Topic 4:

Least Squares Curve Fitting

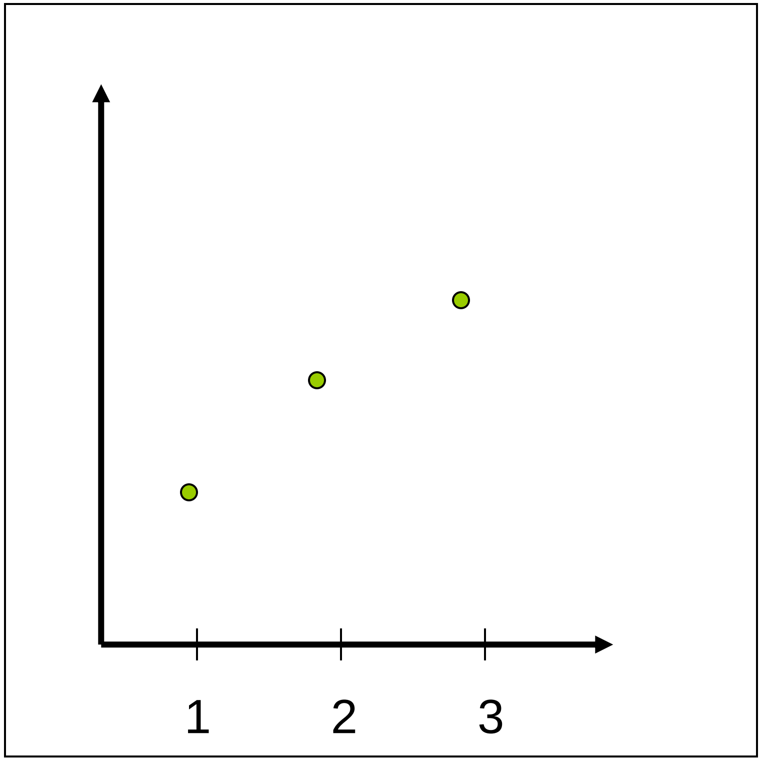
Lectures 18-19:

Lecture 18
Introduction to Least Squares

Motivation

Given a set of experimental data:

x	1	2	3
y	5.1	5.9	6.3



- The relationship between ***x*** and ***y*** may not be clear.
- We want to find an expression for ***f(x)***.

Motivation - Model Building

- In engineering, two types of applications are encountered:
  - ***Trend analysis:*** Predicting values of dependent variable, may include extrapolation beyond data points or interpolation between data points.
  - ***Hypothesis testing:*** Comparing existing mathematical model with measured data.

1. What is the best mathematical model (function ***f*** , ***y***) that represents the dataset $\boldsymbol{y}_i$?
2. What is the best criterion to assess the fitting of the function ***y*** to the data?

Motivation - Curve Fitting

Given a set of tabulated data, find a curve or a function that *best represents the data*.

Given:

1. The tabulated **data**
2. The **form** of the function
3. The curve fitting **criteria**

Find the **unknown** coefficients

Least Squares Regression

Linear Regression

- Fitting a straight line to a set of paired observations:

  $(x_1, y_1), (x_2, y_2),\ldots,(x_n, y_n)$.

  $y=a_0+a_1x+e$

  a_1-slope.

  a_0-intercept.

  e-error, or residual, between the model and the observations.

Selection of the Functions

$$Linear \qquad f(x) = a + bx$$

$$Quadratic \qquad f(x) = a + bx + cx^2$$

$$Polynomial \qquad f(x) = \sum_{k=0}^{n} a_k \ x^k$$

$$General \qquad f(x) = \sum_{k=0}^{m} a_k \ g_k(x)$$

$g_k(x)$ are known.

Decide on the Criterion

$$1.\ \text{Least Squares}:$$

$$\min_{a,b} \sum_{i=0}^{n} \left| f(x_i) - f_i \right|^2$$

Chapter 17

$$2.\ \text{Exact Matching (Interpolation)}:$$

$$f_i = f(x_i)$$

Chapter 18

Least Squares

Given:

x_i	x_1	x_2	....	x_n
y_i	y_1	y_2	....	y_n

The form of the function is assumed to be known but the coefficients are unknown.

$$y_i = f(x_i) + e_i$$

The difference is assumed to be the result of experimental error.

Determine the Unknowns

We want to find a and b to minimize:

$$\Phi(a,b) = \sum_{i=0}^{n} \left| f_i - (a + bx_i) \right|^2$$

How do we obtain a and b to minimize: $\Phi(a,b)$?

Determine the Unknowns

Necessary condition for the minimum:

$$\boxed{\begin{aligned} \frac{\partial\, \Phi(a,b)}{\partial\, a} &= 0 \\ \frac{\partial\, \Phi(a,b)}{\partial\, b} &= 0 \end{aligned}}$$

Example 1

x	1	2	3
y	5.1	5.9	6.3

Assume :

$f(x) = a + bx$

Necessary condition for the minimum :

$$\frac{\partial\ \Phi(a,b)}{\partial\ a} = 0$$

$$\frac{\partial\ \Phi(a,b)}{\partial\ b} = 0$$

Remember

$$\frac{d}{dx}\left(\sum_{k=1}^{n} a_i \;\; x\right) = \sum_{k=1}^{n} a_i$$

$$\frac{\partial}{\partial a}\left(\sum_{k=1}^{n} g_i(x) \;\; a\right) = \sum_{k=1}^{n} g_i(x)$$

