Anthony Nardiello, Norhan Abbas, Robert Lewis
CSCI 447
9/26/19

For the second assignment, we have been asked to design an experiment running five algorithms over 6 data sets. During this assignment, we have also been tasked with designing an experiment to explain our results from these different algorithms. With these restrictions in mind, the group has come up with a few different ideas about how we will go about this experiment. In the following document, we will be discussing some of the ideas, including our tuning of the hyper parameters, working basic data collection, and comparison across different problems.

First, we need to discuss the ideas of the hyperparameters, or the K-values, from which our algorithm draws its ability to operate. K Nearest Neighbor, or KNN for short, relies on this K-value to influence the decisions made by our algorithm in choosing a nearest neighbor. The decision of the group has been to tune the hyper parameters through an implementation of min-max randomization. For further detail, we will be setting a minimum and maximum from which we will draw out the hyperparameters. We will then discreetly choose one value from that range for each experimental runthrough of our algorithm. The hyperparameter process will be thoroughly tested to make sure that our discrete selection process is not selecting similar values for each test run.

Next, the discussion about how we will collect and run each experimental runthrough must be addressed. For this section, we will be working through the problems as a divide and conquer approach. We will work through the experimental procedure for one of the two dataset types that we have, then work through the other. With that said, our first step will be working through the classification data set first. This way, we can work through any issues that may come up with the algorithms and our implementation with evaluation functions with which we have a better understanding. Then, we can move towards gathering and recording the results of the experimental procedure from the regression data sets. With the data collection completed in two distinct phases, we will have a better understanding of how this algorithm performs.

Finally, we move to the comparison of each algorithm's result, and making the data into a more coherent form. Naturally, any conclusion we can draw will be available at this time, as we will finally have the data to compare to our hypothesis. At this stage, we can take a look at the data, and see if we have any data sets that did not perform as anticipated. Similarly, we can look at how the data will reinforce our assumptions made when coding the algorithm. This will also be an excellent chance to refine our hyperparameters if necessary, especially before we begin trial 2 and 3.