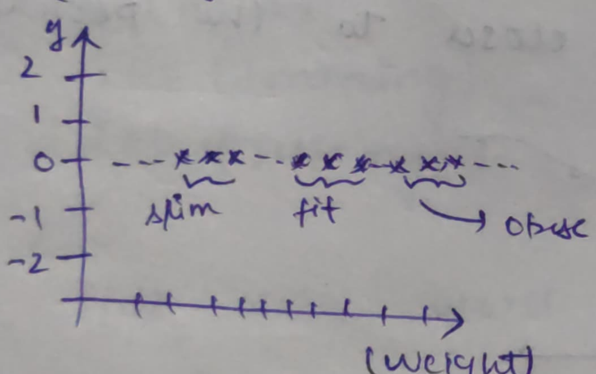
2).

1). <u>univarate analysis</u> :-

We pick up 1 feature & then we determine what the output is.

Height	weight	O/P
180	90	obese
160	50	slim
170	78	fit
190	90	fit.

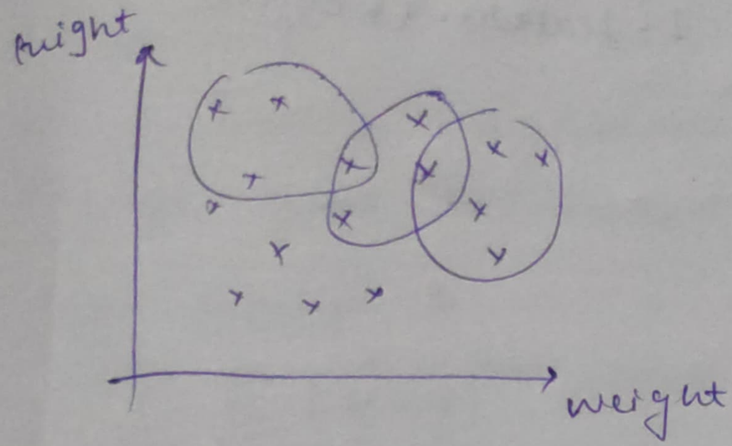
(Taking Just 1 feature) (weight)



But sometimes (fit / slim might overlap or fit / obese might overlap)

due to this we cannot classify the O/P category just by 1 feature.

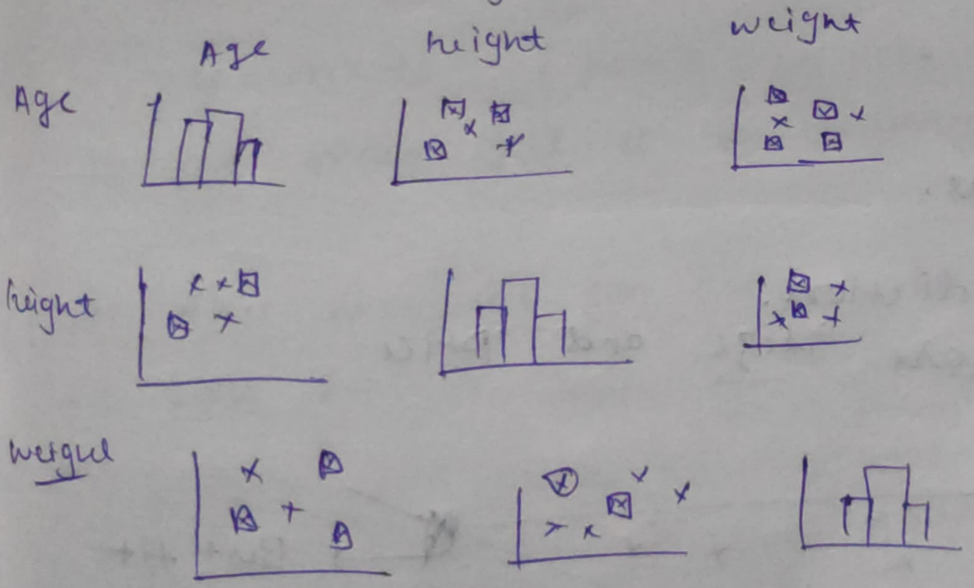
Bivariate Analysis :-



After categorizing them we will apply an ML algorithm & determine the O/P.

Multivariate analysis :-

If there's an age column we'll apply this.



