Simple Linear Regression

A linear regression problem is a type of statistical modeling and machine learning problem where the goal is to establish a linear relationship between an independent variable and a dependent variable

Supervised ML $\longrightarrow$ Regression

TRAIN

$\downarrow$

New weight $\longrightarrow$ Model $\longrightarrow$ Height

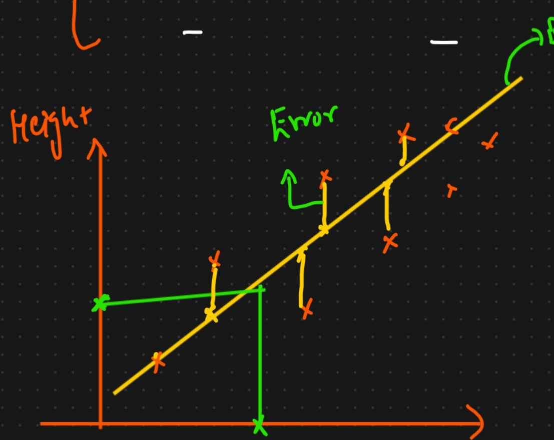
$\uparrow$

Simple Linear Regression

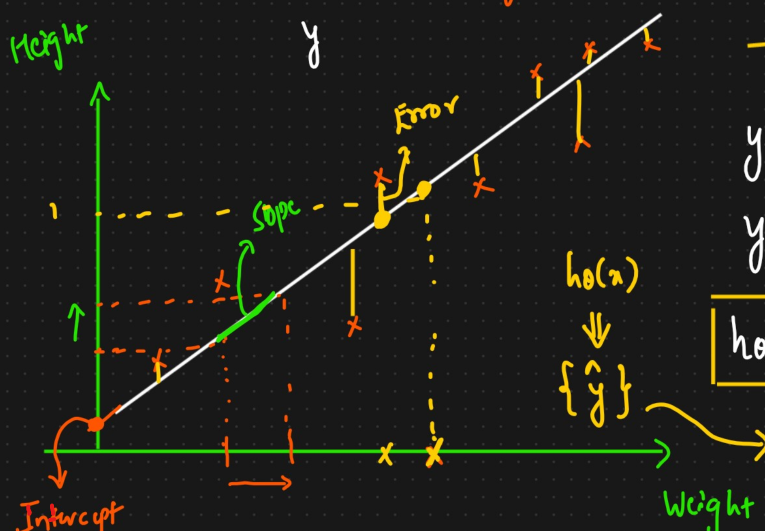
Dataset

I/P features $\uparrow$ x   y $\uparrow$ O/P or dependent feature

Weight	Height
74	170cm
80	180cm
75	175.5cm
-	-



Minimal



θ_0 & θ_1 Minimal Error

Dataset

$h_\theta(x)$

$\Downarrow$

$\{\hat{y}\}$ $\longrightarrow$ predicted point

$y = mx + c$

$y = \beta_0 + \beta_1 x$

$$\boxed{h_\theta(x) = \theta_0 + \theta_1 x}$$

$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$

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

x = I/p Feature

θ_0 = Intercept

θ_1 = slope or coefficient

if $x = 0$

$h_\theta(x) = \theta_0$

Error $(y - \hat{y})$

Cost function

A cost function is a mathematical function that measures how well a machine learning model's predictions match the actual data. It quantifies the error between predicted and actual values and helps the model learn by minimizing this error during training

predicted ⇑ True O/P ⇑

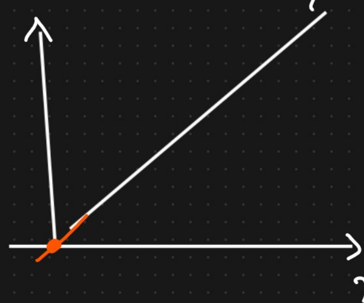
$$J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2 \Rightarrow \text{Mean Squared Error}$$

[redacted]

Final Aim what we need to solve

$$\underset{\theta_0,\theta_1}{\text{Minimize}} \quad J(\theta_0,\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{i} - y^{(i)}\right)^2 \downarrow\downarrow\downarrow$$

