CSC 401, Weekly Coding Assignment #3, 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. Calculate a single-variable regression using the normal equation. You can use the Orange dataset to predict circumference from age. Construct the covariance matrix.

2. Use the mtcars dataset to complete the following regression examples using the normal equation. Use the other variables in the data set to model mpg. In the following exercise, treat $x_1 = cyl, x_2 = disp, x_3 = hp, x_4 = drat, x_5 = wt, x_6 = qsec, x_7 = vs, x_8 = am, x_9 = gear, x_{10} = carb$ $(and\ y = mpg)$.
   a. $y = \beta_0 + \beta_1 x_5 + \beta_2 x_5^2$
   b. $y = \beta_1 x_1 + \beta_2 x_1^2 + \beta_2 x_2 x_{10}$
   c. $y = \beta_0 + \beta_1 \ln x_5 + \beta_2 x_2 + \beta_3 \sqrt{x_3}$
   d. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10}$

   In each case, after determining the coefficients, write the equation of your model. Plot the predictions vs. the true values.

3. Construct a linear model, a LOESS model, a gaussian process model of mpg using hp from mtcars.
   a. Find the residuals of each model
   b. Conduct a thorough residual analysis of each model
   c. Calculate the following metrics for each model
      i. Adjusted R^2
      ii. Mean absolute error (MAE)
      iii. Mean squared log error (MSLE)
      iv. C_p
   d. Identify any outliers
   e. Compare your metrics, where possible, to built-in functions for these metrics, if they exist, and if they work on your particular model type. Where they exist, make sure your outcomes are the same. Describe where they fail to exist, or fail to work.

4. Calculate a ridge regression model using the variables in problem 2d (above). Then repeat the calculations in problem 3 for this model.