Homework 4

Chapter 3 section 3.7 (pages 120-125):

Exercises 1, 2, 3, 7, and 13. Each exercise deserves 5 points.

Write down your answers to the exercises in Word document and submit the Word document to Canvas. Include plots in the Word document for exercise 13 (d) and (f).

For exercise 13, also submit your code in R Markdown and the HTML report directly generated from the R Markdown.

Notes:

1. To get document for the rnorm() function in R, type "?rnorm".

2. Please submit all individual documents. Do not zip the documents.

1. Describe the null hypotheses to which the p-values given in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.
   1. The p-values of TV and Radio both have p < 0.0001 and shows that we can reject the null hypothesis and we can conclude that TV and Radio both have an impact on sales. The p value of newspaper shows that newspaper does not have an impact on sales
2. Carefully explain the differences between the KNN classifier and KNN regression methods.
   1. KNN classifier classifies a result into a qualitative group based on using the most common group found within the K nearest neighbors. While the KNN regression makes a quantitative estimate by averaging the result of the K nearest neighbor.
3. Suppose we have a data set with five predictors, X1 = GPA, X2 = IQ, X3 = Gender (1 for Female and 0 for Male), X4 = Interaction between GPA and IQ, and X5 = Interaction between GPA and Gender. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to fit the model, and get βˆ0 = 50, βˆ1 = 2 0 , βˆ 2 = 0 . 0 7 , βˆ 3 = 3 5 , βˆ 4 = 0 . 0 1 , βˆ 5 = − 1 0 .
   1. 3. For a fixes value of IQ and GPA, males earn more on average than females provided that the GPA is high enough. The starting salary for males is higher than for females on average because 50 + 20GPA ≥ 85 + 10GPA ~ GPA ≥ 3.5

7. It is claimed in the text that in the case of simple linear regression of Y onto X, the R2 statistic (3.17) is equal to the square of the correlation between X and Y (3.18). Prove that this is the case. For simplicity, you may assume that x ̄ = y ̄ = 0.

a. After simplifying down the equation I got R^2 = Cor(x,y)^2

13. In this exercise you will create some simulated data and will fit simple linear regression models to it. Make sure to use set.seed(1) prior to starting part (a) to ensure consistent results.

d. Chart, scatter chart

Description automatically generated

f.Chart, scatter chart

Description automatically generated