Homework 2 – Report

Description:

The program takes two files as an input, a training file and a test file and implements the k nearest neighbors and predicts the values based on the Euclidean distance similarity measure, it finds neighbors when k=1, k=3 and k=5.

Compilation instructions:

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

   Found Accuracy:
   Accuracy on training instances k1 (800): 79.75%
   Accuracy on training instances k3 (800): 86.38%
   Accuracy on training instances k5 (800): 89.38%

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 for k1 (200): 73.89%
   Accuracy on test instances for k3 (200): 84.24%
   Accuracy on test instances for k5 (200): 87.19%
   - Based on the accuracies obtained, I would choose k =5 for this dataset since it generates better accuracy.

Accuracy on other datasets:

1. Accuracy on Train2 and Test2 for k1 (100): 66.0%
   Accuracy on Train2 and Test2 for k3 (100): 78.0%
   Accuracy on Train2 and Test2 for k5 (100): 80.0%
2. Accuracy on Train2 and Train2 for k1 (100): 99.50%
   Accuracy on Train2 and Train2 for k3 (100): 90.00%
   Accuracy on Train2 and Train2 for k5 (100): 87.50%

Name: Sri Sravya Tirupachur Comerica

ID: 11259523