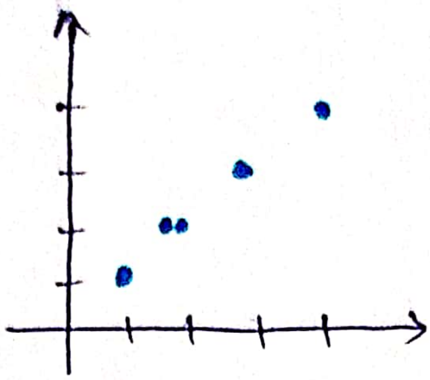
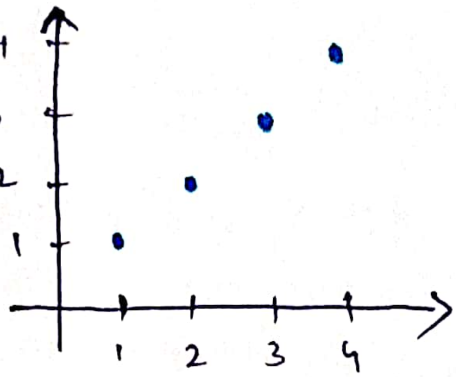
let our dataset is.

X	Y
1	1
2	2
3	3
4	4

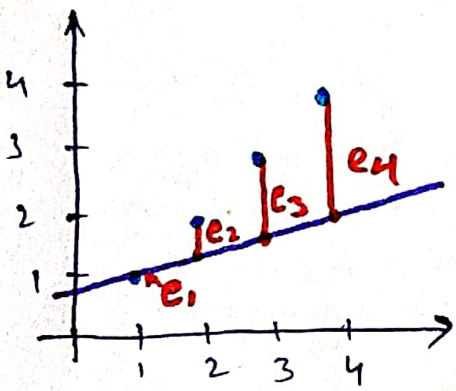


let our dataset is.

	X	Y	
$x_1 \rightarrow$	1	1	$\leftarrow y_1$
$x_2 \rightarrow$	2	2	$\leftarrow y_2$
$x_3 \rightarrow$	3	3	$\leftarrow y_3$
$x_4 \rightarrow$	4	4	$\leftarrow y_4$



We can have many line passing through this dataset but we have to choose the best one which give us lowest error. This error we calculate by cost f^n. i.e we will try to minimize cost f^n.



~~Cost f^n = 8 Σ of all error~~ squo square

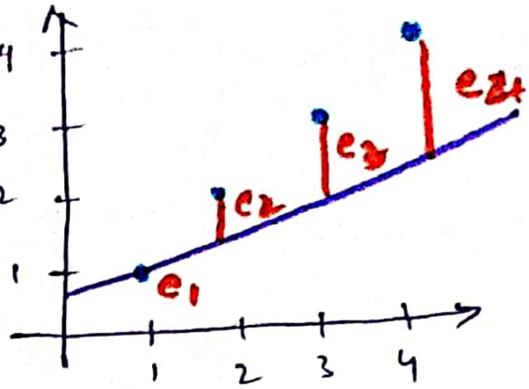
$$\text{Cost } f^n = \frac{(\Sigma \text{ of square of all error})}{2m}$$

$$= \frac{e_1^2 + e_2^2 + e_3^2 + e_4^2}{2 \times (4)} \qquad \boxed{m=4}$$

No. of data points

So let our line be $\hat{y} = mx + b$

$\hat{y}_1 = mx_1 + b$
$\hat{y}_2 = mx_2 + b$
$\hat{y}_3 = mx_3 + b$
$\hat{y}_4 = mx_4 + b$



$e_1^2 = (y_1 - \hat{y}_1)^2 \qquad e_2^2 = (y_2 - \hat{y}_2)^2$
$e_3^2 = (y_3 - \hat{y}_3)^2 \qquad e_4^2 = (y_4 - \hat{y}_4)^2$

$$\text{Cost } f^n_{(J)} = e_1^2 + e_2^2 + e_3^2 + e_4^2 = \frac{1}{2n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{2n}\sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2$$

So we have to minimize cost f^n. for that we have to choose that value of m and b so that cost f^n (J) is minimum.

So $$J = \frac{1}{2n}\sum_{i=1}^{n} \left(y_i - (mx_i + b)\right)^2$$

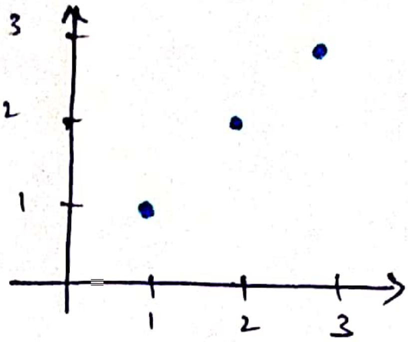
x_i and y_i are fixed. We can change m.

