Gradient Descent

* Means moving down a slope which implies we are minimizaing something
* In machine learning, we minimize a loss function which is the difference between the true and predicted value

L = Ypred - Ytrue

* Now suppose we have set of points and we want to make a model to fit these set of points
* Here the model should represent a line which is controlled by a parameter w
* In order to minimize the loss function we need to find a line which best fits the data. The line is controlled by parameters w so we want to find the best parameter
* So we take derivative of the loss function wrt the paramater and update the parameter in next iteration

W2=w1-dL/dw1

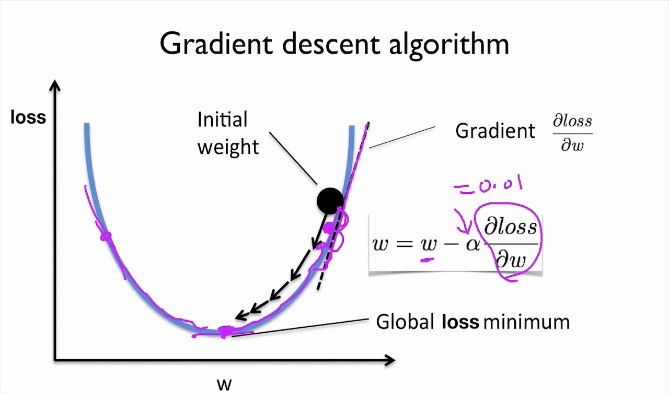
W3= w2-dL/dw2

.

.

.

* Graphically it looks like this



* So we keep on updating the parameter till we reach a point where the loss function cannot be minimized further
* The jump is controlled by learning rate. If the learning rate is too high the updates can oscillate the minimum point, if too low the descent would be too slow.