① $h_\theta(x) = \theta_0 + \theta_1 x$



$\boxed{\theta_0 = 0}$

$\boxed{h_\theta(x) = \theta_1 x}$

DATASET

x	y
1	1
2	2
3	3

$\theta_0 = 0$

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

let $\theta_1 = 1$ {slope}

$h_\theta(x) = 1 \quad x = 1$

$h_\theta(x) = 2 \quad x = 2$

$h_\theta(x) = 3 \quad x = 3$

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

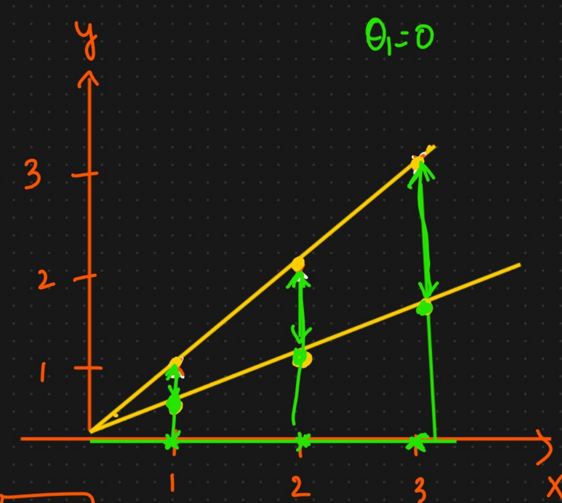
↓

let $\theta_1 = 0.5$

$h_\theta(x) = 0.5$ if $x = 1$

$h_\theta(x) = 1$ if $x = 2$

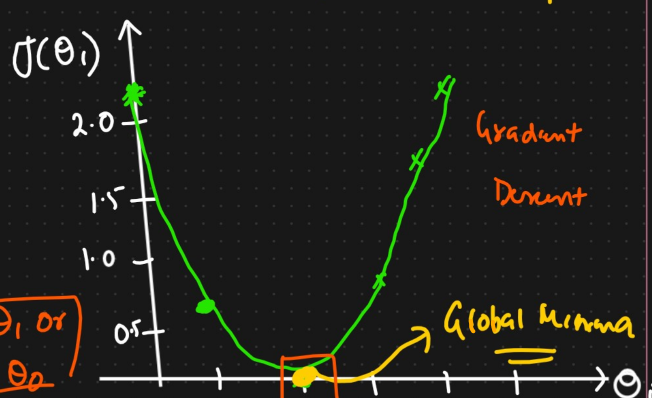
$h_\theta(x) = 1.5$ if $x = 3$



$\boxed{\theta_1 = 1}$

$$J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

$$= \frac{1}{2\times 3}\left[(1-1)^2 + (2-2)^2 + (3-3)^2\right]$$



$\boxed{\theta_1 \text{ or } \theta_0}$

0.5 1 1.5 2.0 2.5

$J(\theta_1) = 0 \leftarrow$ Error has been minimized

$\theta_1 = 0.5$

$$J(\theta_1) = \frac{1}{2\times 3}\left[(0.5-1)^2 + (1-2)^2 + (1.5-3)^2\right]$$

$$J(\theta_1) = \approx 0.58$$

if $\theta_1 = 0$

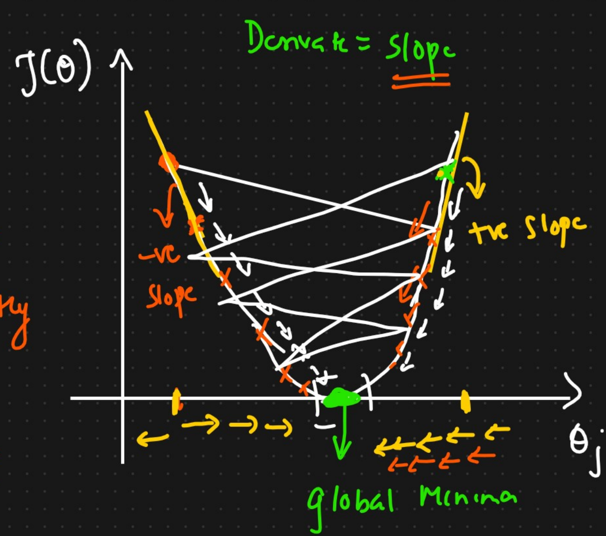
$$J(\theta_1) = \frac{1}{2\times 3}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

$$J(\theta_1) \approx 2.3$$

Gradient Descent :-

Gradient Descent is an optimization algorithm used to find the minimum of a function.