⇒ If age ↑, height ↑ (+ve) correlation

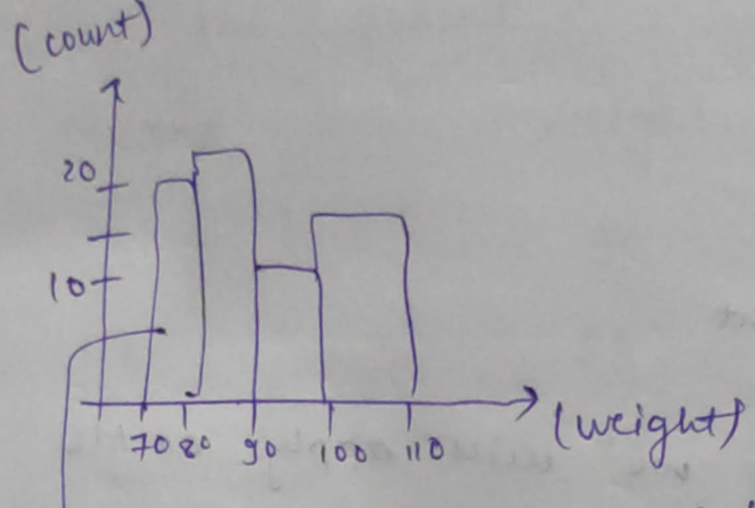
⇒ " " ↓ , " " ↓ (-ve) "

⇒ If evenly increasing / decreasing correlation (0).

(-1, 0, +1)

Histograms EDA :-

Plotting histogram using 1- feature. (weight)



→ This histogram shows that b/w the range of (70-80) there are 10 (counts present)

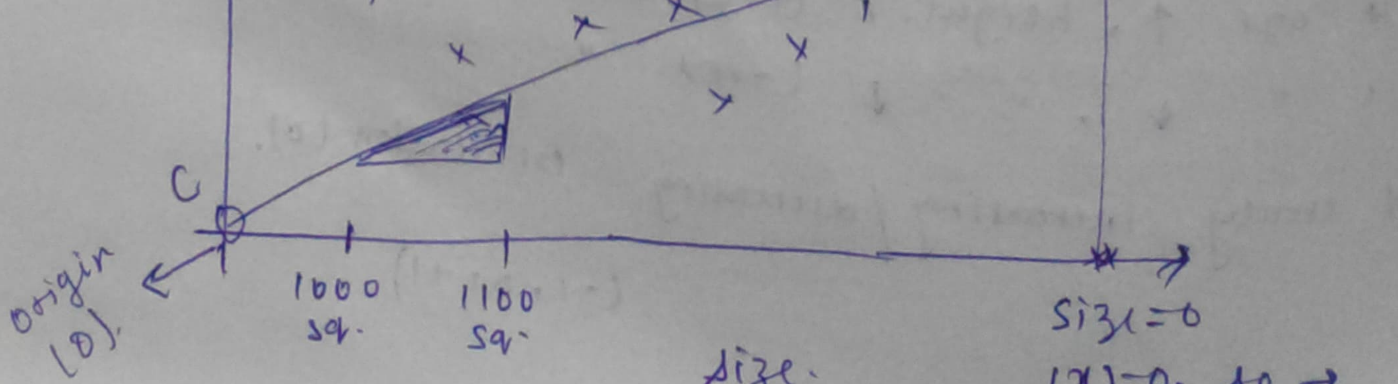
Linear Regression :-

$y = mx + c$

m = slope

c = Interupting line.

→ House price prediction.
The key factors are size and price.



(x) = 0. so →

Please try to go through the next part of the notes, I'll surely improve my handwriting after this Linear Regression notes

$y = m(0)+c$

$y = c.$

m is that, the chang in my x axes what change will be there in my y-axis?

<u>cost function</u> :-

$$\frac{1}{2m}\sum_{i=1}^{m}(\hat{y}-y)^2$$

$\hat{y}$ = Points on the Best fit line.

y = Indicates the real points, points not on the best fit line.

