Daniel Montes

11/5/24

Dan Li

CSCD 429

Homework 2 : Report

How to run the program:

* Open terminal
* cd into the Homework 2 folder
* “ls” and make sure you see “Hw2 Report.docx”, “gene\_files”, “hw2.py”, and “predictions.txt”
* “python3 hw2.py” is the command needed to run the program, and you should see this

// While it makes the predictions

A black background with white text

Description automatically generated

// After predictions are made, if there was not a “predictions.txt” file there it will make it by itself

A computer screen with white text

Description automatically generated

Description of Classification Method:

For homework 2 I used the K-nearest neighbor algorithm which classifies gene localization data. The program first loads, trains, and tests the data which contain the gene attributes. I ended up setting K = 101 nearest neighbor for each testing instance. The hamming method is used to calculate how similar two genes are based on their features. This is the method that was used to calculate the percent of accuracy “accuracy = (correct / total) \* 100.” Any missing values in the data should not affect the accuracy prediction in my code.

// Accuracy being displayed on the bottom of the pycharm ide.

A screen shot of a computer

Description automatically generated