

Homework 1 – Report

Description:

The program takes two files as an input, a training file and a test file and builds a decision tree based on the training data using entropy and information gain as the splitting criterion. It returns the accuracy of the decision tree given a test data set.

Compilation instructions:

1. Open the “id3.py” in a python compiler and run the “id3.py ../data/train.dat ../data/train.dat” in the terminal, to calculate the accuracy on training set.

Found Accuracy:

Accuracy on training instances (800): 72.25%

2. Open the “id3.py” in a python compiler and run the “id3.py ../data/train.dat ../data/test.dat” in the terminal, to calculate the accuracy on test set.

Found Accuracy:

Accuracy on test instances (200): 68.96%

Accuracy on other datasets:

1. Train.dat (training set) and Train.101.dat (test set): 73.0%
2. Train.dat (training set) and Train.201.dat (test set): 72.0%
3. Train.dat (training set) and Train.301.dat (test set): 72.0%
4. Train.dat (training set) and Train.401.dat (test set): 72.25%
5. Train.dat (training set) and Train.501.dat (test set): 73.2%
6. Train.dat (training set) and Train.601.dat (test set): 72.33%
7. Train.dat (training set) and Train.701.dat (test set): 72.0%
8. Train.dat (training set) and Train.801.dat (test set): 72.25%
9. Train2.dat (training set) and Test2.dat (test set): 68.0%

Name: Sri Sravya Tirupachur Comerica

ID: 11259523