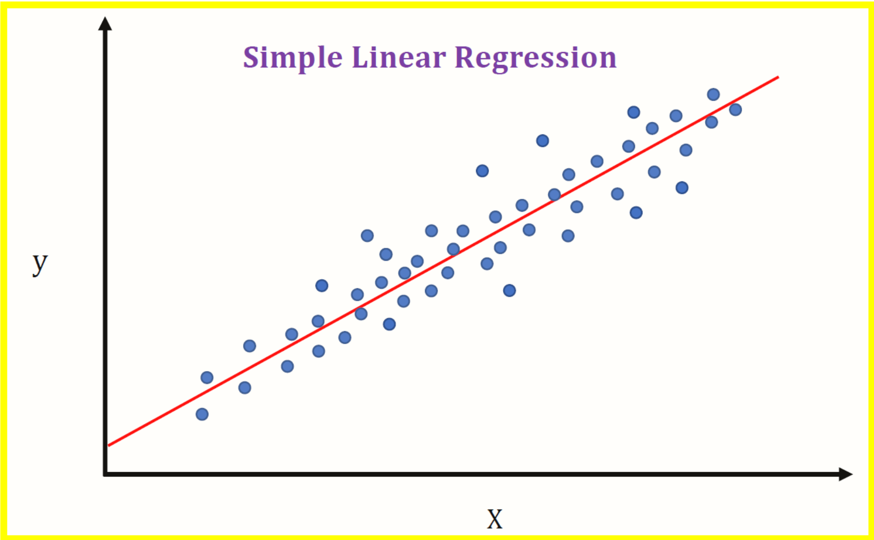
Linear Regression



For Finding the best fit line you use this particular formula:

Y = mx + b

Here both y and x are constant so you need to regulate 'm' and 'b' to get the best-fit line.

Now, you can get 'm' and 'b' by:

- Closed Form Solution
  - Direct Formula[OLS method, used in scikit learn LR model]
- Non Closed Form Solution
  - Gradient Descent [SGDRegressor, Gradient Descent is implemented]

A common question might strike that if we have a direct formula then why we should use anything else?

So basically in Direct Formula when the dimensions start increasing finding the 'm' and 'b' using the formula starts getting difficult.

There the efficient technique becomes Gradient Descent.

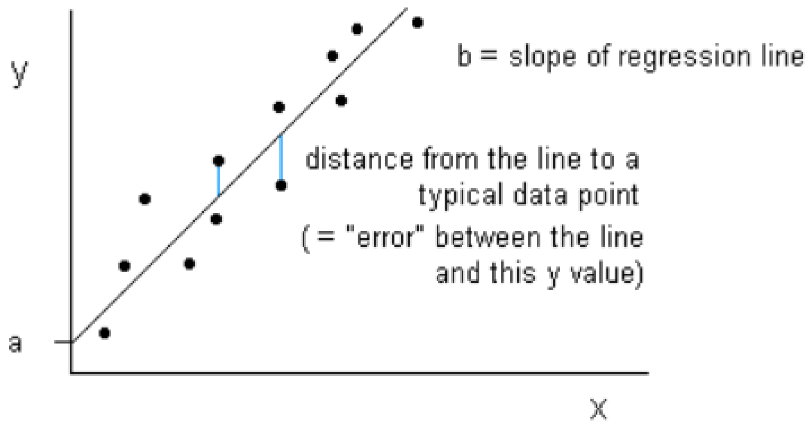
Using OLS finding ‘M’ and ‘B’

$$b = \bar{y} - m\bar{x} \qquad m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\bar{x}, \bar{y} \rightarrow$ mean

The above is the formula for finding b and m, xi and yi represent the value of x and y at ‘i’.

Now lets find the B and M from scratch, like from where it is coming!



Lets take the error of each point from the line as

d1,d2,d3,d4,….,dn

We need to find a best-fit line such that the summation of the error or the total error is minimized.

E = d1 + d2 + d3 + …. +dn

Now as some points may be below the line which will result in -di [where i represents the point number], resulting in canceling the number out. So for that reason, we are squaring all the errors.

