Multiple Linear Regression

Friday, 10 December 2021 7:36 PM

① Hypothesis : $\underset{\hat{y}}{h_\theta(x)} = \theta_0 + \theta_1 x$ → Variables to find

② Error : $J(\theta)$, $L(\theta)$ $= \frac{1}{2m}\sum_{i=1}^{m}\left(y^{(i)} - h_\theta(x^i)\right)^2$ / $\frac{1}{2m}\sum_{i=1}^{m}\left(\underbrace{h_\theta(x^i)}_{\hat{y}^i} - y^{(i)}\right)^2$

G.D ③
- init $\theta()$ randomly
- repeat until converge {

$$\theta_0 = \theta_0 - \alpha \cdot \frac{\partial L}{\partial \theta_0}$$

$$\theta_1 = \theta_1 - \alpha \cdot \frac{\partial L}{\partial \theta_1}$$

}

Derivatives

$$L = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{y^i - \hat{y}^i}_{z})^2$$

$$\frac{\partial L}{\partial \theta_0} \Rightarrow \frac{1}{2m}\sum_{i=1}^{m}(z)^2$$

$$\Rightarrow \frac{\partial L}{\partial z} \cdot \frac{\partial z}{\partial \theta_0}$$

$$\Rightarrow \frac{\cancel{2}}{\cancel{2}m}\sum_{i=1}^{m} z \cdot \frac{\partial\left(\cancel{y^i} - (\theta_0 + \cancel{\theta_1 x^i})\right)}{\partial \theta_0}$$

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m} z \cdot \frac{\partial(-\theta_0)}{\partial \theta_0}$$

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m}(y^i - \hat{y}^i)\boxed{(-1)}$$

$$\frac{\partial L}{\partial \theta_0} \Rightarrow \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^i - y^i)$$

$$\boxed{\Rightarrow \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)}$$

$$\boxed{\frac{\partial L}{\partial \theta_1} \Rightarrow \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i) \cdot x^i} \; -1$$

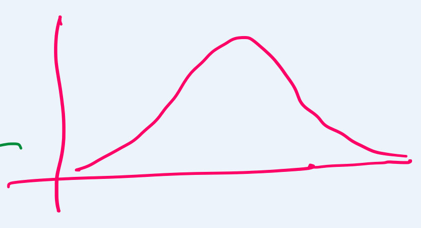
Test Case

$\theta_0 = 100$ $x = 5$

$\theta_1 = 20$ $\Rightarrow 100 + 20 \times 5$

$\Rightarrow 200$

$\underset{\begin{bmatrix} \frac{\partial L}{\partial \theta_0} \\ \frac{\partial L}{\partial \theta_1} \end{bmatrix}}{\text{grad}}$ = gradient(X, Y, theta)



Standardization

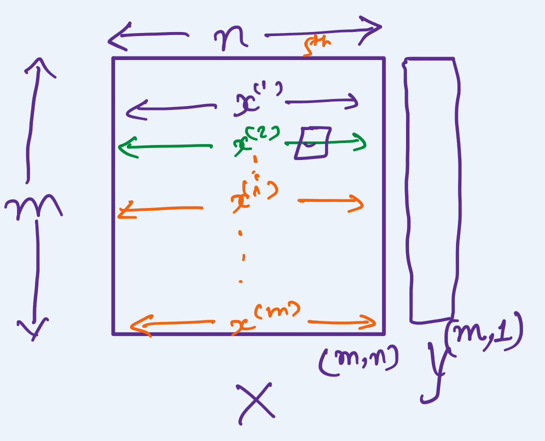
$\mu = 0$

$\sigma = 1$

$$\frac{X - \mu}{\sigma}$$

Terminology:

