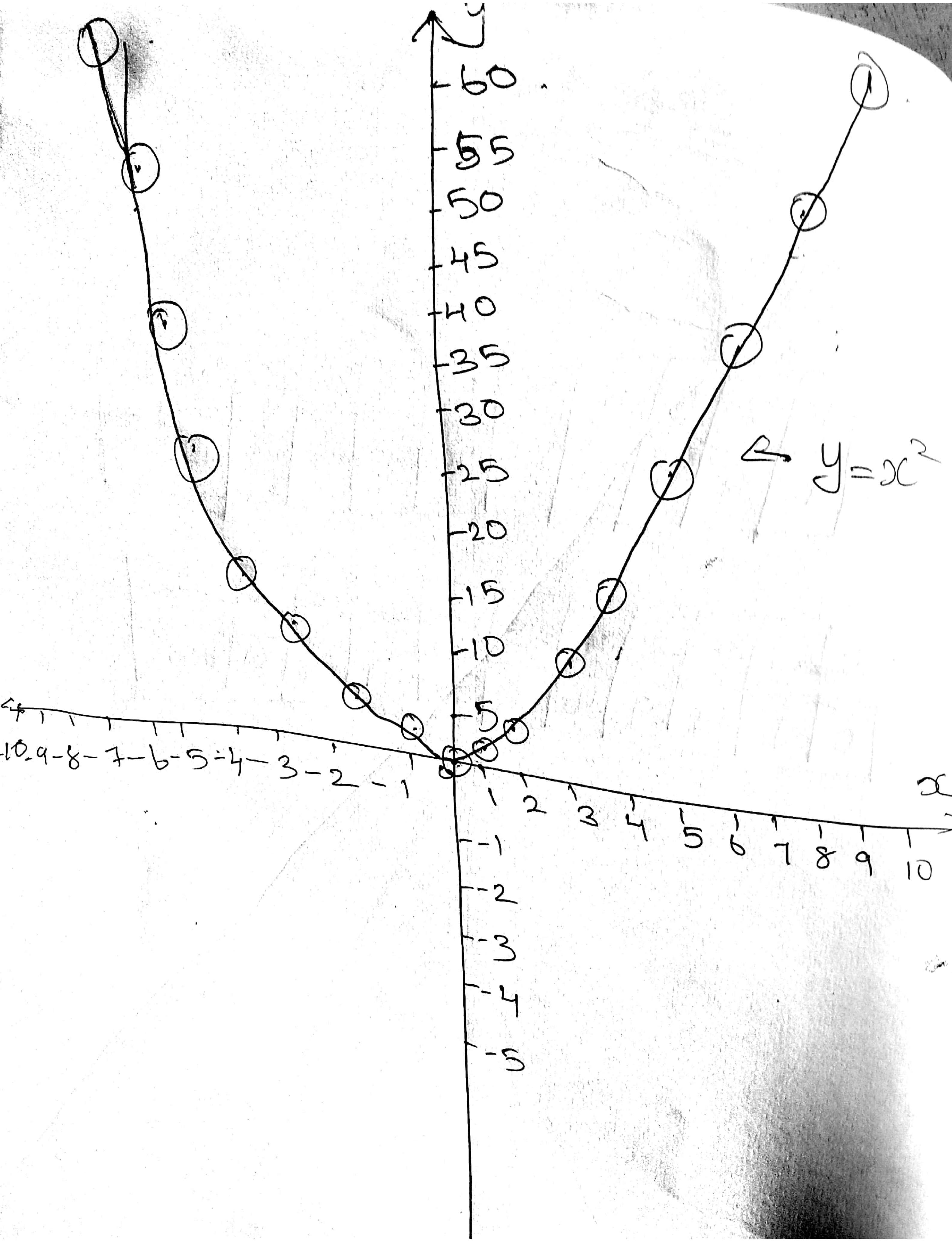
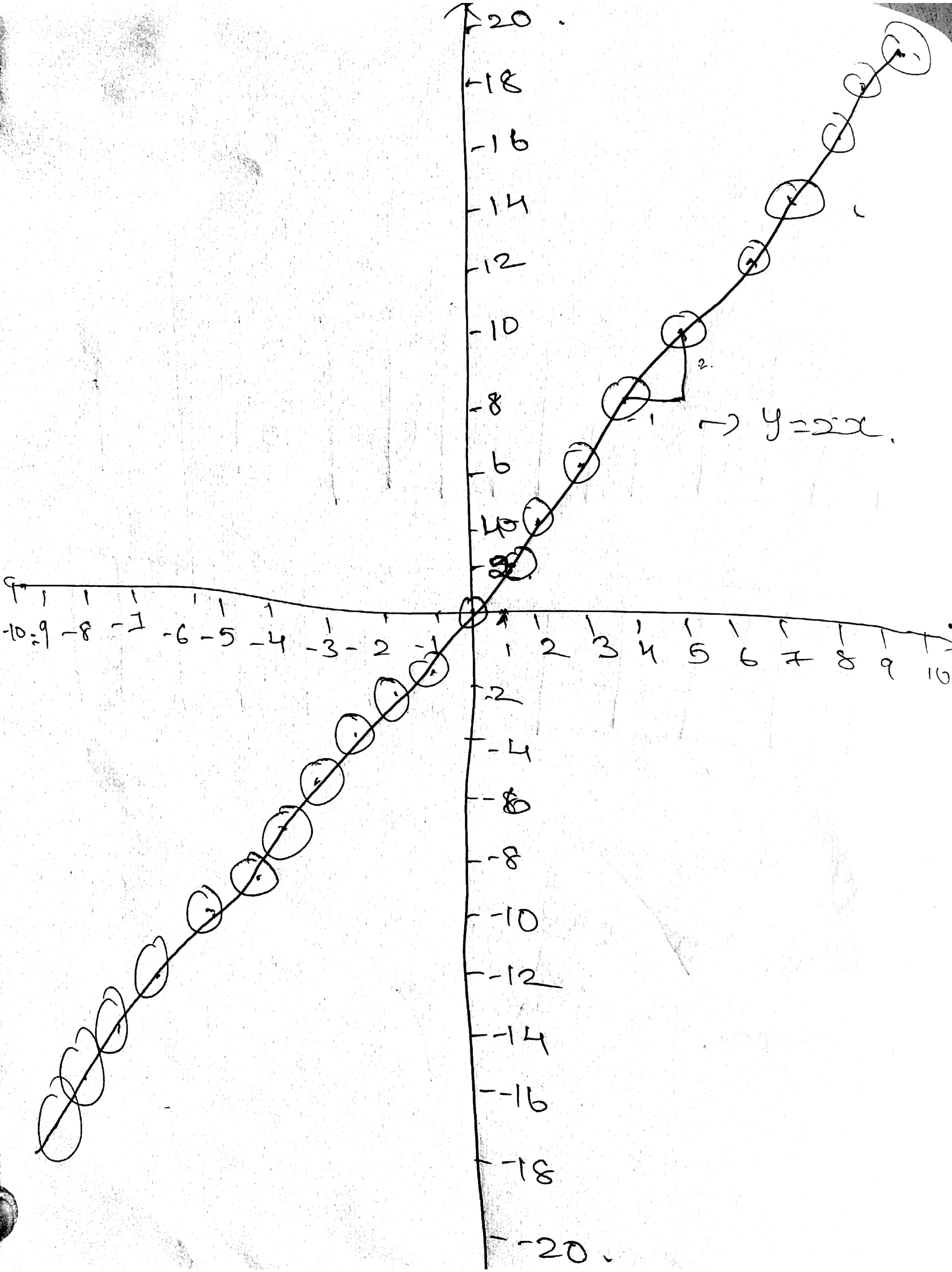
$$y = x^2$$

$$\frac{dy}{dx} = 2x.$$

x	0	1	2	3	4	5	6	7	8	9	10
y	0	2	4	6	8	10	12	14	16	18	20.

