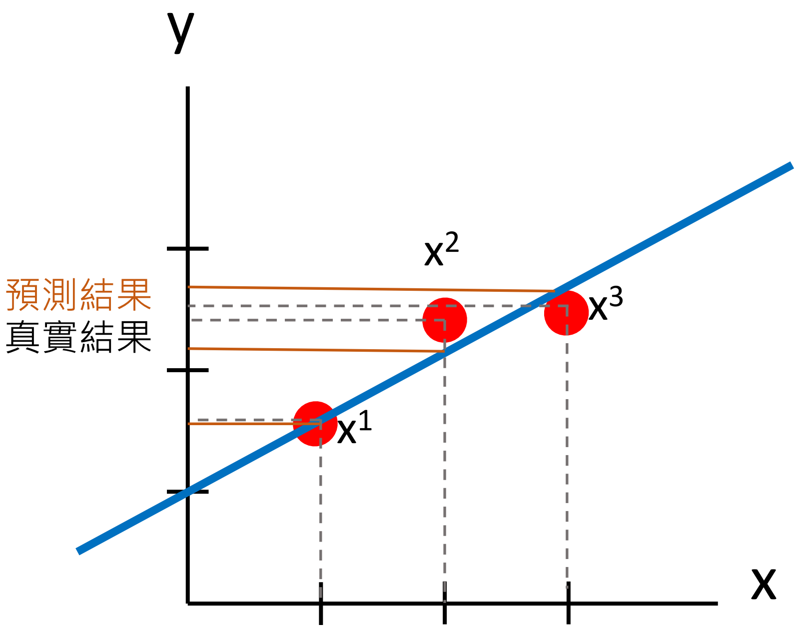
目標函數Cost Function

損失函數（Loss Function ）是定義在單個樣本上的，算的是一個樣本的誤差。
代價函數（Cost Function ）是定義在整個訓練集上的，是所有樣本誤差的平均，也就是損失函數的平均。（也被稱作經驗風險）
目標函數（Object Function）定義為：最終需要優化的函數。等於經驗風險+結構風險（也就是代價函數 + 正則化項）。代價函數最小化，降低經驗風險，正則化項最小化降低。

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

目標函數 loss function



$$y = ax + b \begin{cases} ax_1 + bx_0\text{ , } x_0=1 \\ ax_1\text{, } b=0 \end{cases}$$

$$y = \theta^T x$$

$$y = h\theta(x)$$

函數hθ(x)的執行結果
=>即 x*θ 所得到的結果
=>預測值

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{h_\theta(x_i)}_{\text{預測結果}} - \underbrace{y_i}_{\text{真實結果}})^2$$

特徵(欄位)

每一特徵的θ

預測結果

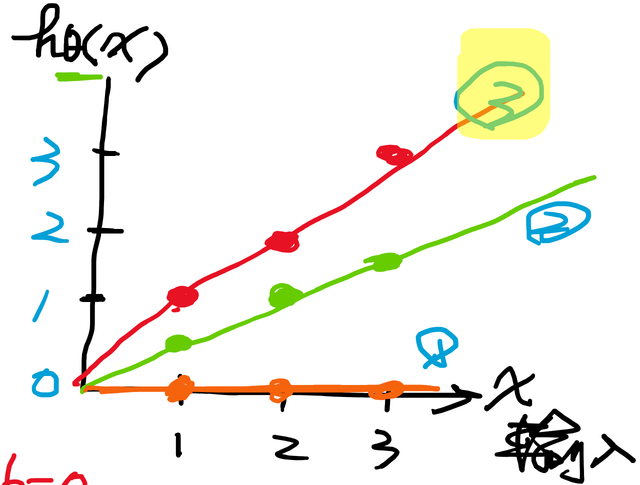
資料列

$$X=\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ x_1^{(i)} & x_2^{(i)} & x_3^{(i)} & \cdots & x_n^{(i)} \end{bmatrix},\ \theta=\begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix},\ y=\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(i)} \end{bmatrix}$$

$$i \times n \qquad\qquad n \times 1$$

結構相同

結果



$b=0$

$x^1=1,\ y^1=1$ ✓

$x^2=2,\ y^2=2$

$x^3=3,\ y^3=3$

最小值

① 设 $\theta=0$

$$J(\theta)\rightarrow J(0)=\frac{1}{2\cdot 3}\left((0\cdot 1-1)^2+(0\cdot 2-2)^2+(0\cdot 3-3)^2\right)$$