E = $d1^2$ + $d2^2$ + $d3^2$ + …. + dn^2

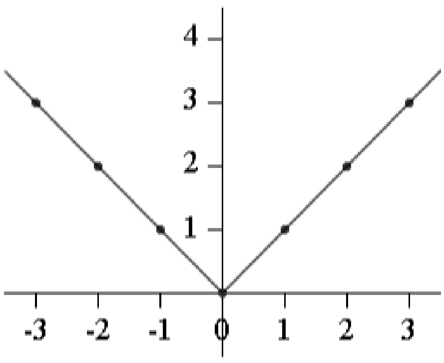
Why are we not using the mod in the above case?

So after a certain point, we need to differentiate E,

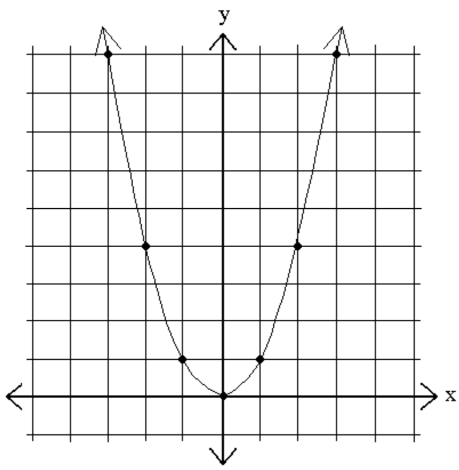
The mod graph if you see it is continuous but not differentiable at origin

Whereas the square graph is differentiable at any point.

Mod Graph:

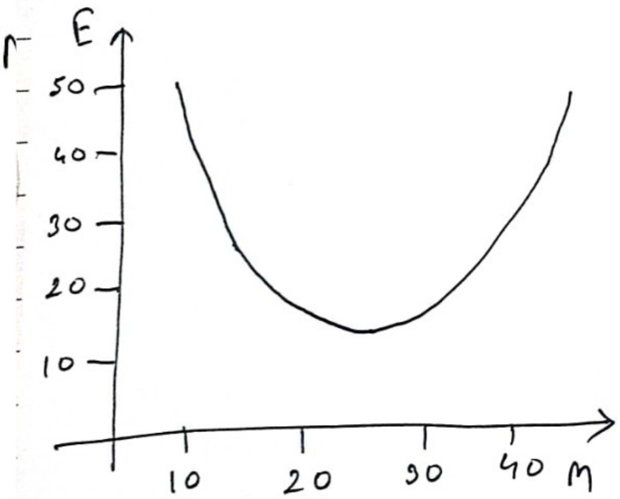


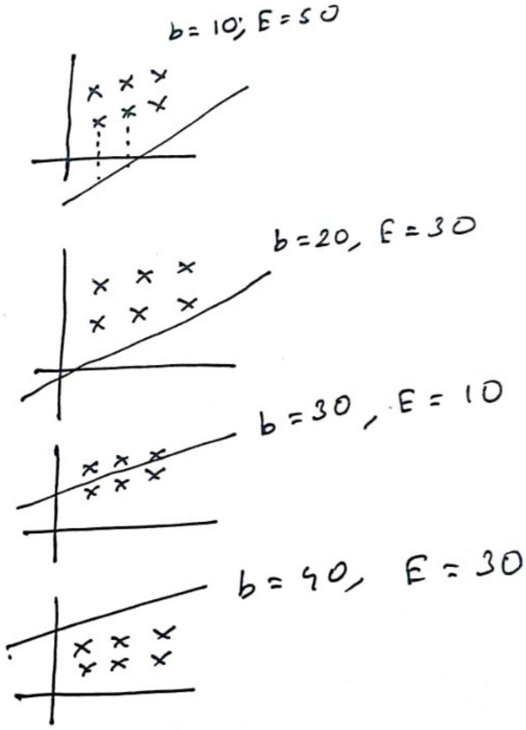
Square Graph:



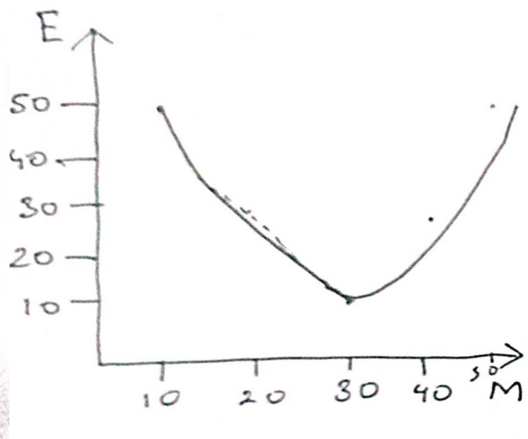
Error Function:

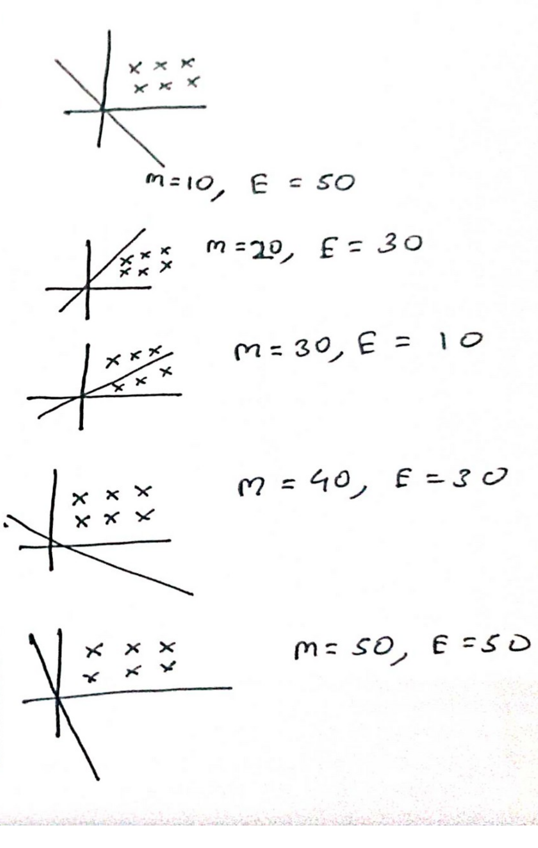
For $\boxed{m=1}$, the graph will be,



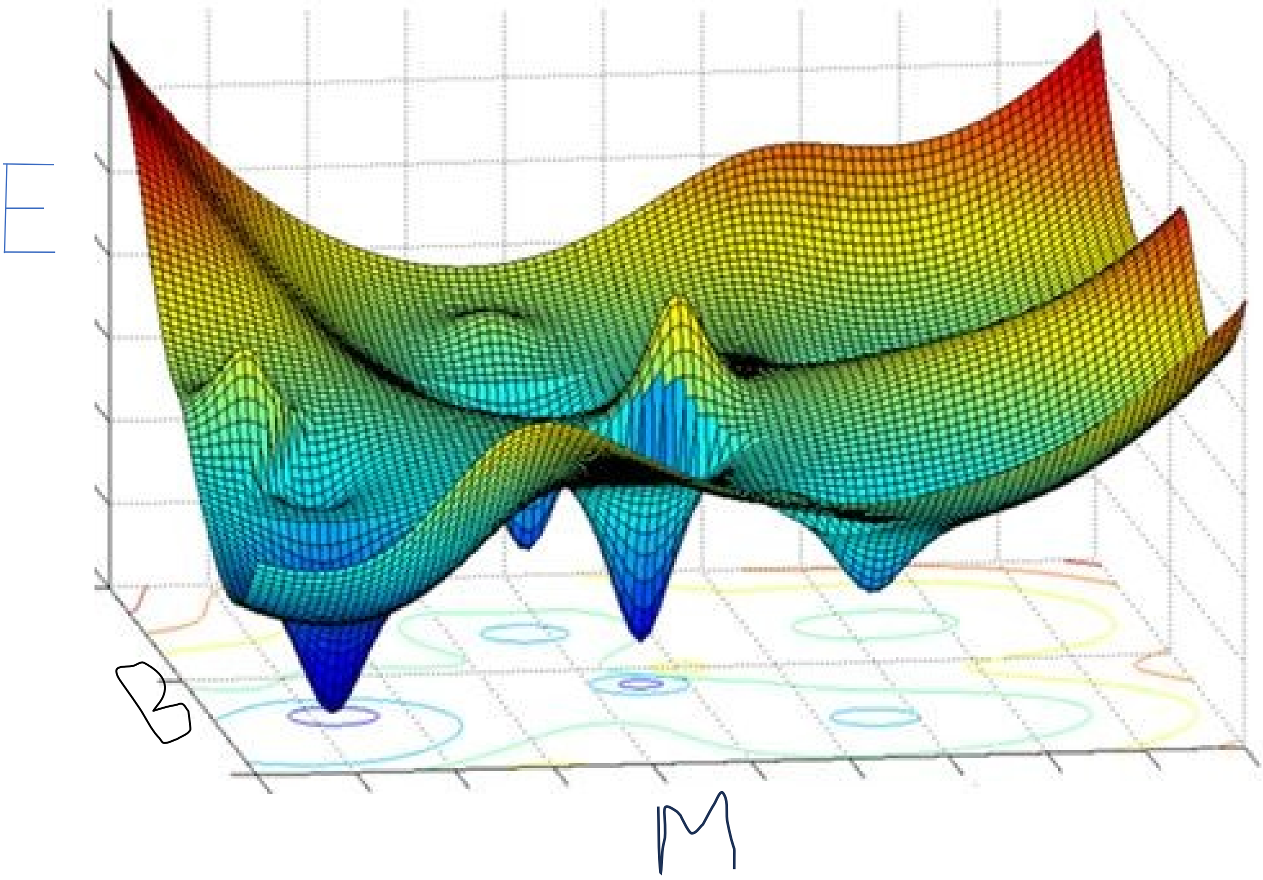


For $\boxed{b=0}$, we will ~~plot the~~ plot the graph.





We got the graph for (m, E) and (b, E) now let's plot a graph in 3D with m,b and E.



Can you see certain low and high dips?

So for certain values of M and B the Error is Min and for certain values Error is Max.

Now when we are at Minimum the slope will be 0.

For finding the slope we need to find the derivative and do it equal to 0.

So, we will find the derivative of our Error function first with respect to m and then with b.

At minimum the slope is 0.

$$\frac{\partial E}{\partial m} = 0 \qquad \frac{\partial E}{\partial b} = 0$$

$$(m, b)$$

$$\frac{\partial E}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - mx_i - b)^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{\partial}{\partial b} (y_i - mx_i - b)^2 = 0$$

By Applying Chain Rule,

$$\Rightarrow \sum_{i=1}^{n} -2 (y_i - mx_i - b) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - mx_i - b) = 0$$