x	0	-1	-2	-3	-4	-5	-6	-7	-8	9	-10
y	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20


y
60
55
50
45
40
35
30
25
20
15
10
5
-1
-2
-3
-4
-5
x
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1
1 2 3 4 5 6 7 8 9 10
y=x²

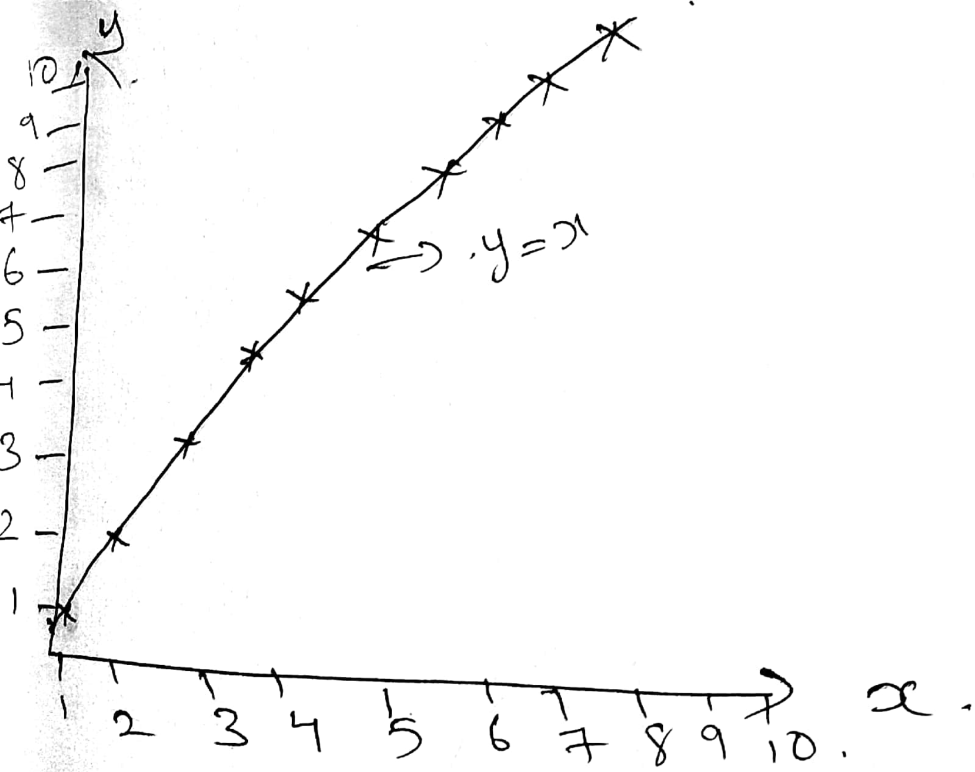


20
18
16
14
12
10
8
6
4
2
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1
1 2 3 4 5 6 7 8 9 10
2
1
y = 2x

Gradient Descent

$y = x$.

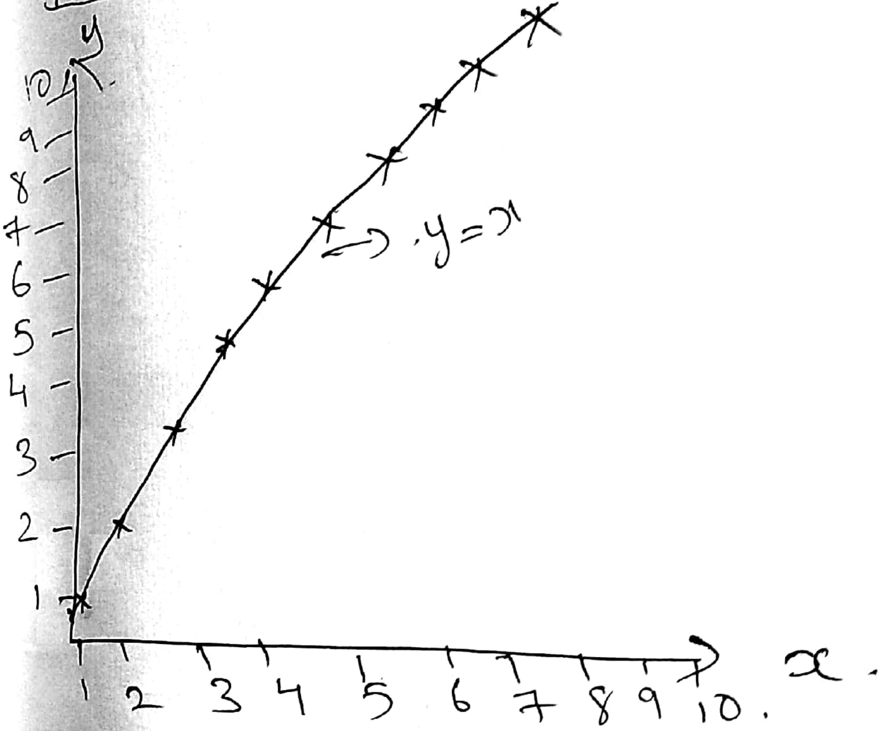
x	0	1	2	3	4	5	6	7	8	9	10.
y	0	1	2	3	4	5	6	7	8	9	10.



Gradient Descent

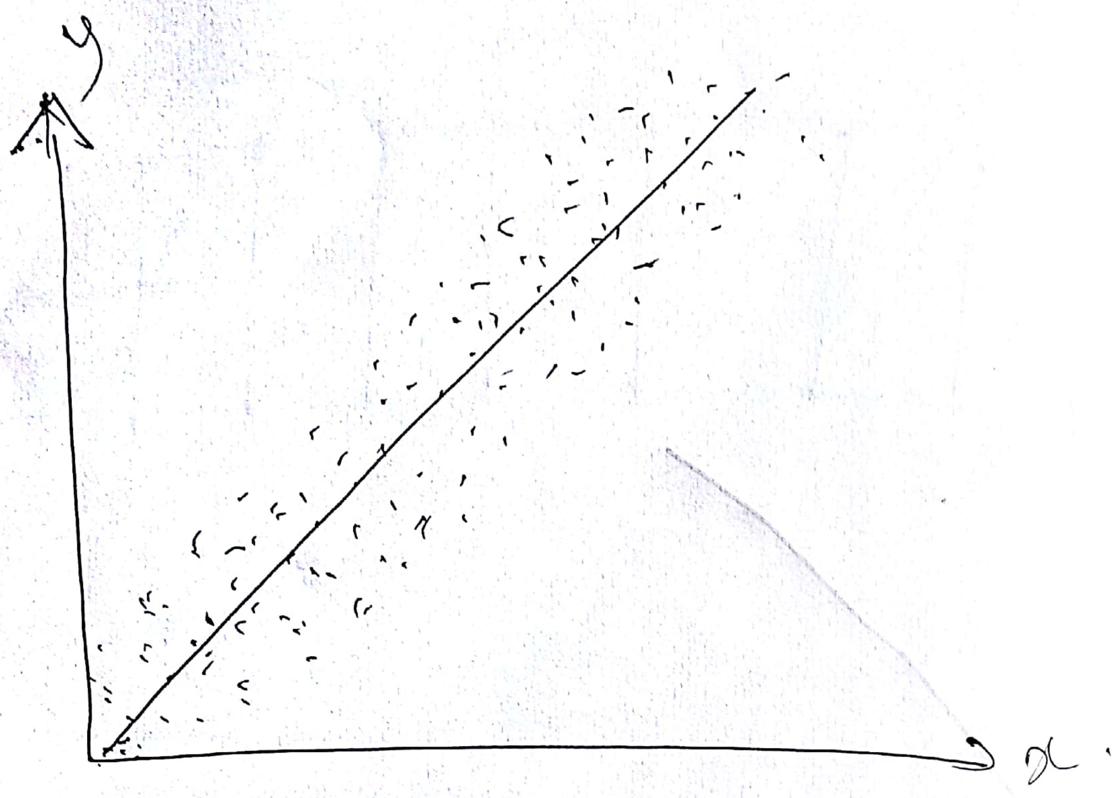
$y = x$.