So J is a function of m and b.
J(m, b) so to find extremities what we do.

$$\boxed{\frac{\partial J(m,b)}{\partial m} = 0} \qquad \boxed{\frac{\partial J(m,b)}{\partial b} = 0}$$

lets calculate J for some values of m for a small data set

X	Y
1	1
2	2
3	3



for the sake of simplicity lets take $b = 0$ and we have $\hat{y} = mx$

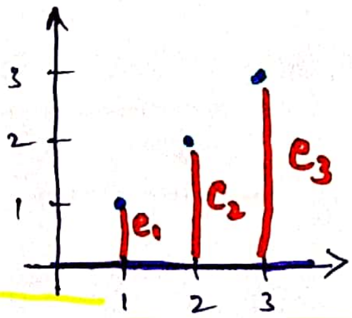
So lets calculate J for $m = [0, 0.5, 1, 1.5]$

(i) $m = 0$

$\hat{y} = 0 \times x = 0$

$e_1 = 1 \quad e_2 = 2 \quad e_3 = 3$

$J = \frac{1^2 + 2^2 + 3^2}{2 \times 3} = \frac{14}{2 \times 3} = 2.33$



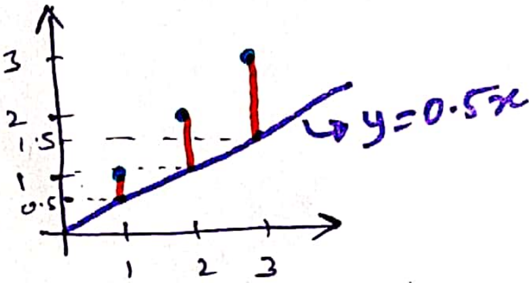
(ii) $m = 0.5$

$\hat{y} = 0.5x$

$e_1 = 1 - 0.5 \quad e_2 = 2 - 1 \quad e_3 = 3 - 1.5$

$J = \frac{(0.5)^2 + 1^2 + (1.5)^2}{2 \times 3} = \frac{0.25 + 1 + 2.25}{6}$

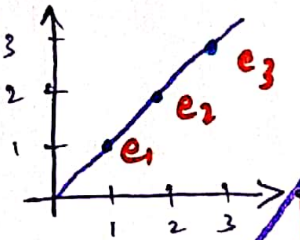
$= \frac{3.5}{6} \approx 0.58$



(iii) $m = 1 \qquad y = x$

$e_1 = 0 \quad e_2 = 0 \quad e_3 = 0$

$J = 0$

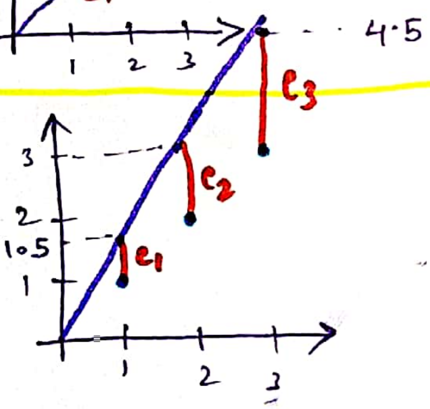


(iv) $m = 1.5 \qquad y = 1.5x$

$e_1 = 1 - 1.5 \quad e_2 = 2 - 3 \quad e_3 = 3 - 4.5$

$J = \frac{e_1^2 + e_2^2 + e_3^2}{2 \times 3} = \frac{(0.5)^2 + (1)^2 + (1.5)^2}{6}$