$$=\frac{1}{6}(1+4+9)=\frac{14}{6}=\frac{7}{3}$$

② 设 $\theta=\frac{1}{2}$

$$J(\theta)\rightarrow J\left(\tfrac{1}{2}\right)=\frac{1}{6}\left(\left(\tfrac{1}{2}\cdot 1-1\right)^2+\left(\tfrac{1}{2}\cdot 2-2\right)^2+\left(\tfrac{1}{2}\cdot 3-3\right)^2\right)$$

$$=\frac{1}{6}\left(\frac{1}{4}+1+\frac{9}{4}\right)$$

$$=\frac{7}{12}$$

③ 设 $\theta=1$

$$J(\theta)\rightarrow J(1)=\frac{1}{6}\left((1\cdot 1-1)^2+(1\cdot 2-2)^2+(1\cdot 3-3)^2\right)$$

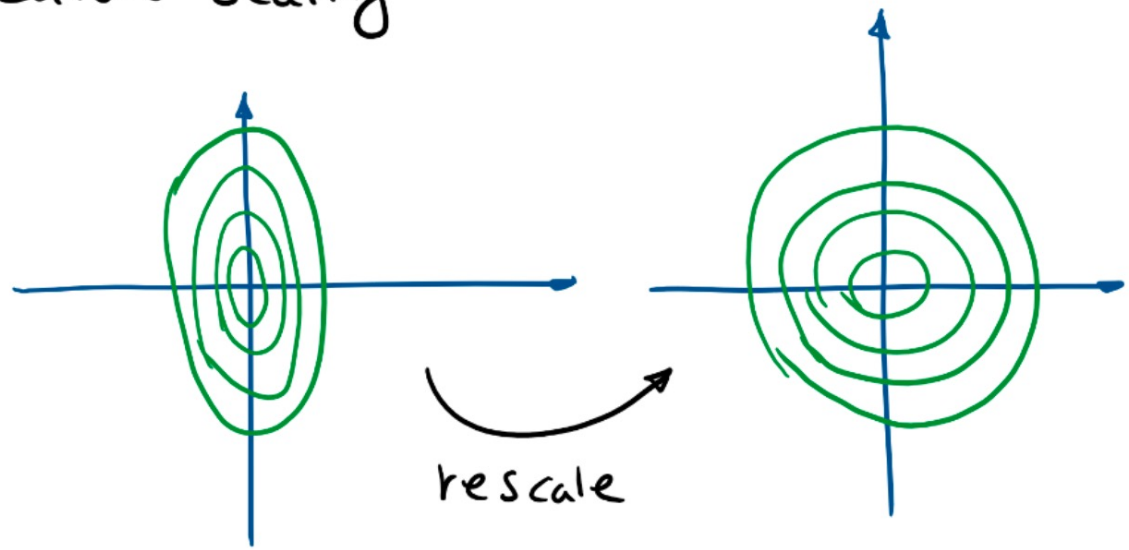
$$=\frac{1}{6}\cdot 0=0$$

X_1 Size ($feet^2$)	X_2 Number of bedrooms	X_3 Number of floors	X_4 Age of home (years)	y_1 Price ($1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

$$+ \theta_0 x_0$$

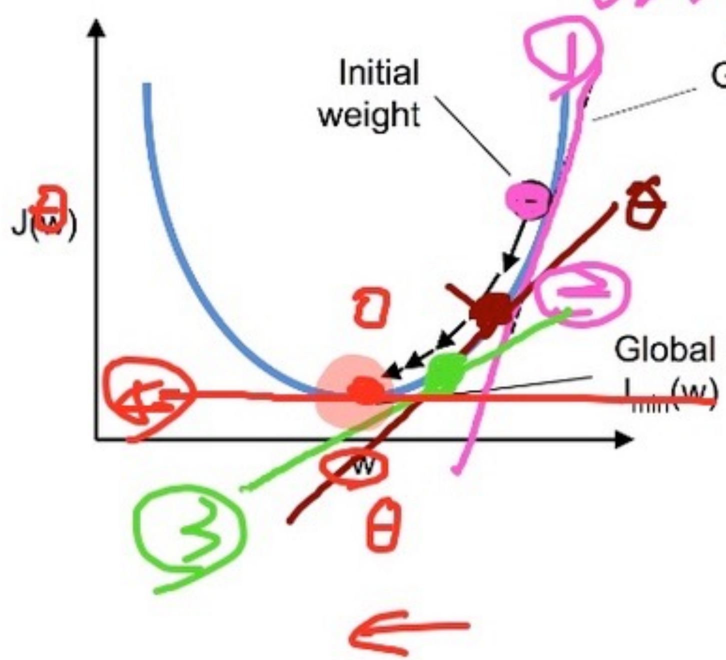
$$h_\theta(x) = \theta^T x = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4$$

