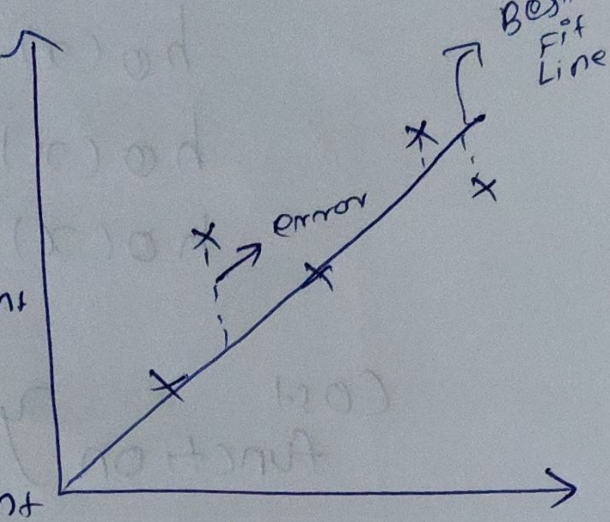
Simple Linear Regression:

Single column with Target value.

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

$h_\theta(x)$ ↑ Dependent

x ↑ Independent



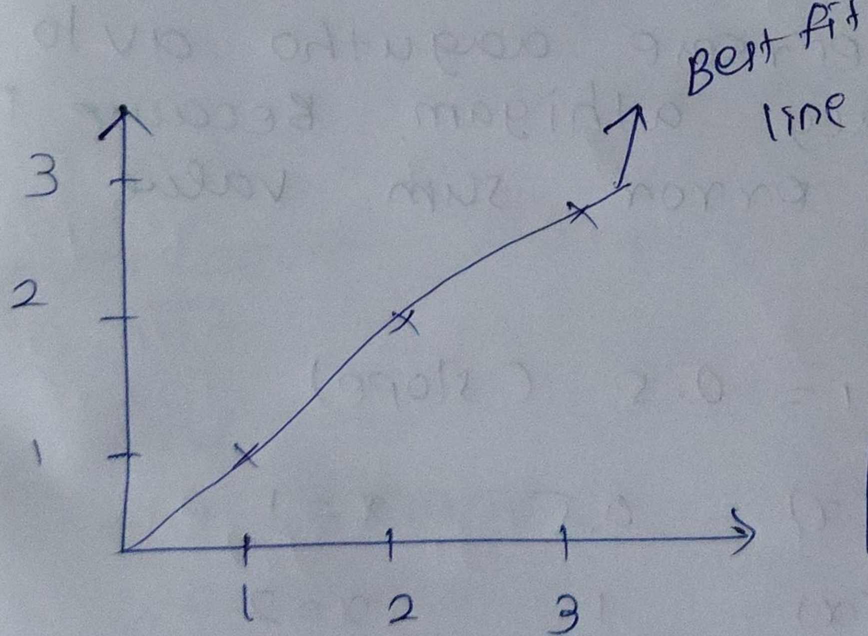
$\theta_0 \rightarrow$ Intercept

$\theta_1 \rightarrow$ Unit movement in the x-axis is change in unit movement in y axis. (slope)

Cost function:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \Big(\underbrace{h_\theta(x)^{(i)}}_{\text{Predicted}} - \underbrace{y^{(i)}}_{\text{Actual value}}\Big)^2$$

eg



Data

x	y
1	1
2	2
3	3

$\theta_0 = 0$ (origin la Incept value 0)

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

Let $\theta_1 = 1$ {slope}

$h_\theta(x) = \theta_1(x) = 1 \times 1 = 1$

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

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

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

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

$m = 3$ (No of data record)

$= \frac{1}{2 \times 3} \sum_{i=1}^{3} (h_\theta(x)^{(i)} - y^{(i)})^2$

$= \frac{1}{6} [(1-1)^2 + (2-2)^2 + (3-3)^2]$

$= \frac{1}{6}(0) \Rightarrow 0 \rightarrow ①$

Note :- ⊛

cost function value evlo decrease aagutho avlo accuracy athigam. Because it is the error sum value

