PC1134 Lecture 10

Topic

Error in Least Square Fitting

Objectives

Learn how to estimate the errors in slope and y-intercept of linear curve fitting.

Relevance

Sometimes, the relationship between two quantities is linear, or can be linearized. Certain parameters can be obtained from the slope and y-intercept of the linear graph.

Curve Fitting

From last lecture, fitting straight line $y = mx + b$ to data points $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$. The slope m and the y-intercept are given by

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

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

We assume that the x values are precise, that all the uncertainty is contained in the y values. m and b are calculated by using these data, there must be also uncertainty in m and b.

Least Square Fitting

Consider m and b are functions of $y_1, y_2, \cdots, y_N$. Then

$$\sigma_m^2 = \sum_{i=1}^{N} \left(\frac{\partial m}{\partial y_i}\right)^2 \sigma_{y_i}^2$$

$$\sigma_b^2 = \sum_{i=1}^{N} \left(\frac{\partial b}{\partial y_i}\right)^2 \sigma_{y_i}^2$$

In obtaining the set of measured values of x and y, one would normally determine the standard deviation for each y value. In the absence of these, the standard deviation of y_i can be replaced by the standard deviation of the distribution $\delta y_i = y_i - (mx_i + b)$ values about the best line,

$$\sigma_{y_i} = \sqrt{\frac{1}{N-2} \sum_{j=1}^{N} (\delta y_i)^2} = \sigma_y$$

σ_y is calculated with $N-2$ instead of N or $N-1$. This is a consequence of applying the definition of the standard deviation to the positioning of a line on a plane.

Least Square Fitting

Since

$$m = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i^2}$$

$$\frac{\partial m}{\partial y_i} = \frac{N x_i - \sum_{j=1}^{N} x_j}{N\sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j\right)^2}$$

$$\sigma_m^2 = \sum_{i=1}^{N} \left(\frac{\partial m}{\partial y_i}\right)^2 \sigma_{y_i}^2 = \sigma_y^2 \sum_{i=1}^{N} \left(\frac{\partial m}{\partial y_i}\right)^2$$

$$= \sigma_y^2 \left[N\sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j\right)^2\right]^{-2} \sum_{i=1}^{N} \left(N x_i - \sum_{j=1}^{N} x_j\right)^2$$

Least Square Fitting

Note that

$$\sum_{i=1}^{N}\left(Nx_i - \sum_{j=1}^{N} x_j\right)^2$$

$$= \sum_{i=1}^{N}\left[N^2 x_i^2 - 2Nx_i \sum_{j=1}^{N} x_j + \left(\sum_{j=1}^{N} x_j\right)^2\right]$$

$$= N^2 \sum_{i=1}^{N} x_i^2 - 2N\left(\sum_{j=1}^{N} x_j\right)^2 + N\left(\sum_{j=1}^{N} x_j\right)^2$$

$$= N\left[N\sum_{i=1}^{N} x_i^2 - \left(\sum_{j=1}^{N} x_j\right)^2\right]$$

$$\sigma_m^2 = \frac{N\sigma_y^2}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{j=1}^{N} x_j\right)^2}$$

Least Square Fitting

Similarly

$$b = \frac{\sum_{j=1}^{N} x_j^2 \sum_{j=1}^{N} y_j - \sum_{j=1}^{N} x_j \sum_{j=1}^{N} x_j y_j}{N \sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j \right)^2}$$

$$\frac{\partial b}{\partial y_i} = \frac{\sum_{j=1}^{N} x_j^2 - x_i \sum_{j=1}^{N} x_j}{N \sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j \right)^2}$$

$$\sigma_b^2 = \sum_{i=1}^{N} \left(\frac{\partial b}{\partial y_i} \right)^2 \sigma_{y_i}^2 = \sigma_y^2 \sum_{i=1}^{N} \left(\frac{\partial b}{\partial y_i} \right)^2$$

$$= \sigma_y^2 \left[N \sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j \right)^2 \right]^{-2} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} x_j^2 - x_i \sum_{j=1}^{N} x_j \right)^2$$

Least Square Fitting

Note that

$$\sum_{i=1}^{N}\left(\sum_{j=1}^{N}x_j^2 - x_i\sum_{j=1}^{N}x_j\right)^2$$

$$= \sum_{i=1}^{N}\left[\left(\sum_{j=1}^{N}x_j^2\right)^2 - 2x_i\left(\sum_{j=1}^{N}x_j^2\right)\left(\sum_{j=1}^{N}x_j\right) + x_i^2\left(\sum_{j=1}^{N}x_j\right)^2\right]$$

$$= N\left(\sum_{j=1}^{N}x_j^2\right)^2 - 2\left(\sum_{j=1}^{N}x_j^2\right)\left(\sum_{j=1}^{N}x_j\right)^2 + \sum_{i=1}^{N}x_i^2\left(\sum_{j=1}^{N}x_j\right)^2$$

$$= N\left(\sum_{j=1}^{N}x_j^2\right)^2 - \left(\sum_{j=1}^{N}x_j^2\right)\left(\sum_{j=1}^{N}x_j\right)^2$$

$$= \left(\sum_{j=1}^{N}x_j^2\right)\left[N\left(\sum_{j=1}^{N}x_j^2\right) - \left(\sum_{j=1}^{N}x_j\right)^2\right]$$

$$\sigma_b^2 = \frac{\sigma_y^2\sum_{i=1}^{N}x_i^2}{N\sum_{i=1}^{N}x_i^2 - \left(\sum_{j=1}^{N}x_j\right)^2}$$

Summary

1. Obtained data set $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$.

2. Compute the m and b corresponding to the best fitted line $y = mx + b$,

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

$$b = \frac{\sum_{j=1}^{N} x_j^2 \sum_{j=1}^{N} y_j - \sum_{j=1}^{N} x_j \sum_{j=1}^{N} x_j y_j}{N \sum_{j=1}^{N} x_j^2 - \left(\sum_{j=1}^{N} x_j \right)^2}$$

3. Calculate the standard deviation in m and b,

$$\sigma_m = \sigma_y \sqrt{\frac{N}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{j=1}^{N} x_j\right)^2}}$$

$$\sigma_b = \sigma_y \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{j=1}^{N} x_j\right)^2}}$$