Assignment 3

Arvind Pawar

June 13, 2019

ALY6020-Predictive Analytics

Instructor: Dr. Marco Montes de Oca

Gradient Boosting

- Gradient boosting is a method to improve the accuracy of a predictive model which helps in achieving the best accuracy that we might not achieve using other algorithms.
- Gradient boost starts with a single tree. Trees at first are weak learners.
- Later the model is improved based on the mistakes made by previous model and the process is repeated to till minimum error is achieved.
- For example, if at first a tree has misclassified some inputs, we train the next tree with those previously generated errors which directly helps in improving the model.
- In this way a model can correctly classify the values and make itself a strong learner for accurate predictions.
- The desired sum of residual is less than or equal to 10.

Advantages

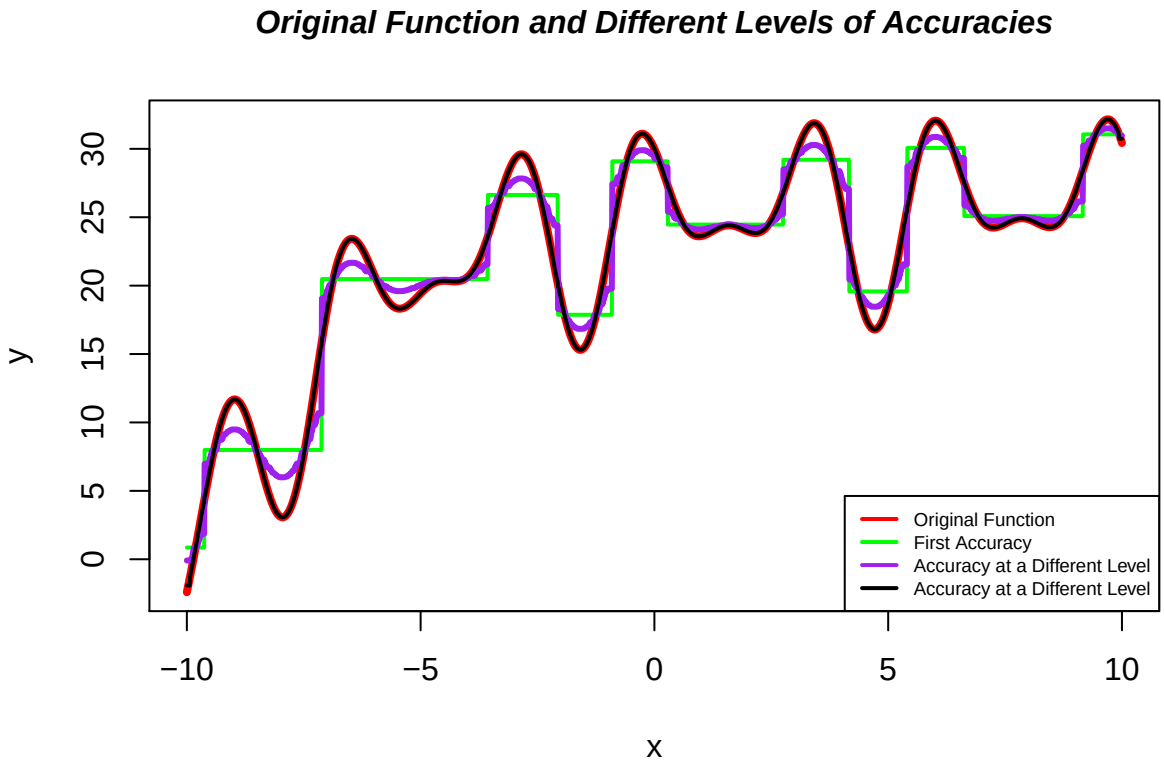
- As we know, one of the advantages of the gradient boost is the accuracy in prediction.
- Optimization of different loss functions is enabled.
- Hyperparameters tuning options are provided which helps in making the functions flexible.
- Missing data is handled.
- Excellent performance is delivered when data is categorical and numerical; hence pre-processing of data is not needed.

Disadvantages

- One of the disadvantages of gradient boosting is that it can lead to overfitting because of the continuous improvement in the model by minimizing all the errors.
- It can be time-consuming sometimes as more than 1000 trees are often required which also may degrade the system performance.
- Behavior of iterations, tree depth, regularization parameters, etc. can be influenced due to the parameter flexibility.

We were asked to generate values(x) between -10 to 10 and approximate function f(x) =sin(x) + 5cos(2x)-3sin(3x) +(1-e-x/3) +25 using gradient boosting algorithm.

- Tree models were generated by using rpart() function along with help of y values.
- Later, residuals were calculated for the first model, which is a weak learner.
- These residuals were then used to train the next or new model.
- The new model is required to adjust the predicted y values so that the error aligns in accordance to the desired value.
- Fractions of the adjustment to be made by the predicted y values to reduce the error is decided by the learning rate known as Alpha.
- When Alpha is equal 0.1, the error value reduces for few iterations and later becomes constant at a certain level.
- Complexity Parameter(cp) is used to overcome this situation.
- Optimum error value is reached faster if the value of cp is smaller.



Conclusion

- Plotted the function values corresponding to the x values ranges from -10 to 10.
- The green line is initially predicted values.
- In a starting few iterations model is a weak learner as it not able to capture the valleys and peaks of the actual function and becomes constant after reaching a certain level.
- Then we plotted the predicted values at a different level to infer that are due to appropriate learning rate and complexity parameter the function is tracing actual y values more accurately than the previously predicted values.
- All these indicates that the gradient boosting is working appropriately to improve the model generated with each iteration.
- At the final iteration, we have reduced the error at such level that if we plot those (improved) predicted values, then it will overlap the original function because our model is now making an accurate prediction.