25/03/20

**Linear Regression and gradient descent**

Our datapoints are dependent variable in terms of an independent variable. With this collection of dependent and independent variables we want to find what will be the dependent variable for some other independent variable. We want our prediction to be most accurate.

Regression means fitting a mathematical curve to the given datapoints, it can be any kind of curve. With this curve we can predict any independent variable.

In Linear regression we fit the data by a straight line. The straight line may not predict the independent variable accurately but it is a simple model to understand regression.

We need to find slope and intercept of the best fit line,

Error term:

This is a function of m and b and we have to minimize f. we take the value of m some value say 0.

we calculate df/dm and assign m as m(old)-df/dm\*constant and b as b(old)- df/db\*constant and we calculate f(m,b). If f is less than previous f we continue with the same constant else we change the constant. The constant is called learning rate.

We continue this process until the f increases after large number of iterations.

The m,b at this point is the required value of slope and intercept. This process is called gradient descent algorithm.

**Stochastic Gradient descent**

Gradient descent involves computation of derivatives in large numbers when the dataset is large, so we use stochastic gradient descent. In this we pick a random point and compute the gradient and multiply with the learning rate to get new m and b rather. We do not do the summation as in gradient descent.

This is very useful when the data is clustered together, we use one data point from each cluster and come up with the solution.

When new data point is given we need not compute m and b from beginning as in gradient descent we just have to update the finally obtained m and b by the new datapoint.