

---

# CS760 Homework 1: Decision Tree

---

Qihong Lu

September 27, 2016

## QUESTION1

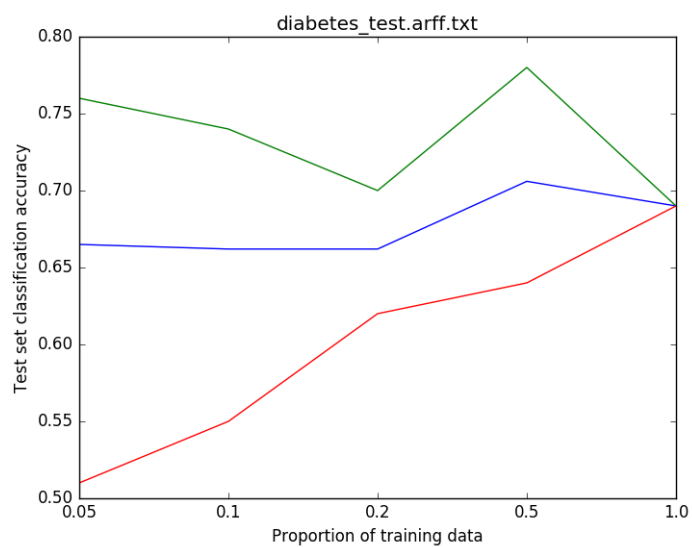
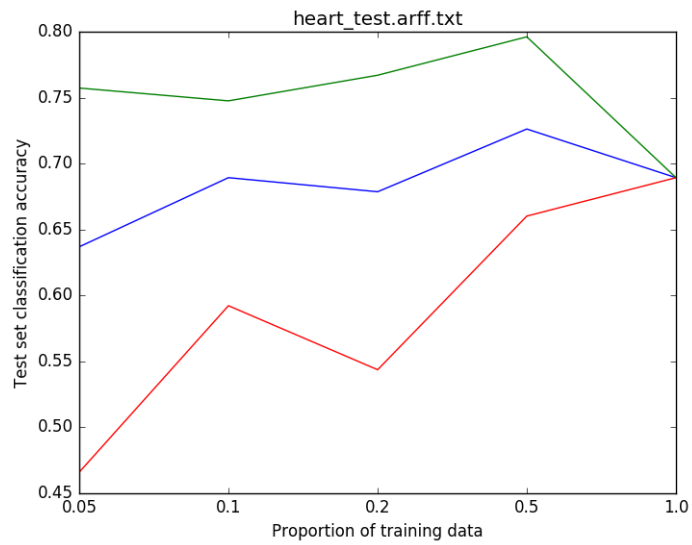
Code:

- dt-learn.py: the main program that implements the ID3 decision tree algorithm
- util.py: the definitions of some constants
- decisionTreeNode.py: the definition of the tree node class

Dependencies:

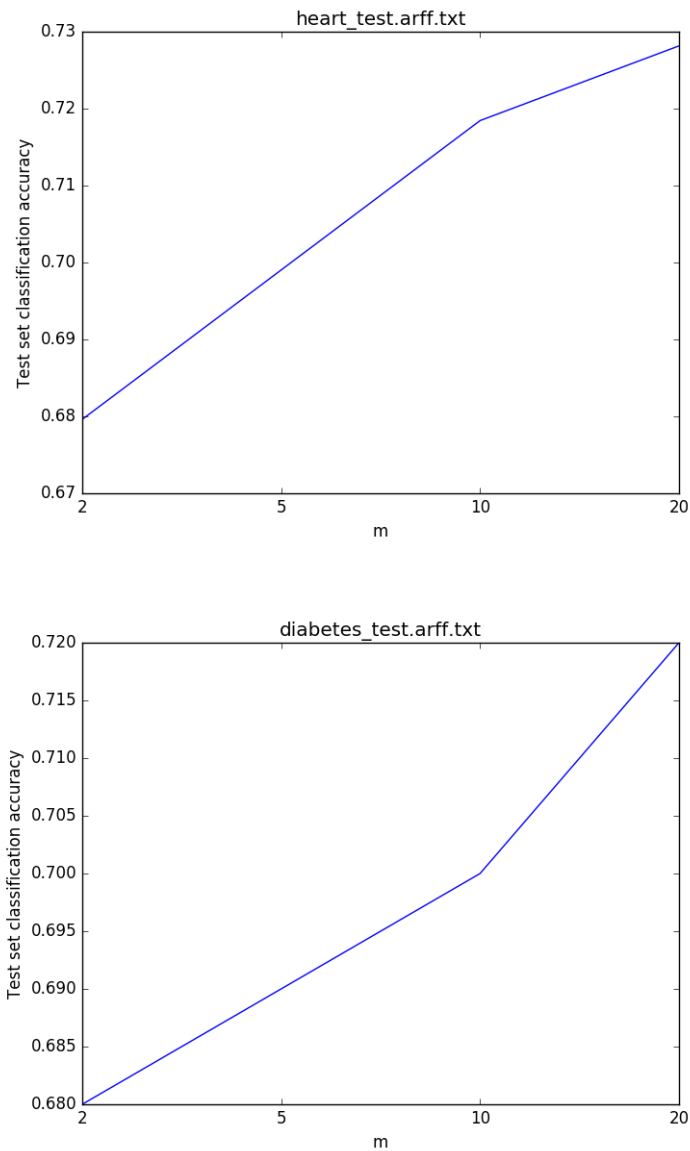
- python 2.7
- numpy
- scipy
- sys

## QUESTION2



Here are the learning curve: the performance, measured in terms of classification accuracy, against the amount of training data used when constructing the decision tree. In general, the performance is increasing as more training data were supplied. This trend holds for both of the data sets we explored.

### QUESTION3



As we increase  $m$ , the classification accuracy increases for both data sets. It shows that, as we had no pruning, the decision tree tends to overfit the training data. And by early stopping (increase  $m$ ), we obtain a simpler tree which tends to ameliorate the overfitting issue.