Simple Linear Regression

Supervised ML → Regression O/p → Continuous

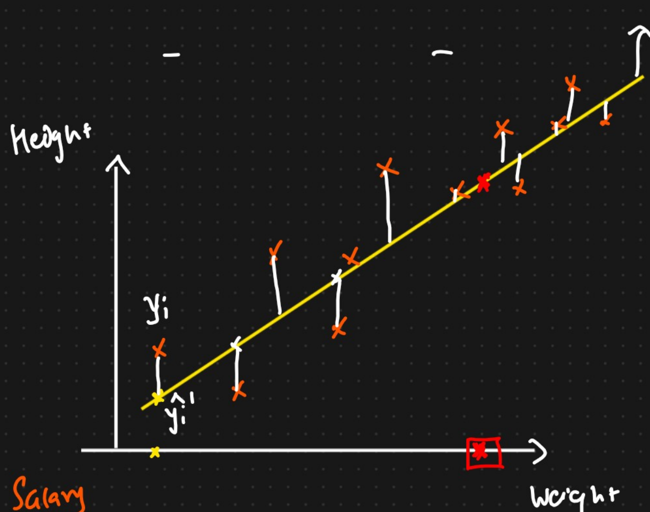
Supervised ML → Classification O/p → Binary, multiclass categories

Dataset

Independent feature   Dependent feature

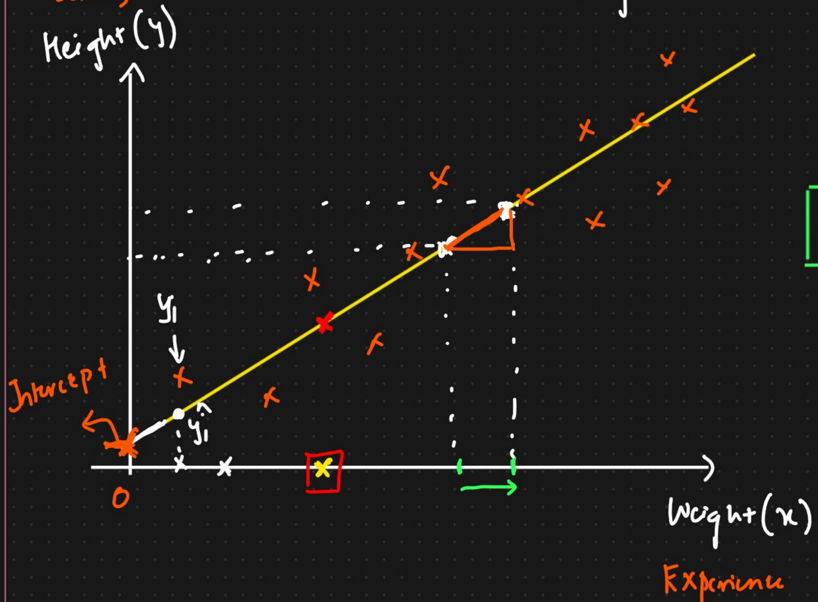
Weight	Height
74	170
80	180cm
75	175.5cm
–	–
–	–

New weight → TRAIN → Model → Height



$\sum$ Error → Minimal

Geometrical Intuition



$$\hat{y} = mx + c$$

$$\hat{y} = \beta_0 + \beta_1 x$$

⇓

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

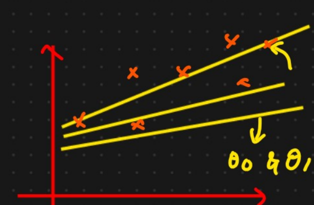
θ_0 = Intercept

θ_1 = Slope or coefficient

x_i ⇒ data points

Predicted

Error $\{y_i - \hat{y}_i\}$



Error ↓ θ_0 & θ_1

⇓

Optimization Process

Cost function

$$J(\theta_0,\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\Big(\underbrace{y_i}_{\text{actual}} - \underbrace{h_\theta(x)}_{\text{prediction}}\Big)^2 \Rightarrow \boxed{\text{Mean Squared Error}}$$