x	0	1	2	3	4	5	6	7	8	9	10
y	0	1	2	3	4	5	6	7	8	9	10.

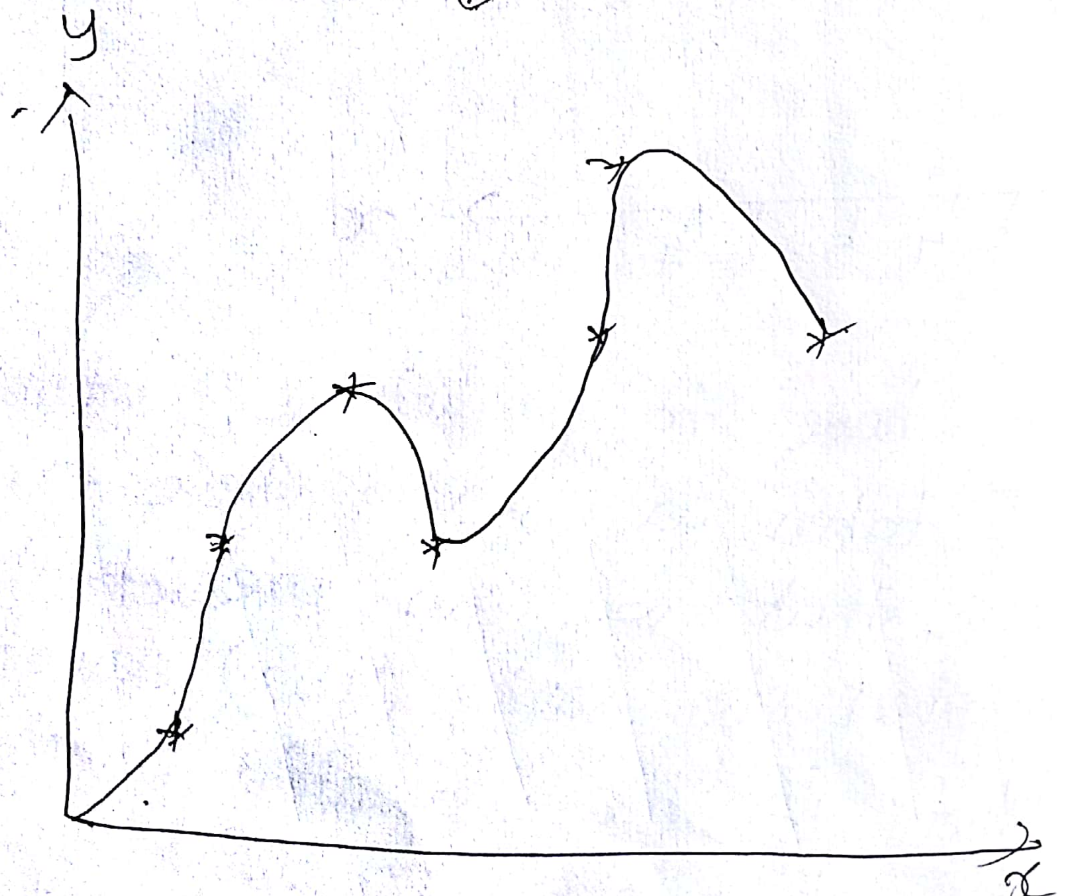


So we have to predict ~~x~~ y when x is given I taken a. ~~a~~ very simple example. of course that we can predict y when x is given. by seeing $(y = x)$ equation. But what if the ~~equation~~ plotting point different ~~a~~ that gives polynomial equation like $y = x^3 + x^2 + x^{56} + x^{36} + 1$.

or figure shown below.



or



for simplification I go with $y=x$ So that we write generalize equation.

$y = \theta x + b$ -> General equation of
or
$y = mx + b$. Linear equation.

So we have to find the θ and b or slope and y-intercept where when x value is given we have to predict the y value for equation $y = x$.

And by seeing that we can say that θ value should be 1 $\theta = 1$ and b value should be 0. But we are going to randomly choose a θ and b value and we have find the θ and b optimum value using Gradient descent. Let's start I'm going to randomly initialise the value of $\theta = 0.5$ & $b = 0.1$ and I'm going to get the θ to 0 and $b = 0$ using Gradient descent. I am using Mean Square Error Cost Function but there are many cost function to choose but again for simplicity I'm going to use. Mean Square Error Cost function.

Mean Square Error $= \frac{1}{2}(\text{Target} - \text{Output})^2$

First I'm going to plot for $\theta = 0.5$ and $b = 0.1$.

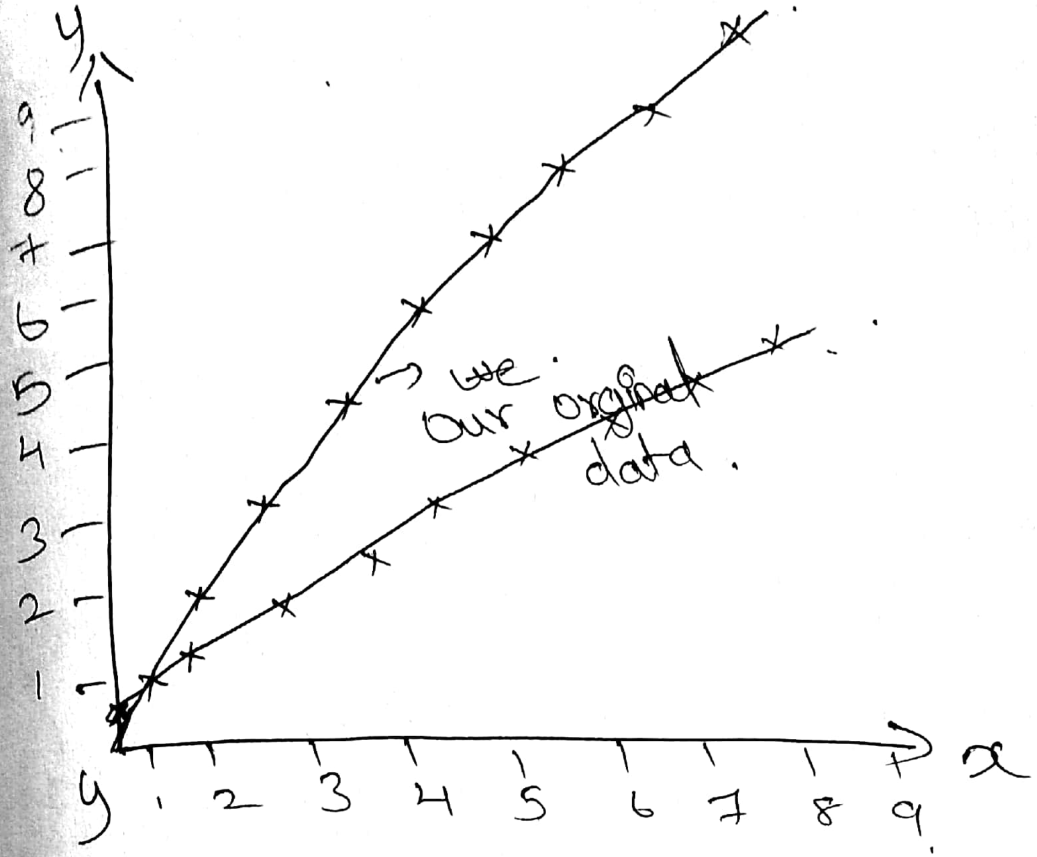
$y = \theta x + b$. $y = (0.5)(x) + 0.1$