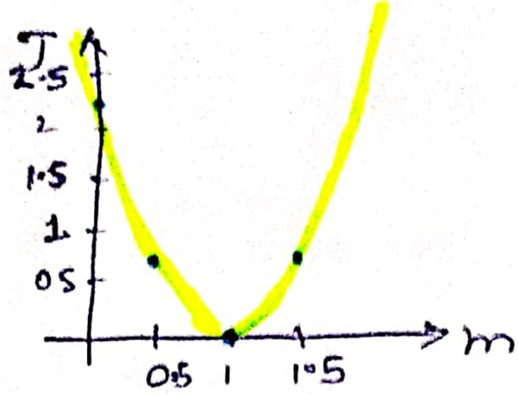
≈ 0.58



Scanned with CamScanner

So lets plot m and J

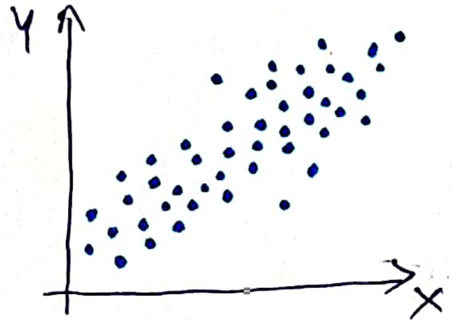
m	J
0	2.33
0.5	0.58
1	0
1.5	0.58

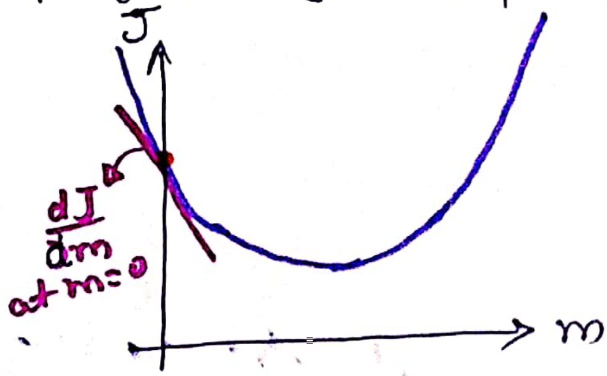


So when we plot m and J we get a curve like Quadratic ~~Eq~~ fn.

· So How can we find the value of m for which J is minimum i.e $\frac{\partial J}{\partial m} = 0$. We have a algorithm called $\boxed{\text{gradient Descent}}$

Assume for some data points X and Y we have : ~~our~~ (We are taking b = 0 for simplicity).





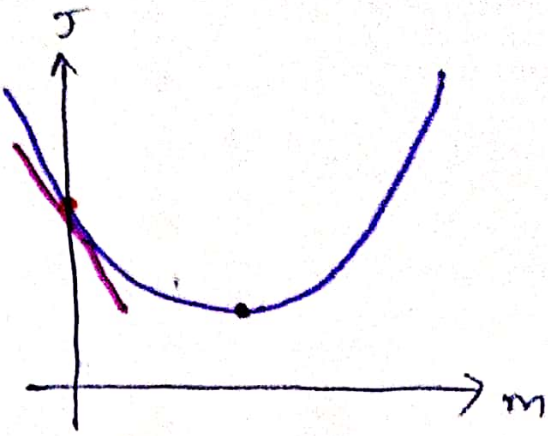
Step 1 : take any random value of m in J v/s m graph . Let m = 0 and find $\frac{\partial J}{\partial m}$ at that point . like

$$J = \frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{2n} \sum_{i=1}^{n} (y_i - (m x_i))^2$$

$$\boxed{\frac{dJ}{dm} = \frac{1}{2n} \sum_{i=1}^{n} 2(y_i - m x_i)\, x_i}$$

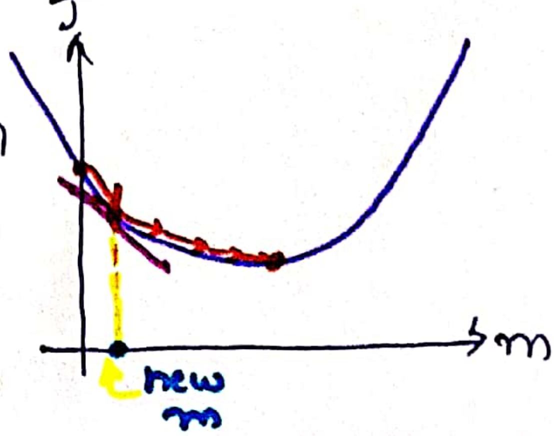
Scanned with CamScanner

Step 2: now calculate the sign of $\frac{dJ}{dm}$. so as per diagram It is (-ve). So to reach $\frac{dJ}{dm} = 0$ we have to add some very small (+ve) number to $\cancel{\frac{dJ}{dm}}$. m.



$\therefore$ Now our new $m = \underset{\text{previous}}{m} +$ (small number)

Step 3: Again do step 1
Calculate $\frac{dJ}{dm}$ for new m and then step 2.
So until $\frac{dJ}{dm} = 0$



So we can do the above 3 step by these equation.

$$\boxed{m = m - \alpha \underbrace{\frac{dJ}{dm}}}$$

α is a very small positive no. called learning rate.

If $\frac{dJ}{dm}$ is (+ve) then this eq^n will decrease m

If $\frac{dJ}{dm}$ is (-ve) this eq^n will increase m

Scanned with CamScanner