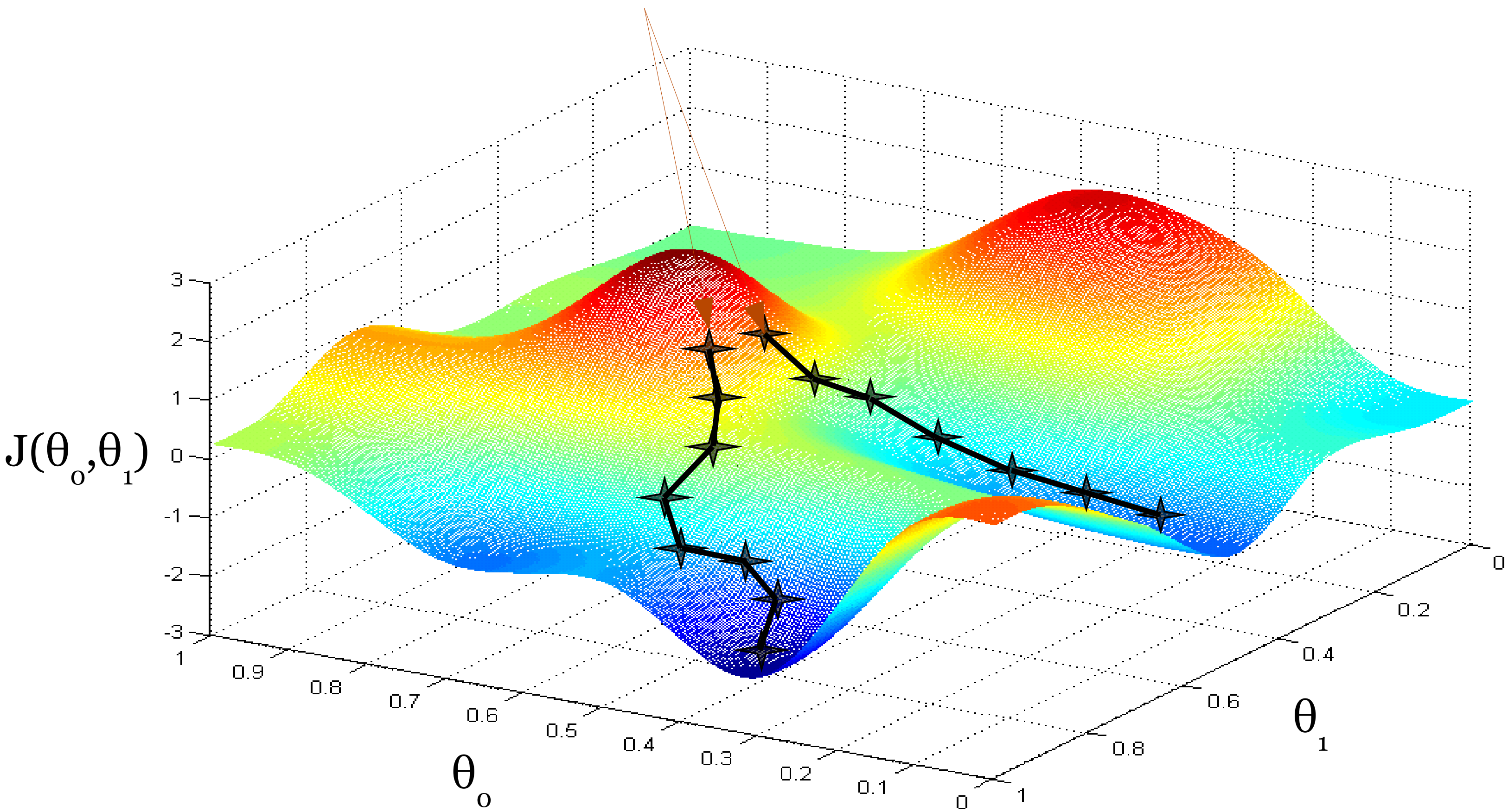
- Changer $\theta = [\theta_0 \ \theta_1]^T$ en espérant que l'on atteigne un minimum.

