Homework 2

STAT 697STA Spring 2023

Due February 22, 2023, 9:40am on Gradescope

1 Reading

- Read Chapter 2 through Section 2.10 of the text.

2 Questions

1. HRW 2.3, a-d, g

   For part (g), if you don't do the extra credit, you may assume the results of (e) and (f).

2. HRW 2.5

3. Please think about the structure of penalized spline fits, as described in the text.
   (a) List at least 3 tuning parameters a modeler using penalized splines has control over.
   (b) Which tuning parameters has the greatest impact on the model fit? Briefly describe how this parameter changes the model fit. Be sure the be clear about what would happen with an extremely large or extremely small value of this parameter.
   (c) What is one way to automatically choose this tuning parameter?

4. The below code creates 3 datasets on the interval $[0, 1]$.

```
set.seed(1) ; n <- 500 ; x <- sort(runif(n))
y_curve <- (6*x + sin(4*pi*x^2))/(5*x+1) + 0.1*rnorm(n)
plot(x,y_curve,bty="l", col="dodgerblue")
y_line <- x + 0.1*rnorm(n)
points(x,y_line, bty="l", col="orange")
y_step <- as.numeric(x>.5) + 0.1*rnorm(n)
points(x,y_step, bty="l", col="darkgreen")
```

   (a) Before fitting spline models, which dataset do you think will benefit the most from a penalized spline fit?
   (b) Fit both a linear model and the same penalized spline model to all 3 of these datasets and plot the resulting fits and residuals. (so you should have 3 plots, each with a set of datapoints and 2 lines, one for the linear and one for the spline fit). *Hint: the following code adds a linear regression line to an existing plot:*
   `abline(lm(y~x), col="orange")`
   (c) Briefly describe what the 3 fits tell you about the usefulness of spline fits for different types of data patterns.
   (d) Compare the effective degrees of freedom of the 3 fits. Briefly discuss what you see. *Hint: the following code finds the EDF of a model fitted using the* `gam()` *function. Note that this function excludes the intercept in the EDF calculation.*
   `summary(fitGAMcurve)$edf`

5. (EXTRA CREDIT points) HRW 2.3 e, and f