Matthew Dreher

Math 400

Machine learning with gradient bosting decision trees

One of the methods use in machine learning is the use of gradient boosting which is a machine learning technique for regression and classification problem. The prediction model is made from an ensemble of weak prediction model to from a strong prediction model normally a decision trees the trees are made in the stepwise fashion

The gradient boosting method is similar to method used for least squares regression with a forwards stepwise logistic regression method where it’s a greedy Algorithms to minimize the loss function which is normally the mean squared error $\frac{1}{n}\sum\_{}^{}(\hat{y\_{i} }-y\_{i})^{2}$of the model plus some function Ω to prevent over fitting, it decreases the MSE by adding one new variable per iteration that best decreases the mean squared error of the mode,

This is done by making a decision tree which is a data str that use binary decisions to determine if it on the left or right side of a breach and repeated until it hit a leaf node which is a node with no children

The process start with spiting the data set into y testing y training x testing x training where x is the decision variable and y is the target variable and training data is use to make the model and the testing data is the data use to test the model to prevent over fitting

Then take the mean value of the y training set and take the MSE with y mean as the model then split the data into xi>j and xi<=j where xi and j are a tree of height one that are pick to minimize the MSE to each half y mean of the tree if the decrease in the MSE plus the Ω function the goes on to the next iteration this repeated adding another tree of height one to the model until the number of depth is hit or the MSE plus the Ω function does not improvement



Then

In the end the gradient boosting model look like the linear regression model which has one dummy variable made up of each of the leaf nodes



The leaf prediction would be the mean of the data that fall in the leaf

Then after you would use your test data to validate test the different method such as MSE, MAPE, ex may want to test the number of time a xi is use as a spit it may show which data is important or show that there is a larking variable

Because the nature of the model to make multiple test there is strong chance that there will be some type 1 and 2 error with the xi use or not used for the model