Idée



Source : Andrew Ng

Initialisation



Source : Andrew Ng

Algorithme de descente de gradient

Tant que convergence {

$$\theta_i = \theta_i - \boxed{\alpha} \boxed{\frac{\partial}{\partial \theta_i} J(\theta)}$$

Dérivée partielle

}

Pas de gradient (learning rate)

$$Pour\ \theta = [\theta_0\ \theta_1]^T$$

$$tmp0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta)$$

$$tmp1 = \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta)$$

$$\theta_0 = tmp0$$

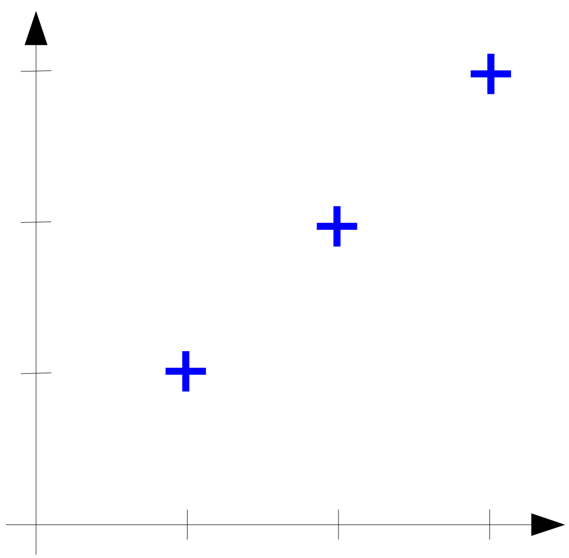
$$\theta_1 = tmp1$$

Simplification

Tant que convergence {

