CSC 401, Weekly Coding Assignment #8, Fall 2024 Name_____

Instructions: For these weekly coding assignments, you will be asked to extend the examples from class to create custom code to answer the questions below. You will create an R code file that uses built-in datasets as the data sources. You will write the code, and an example showing that the code works. Be sure to include any packages in the code that are required for the functions to run (you may want to clear the environment in RStudio before your final check to make sure nothing is missing). The instructor will run the file to ensure that it works with no errors. Clearly label your code so it's clear which question/task is being responded to.

Submission:
A word document with any explanations (if needed), and a clearly labeled R code file.

Tasks/Questions:

1. Modify the code from the lecture notes to compare Euclidean and cosine similarity. Use the iris dataset to compare k=3 to k=8. Calculate the confusion matrix and accuracy. Construct a plot that compares the k levels and distance metrics for the best fit.

2. Use the example from lecture as a model, run a simulation of 500 and create a box plot similar to the one in the example for one of the other tie-breaking methods. How can you use the boxplots to compare stability? Which method is more stable? Random or the alternative you chose?

3. Use the example from lecture as a basis, but retain the setosa variable in the iris dataset and run tests for k=3 to k=10 with random tiebreaking. How does it do when there are three classes versus 2 classes.

4. Modify the basic KNN algorithm for regression to apply alternative means of predicting—such as the median, harmonic mean, etc. (see the lecture notes for other options) Compare the results from the mtcars dataset. Compare the results using a graph and at least three metrics.