Notes and exercises from "An Introduction to Statistical Leaning"" by James, Witten, Hastie & Tibshirani

Nicholas McGlincy

July 21st 2016

Table of Contents

# Chapter 1 Introduction

The introduction outlines three basic tasks in statistical learning:

1. **Regression** - predicting *continuous numerical* values from other continuous values
2. **Classification** - predicting class membership; has a *categorical* or *qualitative* output.

In both these tasks, we are trying to predict an output.

1. **Clustering** - here we are not trying to predict an output variable, but instead are trying to understand how similar/different a set of observations are, usually based on a number of variables.

It is clear that I need to revise matrix multiplication, I didn't really understand the example on p. 12.

The book's website is [here](http://www-bcf.usc.edu/~gareth/ISL/).

# Chapter 2 Statistical Learning

Broadly, the goal of statistical learning is to estimate in the equation:

Where:

is the *response/output/dependent variable*.  
 is one or more *input/independent variables*, or *predictors* or *features*.  
 is a function that relates to , and  
 is a random error term, that is independent of and has mean zero.

We estimate for two ends:

1. **To predict**  -- here we are not primarily concerned with the exact nature of the function, rather how accurately it predicts . Thus our estimate of comes from our estimate of :

* The accuracy of depends on two quantities:
* A. *The reducible error* -- will be a more-a-less imperfect estimate of . This inaccuracy is termed the *reducible error* as it is reducible by determining a better .
* B. *The irreducible error* -- Even if our estimate of was perfect, it would still have error in it, as is also a function of , which by definition cannot be predicted from - hence *irreducible error*.

1. **To infer the relationship between and $X\_1,\dotsc,X\_p$**