HW 1 – 4803, Fall 2019

Instructor: Wenjing Liao

- HW 1 is due on Wednesday September 4 at the beginning of the class.
- 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 (theoretical problems)

2.4 Exercises: 7 (the Bayes decision boundary is defined in Page 37-28)

3.7 Exercises: 3,4,5,6

Linear algebra exercise: Use the least square formula (pseudo-inverse $\beta = (A^T A)^{-1} A^T y$) to derive Equation (3.4) in the textbook.

Part II (programming)

Programming Problem 1: This question involves the use of simple linear regression on the Auto data set, which can be downloaded at `http://faculty.marshall.usc.edu/gareth-james/ISL/data.html`.

(1) Read the Auto data into matlab, python or R.

(2) Use least squares to perform a simple linear regression with "mpg" as the response and "horsepower" as the predictor.

(3) Plot the response and the predictor, and display the least squares regression line.

(4) Does the regression line fit the data well? What is the RSS?

Programming Problem 2: This question involves the use of multiple linear regression on the Auto data set.

(1) Produce a colorplot matrix which includes all of the variables except "name" in the data set.

(2) Compute the matrix of correlations between the variables, and display the matrix. You will need to exclude the "name" variable since it is qualitative.

(3) Use least squares to perform a multiple linear regression with "mpg"as the response and all other variables except "name "as the predictors. What are the coefficients and the RSS?

(4) Compare the result of multiple linear regression with simple linear regression above. Which one is better?

Programming Problem 3: 3.7 Exercises 13 (a-i) If you use matlab or Python, the functions are different from the ones in R.