STAT 4533/5533 - Homework 5 - Due March 17, 2024

Instructions: Everyone must answer Questions 1(a)-(e) and 2. Graduate students are also required to answer Question 1(f) and Question 3. Questions 4-6 are optional and will not be graded. Solutions must be created using R Markdown, and you will upload your HTML file to D2L to submit your answers. Your markdown document must be logically organized and easy to read. Do not print entire data sets in your output.

1. The file *FEV.csv* on D2L contains information related to forced expiratory volume (FEV) tests evaluating pulmonary capacity.

   (a) Create and interpret a plot of height versus FEV treating FEV as the dependent variable.

   (b) Fit a natural cubic spline to predict FEV using height. Select the degrees of freedom using LOOCV choosing between values of 5, 6, 7, and 8. What value is chosen?

   (c) Fit a smoothing spline to predict FEV using height. Use LOOCV to choose the value of the smoothing parameter. Note: you will probably get a warning here - you can ignore this warning.

   (d) Fit a local regression using span of 0.8 to predict FEV using height.

   (e) Add the resulting fits from parts (b) - (d) to your plot from part (a). You only need to plot the best model from part (b).

   (f) GRADUATE STUDENTS ONLY: Create a training and test set and repeat parts (b) - (d) on the training data. Which model produces the lowest test MSE?

2. Chapter 7 Exercise 9 parts (d)-(f) only

3. GRADUATE STUDENTS ONLY: Chapter 7 Exercise 3

   The following question is recommended, but will not be graded:

4. Chapter 7 Exercise 6 (use $d = 1, 2, 3, 4, 5$, use up to 5 cuts, ignore the ANOVA part)

5. Chapter 7 Exercise 8

6. Chapter 7 Exercise 10