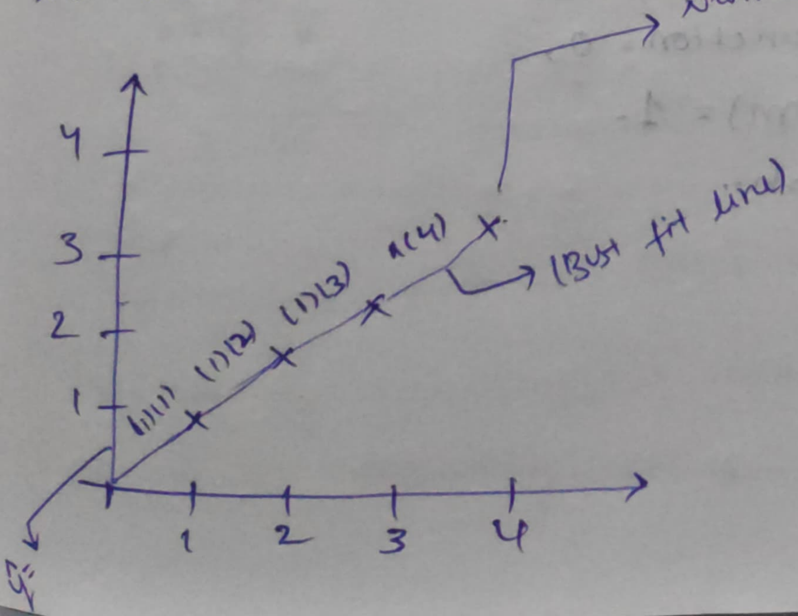
m = All the points in the graph.

so, $\hat{y} = mx+c$ (points on the best fit line)

So our main goal is to minimize $(\underbrace{\hat{y}-y}_{\text{the error}})^2$

like this here we can have a lot of best fit lines,

so lets see a more efficient way.



$y=1, x=1, y=2, x=2$ so,

$\boxed{y=x}$.

⇒ Best fit line :-

$y = mx + c$ → passes o (origin)

$\hat{y} = mx.$

⇒ initializing slope (m=1)

$\hat{y} = (1)(1)$

⇒ $x=2$

$\hat{y} = (1)(2).$

⇒ $x=3$

$\underbrace{\hat{y} = (1)(3)}_{\text{Best fit lines}}$

→ Now we will find cost function →

$$= \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2 \quad \left[\text{There are 3 points so, } m = 3 \right]$$

$$= \frac{1}{2m} \left[(1-1)^2 + (2-2)^2 + (3-3)^2 + (4-4)^2 \right]$$

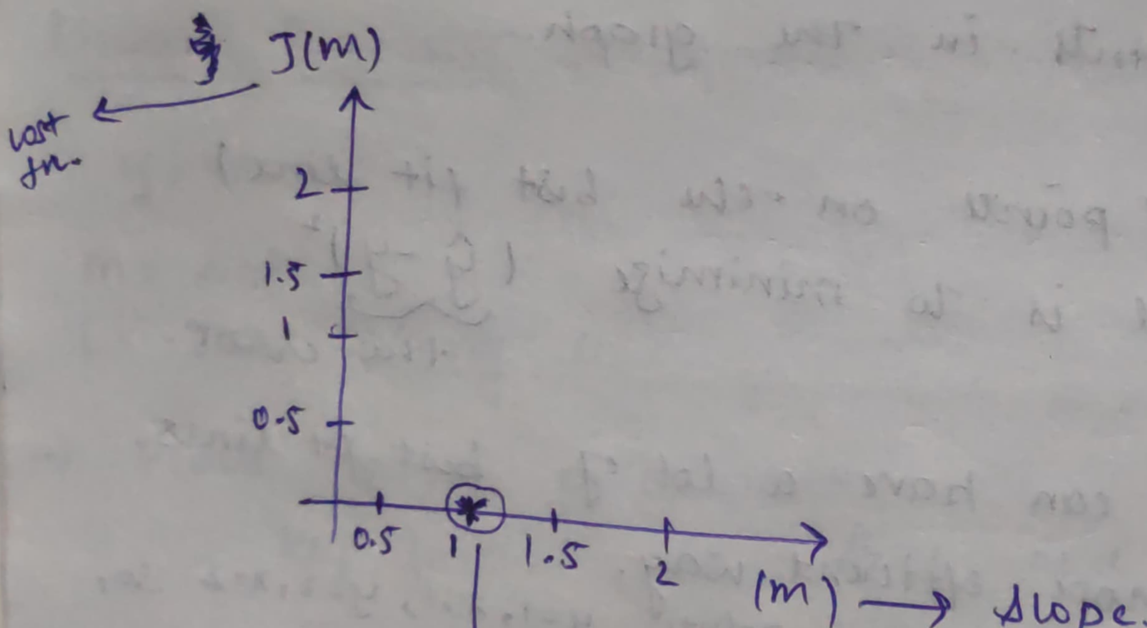
$$\Rightarrow \frac{1}{6}(0) \Rightarrow \underline{\underline{0}}.$$

→ plotting cost function & normal slope

x-axis = cost function

y-axis = slope (m)

So, the graph.



cost function = 0,

slope (m) = 1.

