CS 373 Spring 2019: Homework 2

Youngsik Yoon, 0029846135

I collaborated with Julian Haresco, Harjas Monga,
and Aditya Subramaniam.

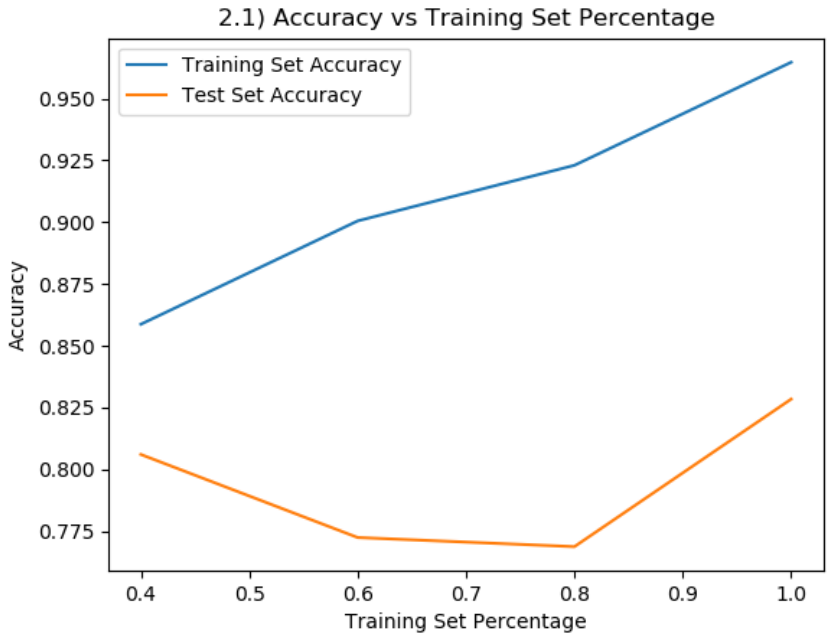
March 4, 2019

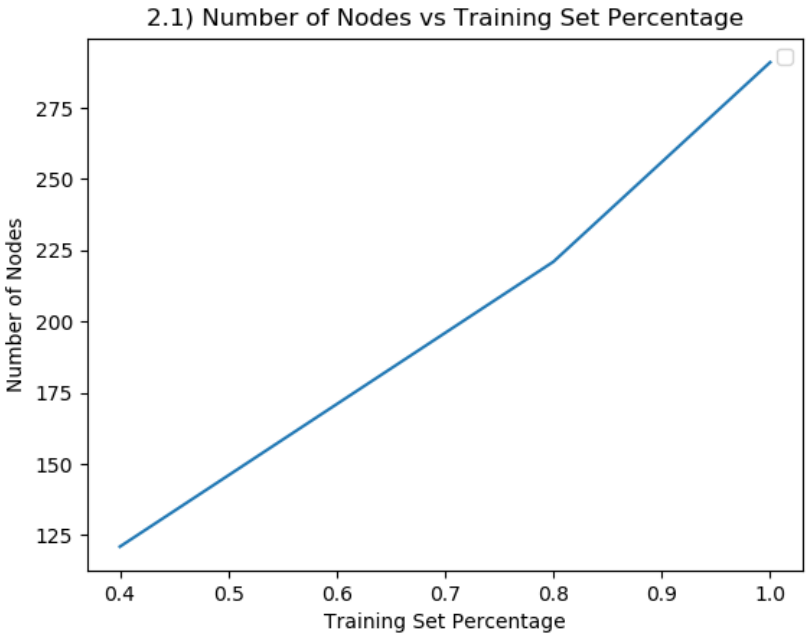
1 Part I: Decision Trees

Code as a separate file.

