STA4241 Interactive Lab 12: Estimating smooth and nonlinear functions

Today we will focus on estimating flexible functions in the model

$$Y_i = f(X_i) + \epsilon_i.$$

We will have one covariate X and a continuous outcome Y. Today's lab will be interactive! I want you to download the training data off of the course website via the following lines of code:

```
data = read.csv(file="TrainingData.csv", header=TRUE)
x = data$x
y = data$y
```

Using this data, I want you to come up with an estimate of $f(\cdot)$ using any of the approaches we have seen so far. To refresh your memory, these include

1. K-nearest neighbors regression

2. Kernel regression

3. Local regression

4. Natural cubic splines

5. Smoothing splines

6. Combinations of these or combinations with penalized regression approaches

7. Anything else you can think of or come up with!

You will then use your function to predict the outcome at a new set of locations given by

```
xnew = seq(-10, 10, by = 0.01)
```

At the end of the lab, I want you to email me your predictions in a .csv file titled Lastname_Firstname.csv. This should have a very specific structure, and a failure to abide by this structure will eliminate your chances of extra credit. Your file should only contain columns with your predictions for the outcome at these locations. You may submit up to 3 predictions and the column names for your predictions should be prediction1, prediction2, and prediction3. If you only have one prediction, that's fine, call it prediction1. Each of these columns should be the same length as xnew. Example code for saving your predictions and writing them to a file is below:

```
finalPredictions = data.frame(prediction1 = prediction1,
                              prediction2 = prediction2,
                              prediction3 = prediction3)
write.csv(finalPredictions, file="~/YourDirectory/Antonelli_Joseph.csv")
```

You will get one bonus point on your midterm exam score for every prediction that has MSE below a certain threshold. The student with the lowest MSE with receive an additional bonus point. You will also get an additional bonus point if any of your predictions have a lower MSE than the one I came up with. Therefore, at most you can get 5 bonus points on your exam score. In your email, you must tell me what methods you used to obtain each prediction. You may not submit 3 predictions from the same approach with different tuning parameters. They must be three distinct types of predictions. Good luck!