The plotted graph.

→ Changing M value to, $m = 0.5$.

equate the point in $\hat{y} = m(x)$

$\hat{y} = (0.5)(1) = 0.5$

$\Rightarrow$

$\hat{y} = (0.5)(2) = 1$

$\hat{y} = (0.5)(3) = 1.5$ etc.

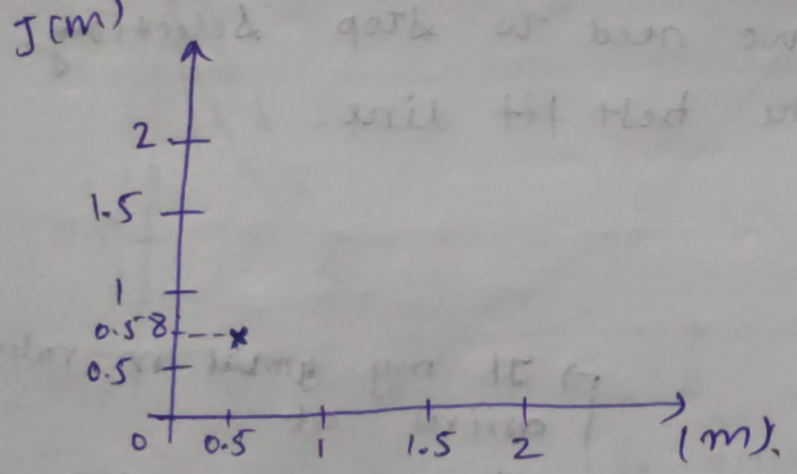
so, now find the cost function,

$= \frac{1}{2m} \sum_{i=1}^{m} (\hat{y} - y)^2$

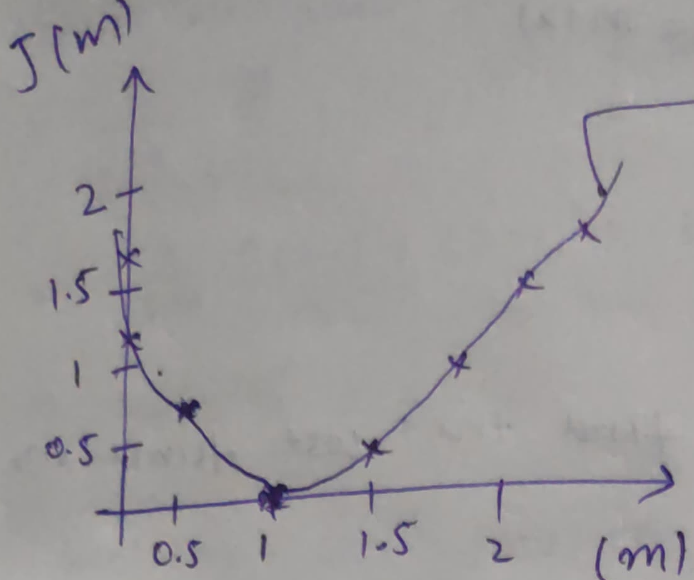
$\Rightarrow \frac{1}{2m} [(0.5-1)^2 + (1-2)^2 + (1.5-3)^2]$

$\Rightarrow \frac{1}{2(3)} [\text{"} \quad \text{"}]$

$\Rightarrow \underline{0.58} \} \rightarrow$ cost function (0.58) when your $m = 0.5$.



so for different (M) values we get different lines of curvature, combining both the diagrams & assuming other M values.



gradient descent.