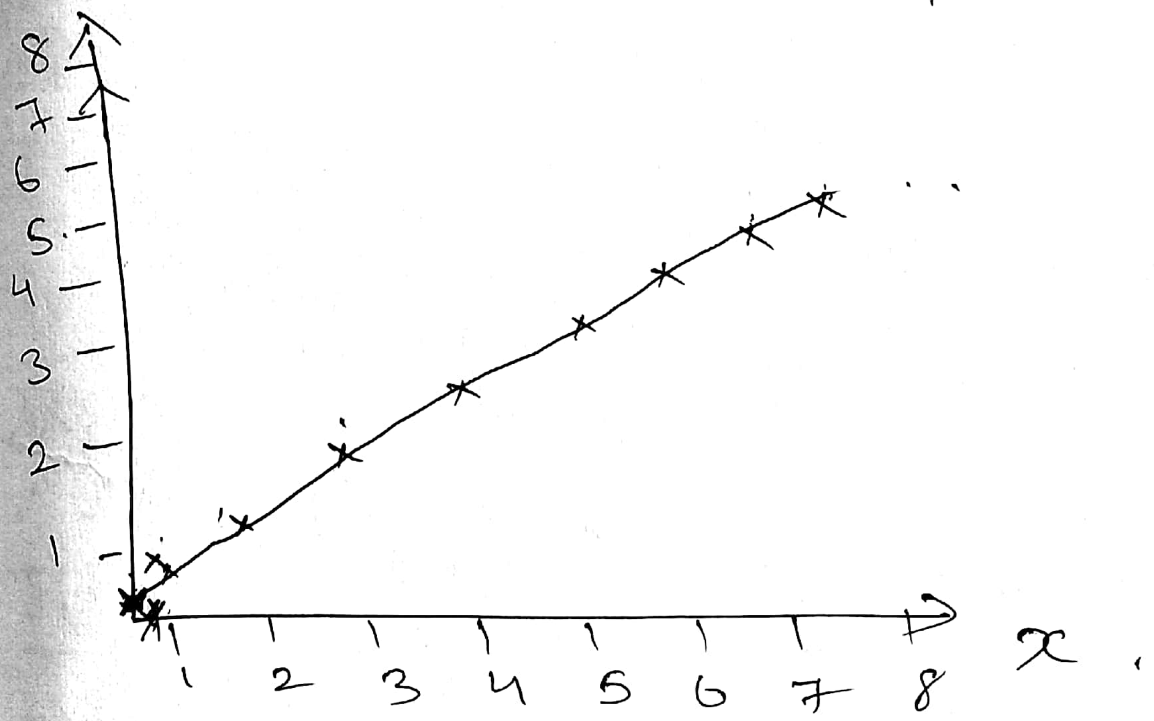
x =	0	1	2	3	4	5	6	7	8
y	0.1	0.6	1.1	1.6	2.1	2.6	3.1	3.6	4.1

But we want.

x	0	1	2	3	4	5	6	7	8
y	0	1	2	3	4	5	6	7	8

So let plot.





By using Gradient descent we have to optimize the ~~problem~~ equation.

$$J(\theta, b) = \text{Cost function} = \frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right)^2$$

$$\frac{\partial J(\theta,b)}{\partial \theta} = \frac{1}{2m} \sum_{i=1}^{m} \left((\theta x^{(i)} + b) - y^{(i)} \right)^2$$

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

$$\frac{\partial J(\theta,b)}{\partial b} = \frac{1}{2m} \sum_{i=1}^{m} \left((\theta x^{(i)} + b) - y^{(i)} \right)^2$$

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

$$= \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right)$$

Gradient descent

$$b = b - \alpha \cancel{\frac{1}{m} \sum_{i=1}^{m}} \left(\frac{\partial J(\theta,b)}{\partial b} \right.$$

$$\theta = \theta - \alpha \frac{\partial J(\theta,b)}{\partial \theta}$$

Scanned by CamScanner

$$b = b - \alpha \frac{\partial J(\theta, b)}{\partial b}.$$

$$\theta = \theta - \alpha \frac{\partial J(\theta, b)}{\partial \theta}.$$

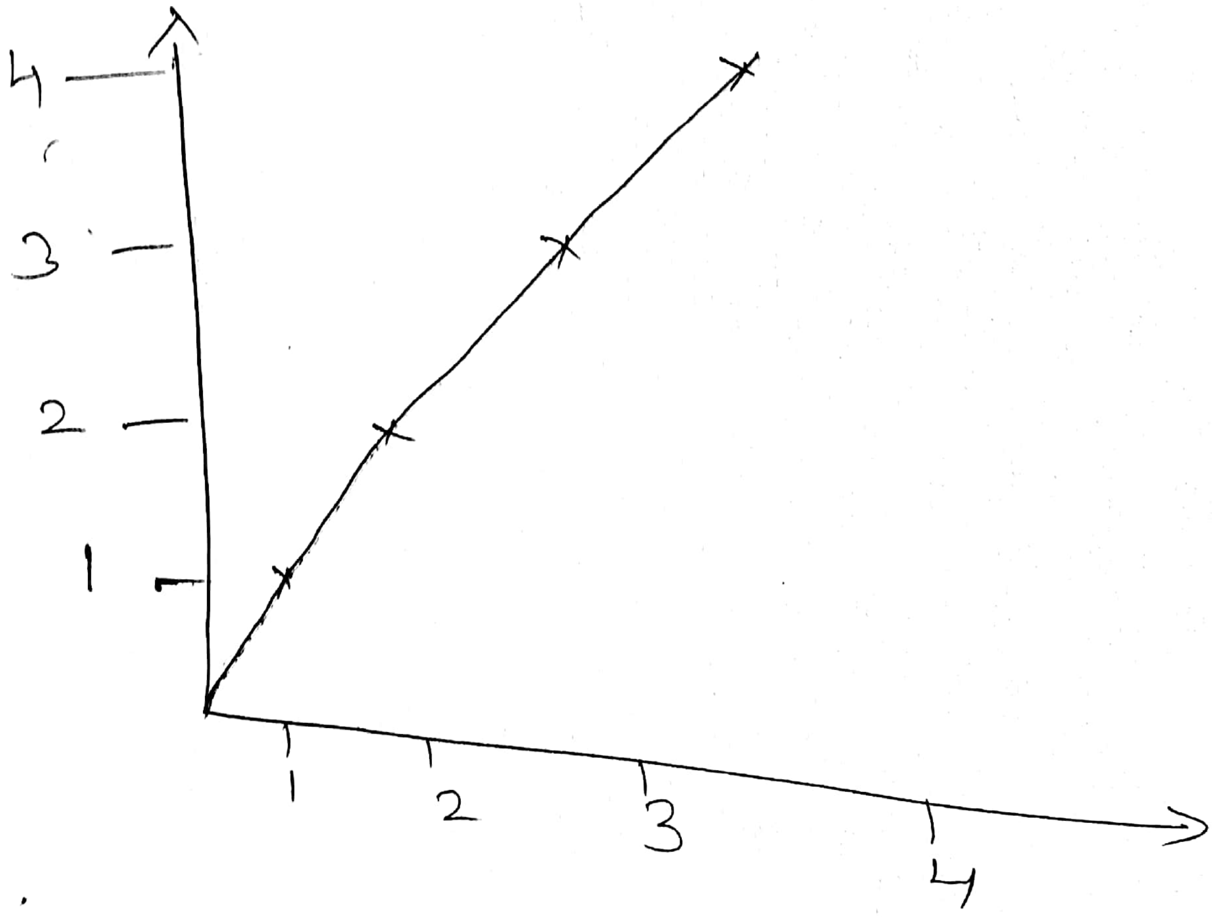
$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right) \cancel{x^{(i)}}$$

