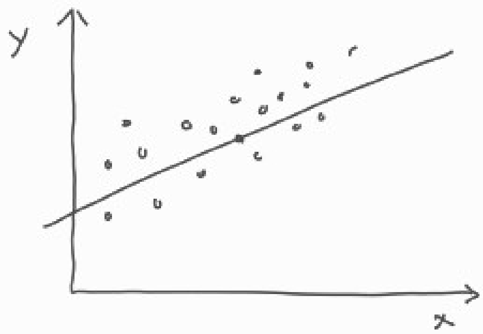
Linear Regression

→ The General idea is to fit a line, to best represent the relation between input & output parameter



$y = b + mX$

y = result
x = input.

b = bias

m = feature weight/slop

We represent the same in program as

$h_\theta = \theta_0 + \theta_1 x$ → Simple Linear Regression
one input parameter/feature

$\theta \Rightarrow$ theta

$h_\theta^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \dots$ → Multi Linear Reg
multiple feature/param

θ_0 = bias

θ_1, θ_2 = feature weight.

Our Goal is to find the bc θ value which will repoent the line.

→ And we do that with some thing called
→ Gradient Decent

And Some of the popular chooises of GD are

1. Batch GD [BGD]
2. Stochastic GD [SGD]
3. Mini - Batch GD → most . commonly used.

Here we will be using Batch Gradient Decent [BGD] to calculate the value of θ or the slop

formula →

$$\theta = \theta - \alpha \frac{\partial J\theta}{\partial \theta}$$

α = Learning rate

$J\theta$ = cost/loss function

Cost function → Half MSE / Cost function {Mean Squared Error}

$$J = \quad J(\theta) = \frac{1}{2m} \Sigma \Big(\underbrace{\overbrace{h_\theta(x^{(i)})}^{h} - y^{(i)}}_{\text{error}} \Big)^2$$

$$\frac{\partial J}{\partial \theta_k} = \frac{1}{2m} \Sigma\, 2\,(h - y)\, \frac{\partial (h - y)}{\partial \theta_k}$$

Since $\frac{\partial (h-y)}{\partial \theta_k} = X_k$

$\frac{d}{dx} x^2 = 2x$

$$\nabla J\theta_k = \frac{\partial J}{\partial \theta_k} = \frac{1}{m} \Sigma\, \underbrace{(h-y)}_{} x_k$$

→ error , we will see

So the final formula is

$$\theta_k = \theta_k - \alpha \frac{1}{m} \Sigma (h - y) X_k$$

For Example let's take

$$\begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ X = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 2 \\ 1 & 3 \end{bmatrix} \end{array} \qquad y = \begin{bmatrix} 6 \\ 8 \\ 12 \\ 8 \end{bmatrix}$$

$m = 4$

Before we start

we will add intercepter in our input vector

$$\begin{array}{c} \begin{array}{ccc} x_0 & x_1 & x_2 \end{array} \\ X = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 1 & 3 \end{bmatrix} \end{array}$$

intercepter

$$\theta = [0 \quad 0 \quad 0]$$

$\alpha = 0.1$ Learning Rate.

Iteration - 1

$$h = \Theta_0 x_0 + \Theta_1 x_1 + \Theta_2 x_2$$

Step-1 Calculate hypotrsis

$i=0, \quad h^{(0)} = 0 + 0 \cdot \overset{1}{x_1^{(0)}} + 0 \cdot \overset{2}{x_2^{(0)}} = 0$

$i=1, \quad h^{(1)} = 0 + 0\, \overset{2}{x_1^{(1)}} + 0\, \overset{1}{x_2^{(1)}} = 0$

$i=2, \quad h^{(2)} = 0$

$i=3, \quad h^{(3)} = 0$

Step-2 Calculate Error

$e^{(0)} = h^{(0)} - y^{(0)} = 0-6 = -6$

$e^{(1)} = h^{(1)} - y^{(1)} = 0-8 = -8$

$e^{(2)} = -12$

$e^{(3)} = -8$

Step-3 Calculate Gradient $\nabla J\Theta_k = \frac{\partial J}{\partial \Theta_k} = \frac{1}{m} \Sigma e_k^T X_k$

$$\Sigma e^{(i)T} X_i = \frac{1}{4} \left[\underset{1\times 4}{\begin{bmatrix} -6 & -8 & -12 & -8 \end{bmatrix}} \cdot \underset{4\times 3}{\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \\ 1 & 1 & 3 \end{bmatrix}} \right]$$