Feature Scaling



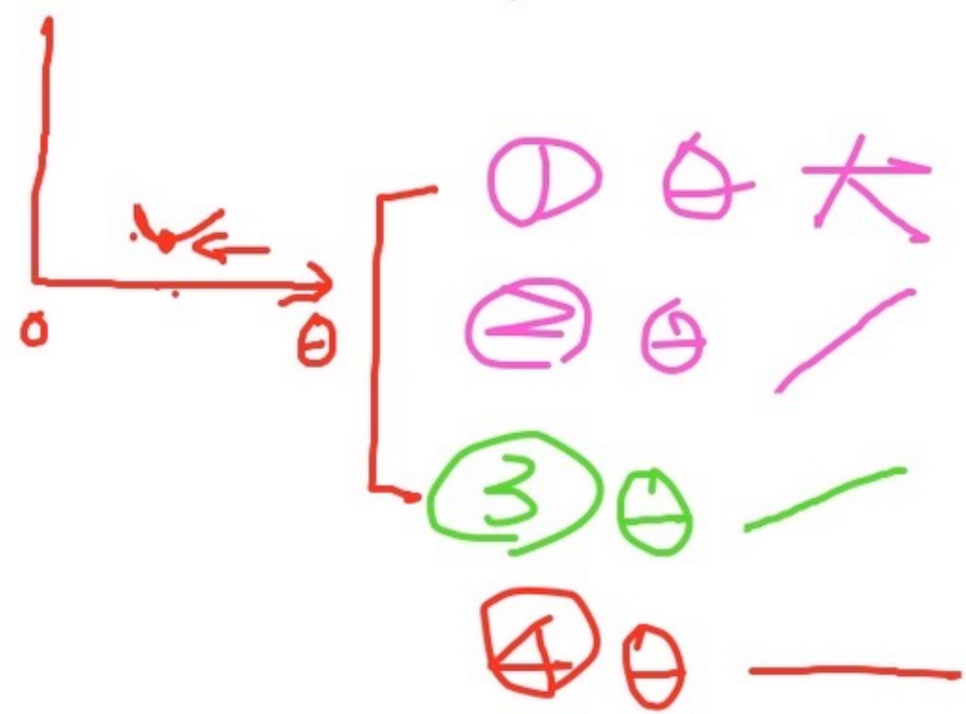
梯度下降Gradient Descent

切线の斜率

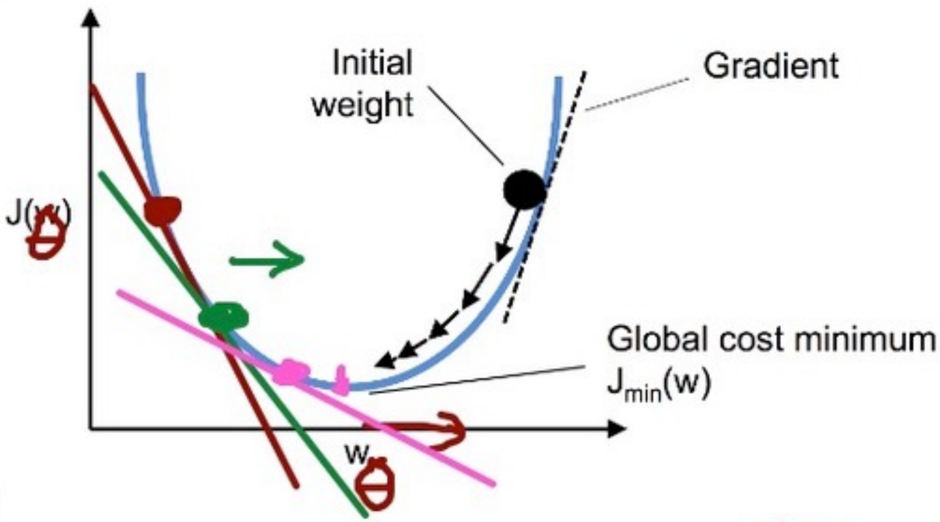


$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}(y^i - h_\theta(x^i))x_j^i$$

