Project 1
k-Nearest Neighbor

Due before midnight on September 16, 2019

Problem Description
Clemson fisheries experts have recently discovered two new species of fish in Lake Hartwell. Species TigerFish1 is a delicious fish that tastes like Bluefin Tuna. TigerFish0 looks almost identical to TigerFish1 but is slightly poisonous and usually makes the person who consumes it very ill. The Clemson Wildlife and Fisheries Biology graduate students have captured, measured and tested hundreds of each species and created a file that contains measurements of the body length and dorsal fin length of each fish, along with its species. You have been hired to create a k-Nearest Neighbor program that, given the body length and dorsal fin length of a fish, will predict if it is TigerFish1 or TigerFish0.

Data File
You will be given a data file of labelled data. The first line contains a single integer indicating how many sets of labelled data you have to work with. Each line after that contains three tab-separated entries. The first is a float representing the body length in centimeters, followed by a float representing the dorsal fin length in centimeters, then an integer identifying the fish as either TigerFish0 (with a 0) or TigerFish1 (with a 1).

Assignment
Develop a k-Nearest Neighbor algorithm that will predict the species of a fish. Use a test set and 5-fold cross validation to determine the best number of neighbors to use in your prediction. Use a confusion matrix to help evaluate your results.

Your write up should follow the example on Canvas and should include:

- Problem Description
- Description of your data set along with a plot of the data.
- A description of your procedure.
- A figure showing how many misidentifications you had for each training/validation set combination for different values of k.
- A plot of average accuracy for different values of k.
- A confusion matrix showing your results on the final test set.
- A description of your final results that includes choice of k, accuracy, precision, recall and F1 values.

What to Upload to Canvas

- Your complete write up as a pdf file (lastname_firstname_P1Report.pdf).
- A single python (lastname_firstname_P1.py) file that prompts for the name of a file containing labelled training data (same format as your initial data file). Then repeatedly prompts for the body length and dorsal fin length (in centimeters) of a candidate fish and prints out the type to the screen (based on your best choice of k from your write up). The program should terminate when you enter zero for both values.