$$\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right) x^{(i)}$$

Desired Output (y desired)

$y = x$

x	0	1	2	3	4
y	0	1	2	3	4.



First Random initialization of θ & b.

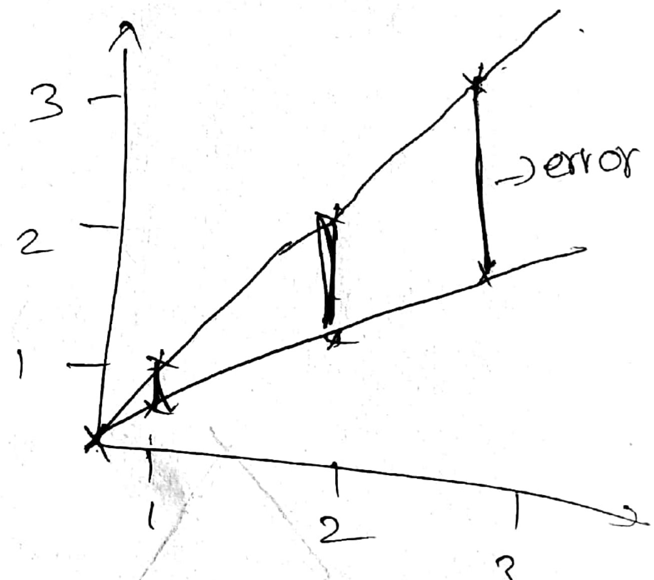
$b = 0$

$\theta = 0.5$

$y = \theta x + b$

$y = 0.5x + 0$.

x	0	1	2
y	0	0.5	1



$$J(\theta, b) = \frac{1}{2m}\left[(0-0)^2 + (0.5-1)^2 + (1-2)^2\right]$$

$m = 3$

$$= \frac{1}{6}\left[0 + 0.25 + 1\right]$$

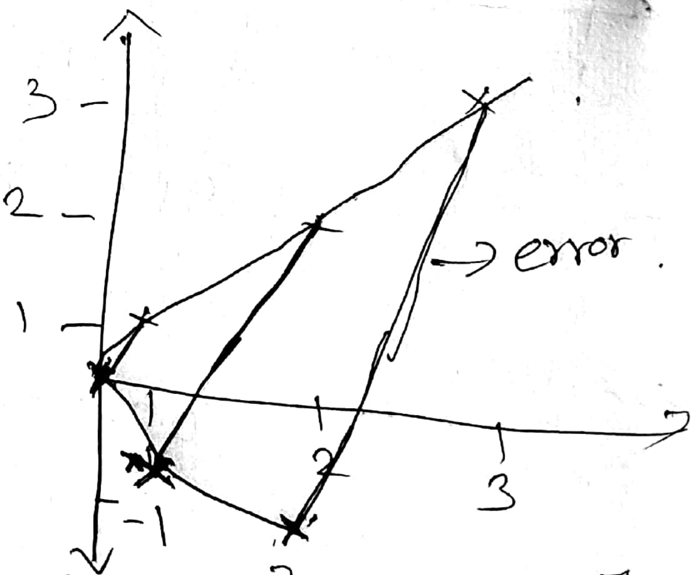
$$= \frac{1.25}{6} \Rightarrow 0.20833$$

$\theta = -0.5$.

$y = \theta x + b$

$y = -0.5x + 0$.

x	0	1	2
y	0	-0.5	-1



$$J(\theta, b) = \frac{1}{2m}\left[(0-0)^2 + (-0.5-1)^2 + (-1-2)^2\right]$$

$$= \frac{1}{6}\left[0 + 2.25 + 9\right]$$

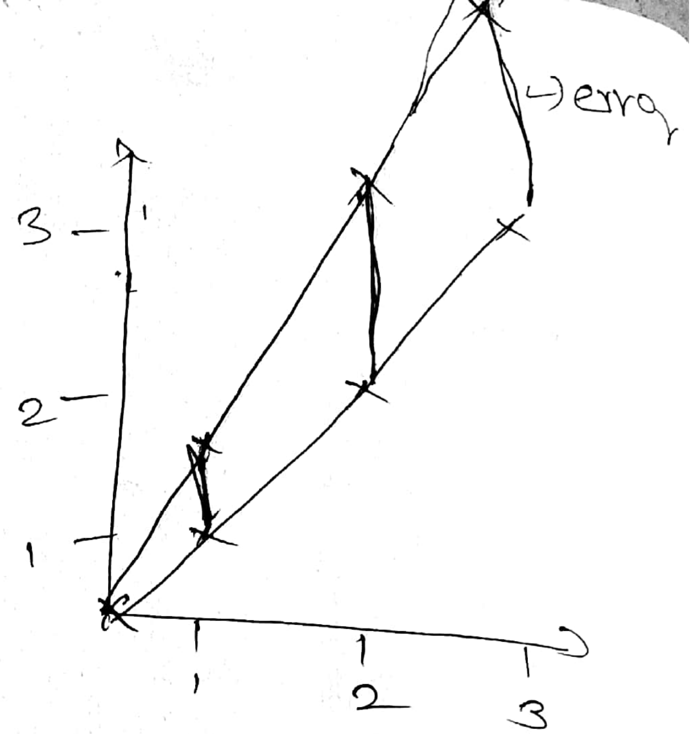
$$= 1.875$$

$\theta = 1.5 \quad b = 0.$

$y = \theta x + b.$

$y = 1.5x + 0.$

x	0	1	2
y	0	1.5	3.