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



梯度下降Gradient Descent

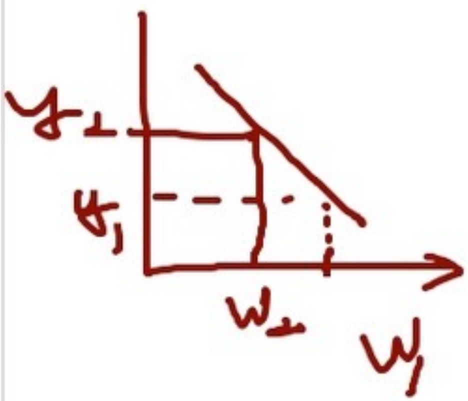


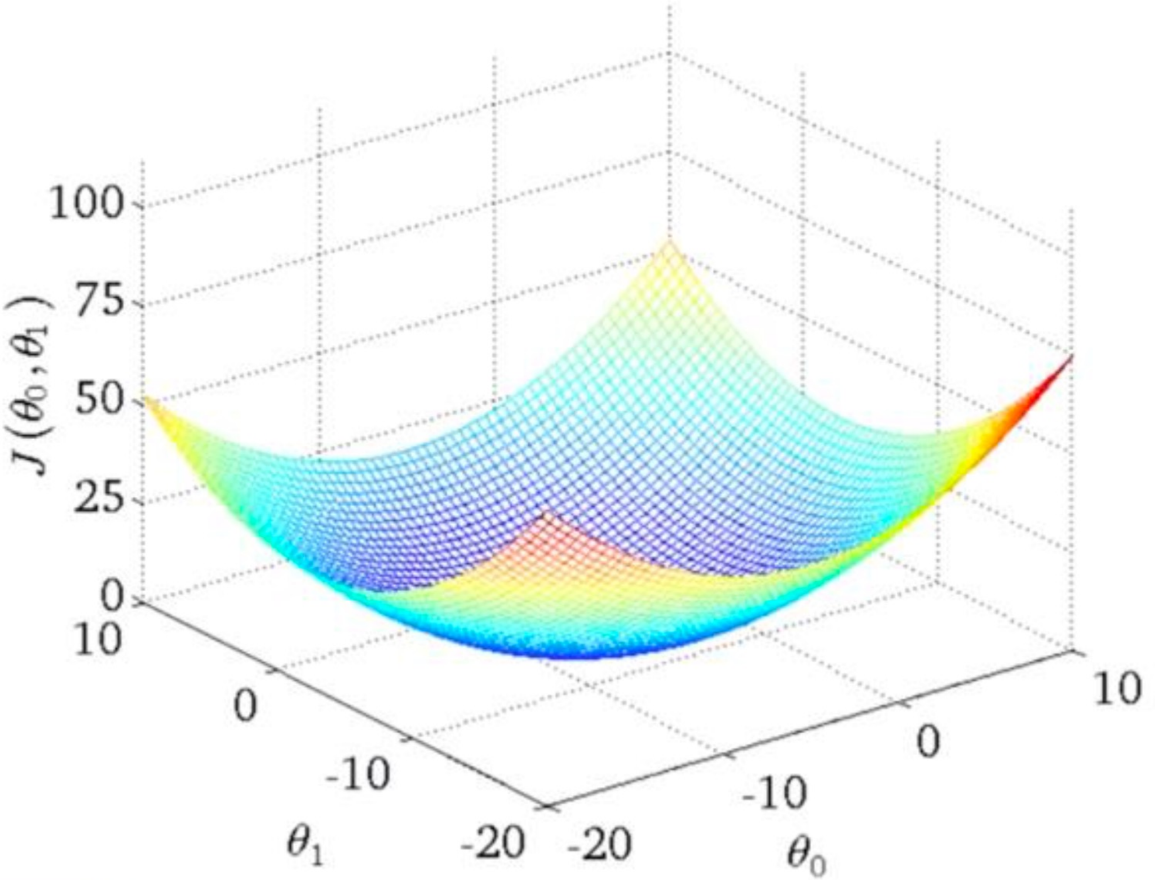
$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}(y^i - h_\theta(x^i))x_j^i$$

负

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

<0 → 负 → 加 → 大





100
75
50
25
0
$J(\theta_0, \theta_1)$
10
0
-10
-20
θ_1
-20
-10
0
10
θ_0