An Introduction to Statistical Learning

(Summary by *Pablo Vivas*)

Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani

January 13, 2020

1. Introduction

Statistical learning refers to a vast set of tools for **understanding data**. These tools can be classified as:

- supervised

- unsupervised

Broadly speaking, supervised statistical learning involves building a statistical model for predicting, or estimating, an output based on one or more inputs. Problems of this nature occur in fields as diverse as business, medicine, astrophysics, and public policy. With unsupervised statistical learning, there are inputs but no supervising output; nevertheless we can learn relationships and structure from such data. The datasets used in this books are:

- Wage

- Stock Market

- Gene Expression

Some history

> Legendre and Gauss published papers on the method of *least squares*.
> Fisher proposed *linear discriminant analysis* in 1936.
> In the 1940s, various authors put forth an alternative approach, *logistic regression*.
> In the early 1970s, Nelder and Wedderburn coined the term *generalized linear models* for an entire class of statistical learning methods that include both linear and logistic regression as special cases.
> In mid 1980s Breiman, Friedman, Olshen and Stone introduced *classification and regression trees*.
> Hastie and Tibshirani coined the term *generalized additive models* in 1986 for a class of non-linear extensions to generalized linear models.

2. **Statistical Learning**

3. **Linear Regression**

4. **Classification**

5. **Resampling Methods**

6. **Linear Model Selection and Regularization**

7. **Moving Beyond Linearity**

8. **Tree-Based Methods**

9. **Support Vector Machines**

10. **Unsupervised Learning**