We can take a matrix dot product

$\Sigma e^0 x_0 = \nabla J \Theta_0 = (-6 \times 1) + (-8 \times 1) + (-12 \times 1) + (-8 \times 1)$

$= -6 - 8 - 12 - 8 = -34 \Rightarrow -34/4 = -8.5$

$\nabla J \Theta_1 = (-6 \times 1) + (-8 \times 2) + (-12 \times 3) + (-8 \times 1)$

$= -6 - 16 - 36 - 8 = -66 \Rightarrow -66/4 = -16.5$

$$\nabla J\theta_2 = -12 - 8 - 24 - 24 = -68 \Rightarrow -68/4 = -17.0$$

$$\nabla J\theta_0 \Leftrightarrow \frac{\partial J}{\partial \theta_0} \Leftrightarrow g_0$$

Step-4 update the rule

$$\theta_j \leftarrow \underline{\theta_j - \alpha g_i}$$

$$\theta_0 = 0 - 0.1(-8.5) = 0.85$$

$$\theta_1 = 0 - 0.1(-16.5) = 1.65$$

$$\theta_2 = 0 - 0.1(-17) = 1.7$$

Step-5 Summary

Cost function before updating θ

$$J\theta = \frac{1}{2m}\Sigma(e^{(i)})^2 \longrightarrow \frac{1}{2m}\Sigma\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{1}{2\times4}\left((-6)^2 + (-8)^2 + (-12)^2 + (-8)^2\right)$$

$$= \frac{1}{8}(36 + 64 + 144 + 64)$$

$$= \frac{308}{8} = \underline{38.5}$$

End with

$$\theta = [0.85 \quad 1.65 \quad 1.7]$$

$$\nabla J\theta = g = \underline{[-8.5, -16.5, -17.0]}$$

$$J = 38.5$$

Iteration -2 start $\theta = (0.85 \quad 1.65 \quad 1.70)$

S1

$$h^0 = 0.85(1) + 1.65(1) + 1.7(2) = 5.90$$

$$h^1 = 0.85(1) + 1.65(2) + 1.7(1) = 5.85$$

$$h^2 = 0.85(1) + 1.65(3) + 1.7(2) = 9.20$$

$$h^3 = 0.85(1) + 1.65(1) + 1.7(3) = 7.60$$

Step-2

$$e^0 = 5.90 - 6 = -0.10$$

$$e^1 = 5.85 - 8 = -2.15$$

$$e^2 = 9.20 - 12 = -2.80$$

$$e^3 = 7.6 - 8 = -0.40$$

Step-3 Gradient decent

$$g_0 = (-0.10 - 2.15 - 2.80 - 0.40)/4 = -1.3625$$

$$g_1 = ((-0.10)1 + (-2.15)2 + (-2.80)\cdot 3 + (-0.40)\cdot 1)/4 = -3.30$$

$$g_2 = ((-0.10)2 + (-2.15)1 + (-2.80)\cdot 2 + (-0.40)\cdot 3)/4 = -2.2875$$

Step-4 update

$$\theta_0 = 0.85 - 0.1(-1.3625) = 0.98625$$

$$\theta_1 = 1.65 - 0.1(-3.30) = 1.98$$

$$\theta_2 = 1.70 - 0.1(-2.2875) = 1.92875$$

Step-5

$$(-0.1)^2 + (-2.15)^2 + (-2.80)^2 + (-0.40)^2 = 12.6325$$

$$J = \frac{12.6325}{8} = 1.5790625$$

$\Theta \simeq (0.98625 \quad 1.98 \quad 1.92875)$

$g \simeq (-1.3625 \quad -3.30 \quad -2.2875)$

$J \approx 1.57906.$

We will continue untill the end of itteration