Homework3

Leanne Lee

October 12, 2016

Abstract

This homework is to reproduce the analysis from Section 3.2 (pages 71 to 82), from the book "An Introduction to Statistical Learning" (by James et al). It includues multiple linear regressin with the predictor variables TV, Radio, Newspaper and the response variable Sales.

Introduction

Given the 3 predictor variables of TV, Radio and Newspaper, we need to find out what is the relationship between the predictor variables and response variable of Sales. The main goal is to find out which type of advertisement is more effective to increase the sales. Therefore, we break down the linear regression from comparing a single variable to multiple variables. By reproducing the result of the regressions, the marketing team can determine which type of advertising should they invest on.

Data

The dataset **Advertising.csv** comes from *"http://www-bcf.usc.edu/~gareth/ISL/Advertising.csv* It consists for TV, Radio, Newspaper and Sales columns. The structure of the columns are stored in numeric vectors.

Methodology

For simple linear regression:

$$Sales = \beta_0 + \beta_1 * TV$$

Simple Linear regression is useful when predicting a response based on one single predictor variable. For multiple linear regression:

$$Sales = \beta_0 + \beta_1 * TV + \beta_2 * Radio + \beta_3 * Newspaper$$

Multiple Linear regression is useful for Advertising data becasue it can determined the relationship between sales and the three types of advertising. We can compare which type of advertisment is more effective and has a stronger association with sales.

Results

Conclusions