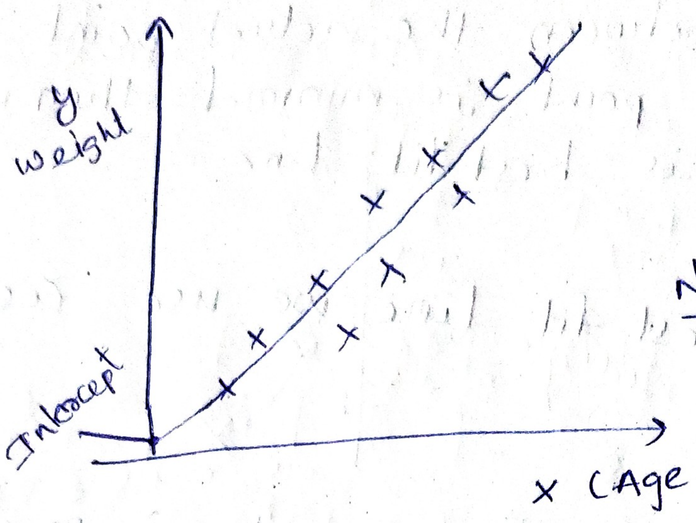
— Sach

linear regression



Train Dataset
↓
model
↓
New Age → [Hypothesis testing] → o/p weight

y is linear function of x

we can represent line in many ways

Equation of straight line → $y = mx + c$

or

$y = \beta_0 + \beta_1 x$

or

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

θ_0 = Intercept

θ_1 = slope (or) coefficient

* When $x = 0$ then $h_\theta(x) = \theta_0$ at what point you are meeting the y-axis is intercept

* with one unit moment in x-axis what is the unit moment in y-axis is known as slope (or) coefficient

our main aim is to create best fit line

→ The distance between the actual point and predicted point is minimal then we should say it is best fit line

For finding best fit line we use Cost function

Cost function

$h_\theta(x)$ = predicted

y → real point

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

→ cost function / Squared error function

we use $\frac{1}{2m}$ instead of $\frac{1}{m}$ to make our derivation much simplier

What we need to solve

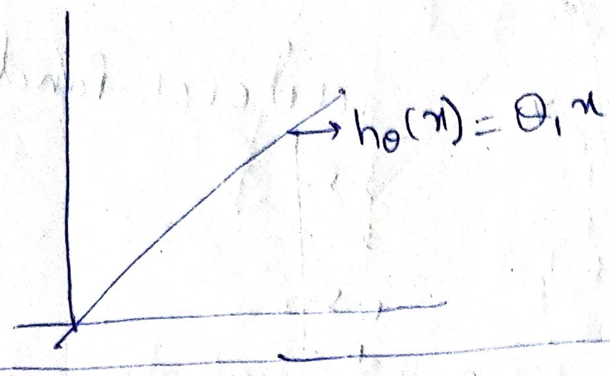
$$\underset{\theta_0, \theta_1}{\text{minimize}} \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\Downarrow$$

$$J(\theta_0, \theta_1)$$

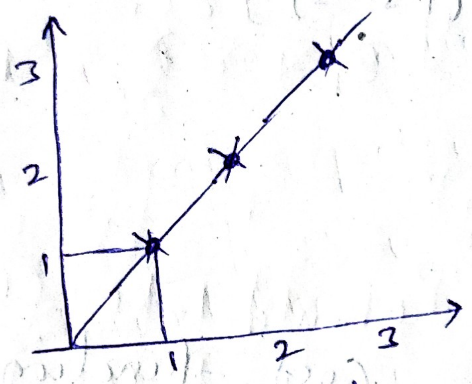
* $h_\theta(x) = \theta_0 + \theta_1 x$

If $\theta_0 = 0$

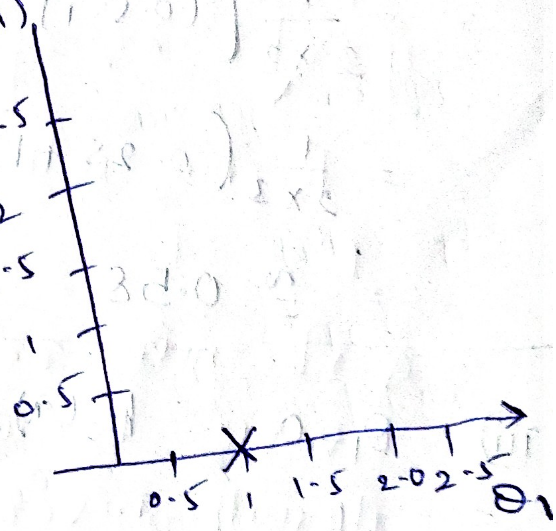


$h_\theta(x) = \theta_1 x^i$ $\quad \theta_1 = 1$

given point (1,1) (2,2) (3,3)