$$\mathcal{D} = \{\underbrace{(x^{(i)}, y^{(i)})}_{\text{Pair}}\}_{i=1}^{m}$$



features → n

Examples → m

$x^{(4)}$ → 4th example input

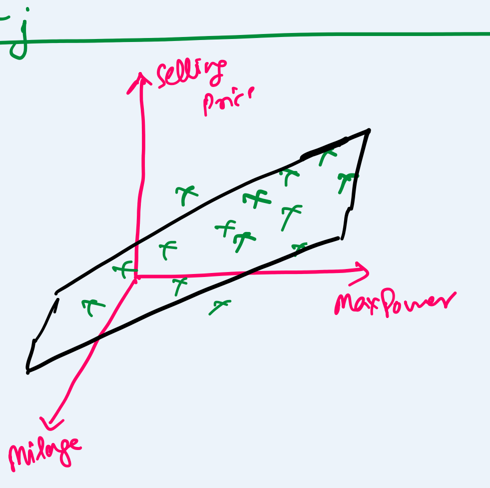
features → in subscript

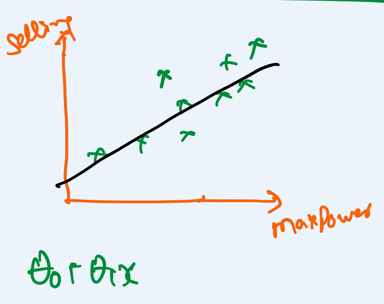
x_5 → entire col^m of 5th feature

x_j → jth feature (entire col^m).

$\boxed{x_4^{(2)}}$ → for 2nd datapoint, 4th feature

$x_j^{(i)}$ $y^{(1)}, y^{(2)}, y^{(i)}$





$\theta_0 + \theta_1 x$

$$\underset{\hat{y}}{h_\theta(x)} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Hypothesis function : $\theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad x = \begin{bmatrix} x_0 = 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \Rightarrow \underbrace{\theta^T x + \theta_0}_{W^T x + w_0}$$

$\boxed{x_0 = 1}$

$$\hat{y} = h_\theta(x) \Rightarrow \boxed{\theta^T x} = \boxed{\sum_{i=0}^{n} \theta_i x_i}$$

Error function

$$L(\theta) = J(\theta) = \boxed{\frac{1}{2m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^2}$$

Scalar (Actual output)

(Predicted output)

How you calculate this function is changed

Gradient Descent

① Init $\theta()$ randomly.
$\theta_0, \theta_1, \theta_2, \dots \theta_n = 0$

② Repeate until Convergence {

$$\theta_0 = \theta_0 - \alpha \frac{\partial L}{\partial \theta_0}$$

$$\vdots$$

$$\theta_j = \theta_j - \alpha \cdot \frac{\partial L}{\partial \theta_j}$$

}

$$\underbrace{\text{theta}}_{\text{vector}} = \underbrace{\text{theta}}_{\text{vector}} - \alpha \cdot \frac{\partial L}{\partial \underbrace{\text{theta}}_{\text{vector}}}$$

gradient vector

Gradients

$$L = \frac{1}{2m}\sum_{i=1}^{m}(\underbrace{y^i - \hat{y}^i}_{z})^2 \Rightarrow L = \frac{1}{2m}\sum_{i=1}^{m}(z)^2$$

$$\frac{\partial L}{\partial \theta_j} = \frac{dL}{dz} \cdot \frac{dz}{d\theta_j}$$

$\theta_0, \theta_1, \theta_j, \dots, \theta_n$

$$\Rightarrow \frac{\cancel{2}}{\cancel{2}m}\sum_{i=1}^{m} z \cdot \frac{d}{d\theta_j}\left(\cancel{y^{(i)}} - (\cancel{\theta_0 x_0^{(i)}} + \cancel{\theta_1 x_1^{(i)}} + \dots \theta_j x_j^{(i)} + \cancel{\theta_n x_n^{(i)}})\right)$$

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m} z \cdot \left[-x_j^{(i)}\right]$$

Coeff of $\theta_j \Rightarrow -x_j^i$

$$\Rightarrow \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)(-1)\, x_j^{(i)}$$

$$\boxed{\frac{\partial L}{\partial \theta_j} \Rightarrow \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right) \cdot x_j^{(i)}}$$

if $\theta_j = \theta_0$

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)\cancel{x_0^{(i)}} \; 1$$

if $\theta_j = \theta_1$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right) \cdot x_1^{(i)}$$

iterating over all examples

1st feature value of ith example