Convergence Algorithm: {Optimize the changes of θ_1 value}



Repeat until Convergence
(Global Minima)

$$\left\{ \theta_j := \theta_j - \alpha \boxed{\frac{d}{d\theta_j} J(\theta_j)} \right\}$$

$\rightarrow$ -ve

θ_1 value much more efficiently

$$\theta_j = \theta_j - \alpha(+ve)$$
$$= \theta_j - (+ve)$$

$$\left\{ \begin{array}{l} \theta_j = \theta_j - \boxed{\alpha}(-ve) \\ \theta_j = \theta_j + (+ve) \end{array} \right\}$$

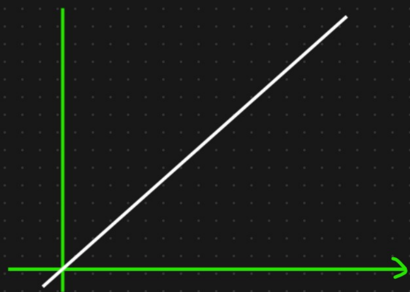
$$\boxed{\alpha = 0.001} \leftarrow$$

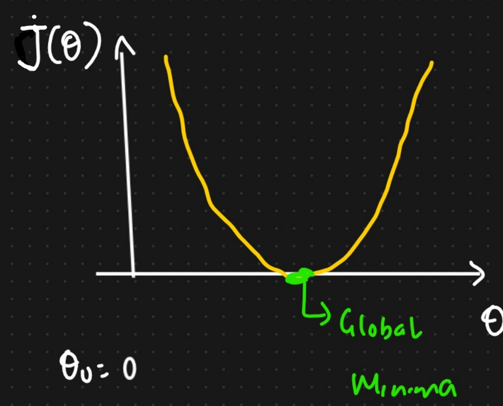
α = Learning Rate

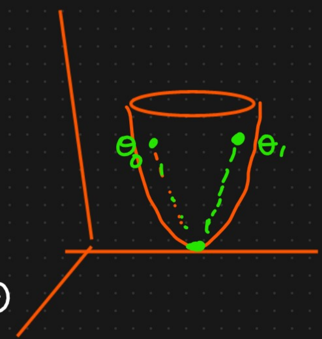
Controls the speed of convergence

Final Conclusion

GRADIENT DESCENT







$\theta_0 = 0$

$\theta_1 =$ Changeable

Convergence Algorithm

repeat until convergence

{

$$\theta_j := \theta_j - \alpha \boxed{\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)}$$

}

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

$j = 0$ and 1

$$\frac{\partial}{\partial x} (x)^2 = 2x$$

$$\frac{\partial}{\partial x} x^n = n x^{n-1}$$

$$\rightarrow \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_j} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$$

if

$$h_\theta(x) = \theta_0 + \boxed{\theta_1 x} \longrightarrow 0$$

$$j = 0 \Rightarrow \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} \frac{1}{2m} \left[\sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right)^2\right]$$

$$= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x)^{(i)} - y^{(i)}\right) \times 1$$

$$j=1 \Rightarrow \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_1} \frac{1}{2m}\left[\sum_{i=1}^{m}\left((\theta_0+\theta_1 x) - y^{(i)}\right)^2\right]$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left((\theta_0+\theta_1 x) - y^{(i)}\right)(x)$$

$$\boxed{\frac{\partial}{\partial \theta_1}\left[\theta_0 + \theta_1 x\right] \Rightarrow x}$$

Repeat until Convergence

$$\{$$

$$\theta_0 := \theta_0 - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x)^{(i)} - y^{(i)}\right) x^{(i)}$$

$$\}$$