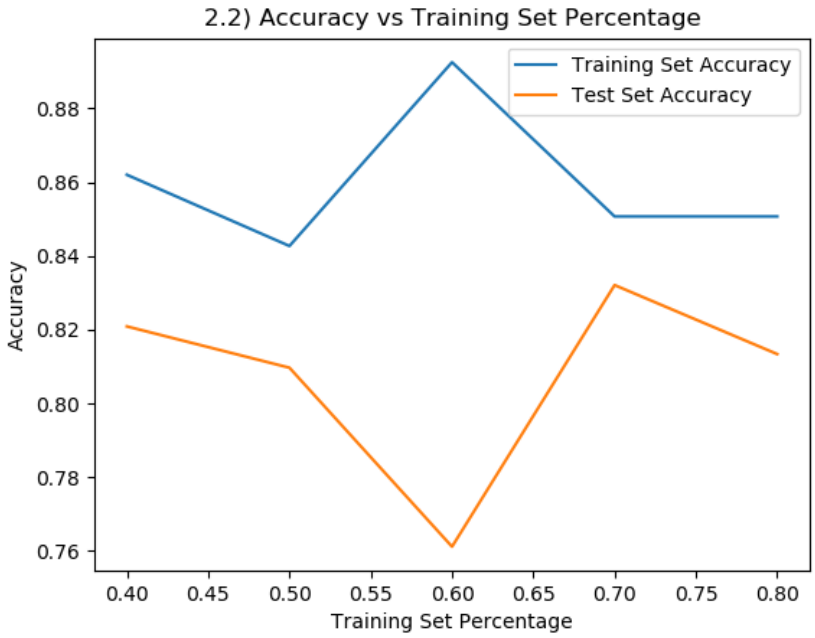
2 Part 2 Analysis

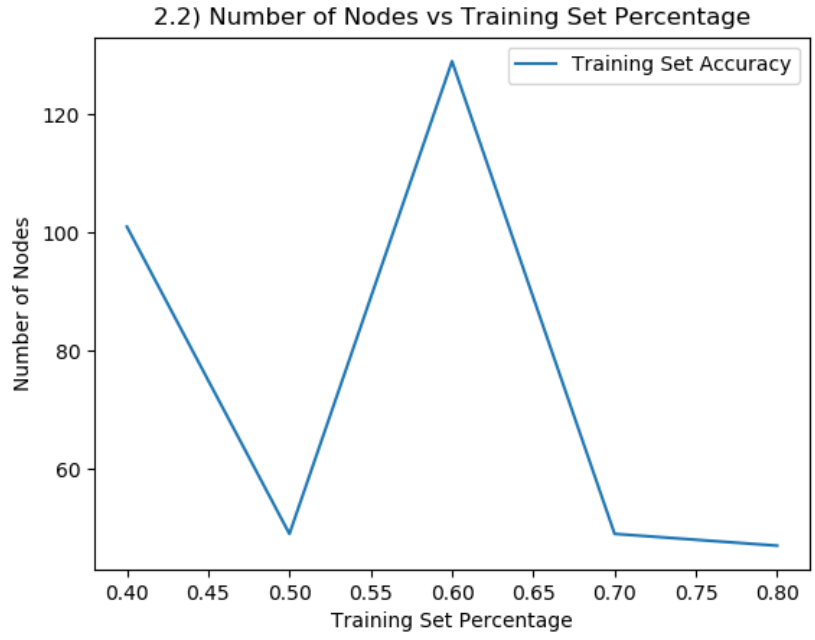
1. The impact of training set size on the accuracy and size of the full decision tree.