Let $\theta_1 = 0.5$ (slope)

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

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

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

cost function $= \frac{1}{2\times3}((0.5-1)^2+(1-2)^2 + (1.5-3)^2)$

$= \frac{1}{6}((-0.5)^2+(-1)^2+(-1.5)^2)$

$= \frac{1}{6}(0.25+1+2.25)$

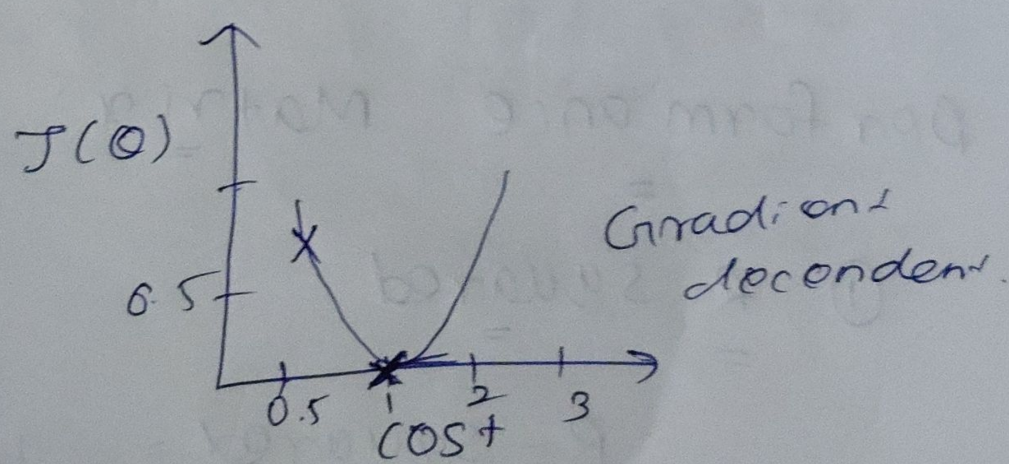
$= 1/6 \times 3.5$

$= 0.58 \rightarrow$ (2).

Comparing (1) & (2)

∴ The value of the best fit line θ is best because it gives the minimal outcome.

Plotting (1) & (2)



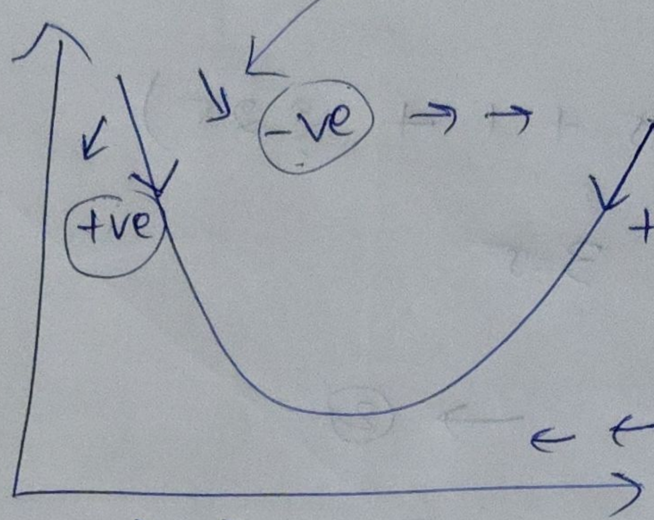
Convergence Algorithm:

It is used to find the slope value for the changing cost function we cannot directly give randomly the slope value for that we using convergence Algorithm

$$\boxed{\theta_j := \theta_j - \underbrace{\alpha}_{\text{(Learning Rate)}} \frac{\partial}{\partial \theta_j} J(\theta_j)}$$

$:=$ → contiously Iterative

$\frac{\partial}{\partial \theta_j} J(\theta_j)$ → find the slope of that particular point



