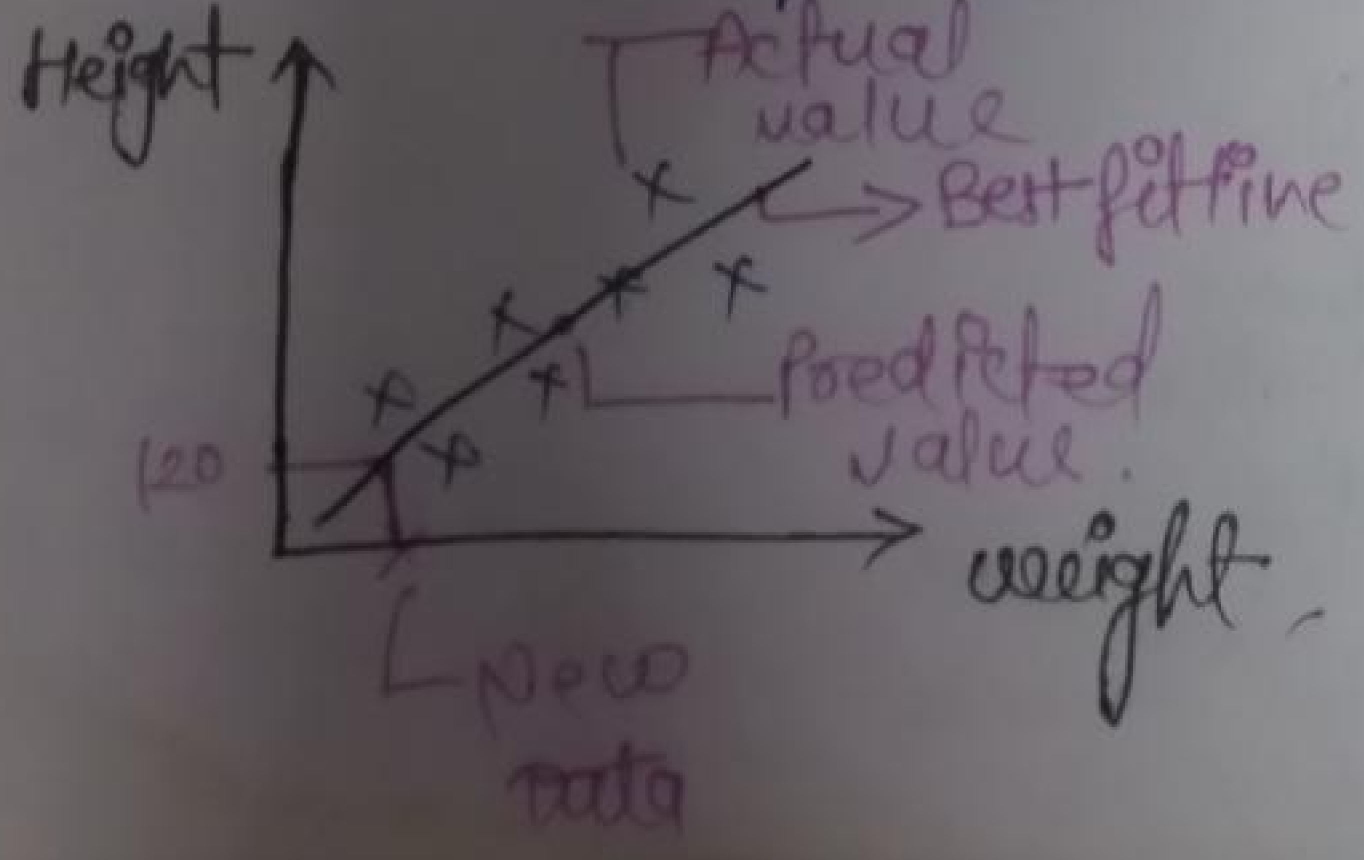
Linear Regression — A supervised learning algorithm used to predict a continuous value by finding the best-fit straight line between input (X) and output (Y)

Eq. — $Y = mX + c$

when —

- m = slope (how much Y changes with x)
- c = intercept (Y when $X = 0$)



Aim — To find Best fit line with minimal error.

weight	Height
73	150
63	155
78	170
82	165

TRAIN

↓

MODEL

↓

weight → hypothesis → Height

Residual Error — The Difference Between Actual and predicted value.

