CPE 695 Homework2 Report

Group 7 Wang Qi, Shiyu Sun, Shaopeng Li

Decision tree learning uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. It is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees. In these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values are called regression trees.

Python language has been the most popular language in the field of data science and machine learning. So we decided to use Python to do the homework.

1. Firstly, we need to format the training data and test data. To do that, we did python list split operation. Because the datas are string type. We also need to convert them to float number so that we could do math calculation.

2. Secondly, we use sklearn to train data. To do that, we need get decision tree classifier object. Then we put training data into the decision tree classifier. After training, we put testing data into predict() function and print out results.
Base on our training, the results are:
['1' '1' '2' '1' '1' '2' '2' '1' '1' '2' '2' '2' '2' '1' '1' '2' '2' '1'
 '1' '1' '1' '1' '1' '1' '2' '2' '1' '1' '2' '2' '2' '1' '2' '1' '2' '1'
 '2' '2' '2' '2']