CS 5565, LAB6 (Moving Beyond Linearity) 50 Points

Name _____

1. View the videos at the following URLs
   https://www.youtube.com/watch?v=uQBnDGu6TYU
   https://www.youtube.com/watch?v=DCn83aXXuHc
   You may download the R Code for Labs and the Data Sets to use from the textbook website.
   http://www-bcf.usc.edu/~gareth/ISL/

2. (20 points total) In this exercise, you will further analyze the `Wage` data set considered throughout this chapter.

   (a) (10 points) Perform polynomial regression to predict `wage` using `age`. Use cross-validation to select the optimal degree d for the polynomial. What degree was chosen, and how does this compare to the results of hypothesis testing using ANOVA? Make a plot of the resulting polynomial fit to the data.

   (b) (10 points) Fit a step function to predict `wage` using `age`, and perform crossvalidation to choose the optimal number of cuts. Make a plot of the fit obtained.

3. (30 points total) This question uses the variables `dis` (the weighted mean of distances to five Boston employment centers) and `nox` (nitrogen oxides concentration in parts per 10 million) from the Boston data. We will treat `dis` as the predictor and `nox` as the response.

   (a) (5 points) Use the `poly()` function to fit a cubic polynomial regression to predict `nox` using `dis`. Report the regression output, and plot the resulting data and polynomial fits.

   (b) (5 points) Plot the polynomial fits for a range of different polynomial degrees (say, from 1 to 10), and report the associated residual sum of squares.

   (c) (5 points) Perform cross-validation or another approach to select the optimal degree for the polynomial, and explain your results.

   (d) (5 points) Use the `bs()` function to fit a regression spline to predict nox using `dis`. Report the output for the fit using four degrees of freedom. How did you choose the knots? Plot the resulting fit.

   (e) (5 points) Now fit a regression spline for a range of degrees of freedom, and plot the resulting fits and report the resulting RSS. Describe the results obtained.

   (f) (5 points) Perform cross-validation or another approach in order to select the best degrees of freedom for a regression spline on this data. Describe your results.