Assignment 2

Your Name

30 September 2023

Question 1: Consider the simple linear regression problem $y = \beta_0 + \beta_1 x + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Assume we are given training data $\{(x_i, y_i)\}_{i=1}^n$ and let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the fitted model. Define $e_i = y_i - \hat{y}_i$. Show the following:

1. $\sum_{i=1}^n e_i = 0$.
2. The regression line always goes through the point $(\bar{x}, \bar{y})$.

Question 2: This question involves the use of multiple linear regression on `Auto` data set. For most of the analysis, you will need to remove the name variable.

1. Produce a scatterplot matrix of all of the variables (you can use the `pair()` command for this.
2. Use the `lm()` function to perform a multiple linear regression with mpg as the response variable. Use the `summary()` command to print the results. Comment on your findings.
3. Use the `ggplot()` package to plot your findings.
4. Based on the correlation matrix and the scatterplots, try transformation of the predictors (e.g, X^2 or $\sqrt{X}$). Comment on your findings.