Joshua Hoshiko

CS 3120

Dr. Jiang

Homework 3 Report

**Program Overview:**

This program uses the KNN Classification model to analyze images of animals and then classify them.

1. The program first loops and loads all of the images from a directory, assigning labels and resizing them
2. The program then reformats the data to be in the correct shape and we use this data to make data vectors
3. The program then slices the program into three parts: training at 60%, validation at 10%, and testing at 20%
4. The program then generates/tests the KNN model with the validation data trying different combinations of K and Manhattan/Euclidean distance metrics
5. The different combinations are scored based on accuracy and the best combination is then used to test using the testing data set

**Program Details:**

1. The best combination determined by this program was K at a value of 5 using Manhattan distance. However, the accuracy was only marginally better than the other combinations (Accuracy was only about 1% better than the others). The poor accuracy can be attributed to KNN’s inability to analyze the images in a way we need it too. Another interesting observation was a drop in accuracy from the validation data vs the testing data (accuracy of 51% dropped to 43%), and talking with other students, the combination is not always determined to be the same, with many other students having different results.

**Program Output:**

