Orthogonal Regression

Alexander Paul Torres

July 24, 2022

The slope-intercept equation for our line of best fit is:

$$y(x) = m \cdot x + b$$

A line which is perpendicular to our line of best fit and passes through a data point (x_i, y_i) is described by:

$$m' = -\frac{1}{m}, \quad b' = y_i - m' \cdot x_i = \frac{x_i}{m} + y_i$$
$$\therefore y'(x) = -\frac{1}{m} \cdot x + \left(\frac{x_i}{m} + y_i\right)$$

This perpendicular line passing through data point (x_i, y_i) will intersect our line of best fit at point some point (x_i^*, y_i^*) given by:

$$y(x_i^*) = y'(x_i^*), \quad \to m \cdot x_i^* + b = -\frac{1}{m} \cdot x_i^* + \left(\frac{x_i}{m} + y_i\right)$$
$$\therefore x_i^* = \frac{m}{m^2+1}\left(\frac{x_i}{m} + y_i - b\right), \quad y_i^* = m \cdot x_i^* + b$$

So the orthogonal distance from a given data point (x_i, y_i) through its point of intersection with our line of best fit is:

$$d_i = \sqrt{(x_i - x_i^*)^2 + (y_i - y_i^*)^2}$$

And thus our function to minimize in order to find the slope and intercept of our line of best fit is:

$$\begin{aligned} S[m, b] &= \sum_{i=1}^{N} d_i \\ &= \sum_{i=1}^{N} \sqrt{(x_i - x_i^*)^2 + (y_i - y_i^*)^2} \\ &= \sum_{i=1}^{N} \frac{|y_i - (m \cdot x_i + b)|}{\sqrt{m^2+1}} \end{aligned}$$

Plots below were generated using the Broyden–Fletcher–Goldfarb–Shanno iterative algorithm to minimize our function $S[m, b]$ to find our best fit values of m and b. Data being plotted included randomly generated Gaussian noise in the y-values.

