

HW 3 – 4803, Fall 2019

Instructor: Wenjing Liao

- HW 3 Part I is due on Wednesday October 2 at the beginning of the class.
- HW 3 Part II is due at the end of Friday October 11 by email.
- You are strongly encouraged to type out your solutions using latex.
- Please write your solutions independently, and include your code at the end of your solutions.

Part I (Conceptual questions)

5.4 Exercises: 3 (not graded)

7.9 Exercises: 1,2,3,4,5

Part II (Programming)

Programming Problem 1: 5.4 Exercise 8. Notice that (g) and (h) are added by the instructor.

(a): x is a vector of length 100 with each entry sampled from the normal distribution of mean 0 and variance 1. The noise is also sampled from the normal distribution of mean 0 and variance 1.

(f): This part asks how accurately the coefficients are estimated in each model.

(g): Use the validation set approach. In order to do this, you must perform the following steps:

- (1) Split the sample set into a training set and a validation set.
- (2) Fit the four models using the training set.
- (3) Compute the validation set error for the four models respectively, and compare the errors. Which model gives the least error?
- (4) Repeat the process above three times, using three different splits of the data into a training set a validation set. Comment on the results obtained.

(h): Compare LOOCV and the validation set approach through the experiments above.

Programming Problem 2: 7.9 Exercise 9. The Boston data set was used in HW 2.

(d): You can try cubic or natural cubic splines. The degree of freedom refers to the number of knots. If the package you use helps to choose the location of the knots, you can use it; if not, you can simply use equally spaced knots.