# Week-3-Prep

September 9, 2020

## 1 Week-3 Prep and Homework-1

### 1.1 Chapter-3 of Introduction to Statistical Learning: Linear Regression- Until 3.5 (p 105)

As you are reading please keep these questions in your mind:

---

- Understand what is slope and intercept in simple linear regression.
- What are the coefficient parameters?
- What *the least squares* method measures?
- What is $e_i$ standing for in the context of 'advertising' dataset?
- **You can skip the derivations of $\beta_1, \beta_2$ in your first read**
- **You don't need to memorize their formula!**
- How would you explain the notion of biased and unbiased estimators?
- **You can skip pages 66, 67**
- How do you interpret $R^2$ statistics?
- How do you go from simple linear regression and multiple linear regression?
- Can you explain why "newspaper" variable was "significant" in the simple linear regression model but not in the multiple linear regression model?
- What are the some tools that helps us to choose between different models?
- Explain Forward Selection.
- What is interaction effect?
- How do you make predictions once you estimated the coefficients $\beta_i$'s?
- How do you deal with qualitative data in linear models?
- What is the effect of removing additive assumption on your linear model?
- What is *Hierarchical principal*?
- What are the potential problems you might encounter in a regression model?
- How do you detect these potential problems?

## 1.2 The first homework details

Recall that the first homework is due to week-4 lecture (09/22).Your task will be applying a linear regression model to a dataset of your choice.

### 1.2.1 Directions:

- You cannot work with datasets that is explored in our textbooks or in our classes.
- The Dataset should contain more than 200 observations (rows) and more than 4 columns.
- Data can be relatively clean but you should be doing the preprocessing procedures on your own.
- As it is explained in Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow (2nd edition) chapter-2, you should be able to explain the business goal of your project very clearly.
- If you are applying a transformation to some of your features you should explain the rationale of such decisions.
- You should use visualization techniques to explore the data and to support your results.
- You should be able to assess the model fitness and you should do model diagnosis as it is described in ISLR 3.3.3.

## 1.3 Deliverables:

As I mentioned in the first class, I want you to upload your project into Github and share it with me the github link. Github folder should contain three main ingredient:

1. ReadMe
2. Code + Notebooks
3. Summary + Report.

I will explain more on what I will be looking for in each of these items later in the class.

Here is a good example of a very good project done by one of my previous students: Alphonso Woodbury - Project