**Gradient boosting**

Inputs:



In this equation, xi refers to each row of the measurement that we are using to predict.On the other hand yi refers to the value that we need to predict.The dataset contains n number of samples so the input goes from n = 1 to n, getting all the data present in the dataset.

With this a differentiable loss function is also given as a input. A loss function is term that is just used to evluate the prediction level.In case of gradient boost the loss function is calculated by taking half of the squared difference between observed and predicted values.This equation is just same as the residual calculated for simple linear regression except the part of multiplying it by ½. Taking half of the Residual doesn’t make any difference, it is just added to the equation to simplfy the diffrentiation of the equation with respect to the predicted value.

**Loss function :** L(y,F(x))

Here F(x) is the function which gives the predicted values.

**Step 1:**

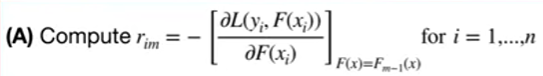


This equation just finds the mean of all the predicted value from the data.This value is given as the root node of the tree.

**Step 2:**

For m = 1 to M

In step 2 a simple loop is initiated to make all the trees.So in this loop we are making ‘M’ number of tree. Mostly the number of tree are set to 100. The small m (‘m’) in the equation refers to a integer which keeps on increasing by a unit after a tree is made.



This part (A) is just the derivative of the loss function which is multiplyed by minus one(-1). The deric=vative just comes out to be the difference between the observed and predicted value i.e. residual.The equation F(x)=Fm-1(x) is just the term to bring the previous error to the equation for makin the tree.So in this way we residual (rim) is calculated where i refers the smaple number.