Final Project Proposal

Aidan Garton

April 19, 2022

1 Team

Currently I am not part of a group and was thinking of working alone. Although, I am not opposed to working in a group and have reached out to some classmates to see if they would be interested in working with me. Regardless, I outline below what I am interested in doing for the final project.

2 Summary

For my final project I am interested in implementing a boosted decision tree model in Java. I would like to implement two different boosting algorithms (Ada boosting and gradient boosting) and compare the resulting models on performance: train accuracy, over-fitting, and train / classify run time. I will use the following data set, available on the UCI machine learning repository website, to compare the performance of the two boosting algorithms: (https://archive.ics.uci.edu/ml/datasets/student%2Bperformance). In order to do this I will need to clean and pre-process the data and turn it into a java readable csv file.

Additionally I would like to experiment on the best hyper-parameters for each boosting algorithm including what loss function to use, what learning rate to use, how many trees to include, ... I will vary these in an attempt to find the optimal hyper-parameter types for this data set and will include graphs of my results

3 Resources

1. Data set: https://archive.ics.uci.edu/ml/datasets/student
2. Boosted decision trees overview: https://indico.fnal.gov/event/15356/contributions/31377/attachments/19671/24560/DecisionTrees.pdf
3. Decision tree model we programmed in week 2 of the class
4. Ada boosting paper: https://www.sciencedirect.com/science/article/pii/S002200009791504X

5. Gradient boosting paper: https://jerryfriedman.su.domains/ftp/trebst.pdf