Lab Exercise #3: Decision Tree Learning using RWeka

CIS400/600
Fundamentals of Data and Knowledge Mining

Problem 1

Using the `adult_train` dataset, train **three** decision trees - you can use different hyperparameters or/and learning algorithms.

1. Compare the size and depth of the three decision trees.

2. Compare the training errors of the three decision trees.

Problem 2

Select any two out of the tree decision trees. Perform k-fold cross validation on both of them. Compare the results. Which one is more likely to perform better on the test dataset?

Problem 3

Perform holdout cross validations on the two classifiers selected in Problem 2.

1. Compare the results between the two classifiers.

2. Does the results from holdout cross validation agree with k-fold cross validation?

Problem 4

Load the `adult_test` dataset, and predict the classes in the testing dataset. Compare the performance of the two classifiers. Present the ROC curve and confusion matrix.