$$\Rightarrow \frac{\Sigma y_i}{n} - \frac{\Sigma m x_i}{n} - \frac{\Sigma b}{n} = 0$$

$$\Rightarrow \bar{y} - m\bar{x} - \frac{\cancel{n} b}{\cancel{n}} = 0$$

$$\Rightarrow \boxed{\bar{y} - m\bar{x} = b}$$

So you can see we are thereby finding the value of b.

$b = \bar{y} - m\bar{x}$

Now lets differentiate E with respect to ‘m’ .

$$E = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2$$

$$\frac{\partial E}{\partial m} = \sum_{i=1}^{n} \frac{\partial}{\partial m} (y_i - mx_i - \bar{y} + m\bar{x})^2 = 0 \qquad \text{[Using Chain Rule]}$$

$$\Rightarrow \sum_{i=1}^{n} 2(y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} -2(y_i - mx_i - \bar{y} + m\bar{x})(x_i - \bar{x}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[(y_i - \bar{y}) - m(x_i - \bar{x})\right](x_i - \bar{x}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} \left[(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2\right] = 0$$

$$\Rightarrow \Sigma (y_i - \bar{y})(x_i - \bar{x}) = m\Sigma(x_i - \bar{x})^2$$

$$\boxed{m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

So as it is evident we have calculated the m and b from scratch.