# GBM

Mannem Charan AI21BTECH11019

**Abstract**

This report consists of my basic understanding of one of the popular Ml methods ”GBM”.

## 1 GBM

GBM, Gradient Boosting Machine is a supervised learning algorithm which used to classification and regression technique.It comes under ensemble learning where we combine weak learners to form a strong learner to solve complex problems.As we discussed random forest earlier which is also a ensemble learning algorithm it follows bagging type of method whereas GBM follows boosting method.When we say boosting we mean that the weak learners in the model learn the training data sequentially.Each learner is meant to compensate the weakness of it's predecessor.

## 2 Understanding GBM

In GBM,we will add a new model which compensates the errors of the existing model.Depending on the task,GBM is divided into two types,

1) Gradient Boosting Regressor
2) Gradient Boosting Classifier

And these two methods are differ from their two loss functions which is used to minimise the error.

## 3 Gradient Boosting Regressor

Since we are trying to predict a continuous value the generally used loss functions are $L_1$ loss function (absolute sum of residuals),$L_2$ loss function(Mean squared error) and etc.

1) In Step-1 we will try to build a base model that predicts the output of the observations.We will take mean of the target column as our first prediction.

$$F_0(x) = \bar{y} \tag{3.1}$$

where, $F_0$ is the model at stage 0 and $\bar{y}$ is the mean of output values of data. We are taking mean here because it follows from minimising the loss function,

$$L = \frac{1}{n}\sum_{i=0}^{n}(y_i - \gamma_i) \tag{3.2}$$

$$\frac{dL}{d\gamma} = -\sum_{i=0}^{n}(y_i - \gamma_i) = 0 \tag{3.3}$$

where $y_i$ are the continuous output values and $\gamma_i$ are the predicted output value. And solving above equation gives that $\gamma_i$ is mean of output values.

2) In Step-2, we calculate the residuals from these predicted values.This step can be written as,

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \qquad \text{for i=1,2 ...., n} \tag{3.4}$$

where $m$ is the no. of decision trees in the model and $F(x_i)$ is the previous model.

3) In Step-3, we will build a model on these residuals and make predictions.Because we want to minimize these residuals and minimizing the residuals will eventually improve our model accuracy.Let say $h_m(x)$ the decision tree made on residuals.
4) In this step we will find the output values of each leaves.The predicted value will be nothing but average of residuals in each leaf and the same thing mathematically can be represented as,

$$\gamma_m = \text{arg min}_\gamma \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{3.5}$$

where L is the loss function and $\gamma_m$ is predicted output value for $m^t h$ DT.

5) And finally we will update the predictions of previous model as follows,

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x). \tag{3.6}$$

This procedure is repeated until there is no improvement in the model further.

## 4 Gradient Boosting Classifier

Since the label here will be categorical we will try to use log loss function(similar to logistic Regression) as the cost function.And the steps followed will same as earlier and only the difference will be in loss function.
In both the tasks the prediction is updated as follows,

$$Newprediction = Oldprediction + (learning_rate) * (treemadeonresiduals) \tag{4.1}$$