Computer Science and Engineering

CS 422 Report Project 1 - Fall 2022

Machine Learning Project 1

Andrew Ramirez

Contents

1 Decision Trees 3

2 Random Forests 3

3 Overall Feelings 3

1 Decision Trees

Did you implement your decision tree functions iteratively or recursively? Which data structure did you choose and were you happy with that choice? If you were unhappy with that choice which other data structure would have built the model with?

I decided to implement the Decision Tree functions recursively. The main reason for this kind of implementation is that it is easier for me to think recursively in terms of building a binary decision tree. In conjugation with this I used a nested list two make the Decision Tree, where the 0^{th} index was feature index, and the 1^{st} and 2^{nd} index where the two list branches. I would have liked to use a Node class but I am not that familiar with object oriented programming with python.

2 Random Forests

Why might the individual trees have such variance in their accuracy? How would you reduce this variance and potentially improve accuracy?

The individual trees might have variance in their accuracy due to the number of accessible samples. A low sample would make the accuracy for each individual tree susceptible to noise. Of course the max-depth puts a limit on the accuracy of the tree. Another hyper-parameter we could change is the number of sub-samples per tree(.10 in this implementation)

Why is it beneficial for the random forest to use an odd number of individual trees?

Random Forests using the "majority vote" to decide on the prediction for an input. Thus, using an odd number of trees will eliminate the possibility of ties.

3 Overall Feelings

Overall, if you are still feeling uncomfortable working with python, what aspect of the coding lan- guage do you feel you are struggling with the most? If you do feel

comfortable, what part of python do you feel you should continue practicing?

In my experience, Python has had both advantages and disadvantages in this assignment. Many of the in-built Python Methods worked great and made the assignment easier. On the other hand, debugging was a nightmare and I felt like there was a lot of unnecessary exception handling. To prepare for the next project, I'll learn more about Python classes to increase my ideas for Data Structures.