$$J(\theta, b) = \frac{1}{2m}\left[(0-0)^2 + (1.5-1)^2 + (3-2)^2\right]$$

$m = 3$

$$= \frac{1}{6}\left[0 + 0.25 + 1\right]$$

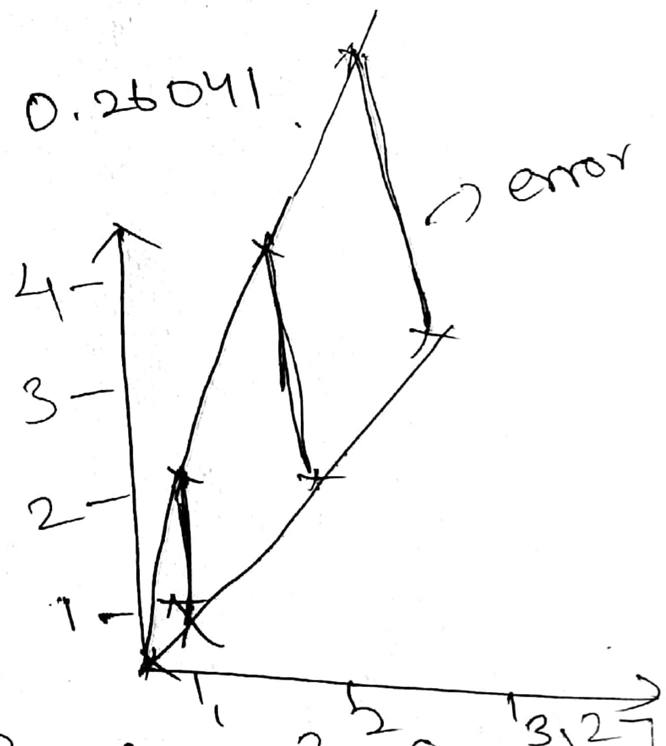
$$= \frac{1.25}{6} \Rightarrow 0.26041.$$

$\theta = 2. \quad b = 0.$

$y = \theta x + b$

$y = 2x + 0.$

x	0	1	2
y	0	2	4



$$J(\theta, b) = \frac{1}{2m}\left[(0-0)^2 + (2-1)^2 + (4-2)^2\right]$$

$$= \frac{1}{6}\left[0 + 1 + 4\right]$$

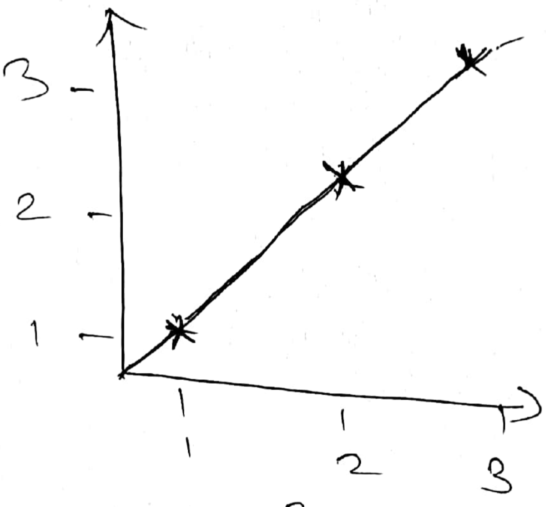
$$= \frac{5}{6} \Rightarrow 0.833$$

$\theta = 1 \qquad b = 0.$

$y = \theta x + b$

$y = (1)x + 0.$

x	0	1	2
y	0	1	2

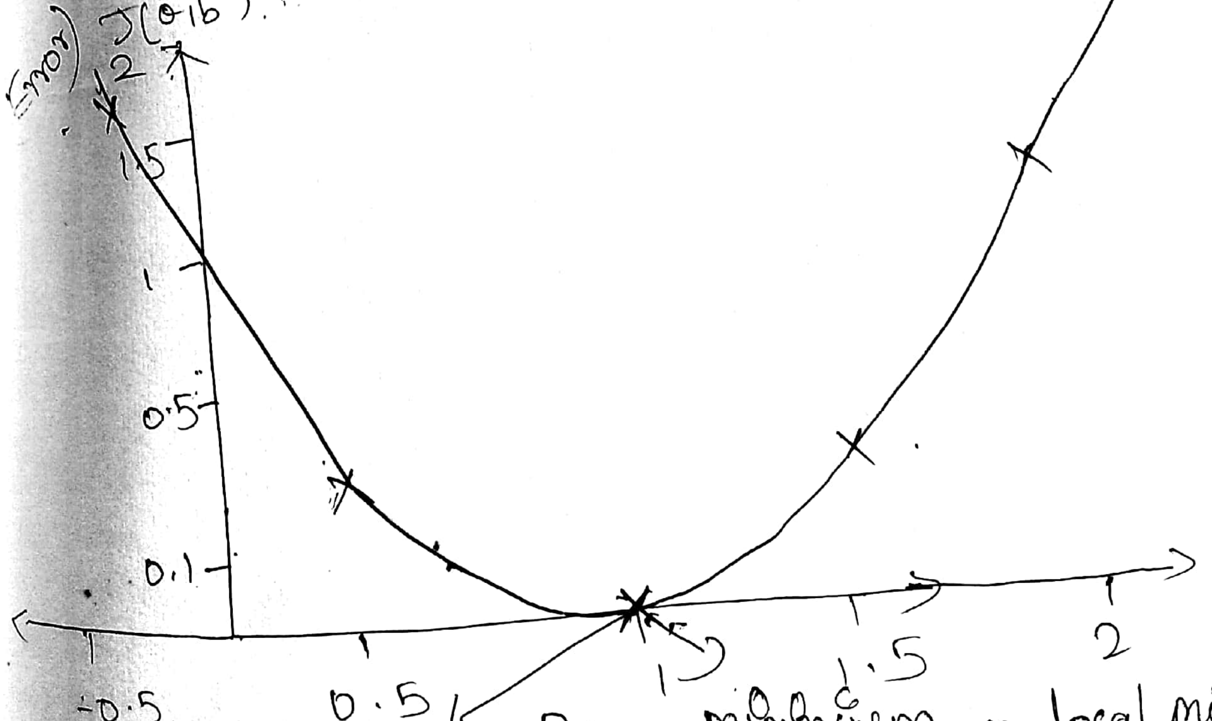


$$J(\theta_1 b) = \frac{1}{2m}\left[(0-0)^2 + (1-1)^2 + (2-2)^2\right]$$

$$= \frac{1}{6}\left[0 + 0 + 0\right]$$

$$= 0$$

When we plot $J(\theta_1 b)$ vs θ_1.



minimum or local minima.

Error is 0 when $\theta_1 = 1$ & $b = 0$.

Gradient descent

$$b = b - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right)$$

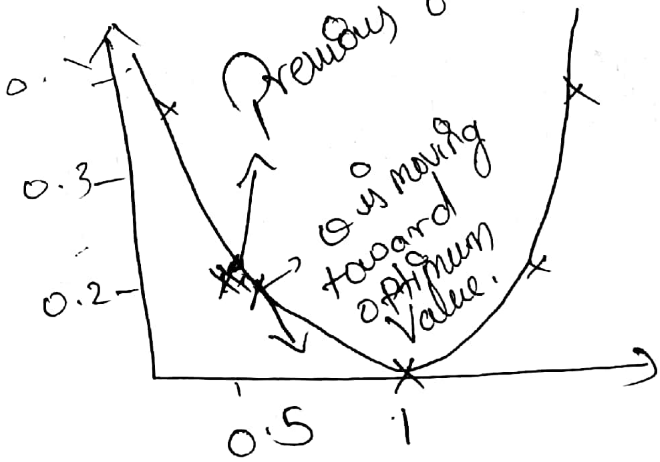
$$\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right) x^{(i)}$$

Here b is already in optimum value. So I'm not going to derive for b.

