Contents

Introduction

My notes and solutions/attempts for G. James et al., *An Introduction to Statistical Learning: with Applications in R*

Book website: http://www-bcf.usc.edu/~gareth/ISL/

GitHub repo with solutions by others: https://github.com/asadoughi/stat-learning

Notes

- `Wage` dataset, plenty to learn from a couple simple plots (scatters + boxplot) see book Figure 1.1, regression problems
- `Smarket` dataset, to be used for classification
- `NCI60` gene expression dataset, used for clustering
- History of machine learning interesting. . . would like to learn more. From Gauss through Fisher, Nelder and Wedderburn then 1980s+ non-linearity, regression trees etc. Breiman, Friedman, Hastie, Tibshirani. . .
- Intended audience: people interested in using techniques, but not PhDs/advanced degree holders in stats, ML, etc. less technical than *Elements of Statistical Learning* (ESL)
- "The purpose of ISLR is to facilitate the transition of statistical learning from an academic to mainstream field"
- ISLR premises:
    - Utility of statistical learning beyond just statistics (and beyond academia)
    - "Statistical learning should not be viewed as a series of black boxes" – understand what is happening in the methods/models!
    - You need to understand what's happening in the box. . . not necessarily be able to build the box. – again, the emphasis on less technical nature of ESL
    - Application to real-world problems (via R)
- Need to get `ISLR` package

Exercises

None, just get `ISLR`, should already have `MASS`

```r
install.packages('ISLR')
```