Convergence Algorithm:

We were taking randomly previously which is not efficient. Instead we will be using convergence algorithm that will help initialize value and then based on the gradient descent, it will be able to increase or decrease the value.

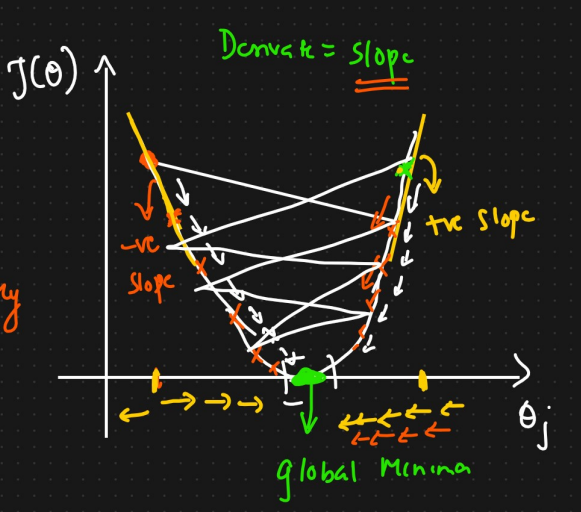
Convergence Algorithm:

Aim: To optimize the changes of values which is the slope.

Principle - Repeat until convergence, i.e, meets the global minima point.

Equation:

:= means it is an iterative process



is nothing but the slope.

Slope = y2 – y1 / x2 – x1

Since y value is decreasing on the left hand side, the slope will be negative. This will add some value to the .

Since y value is increasing on the right hand side, the slope will be positive. This will subtract some value from the .

This is an iterative process.

After repeating this process, after some time, will converge to the global minima.

α is the Learning rate.

It is usually initialized with a smaller variable for example 0.001

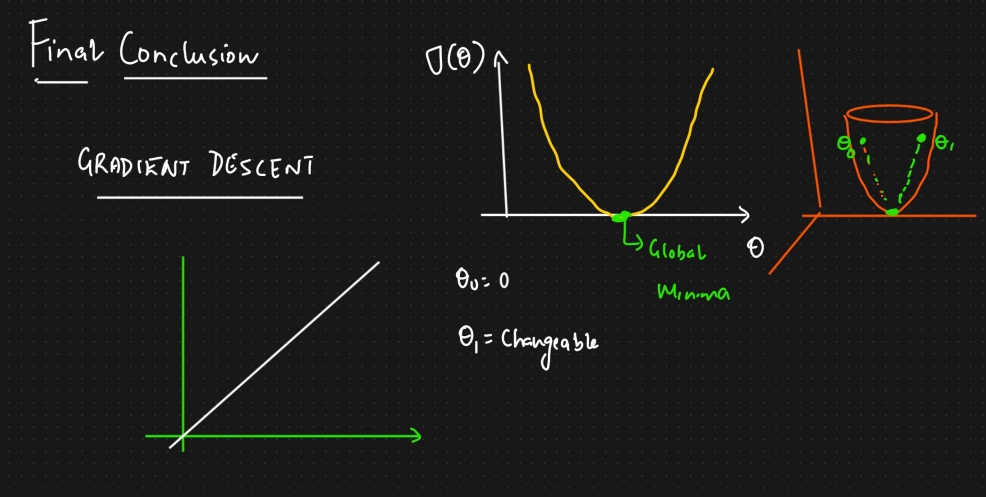
Learning rate controls the speed at which convergence happens.

If it is a very small value, it will take more time to converge.

If it is a very big value, it may never converge.

Should be a optimum value.

Conclusion:



We took as 0 so that we could represent the diagram in 2d.

If is not 0, we will have the diagram shown and both and are there and our aim is to converge towards global minima.

Convergence algorithm:

Repeat until convergence.

Updated equation:

Where

M is the number of data points we have.

So the derivative,

Becomes

We have to find the derivative for j = 0 and j = 1

If j = 0,

=

Where is constant so becomes 0

If j = 1,

=

Here becomes constant

=

Repeat until convergence {

}