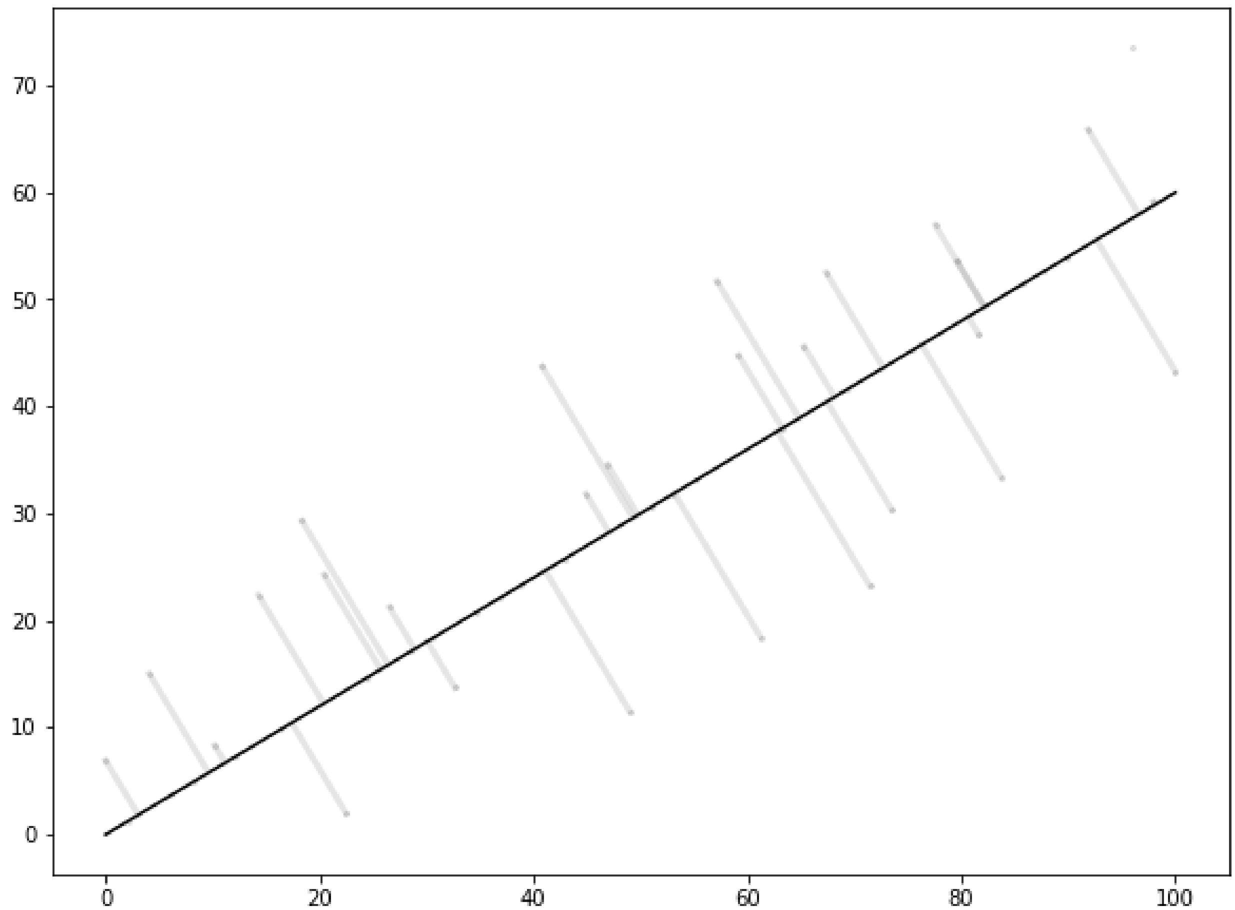

Figure 1: Demonstration of orthogonal regression.