Cost function



actual point and predicted point become same

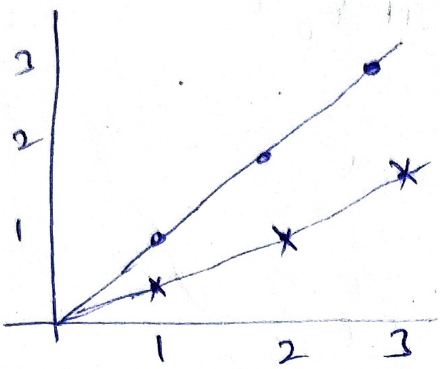
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{3} \left(h_\theta(x^i) - y^{(i)}\right)^2$$

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

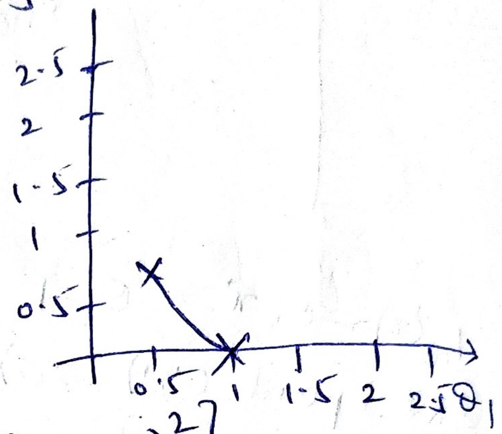
$$J(\theta_1) = 0$$

when $\theta_1 = 1$ then $J(\theta_1) = 0$

ii) $\theta_1 = 0.5$ $\qquad h_\theta(x) = \theta_1 x$



$J(\theta_1)$ cost function



$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{3} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

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

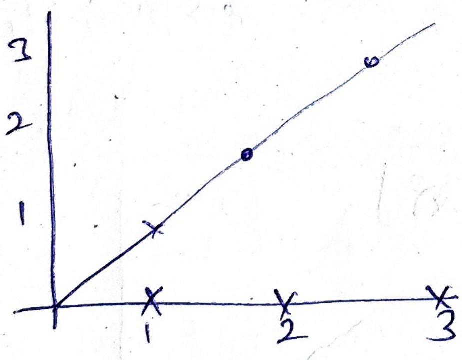
$$= \frac{1}{2 \times 3} (0.25 + 1 + 2.25)$$

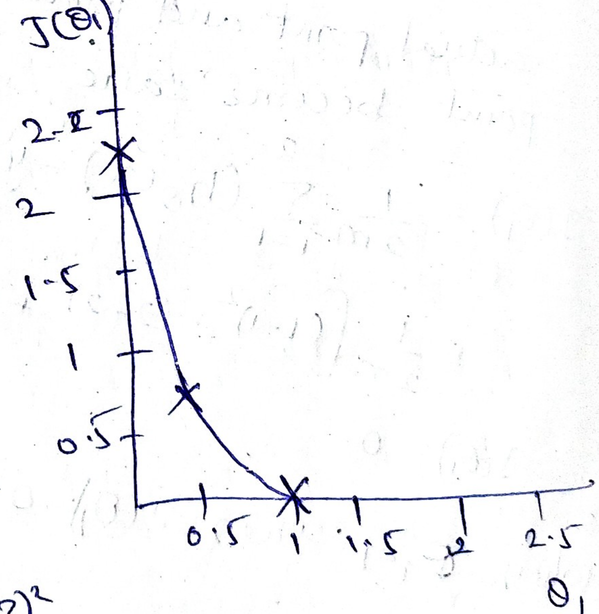
$$= \cong 0.58$$

---

iii) $\theta_1 = 0$, $\qquad h_\theta(x) = \theta_1 x$

Cost function





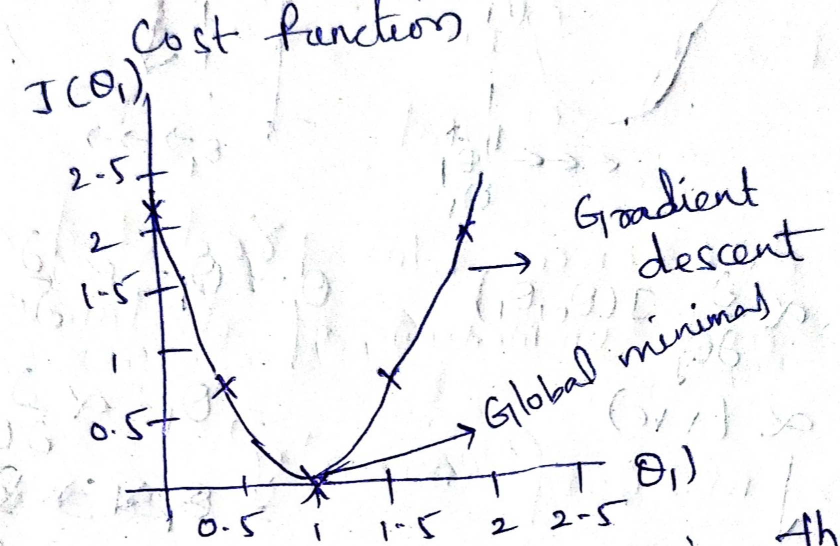
$$J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{2} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

