DATA130013: Homework 3

Due in class on April 12, 2018

1. Shumway's book (4th ed.) Problems 2.2, 2.3, 2.6 2.7, and 2.8.

2. (Geometric interpretation of R square) Consider the simple linear regression model (refer to our last homework): i.i.d. observations $\{(x_i, y_i)\}_{i=1}^n$ follow a linear relationship
$$y_i = \beta_0 + \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, n.$$
The coefficient of determination is defined as
$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$
the ratio of the regression sum of squares and the total sum of squares.
Derive that
$$R^2 = \left\{ \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \right\}^2,$$
i.e., R^2 equals the squared correlation between covariate vector $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and response vector $\mathbf{y} = (y_1, \ldots, y_n)^\top$. Comment this result from the geometric perspective.