METHOD OF LEAST SQUARES FOR A LINE

Anup V Kanale
February 2, 2017

Problem Statement

Given a set of N ordered pairs (x_i, y_i), find a line through these points such that the square of the error is minimized.

Derivation

For every given x_i, let the fitted y_i be obtained from the general equation of the line

$$y_{fit} = ax_i + b \tag{1}$$

where a is the slope of the line and b is the y-intercept.

Now, the error is given by,

$$E = \sum_{i=1}^{N} (y_{fit} - y_i)^2 \tag{2}$$

$$= \sum_{i=1}^{N} (ax_i + b - y_i)^2 \tag{3}$$

For a given set of x_i and y_i, we see that the error is a function of a and b, i.e., $E \equiv E(a, b)$.

The extremum (maximum or minimum) of a function occurs when it's slope is zero, or mathematically put, when the first derivative of the function if zero. Therefore,

$$dE = \frac{\partial E}{\partial a} da + \frac{\partial E}{\partial b} db = 0 \tag{4}$$

$$\therefore \frac{\partial E}{\partial a} = 0, \quad and \quad \frac{\partial E}{\partial b} = 0 \tag{5}$$

Now,

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{N} 2(ax_i + b - y_i) \cdot x_i = 0 \tag{6}$$

$$\implies \left(\sum_{i=1}^{N} x_i^2 \right) a + \left(\sum_{i=1}^{N} x_i \right) b = \sum_{i=1}^{N} x_i y_i \tag{7}$$

Similarly,

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{N} 2(ax_i + b - y_i) \tag{8}$$

$$\implies \left(\sum_{i=1}^{N} x_i \right) a + (N)\, b = \sum_{i=1}^{N} y_i \tag{9}$$

This is a system of linear equations which can be solved for a and b

$$a = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2} \tag{10}$$

$$b = \frac{\sum_{i=1}^{N} y_i - \left(\sum_{i=1}^{N} x_i \right) a}{N} \tag{11}$$

There is an accompanying piece of code written in matlab which calculates and plots the fitted line and the given N ordered pairs.