$$\theta_i = \theta_i - \alpha \frac{\partial}{\partial \theta_i} \; J(\theta)$$

}



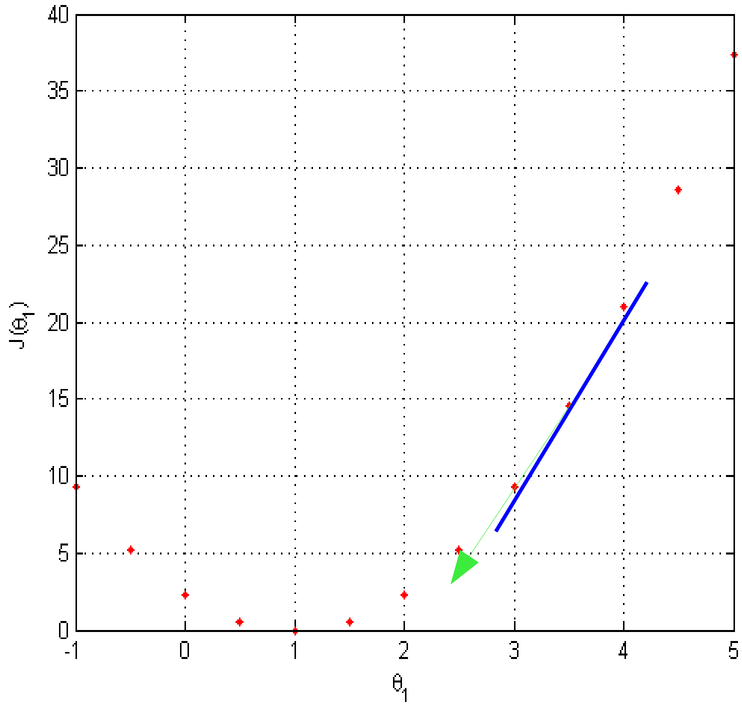
$$\theta_0 = 0$$

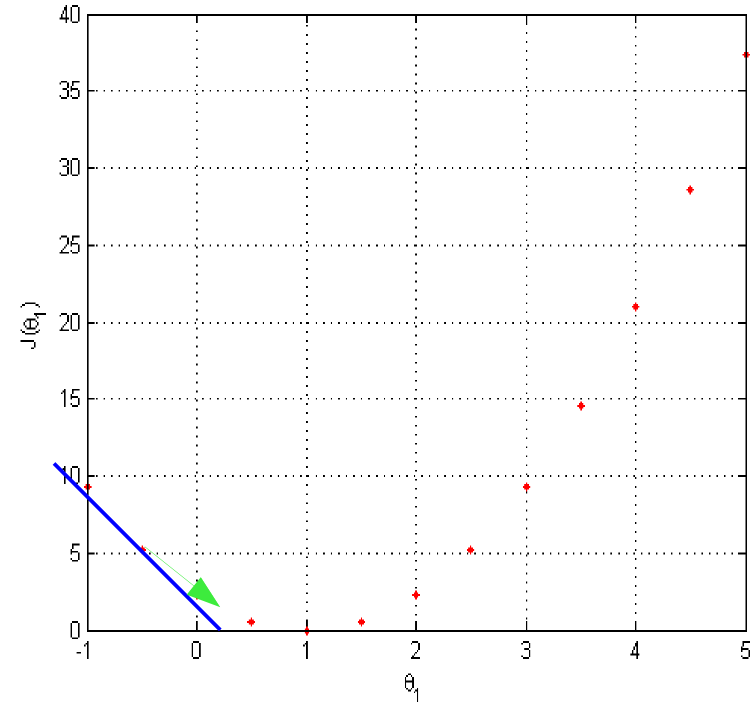
$$h_\theta(X) = \theta_1 \times x$$

$$J(\theta_1) = \; \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(X_i) - Y_i)^2$$

$$\min_{\theta_1} \; J(\theta_1)$$

$J(\theta_1)$





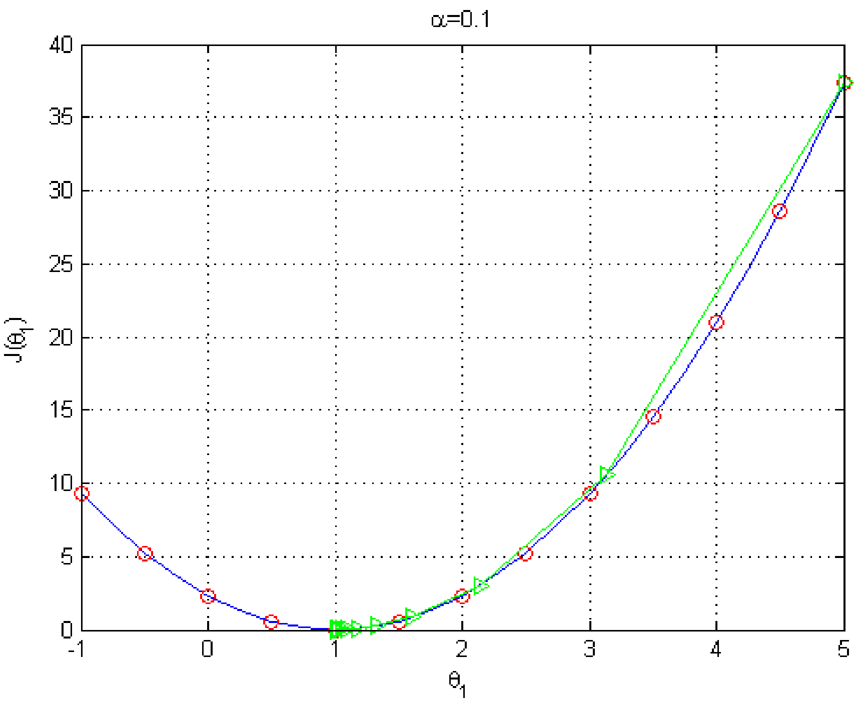
$$\theta_i = \theta_i - \alpha \boxed{\frac{\partial}{\partial \theta_i} J(\theta)}$$

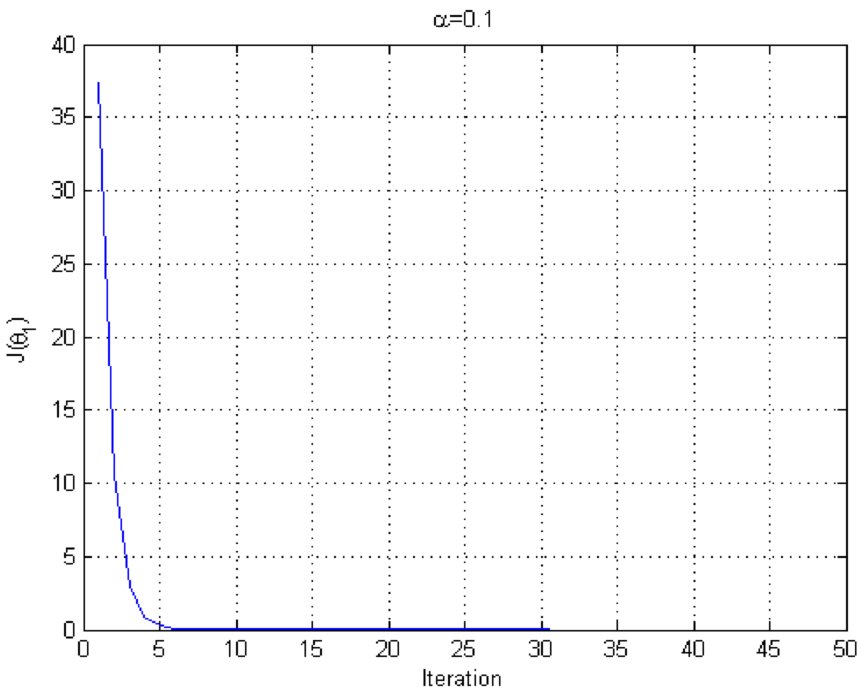
$$\boxed{> 0}$$

$$\theta_i = \theta_i - \alpha \boxed{\frac{\partial}{\partial \theta_i} J(\theta)}$$