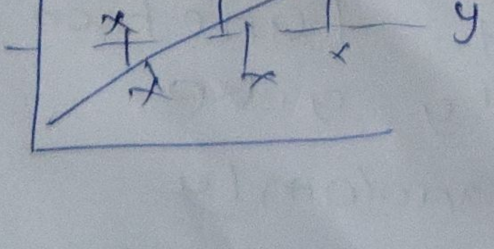
$$\left\{ \begin{aligned} \theta_j &= \theta_j - \alpha(-ve) \\ \theta_j &= \theta_j + \alpha \cdot (+ve) \end{aligned} \right\}$$

Learning Rate → To converge the data in to the Global minima. $\boxed{\alpha = 0.001}$

Performance Matrix:

① R squared:

$$R\text{-Squared} = 1 - \frac{\text{Sum of residuals}}{\text{Sum of square Total}}$$



$$= 1 - \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y})^2}$$

$\bar{y}$ → mean of all result value

to find out the model accuracy If Increase the value of R^2 The model Accuracy also Increases.

Adjusted R^2:-

When the two variable does not correlated with each other when we use the Adjusted R^2 because the normal R^2 Increase accuracy poiya. Increase pannunga.

$$\text{Adjusted } R^2 \; \bar{y} = 1 - \frac{(1-R^2)(N-1)}{N-P-1}$$

$N \rightarrow$ No of data points

$P \rightarrow$ No of Independent features

Types of cost functions

- MSE
- MAE
- RMSE

MSE:- (Mean Squared Error)

$$MSE = \frac{\Sigma (y - \hat{y})^2}{n}.$$

adv

1. Differenciable
2. It has one local and one global minima
3. converges faster

Dis,

1. Not Robust to outliers

MAE :- (Mean Absolute error)

$$\boxed{MAE = \frac{1}{n} \sum_{i=1}^{n} |y - \bar{y}|}$$

Adv

1. Robust to outliers
2. It will be in the same unit because do not square the value

Dis

1. convergence take more time
2. Time consuming

RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{MSE}$$

Adv

1. same unit
2. differenciatiable

Dis

1. Not Robust to outliers

Overfitting & Underfitting:

① Training data set

② Testing data set

③ Validation data set

Data set → Training (70%) → Train, Validate (Hyper Parameter Tuning)

Data set → Testing (30%)

Train	very good Accuracy (High bias)	very Good Accuracy [Low Bias]
Test	very good Accuracy (Low variance)	very bad Accuracy [High variance]
	↓ Generalized model.	↓ Model overfitting

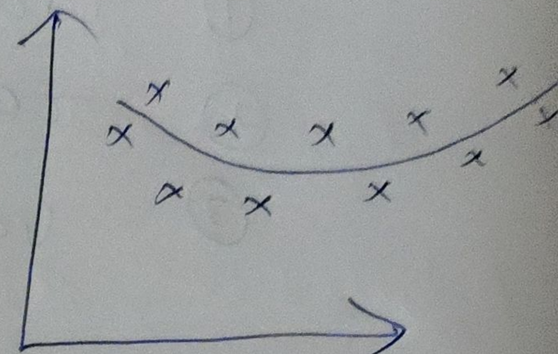
Train	Model Accuracy low (High bias)
test	Accuracy low (High variance)

↓

Underfitting

Polynomial Regression:-

It depends upon the polynomial Degree.



→ degree 0

$$h_\theta(x) = \beta_0 x^0 \rightarrow \text{Constant value}$$

→ degree 1

$$h_\theta(x) = \beta_0 x^0 + \beta_1 x^1 \rightarrow \text{Simple Linear Regression}$$

→ degree 2

$$h_\theta(x) = \beta_0 x^0 + \beta_1 x^{(1)} + \beta_2 x^{(2)}$$

→ degree n

$$h_\theta(x) = \beta_0 x^0 + \beta_1 x^1 + \ldots + \beta_n x^n$$