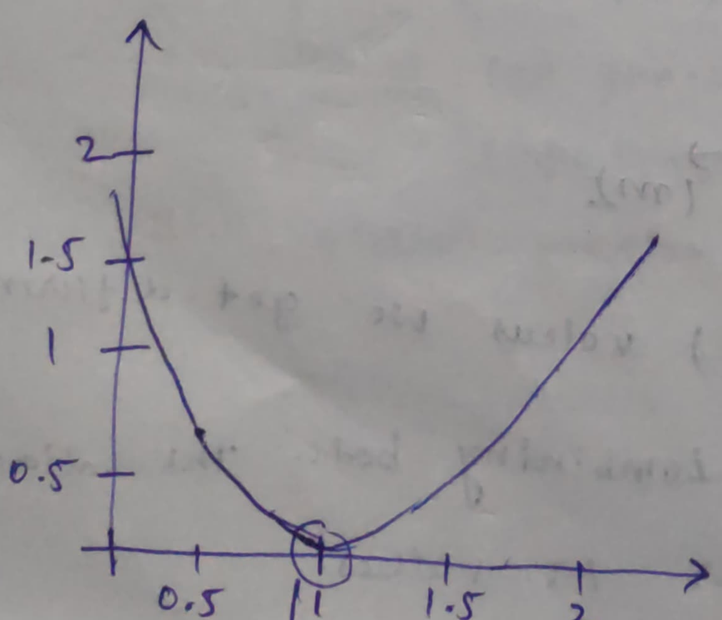
i). J(m) = 0, m = 1

ii). J(m) = 0.58, m = 0.5

assuming for other points (slope (m) randoms)

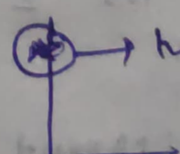
⇒ once we get the gradient descent when should we know, that we need to stop selecting m value, for the best fit line.

⇒ gradient descent:



global minima

→ If my global m value arrives at



→ here (we need to go down the graph)

so now we'll use convergence theorem

formula (convergence theorem)

$$m = m - \left(\frac{\Delta m}{d(m)}\right) \times \alpha$$ (Learning rate alpha).

(Derivative of M w.r.t. M)

($\Delta(m)$ is my slope)

→ Let's see how to arrive at the global minima point.

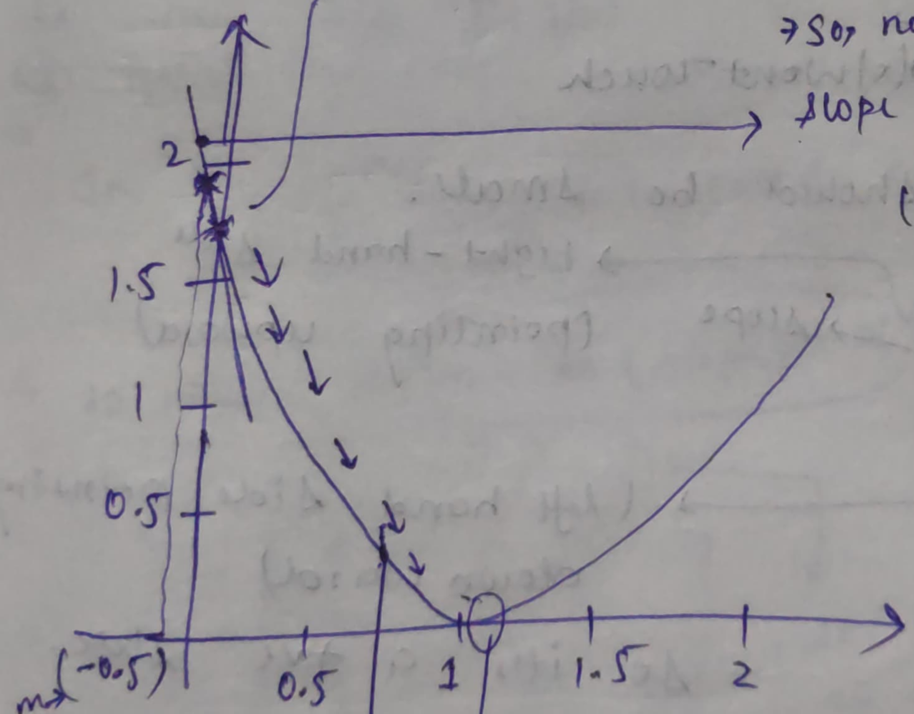
Based on some (M value) that's the initial point.

→ So, now we've to move down↓

slope $\left(\frac{\Delta m}{d(m)}\right)$

(So now we've to find whether it's a positive (or) Negative slope) by focusing on right & left hand of slope.

