Honors by Contract Homework - MATH 425

Owen Queen

April 19, 2021

1 Least-Squares Linear Optimization Proof

We are tasked with proving the following theorem: Given n points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, the straight line $y = a + bx$ minimizing

$$L = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

has slope

$$b = \frac{n \sum\limits_{i=1}^{n} x_i y_i - \left(\sum\limits_{i=1}^{n} x_i\right)\left(\sum\limits_{i=1}^{n} y_i\right)}{n \sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2}$$

and y-intercept

$$a = \frac{\sum\limits_{i=1}^{n} y_i - b \sum\limits_{i=1}^{n} x_i}{n}$$

1.1 Proof

Assume we have n points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$, and let

$$L = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

We know that

$$\frac{\partial L}{\partial a} = 0 = -2 \sum_{i=1}^{n} [y_i - (a + bx_i)]$$

and

$$\frac{\partial L}{\partial b} = 0 = -2\sum_{i=1}^{n} x_i[y_i - (a + bx_i)]$$

Rearrange $\frac{\partial L}{\partial a}$ to obtain

$$0 = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - \sum_{i=1}^{n} bx_i$$

$$\sum_{i=1}^{n} a = na = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} bx_i$$

Therefore,

$$a = \frac{\sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} bx_i}{n}$$

Rearrange $\frac{\partial L}{\partial b}$ to obtain

$$a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} bx_i^2$$

$$a = \frac{\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} bx_i^2}{\sum\limits_{i=1}^{n} x_i}$$

Plugging in the a value:

$$\frac{\sum\limits_{i=1}^{n} y_i - \sum\limits_{i=1}^{n} bx_i}{n} = \frac{\sum\limits_{i=1}^{n} x_i y_i - \sum\limits_{i=1}^{n} bx_i^2}{\sum\limits_{i=1}^{n} x_i}$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} bx_i = \frac{n\sum\limits_{i=1}^{n} x_i y_i - n\sum\limits_{i=1}^{n} bx_i^2}{\sum\limits_{i=1}^{n} x_i}$$

More rearranging:

$$bn\sum_{i=1}^{n} x_i^2 - b\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right) = n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)$$

Therefore:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

■

2 Least-Squares Linear Fit on Example Data

The data for this section is drawn from Case Study 11.2.1 in the Larsen and Marx textbook. For this portion, I simply plugged in the values from the data into the following equation for a linear model:

$$y = a + bx$$

where,

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}$$

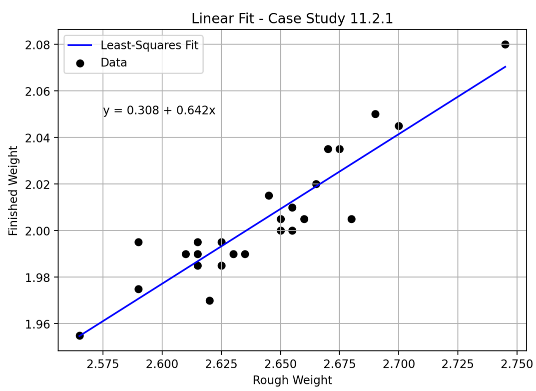

Figure 1: The least-squares fit for the data on a linear model seems to match the data from a visual standpoint.

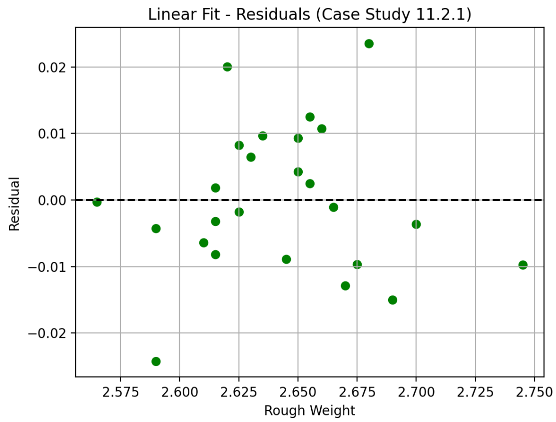

Figure 2: Residual plot for the data. All residuals are contained within a roughly $[-0.03, 0.03]$ range.

3 Least-Squares Non-Linear Fit on Example Data

For this portion, I simply plugged the values of the data in section 11.2.4 into the following nonlinear model:

$$y = ae^{bx}$$

where,

$$b = \frac{n\sum_{i=1}^{n} x_i \ln y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} \ln y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\ln a = \frac{\sum_{i=1}^{n} \ln y_i - b\sum_{i=1}^{n} x_i}{n}$$

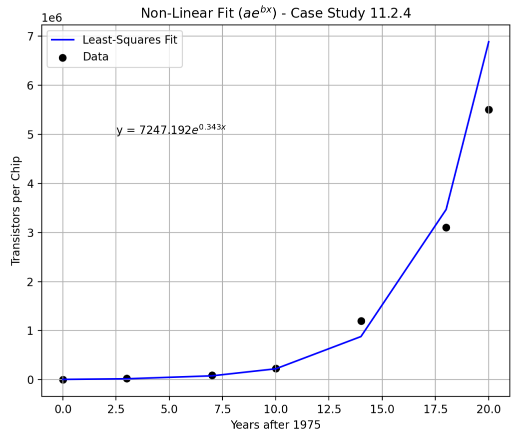

Figure 3: Exponential curve seems to fit the data very well. Note the y-axis has a $1\dot{1}0^6$ scale.

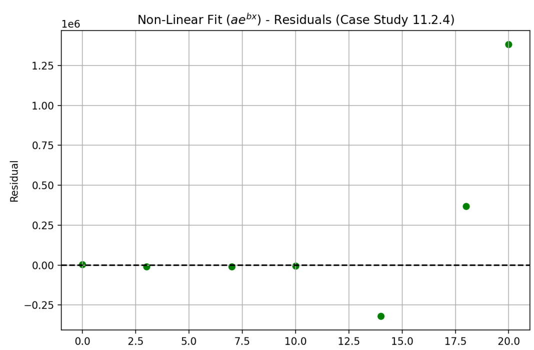

Figure 4: Residuals seem to increase with later years, which is expected since the scale grows exponentially.