CSC 401, Weekly Coding Assignment #4, 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. Use the entire mtcars dataset (10 variables) to predict mpg with a) Ridge Regression, b) LASSO Regression, c) Elastic Net Regression. Test w/o rescaling vs. with rescaling. How does the performance differ? How do the coefficients differ? Use the equations you obtain to predict y (mpg) and calculate at least three regression metrics (of your choice) per model to compare the results. How did scaling affect the quality of your model?

2. Repeat problem 1, with a version of LASSO using a power for the penalty: $\lambda \sum_{i=1}^{n} |\beta_i|^p$ (note that if p=1, you have traditional LASSO, and if p=2, you have Ridge regression). Test: a) $0 < p < 1$, b) $1 < p < 2$, c) $p > 2$. Describe the behaviour you see.

3. Apply a one-variable penalized B-spline to model mpg from hp. Experiment with different λs and check your metrics on at least five values to determine which works the best for this data.

4. Using qsec and disp (rescale both), model mpg with a multivariable spline. Consider both additive splines and the tensor spline. Visualize the results. Check at least one metric to determine which is the better predictor.

5. Perform a grid search on a single variable spline model (say for mpg by hp) to select the optimal number of knots and optimal λ. Calculate regression metrics (at least three) and plot the models for comparison.