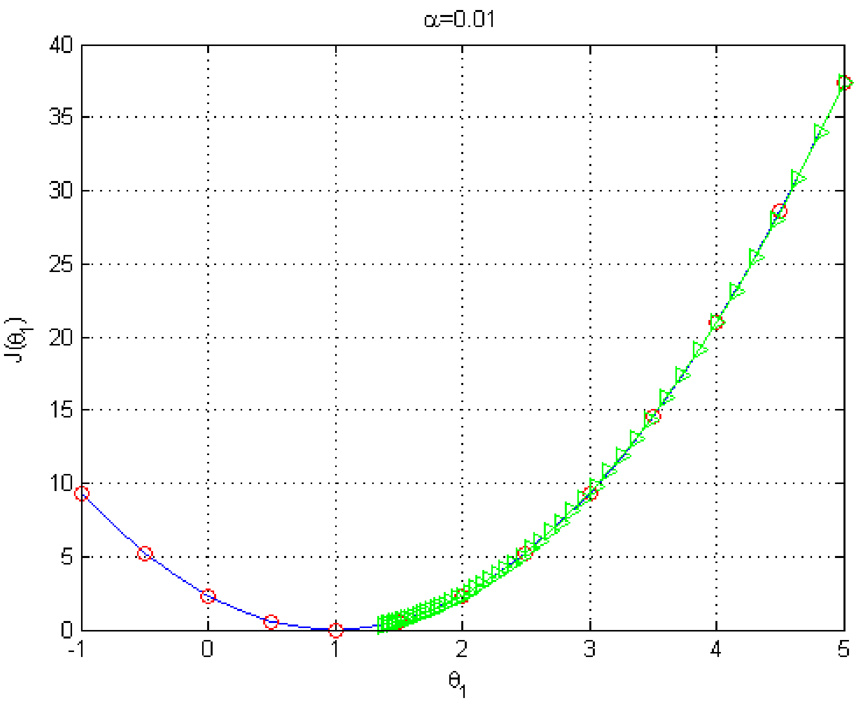
$$\boxed{< 0}$$

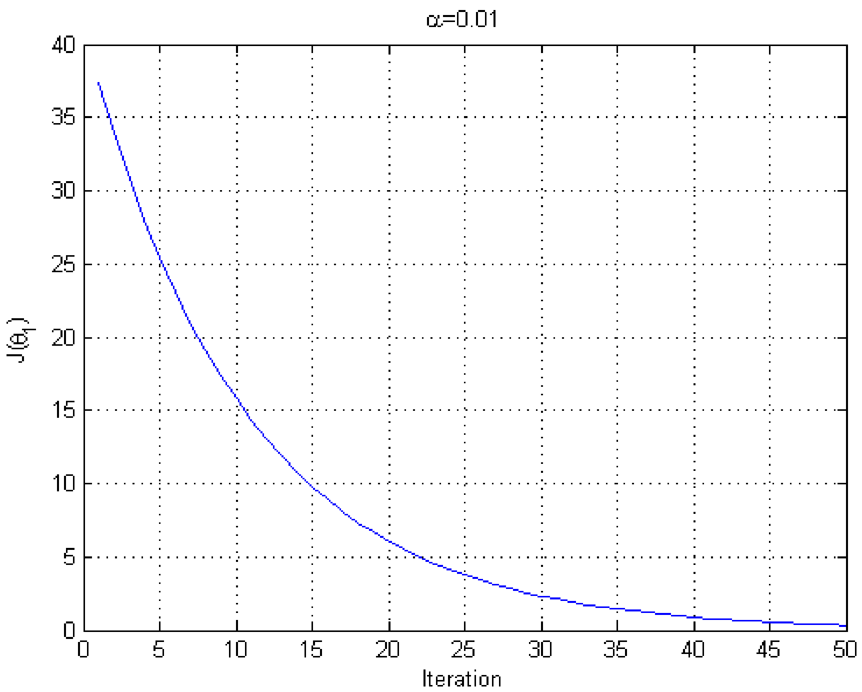
Alpha



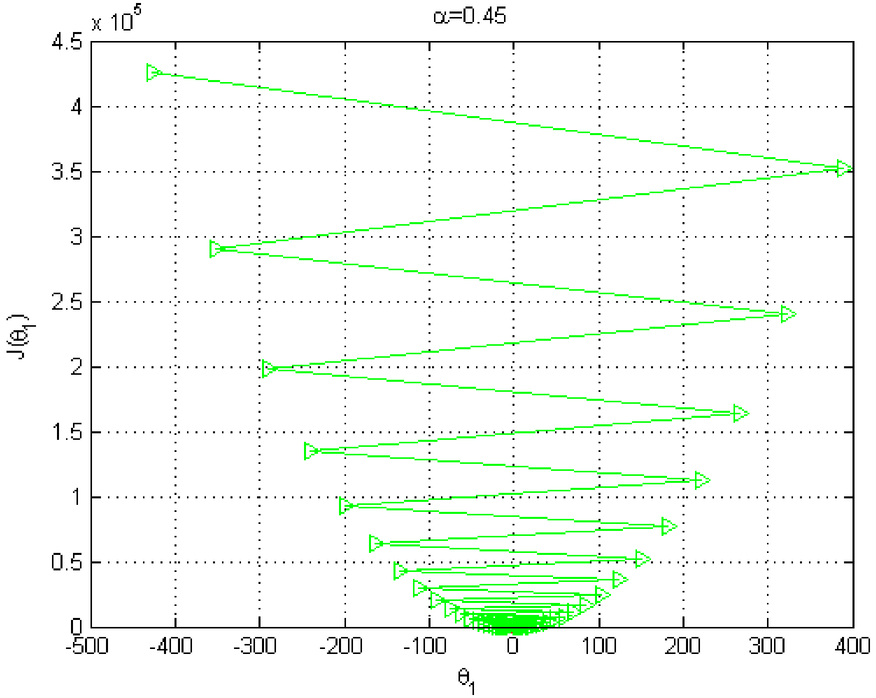


Alpha

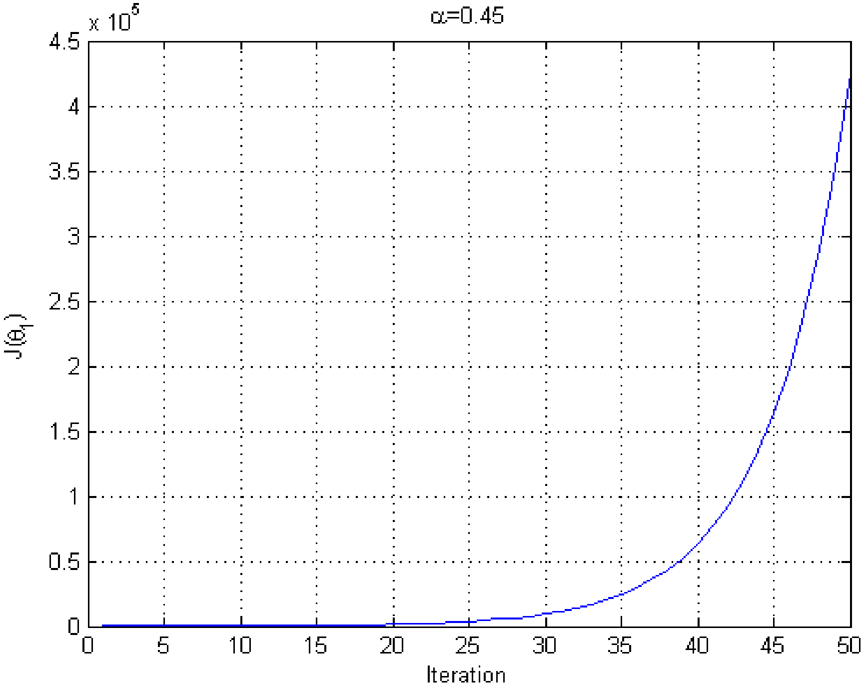




Alpha



α=0.45
x 10^5
J(θ1)
θ1
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
-500
-400
-300
-200
-100
0
100
200
300
400



α=0.45
x 10^5
J(θ1)
Iteration
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
5
10
15
20
25
30
35
40
45
50

Merci de votre Attention

Questions