Eq of Best fit line —

$y = mx + c$

m = Slope or Cofficient (x अक्ष पर value change होने पर y के कितना change होगा।)

C = Intercept.

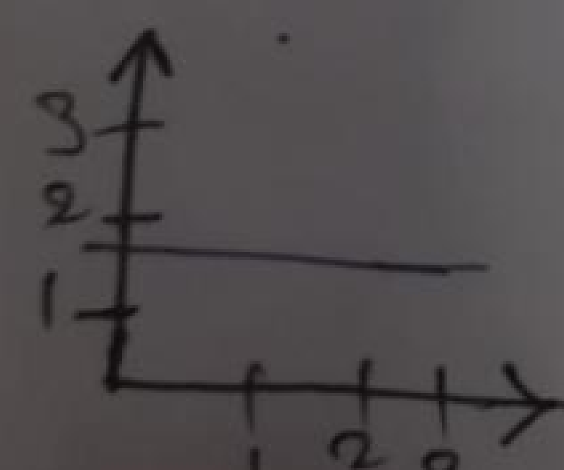
($x=0$ पर y के किस point पर Match होगा.)

y = Actual. Predicted value.

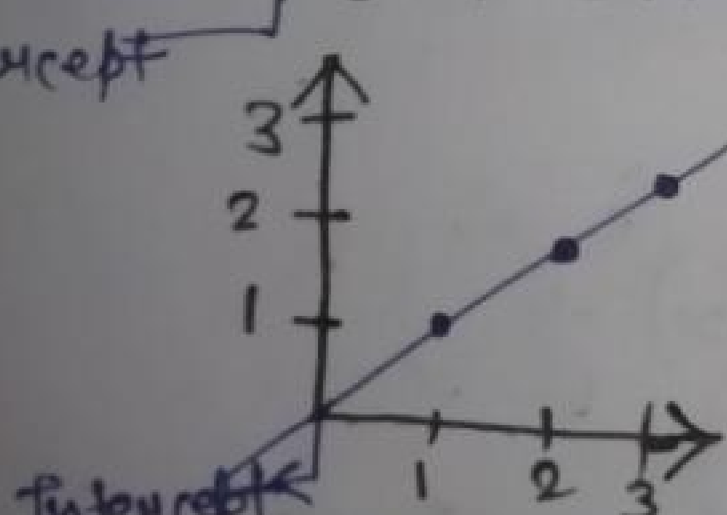
Another Notation — $h_\theta(x) = \theta_0 + \theta_1 x_1$

More than one value. $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots \theta_n x_n$

Hypothesis — $h_\theta(x) = \underbrace{\theta_0}_{\text{Intercept}} + \overbrace{\theta_1}^{\text{Cofficient}} x_1$



$\theta_0 = 1.5$
$\theta_1 = 0$



$0 + (0.5) \times 1 = 0.5$
$0 + (0.5) \times 2 = 1.0$
$0 + (0.5) \times 3 = 1.5$

$\theta_0 = 0$ - Intercept
$\theta_1 = 0.5$.

Some Cost function.

Predicted $\left(h_\theta(x)^{(i)} - y^{(i)}\right)^2$ → Predicted value (-) में हो range होतो, (-) value को (+) करने मे help करता।

$y^{(i)}$ → Actual value

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

m = No. of datapoint

Data point का Distance Calculate

$\theta_0 \theta_1$ को value change करके Cost func. को Minimize करना है/

2 - Differentiation के लिए or slope find कर पाते Mathematical Calculation easy हो जाए

$\frac{1}{m}$ = Difference का Summation from 1 to m value को Average find out करना है इसलिए use किए.