→ If the right side of slope pointing (↓) -ve (slope)



how to arrive?

$$so,\ m = m - \left(\begin{array}{c}\text{-ve slope} \\ \text{derivative}\end{array}\right) \times \alpha$$ (smaller value)

is a -ve value

→ $m = m - (\text{-ve value}) \times \alpha$

$m = m + (\text{+ve}) \times \alpha$ →

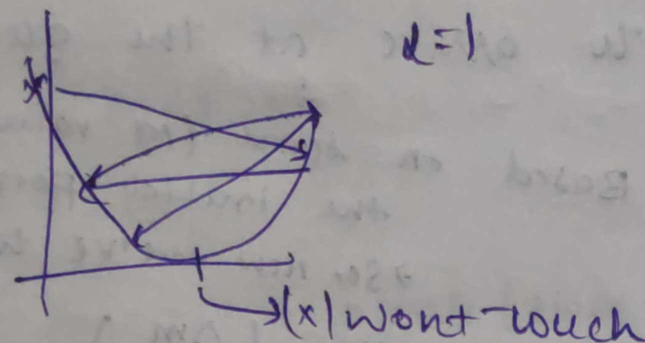
so now the ~~..~~ as will come close to 1.

→ small downward steps.

so when iteration & different M values gets selected, it'll move towards global minima.

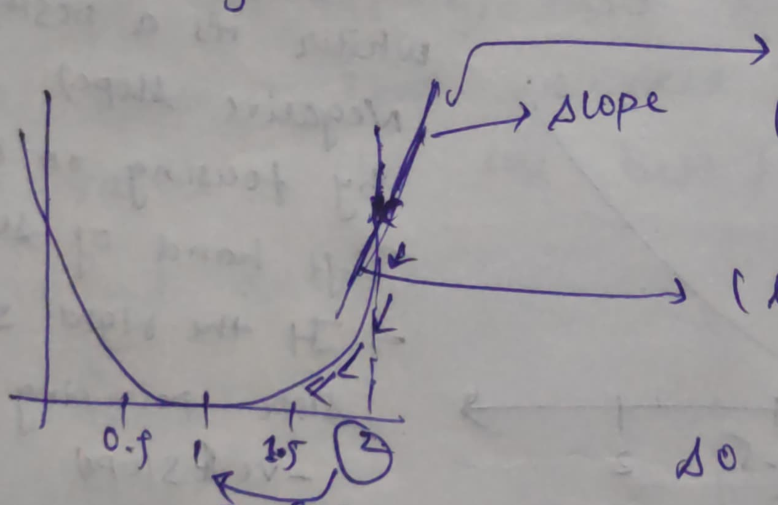
if we take α (large value) then point i.e. ① the point will jump at high spaces & might never touch global minima.

α=1



Thats why α should be small.

Eg:2



so its a +ve slope.

Convergence:

derivative of +ve slope,

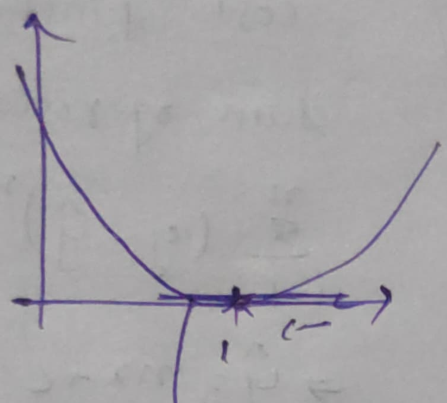
$$m = m - (+ve) \times \alpha$$

$$\Rightarrow m - \alpha.$$

so the m value(2) will subtract with smaller value & reach global minima (1).

4 when the global ~~minima~~ ~~slope~~ m value slowly reaches 10, the ~~m~~ slop will be 0.

↓



slope 0.

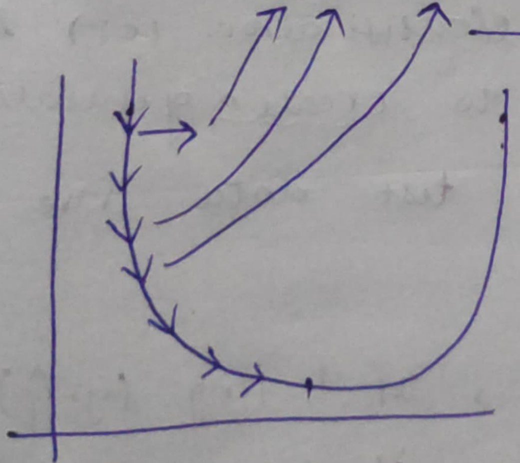
The M value will specify that this should be the best fit line.

~~[illegible]~~

→ In $y = mx + c$ → also called as an intercept.

m → also called as an coefficient.

→ so what is $\boxed{m - \cancel{m}\,(\text{slope}) \times \alpha}$ → ?



so in every step as the arrow ends up starts again that particular equation executes to reach the global minima point.