DATS 7860 Statistical and Machine Learning for Big Data

Homework 1

Textbook (required): *An Introduction to Statistical Learning with Applications in R*, 2nd Edition (ISBN: 978-1071614174), by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2021), Springer (**Referred as ISL in this course**).
The PDF version of the book, R codes for labs and datasets are publicly available from the book's website: https://www.statlearning.com/
A PDF version of ISL was also uploaded to D2L "textbook" folder.

Problems:

1. Problem 7 from Chapter 2 of ISL (Page 54)

2. Problem 3 from Chapter 3 of ISL (Page 121).

3. Problem 10 from Chapter 3 of ISL (Page 124).

Some specific notes for this problem:

- Please submit your R codes for this problem.

- The Carseats dataset is available from the R package ISLR2 associated with the textbook. You can use the following R codes to install the package and load the dataset:

```
install.packages('ISLR2')

require(ISLR2)

data(Carseats)
```

- For part (a), please show the summary of the model (output from summary function in R) in your answer sheet.

- For part (e), please show the summary of the model (output from summary function in R) in your answer sheet.

- For part (f), hint: you can consider using the statistics *RSE* and R^2 to answer this question.

- For part (h), please first read Section 3.3.3 of ISL before answering this question. You may use some plots to justify your answer.

4. Problem 13 from Chapter 3 of ISL (Page 126)

Some specific notes for this problem:

- Please submit your R codes for this problem.

- For part (e), please also show the summary of the model (output from summary function in R) in your answer sheet.

- For part (g), please also show the summary of the model (output from summary function in R) in your answer sheet.

- For part (h): you may want to try several different values for the standard deviation of ε, but showing the results for one of them will be sufficient. Also, do not use too small values for the standard deviation, which will make the data almost a perfect line.

- For part (i): same requirement as (h). Also, do not use too large values for the standard deviation.