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

↳ Square ~~Root~~ error function

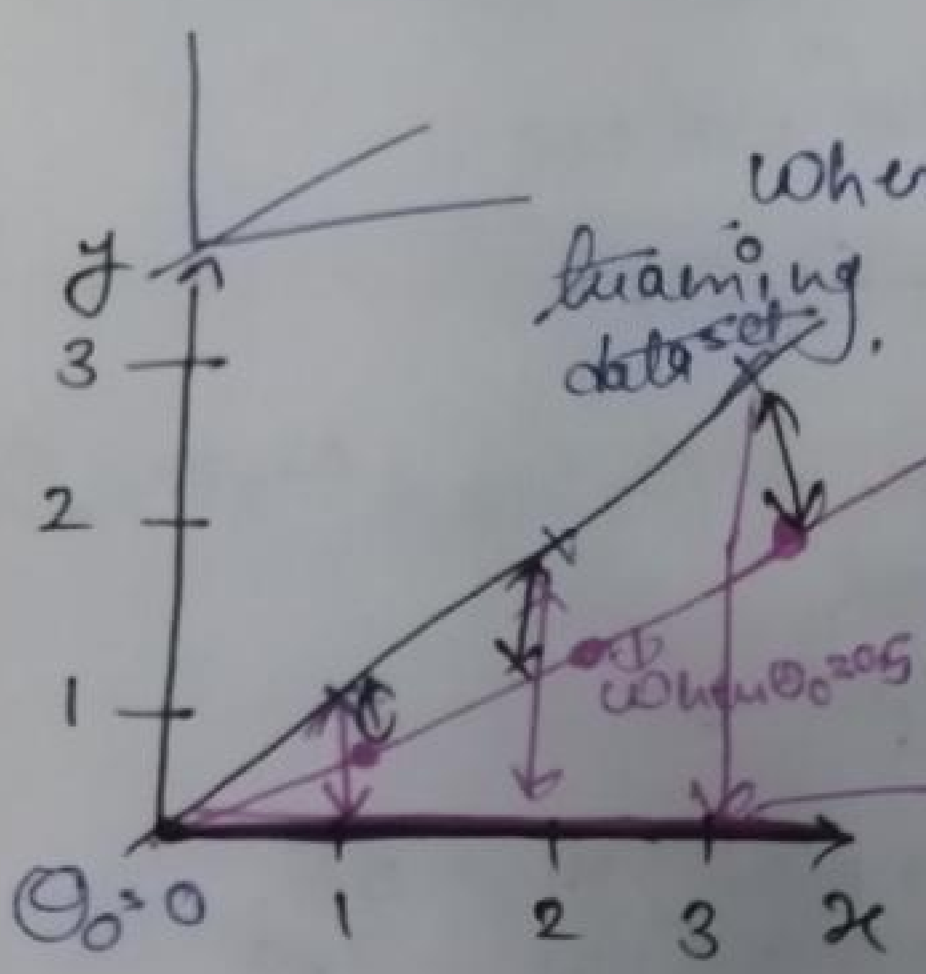
Hypothesis

$h_\theta(x) = \theta_0 + \theta_1 x_1$

Let $\theta_0 = 0$

Cost function

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



when $\theta_1 = 1$

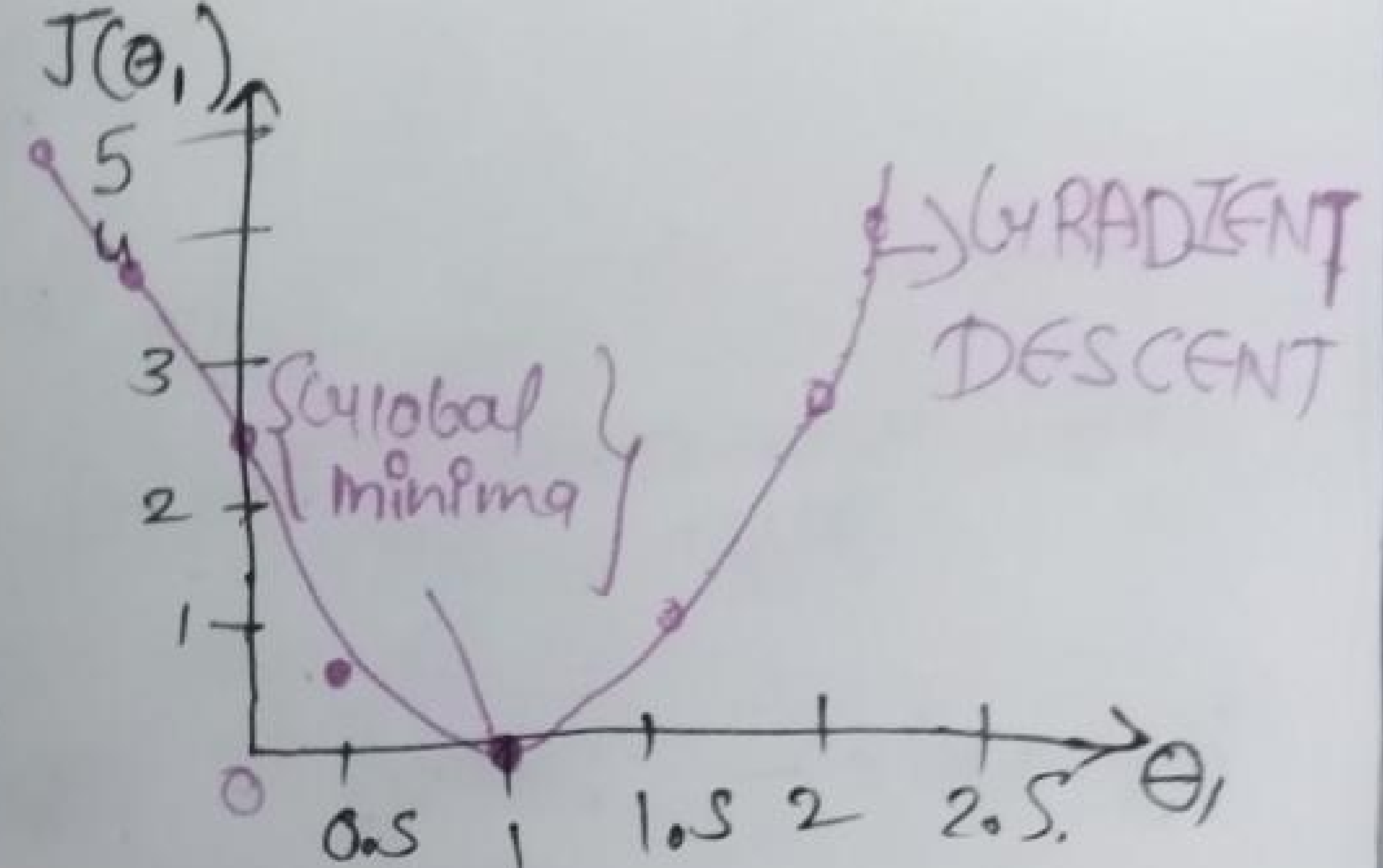
$\begin{Bmatrix} 1, 1 \\ 2, 2 \\ 3, 3 \end{Bmatrix}$

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

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

$h_\theta(x) = 1 \times 3$

when $\theta_0 = 0$



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

$$\frac{1}{6} \sum_{i=1}^{m} \left[(0)^2 + (0)^2 + (0)^2\right]$$

when $\theta_1 = 1$
No diff. Actual and Prediction $= 0$

(2) when $\theta_0 = 0.5$

how in graph

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

$h_\theta(x) = 0.5 \times 2 = 1.0$

$h_\theta(x) = 0.5 \times 3 \quad 1.5$

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

$$\frac{1}{6}\left[(0.5-1)^2 + (1-2)^2 (1.5-3)^2\right]$$

$$\frac{1}{6}\, 3.5 \approx 0.58$$

when $\theta_0 = 0$.