$\theta = 0.5 \quad b = 0$

$y = \theta x + b$

$y = 0.5x + 0.$



x	0	1	2
y	0	0.5	1

$\alpha = 0.1$

$$\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right) x^{(i)}$$

$$= \theta - (0.1)\frac{1}{3}\left[\underbrace{(0-0)^{(0)} + (0.5-1)^{(1)} + (1-2)^{(2)}}_{-0.8333} \right]$$

$$= 0.5 - (-0.0833) = \cancel{0.5 + 0.05}$$

$$\cancel{= 0.55} \rightarrow \theta \text{ value is increasing.}$$

$$= 0.5833$$

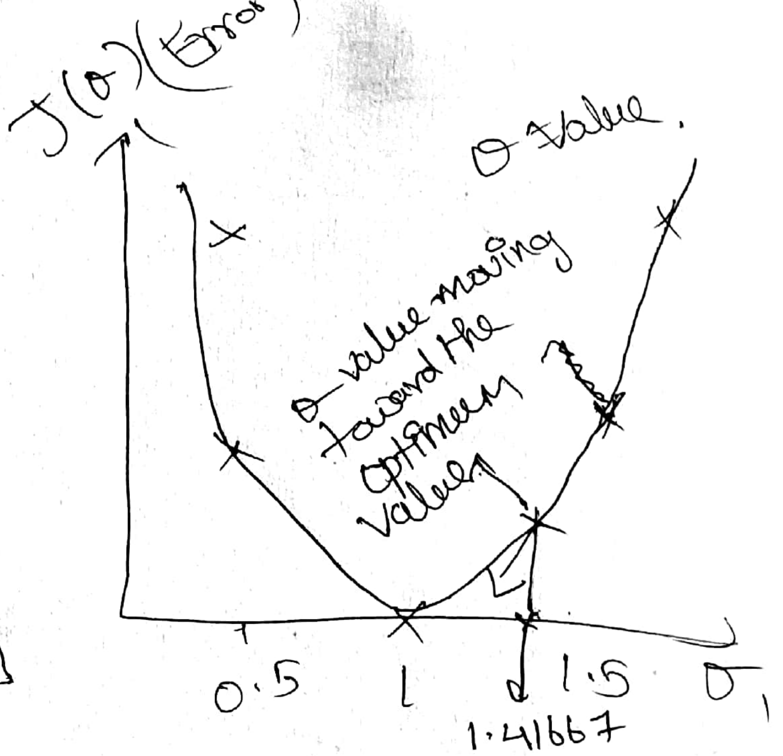
if

$\theta = 1.5.$ $b = 0$

$y = \theta x + b$

$y = 1.5x + 0.$

x	0	1	2
y	0	1.5	3



$\alpha = 0.1$

$$\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired}\right) x^{(i)}$$

$$= \theta - (0.1)\frac{1}{3}\left[(0-0)(0) + (1.5-1)(1) + (3-2)(2)\right]$$

$$= \theta - (0.1)\frac{1}{3}\left[0 + 0.5 + 2\right]$$

$$= \theta - (0.1)(0.8333)$$

$$= 1.5 - 0.08333$$

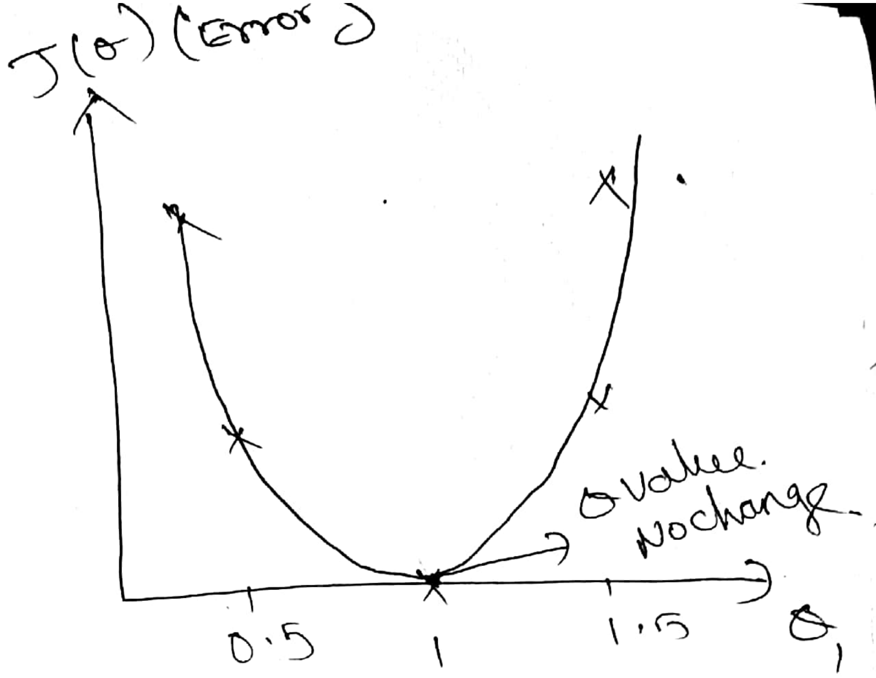
$$= 1.41667 \rightarrow \text{decreasing.}$$

Scanned by CamScanner

$\theta = 1 \qquad b = 0.$

$y = \theta x + b$

$y^{(1)} = 1x + 0.$



x	0	1	2
y	0	1	2

$\alpha = 0.1$

$$\theta = \theta - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)}_{output} - y^{(i)}_{desired} \right) x^{(i)}$$

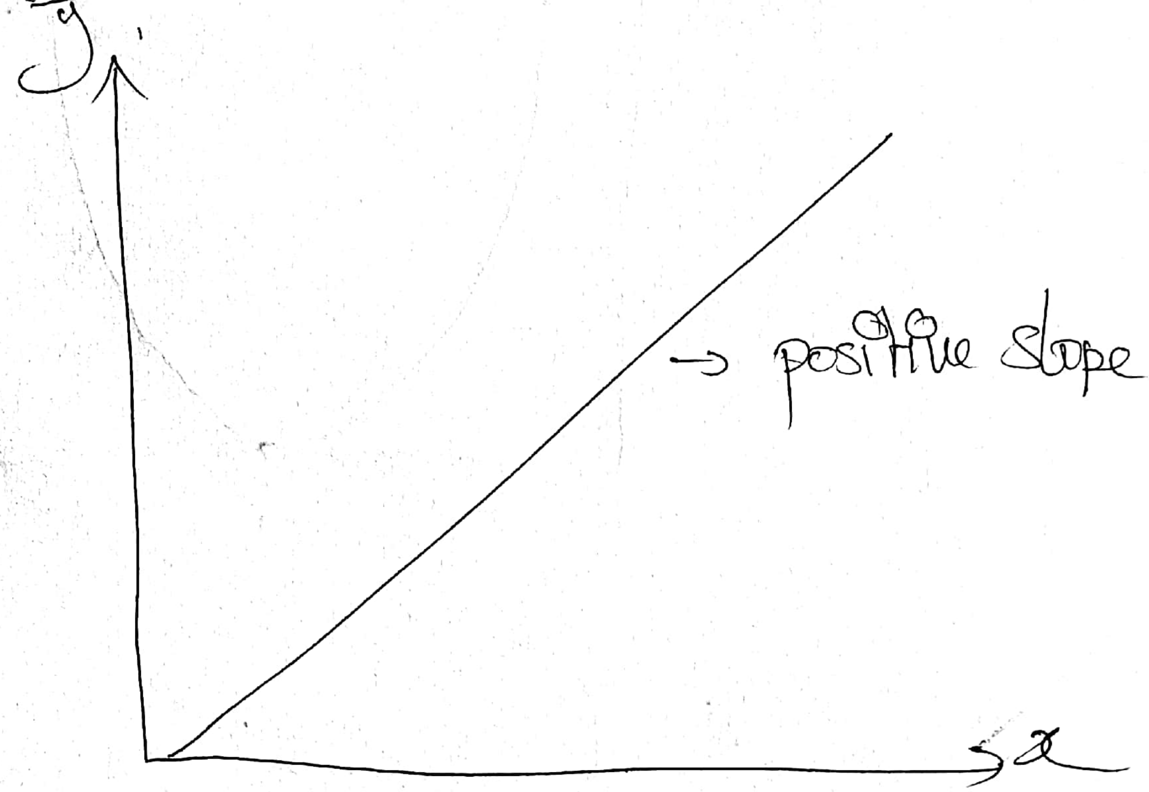
$$= \theta - (0.1) \frac{1}{3} \left[(0-0) + (1-1)(1) + (2-2)(2) \right]$$