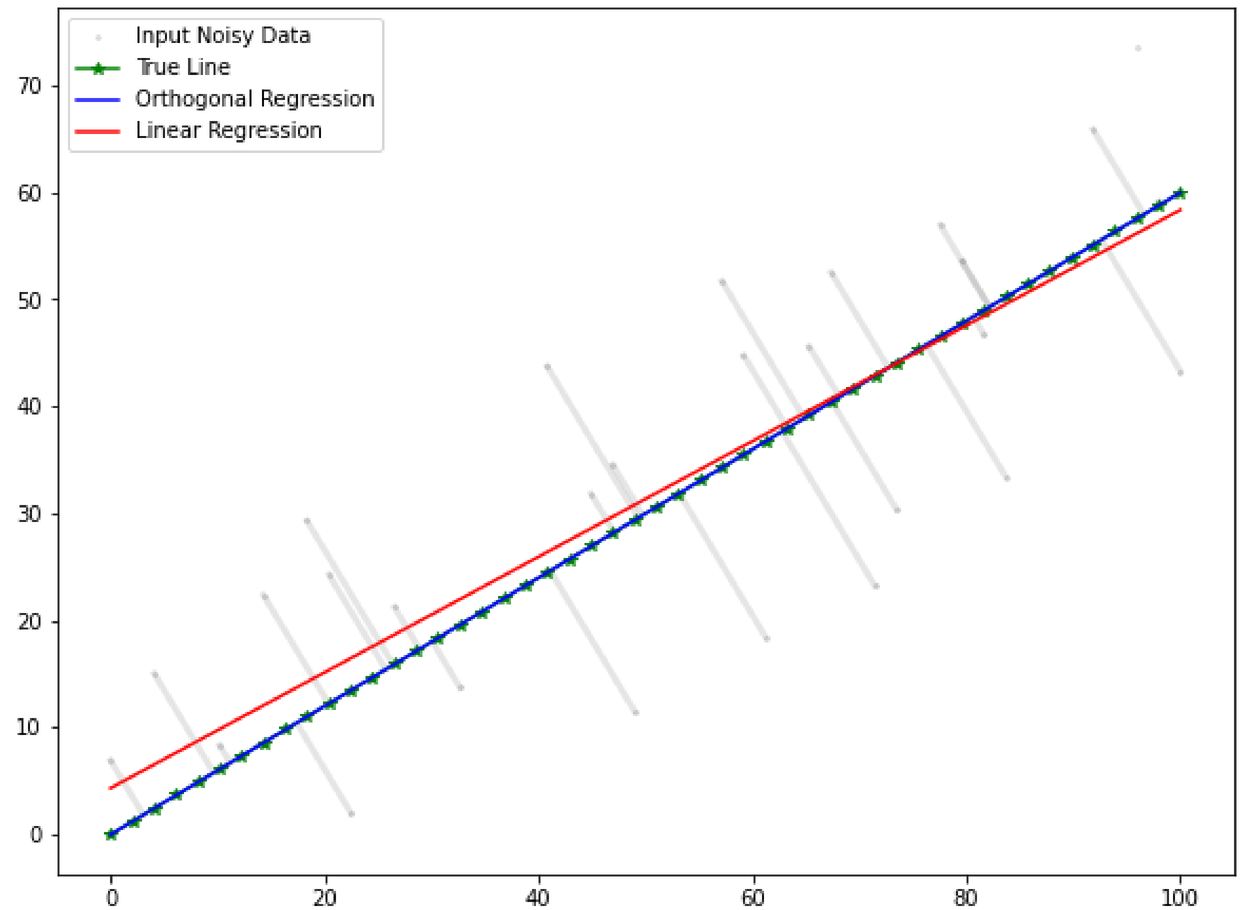

Figure 2: Example comparison between fit from linear regression and the fit from orthogonal regression. The green line with asterisks being overlapped by the blue line represents the data before noise was introduced. The black points represent the data after noise was introduced and which was used to perform both the linear and orthogonal regressions.

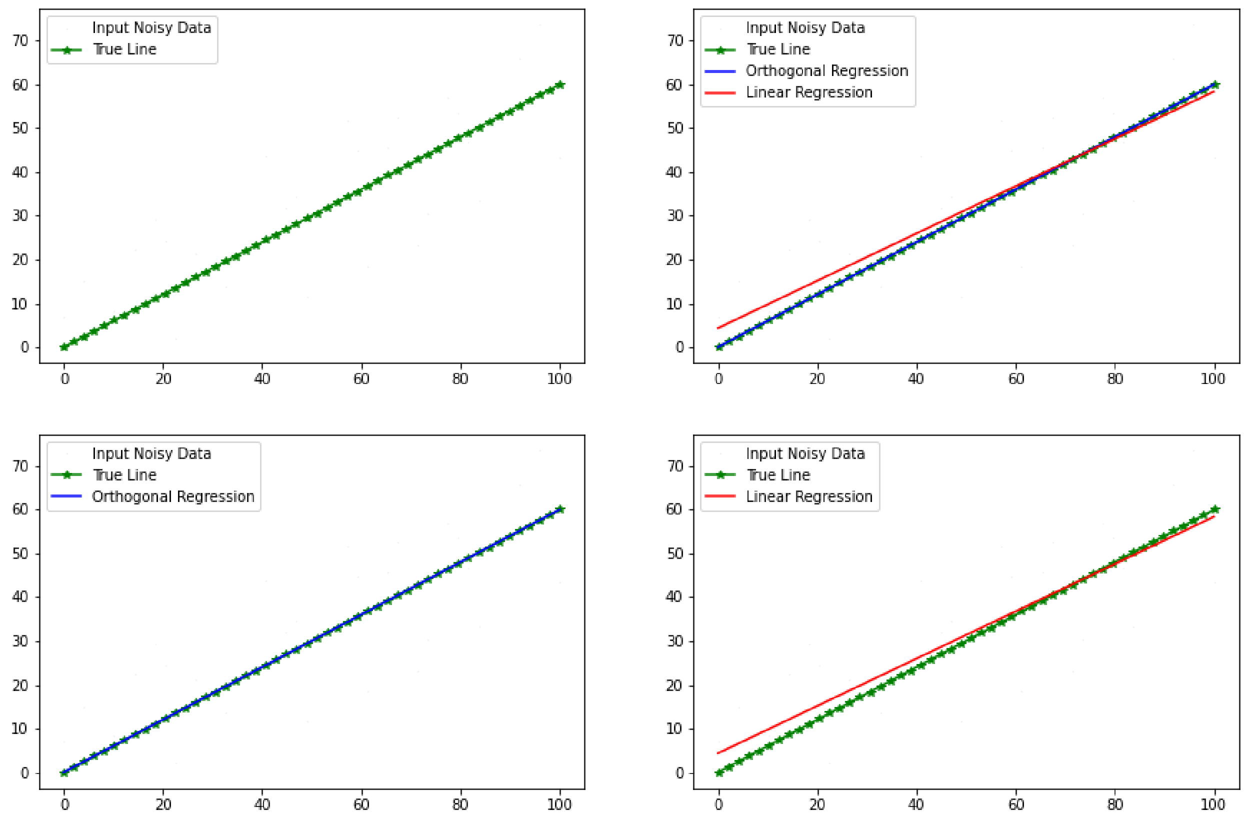

Figure 3: Example comparisons between linear and orthogonal regression.