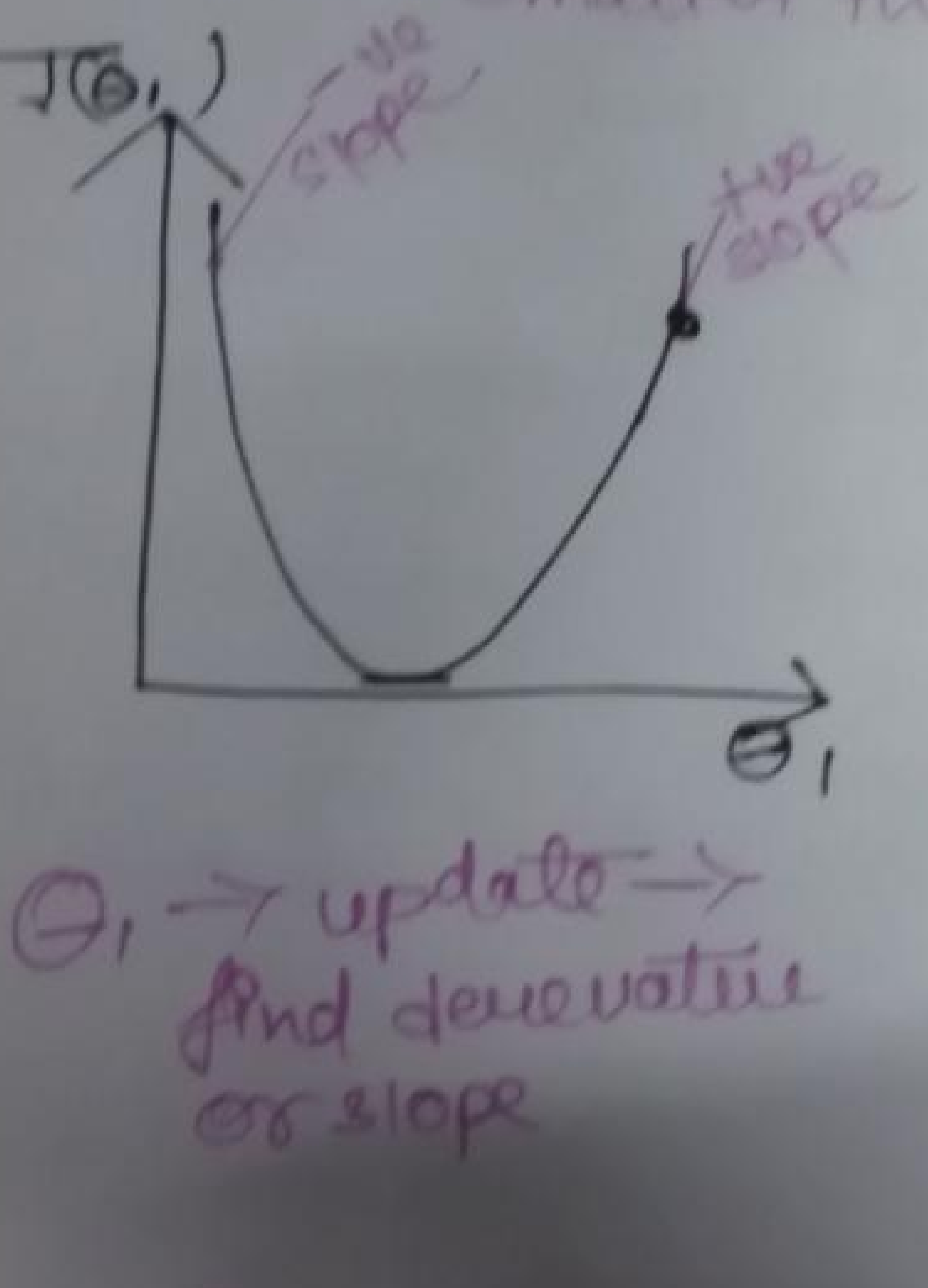
$$J(\theta_0) = \frac{1}{6}\left[(0-1)^2 + (0-2)^2 + (0-3)^2\right]$$

$$\frac{14}{6} \approx 2.3$$

GRADINT DESCENT also called optimizer it automatically change θ_1's value

Global minima — Change the θ_1's value neither it come to Global minima,
- The main focus of global linear regression to come near the global minima.

↳ mean the lowest point on the entire curve of function — where the function value is smaller than at any other point.



+ive slope α value increase hone par y value bhi increase hota hai

- ve slope α value increase hone par y value decrease hota hai

θ_1 → update → find derivative or slope

Repeat Conversion theorem.

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

Learning rate   slope

Learning Rate: — rate control how much we adjust the model's weight with respect to the loss gradient each iteration of training

$\theta_1 := \theta_1 - \alpha(+ve)$ decrement
$\theta_1 := \theta_1 - \alpha(-ve)$ Incremental

इन दोनों को Repeat Conversion Theorem Solve करता है

Small LR - slow but steady
Large LR - might jump around
Optimal learning rate - fast and smooth

always start with small value like = 0.01 or 0.001

outline
- Start with θ_0 and θ_1
- keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$
- until we reach global minima
- conversion theorem

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j}(J(\theta_0, \theta_1)) \quad j = 0 \text{ and } 1$$