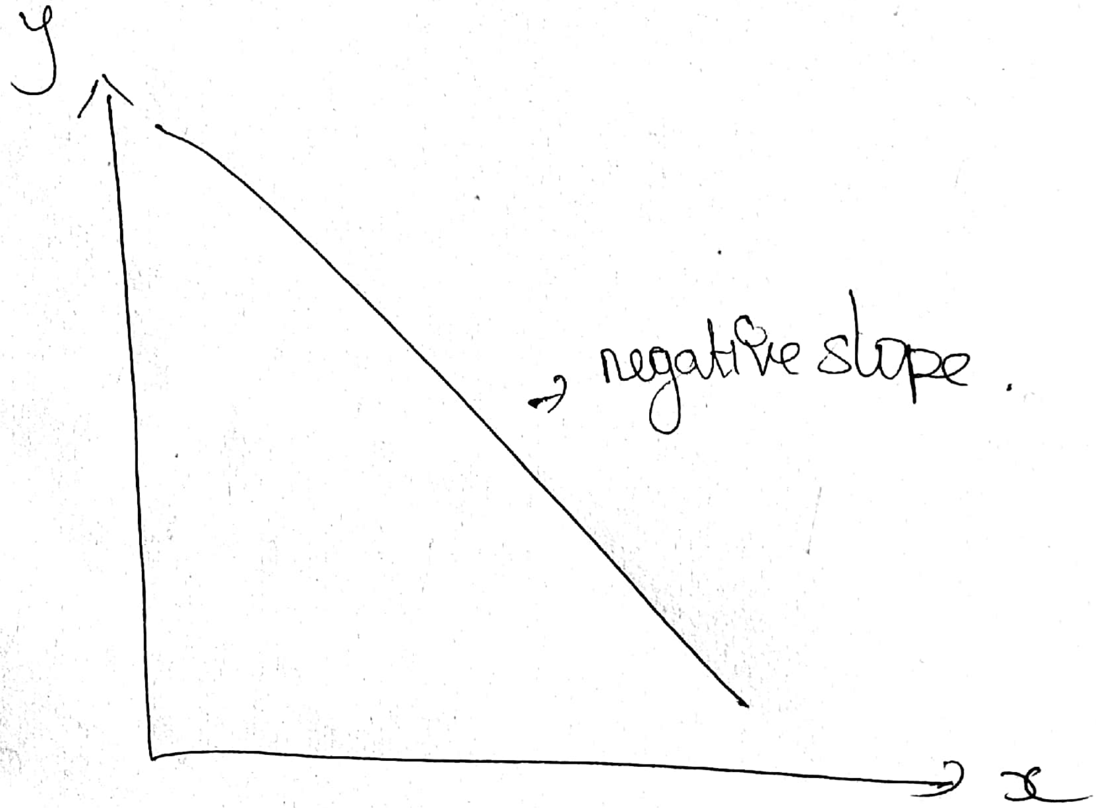
$$= \theta - 0$$

$= 1 - 0 \Rightarrow 1 \rightarrow$ No change because the θ reached the optimum value.

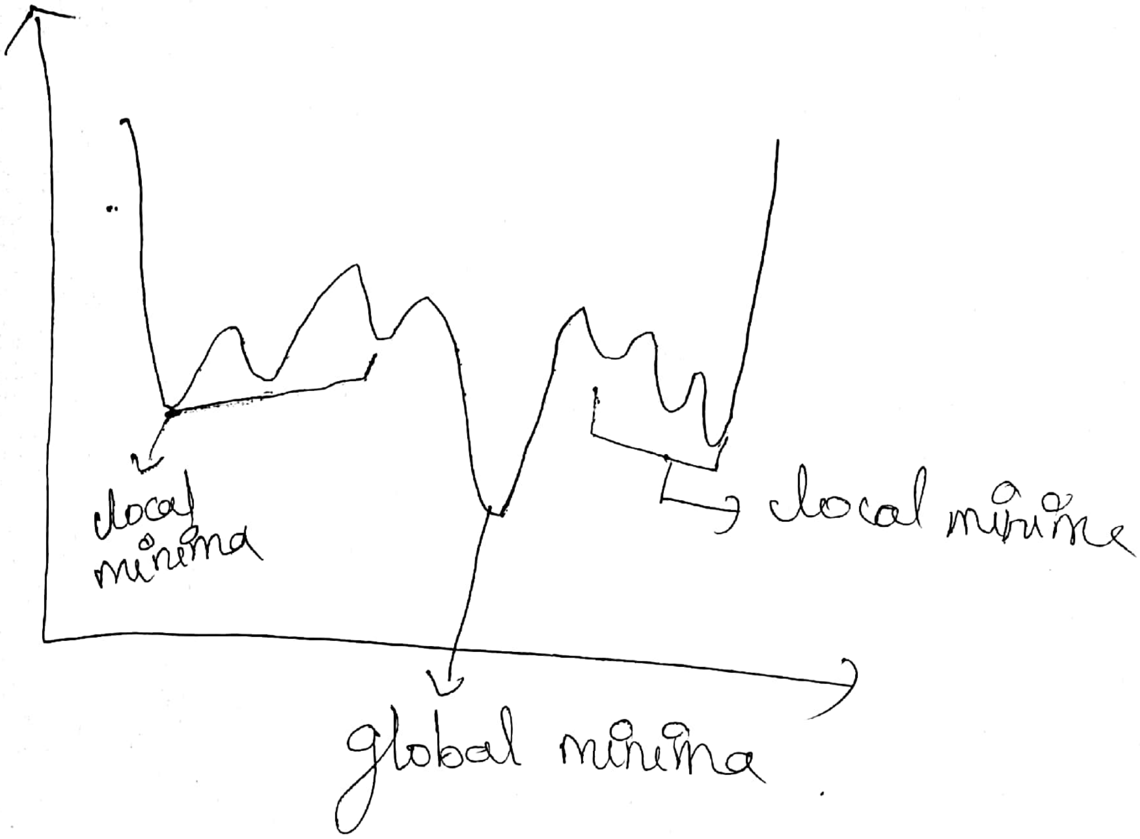
So if you choose $\theta = 1.5$ it will keep on decreasing until it reaches the optimum value of θ in our case $\theta = 1$ for ($y = x$) equation and on other hand if $\theta = 0.5$ it will keep on increasing until it reaches the optimum value of θ in our case $\theta = 1$ for $y = x$.

Slope





Local minima & Global minima



Local Maxima & Global maxima