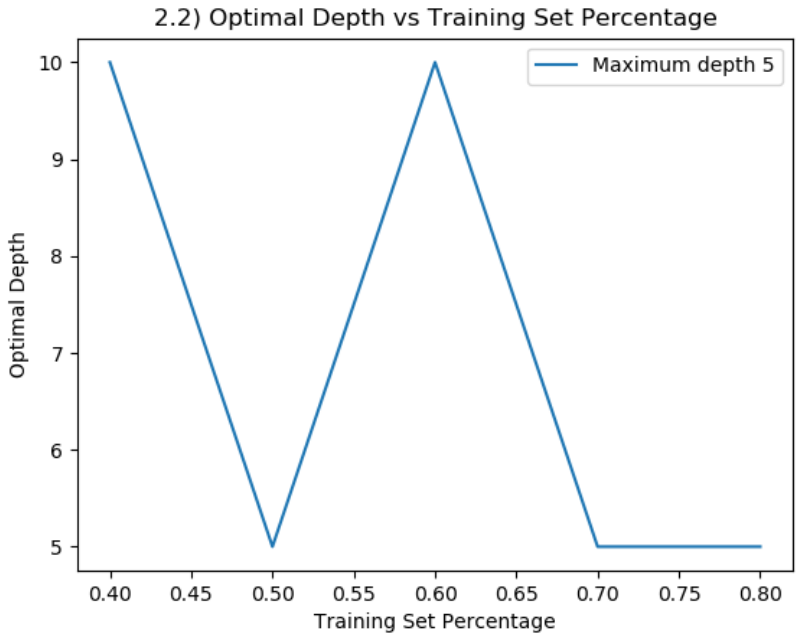




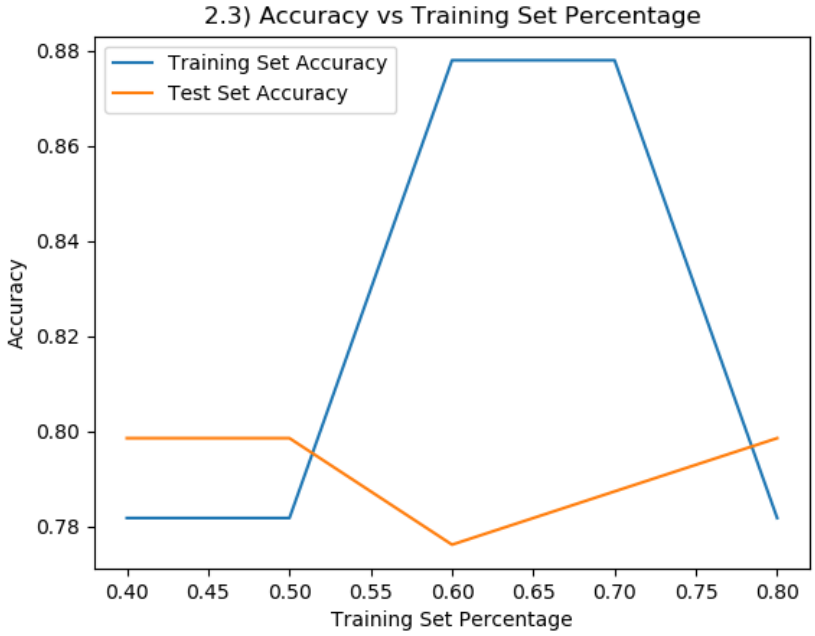
2. The impact of training set size on the accuracy and size of the static-depth case.

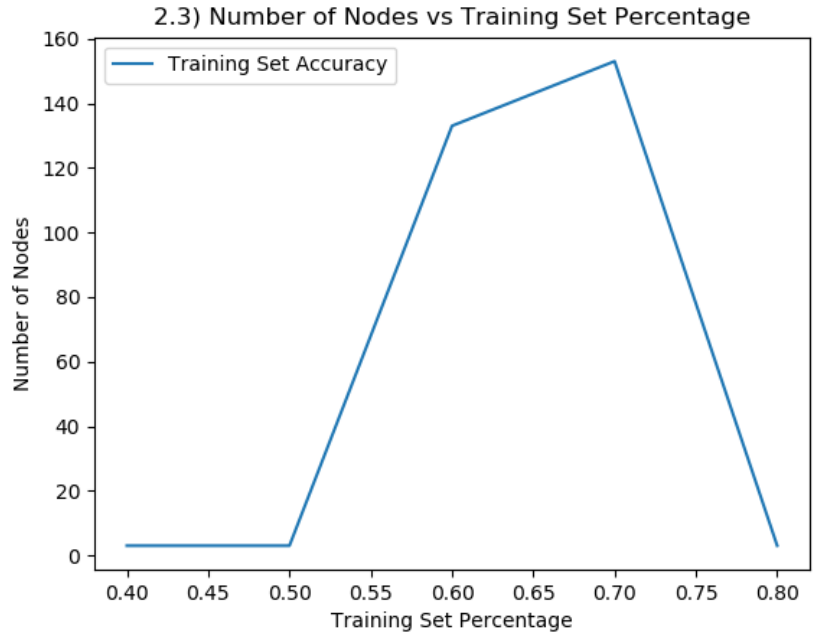

2.2) Accuracy vs Training Set Percentage
Training Set Accuracy
Test Set Accuracy
Accuracy
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
Training Set Percentage


2.2) Number of Nodes vs Training Set Percentage
Training Set Accuracy
Number of Nodes
120
100
80
60
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
Training Set Percentage



3. The impact of training set size on the accuracy and size of the pruning case.


2.3) Accuracy vs Training Set Percentage
Training Set Accuracy
Test Set Accuracy
Accuracy
0.88
0.86
0.84
0.82
0.80
0.78
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
Training Set Percentage


2.3) Number of Nodes vs Training Set Percentage
Training Set Accuracy
Number of Nodes
160
140
120
100
80
60
40
20
0
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
Training Set Percentage

4. By pruning directly on the test set, the test set no longer becomes a "test" set and becomes a "training" set. The purpose of the test set is to check the accuracy of the algorithm. If a separate validation set isn't used, there is no way to check the accuracy of the model created.

5. To convert the decision tree from a class from a classification model to a ranking model, the output would need to be non-binary. In this case, the output is only dead or alive. If the output had more options such as health conditions, then the decision tree could output a ranking over different health conditions instead of a single class label such as dead or alive.