n = no. of datapoints

y_i = Actual value

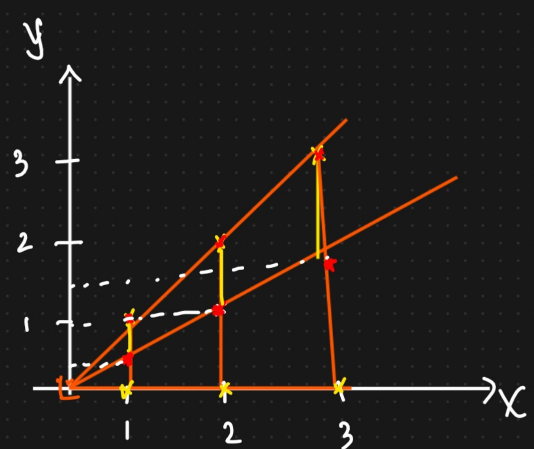
$h_\theta(x)$ = predicted value

Final Aim {To get Best fit line}

Minimize $J(\theta_0,\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2 \downarrow\downarrow\downarrow$.

θ_0,θ_1

Optimization {Minimize the cost function}



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

$\theta_0 = 0$

$\boxed{h_\theta(x) = \theta_1 x_1}$

Dataset

X	y	$h_\theta(x)$
→ 1	1	1
2	2	2
3	3	3

	let $\theta_1 = 1$	let $\theta_1 = 0.5$	let $\theta_1 = 0$
$x=1$	$\Rightarrow h_\theta(x) = 0 + 1(1)$ $h_\theta(x) = 1$	$h_\theta(x) = 0 + 0.5(1)$ $= 0.5$	$h_\theta(x) = 0 + 0(1)$ $= 0$
$x=2$	$h_\theta(x) = 0 + 1(2)$ $h_\theta(x) = 2$	$h_\theta(x) = 0 + 0.5(2)$ $= 1$	$h_\theta(x) = 0 + 0(2)$ $= 0$
$x=3$	$h_\theta(x) = 3.$	$h_\theta(x) = 0 + 0.5(3)$ $= 1.5$	$h_\theta(x) = 0.$

*

Cost function

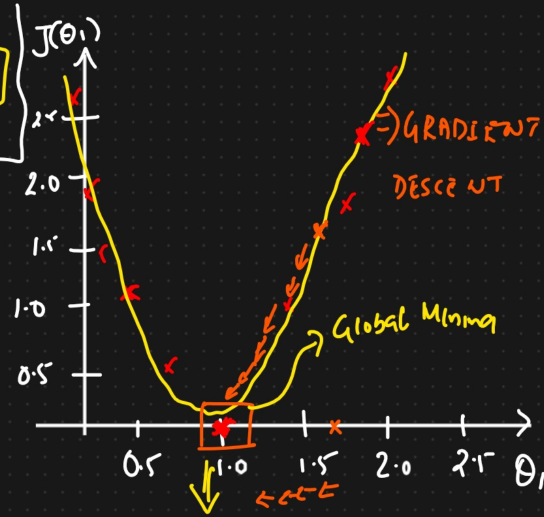
$\theta_0 = 0 \Rightarrow h_\theta(x) = \theta_1 x_i \quad \boxed{\theta_1 = 1}$

$$J(\theta_1) = \frac{1}{n} \sum_{i=1}^{n} (\underbrace{y_i - h_\theta(x)}_{\text{Error}})^2$$

