

Homework 6

Christopher Williams

October 28, 2025

1 EDA

Feature Selection & Prediction

2 Backward Stepwise Regression

3 PCA

4 Random Forest Analysis

5 Table of Metrics

6 Prediction Interval

7 Polynomial Regression and Grid Search

8 Simple Linear Regression

In a simple linear regression with n observations:

$$y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (1)$$

Prove the following:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (3)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ are the sample mean.