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

$$= \frac{1}{2 \times 3} (1 + 4 + 9)$$

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

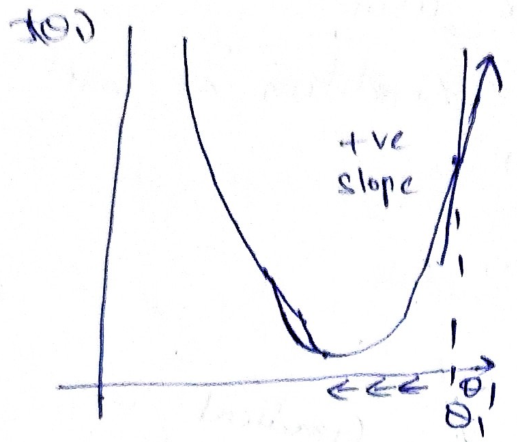
Similarly we solve using
$\theta_1 = 1.5,\ \theta_1 = 2;\ \theta_1 = 2.5$ Condition at last
we get cost function

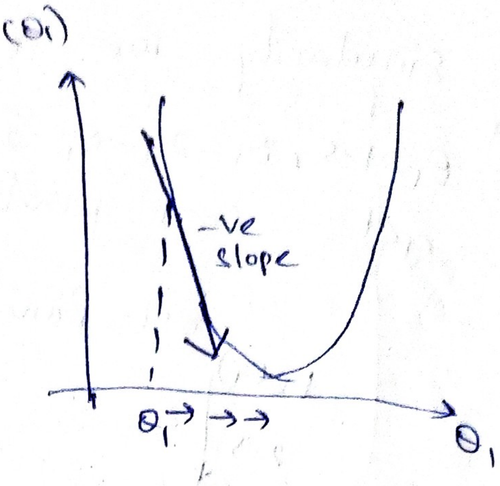


we get above graph by using those
Condition
In that $\theta_1 = 1$ have global minimal so
at $\theta_1 = 1$ is the best fit line
→ like this we do by updated until we get global minimal
We use Converge algorithm

Repeat until Converge

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





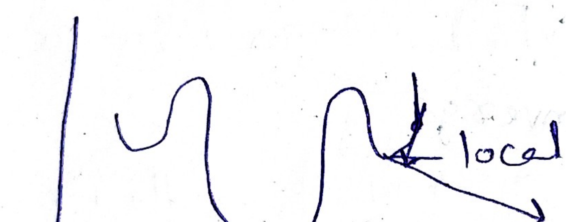
$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$
$$= \theta_1 - \alpha\,(+ve)$$

↓

learning rate

$$\theta_1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$
$$= \theta_1 - \alpha \frac{\partial}{\partial \theta_1} (-ve)$$
$$= \cancel{\theta_1 - \alpha \frac{\partial}{\partial \theta}} \doteq \theta_1 + \alpha(+ve)$$

$\alpha \rightarrow$ learning rate we should take small value because if we take large value it will never reach global minima



$$\theta_1 := \theta_1 - \alpha(\cancel{+ve})$$
$$\theta_1 := \theta_1 - \alpha(0)$$
$$= \theta_1 .$$

we will stake in local minima

we used RMS Prop, Adam optimizer are used to solve local minima problem this are ~~manu~~ mainly occured in Deep ~~learning~~ learning

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

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

If $j=0$

$$\frac{\partial}{\partial\theta_0} J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_0}\frac{1}{2m}\sum_{i=1}^{m} ((\theta_0+\theta_1 x^{(i)}) - y^{(i)})^2$$

$$= \frac{\cancel{2}}{\cancel{2}m}\sum_{i=1}^{m} ((\theta_0+\theta_1 x^{(i)}) - y^{(i)})$$

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

If $j=1$

$$\frac{\partial}{\partial\theta_1} J(\theta_0,\theta_1) = \frac{\partial}{\partial\theta_1}\frac{1}{2m}\sum_{i=1}^{m} ((\theta_0+\theta_1 x^{(i)}) - y^{(i)})^2$$

$$= \frac{\cancel{2}}{\cancel{2}m}\sum_{i=1}^{m} ((\theta_0+\theta_1 x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

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

Repeat until converge

$\{$

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

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

$\}$

Performance metrices

R^2 and Adjusted R^2

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

high

$\hat{y}_i$ = difference b/w predicted and actual and summation of distance

$\bar{y}$ => mean of ~~mean~~ 'y'

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

↓ small number

location Price = Big number

$\to R^2 = 85\%$

→ when we add another feature Bedroom then $R^2 = 90\%$.

→ Added another feature Gender then R^2 become 91% ~~bad~~ but Gender is not correlated with Price but R^2 value increases to prevent such problem we use 'Adjusted R-Square'

Adjusted R^2 P = features (or) predictors

$$= 1 - \frac{(1-R^2)(N-1)}{N-P-1}$$

$P=2 \Rightarrow R^2 = 90\%$ R^2 adjusted $= 86\%$ ↓ decreas

$P=3 \Rightarrow R^2 = 91\%$ ↓inc R^2 adjusted $= 82\%$

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

N = No. of datapoint

P = No. of predictor

$P=2 > \boxed{N-P-1} > P=3$