$n = 3$

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

$J(\theta_1) = 0$



Cost function will

Error will be Minimized

Cost fn $\quad \theta_1 = 0.5$

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

$$= \frac{1}{3}\left[0.25 + 1 + 2.25\right]$$

$J(\theta_1) = 1.16$

Cost fn : $\quad \theta_1 = 0.$

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

$$= \frac{1}{3}\left[1 + 4 + 9\right]$$

$$J(\theta_1) = \frac{14}{3} = 4.66$$

Convergence Algorithm

{optimize the change of θ_1}

Repeat until Convergence

$j = 0, 1$

{

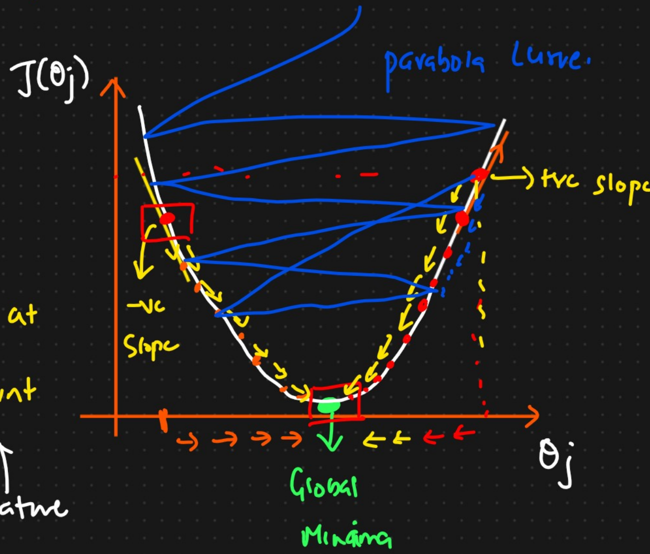
$$\theta_j := \theta_j - \underbrace{\alpha}_{\text{Learning Rate}} \boxed{\frac{\partial J(\theta_j)}{\partial \theta_j}}$$

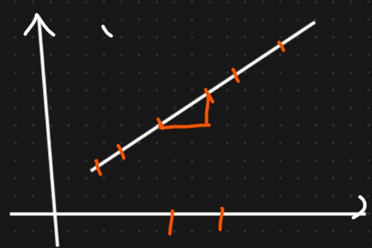
=> Slope at a point

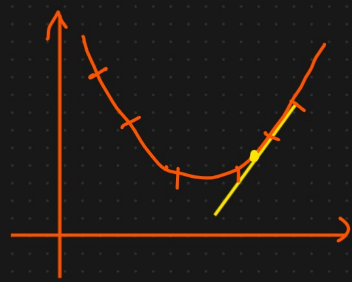
↳ Derivative

}

$\alpha = 0.01$







=> Derivative Slope of a point

$$\theta_1 = \theta_1 - \alpha\,(+ve) \Leftarrow \text{Slope}$$

$$\theta_{1new} = \theta_{1old} - \alpha\,(+ve)$$

$$\boxed{\theta_{1new} << \theta_{1old}}$$ => θ_1 is getting Reduced

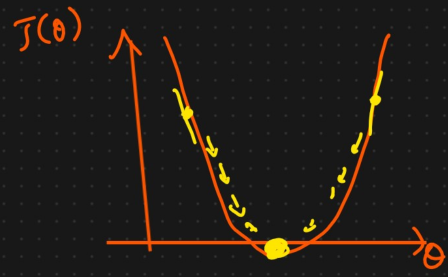
$$\theta_{1new} = \theta_{1old} - \alpha\,(-ve)$$

Learning Rate decides the convergence speed

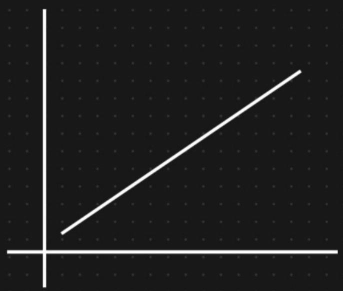
$$\theta_{1new} = \theta_{1old} + (+ve)$$

$$\boxed{\theta_{1new} >> \theta_{1old}}$$

Conclusion



GRADIENT DESCENT



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

Convergence Algorithm

$$\boxed{J(\theta_0, \theta_1) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - h_\theta(x_i)\right)^2}$$

Repeat until convergence

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