Example 1

$$\frac{\partial\ \Phi(a,b)}{\partial\ a} = 0 = \sum_{k=1}^{N} 2\left(a + b x_k - y_k\right)$$

$$\frac{\partial\ \Phi(a,b)}{\partial\ b} = 0 = \sum_{k=1}^{N} 2\left(a + b x_k - y_k\right) x_k$$

$$\text{Normal Equations:}$$

$$N\,a + \left(\sum_{k=1}^{N} x_k\right) b = \left(\sum_{k=1}^{N} y_k\right)$$

$$\left(\sum_{k=1}^{N} x_k\right) a + \left(\sum_{k=1}^{N} x_k^2\right) b = \left(\sum_{k=1}^{N} x_k\ y_k\right)$$

Example 1

Solving the Normal Equations gives:

$$b = \frac{N\left(\sum_{k=1}^{N} x_k\, y_k\right) - \left(\sum_{k=1}^{N} x_k\right)\left(\sum_{k=1}^{N} y_k\right)}{N\left(\sum_{k=1}^{N} x_k^2\right) - \left(\sum_{k=1}^{N} x_k\right)^2}$$

$$a = \frac{1}{N}\left(\left(\sum_{k=1}^{N} y_k\right) - b\left(\sum_{k=1}^{N} x_k\right)\right)$$

Example 1

i	1	2	3	sum
x_i	1	2	3	6
y_i	5.1	5.9	6.3	17.3
x_i^2	1	4	9	14
$x_i\ y_i$	5.1	11.8	18.9	35.8

Example 1

Normal Equations:

$$N\,a + \left(\sum_{k=1}^{N} x_k\right) b = \left(\sum_{k=1}^{N} y_k\right)$$

$$\left(\sum_{k=1}^{N} x_k\right) a + \left(\sum_{k=1}^{N} x_k^2\right) b = \left(\sum_{k=1}^{N} x_k\, y_k\right)$$

$$3a + \;6b = 17.3$$

$$6a + 14b = 35.8$$

$$Solving \Rightarrow \quad a = 4.5667 \qquad b = 0.60$$

Example 2
- Fitting with Nonlinear Functions -

x	0.24	0.65	0.95	1.24	1.73	2.01	2.23	2.52
y	0.23	-0.23	-1.1	-0.45	0.27	0.1	-0.29	0.24

It is required to find a function of the form:

$$f(x) = a \ln(x) + b \cos(x) + c\, e^{x}$$

to fit the data.

Example 2

Necessary condition for the minimum:

$$\left.\begin{aligned} \frac{\partial\, \Phi(a,b,c)}{\partial\, a} &= 0 \\ \frac{\partial\, \Phi(a,b,c)}{\partial\, b} &= 0 \\ \frac{\partial\, \Phi(a,b,c)}{\partial\, c} &= 0 \end{aligned}\right\} \Rightarrow \textit{Normal Equations}$$

Example 2

$$a\sum_{k=1}^{8}(\ln x_k)^2 + b\sum_{k=1}^{8}(\ln x_k)(\cos x_k) + c\sum_{k=1}^{8}(\ln x_k)(e^{x_k}) = \sum_{k=1}^{8} y_k(\ln x_k)$$

$$a\sum_{k=1}^{8}(\ln x_k)(\cos x_k) + b\sum_{k=1}^{8}(\cos x_k)^2 + c\sum_{k=1}^{8}(\cos x_k)(e^{x_k}) = \sum_{k=1}^{8} y_k(\cos x_k)$$

$$a\sum_{k=1}^{8}(\ln x_k)(e^{x_k}) + b\sum_{k=1}^{8}(\cos x_k)(e^{x_k}) + c\sum_{k=1}^{8}(e^{x_k})^2 = \sum_{k=1}^{8} y_k(e^{x_k})$$

Evaluate the sums and solve the normal equations.

How Do You Judge Performance?

Given two or more functions to fit the data,
How do you select the best?

Answer:
Determine the parameters for each function, then compute Φ for each one. The function resulting in smaller Φ is the best in the least square sense.

Multiple Regression

Example:

Given the following data:

t	0	1	2	3
x	0.1	0.4	0.2	0.2
f(x,t)	3	2	1	2

It is required to determine a function of two variables:

$$f(x,t) = a + b\,x + c\,t$$

to explain the data that is best in the least square sense.

Solution of Multiple Regression

Construct Φ , the sum of the square of the error and derive the necessary conditions by equating the partial derivatives with respect to the unknown parameters to zero, then solve the equations.

t	0	1	2	3
x	0.1	0.4	0.2	0.2
f(x,t)	3	2	1	2

Solution of Multiple Regression

$$f(x,t) = a + bx + ct$$

$$\Phi(a,b,c) = \sum_{i=1}^{4} \left(a + bx_i + ct_i - f_i\right)^2$$

Necessary conditions:

$$\frac{\partial \Phi(a,b,c)}{\partial a} = 2\sum_{i=1}^{4} \left(a + bx_i + ct_i - f_i\right) \quad = 0$$

$$\frac{\partial \Phi(a,b,c)}{\partial b} = 2\sum_{i=1}^{4} \left(a + bx_i + ct_i - f_i\right)x_i = 0$$

$$\frac{\partial \Phi(a,b,c)}{\partial c} = 2\sum_{i=1}^{4} \left(a + bx_i + ct_i - f_i\right)t_i = 0$$

Lecture 19
Nonlinear Least Squares Problems + More

- Examples of Nonlinear Least Squares
- Solution of Inconsistent Equations
- Continuous Least Square Problems

Nonlinear Problem

Given:

x	1	2	3
y	2.4	5	9

Find a function of the form ae^{bx} that best fits the data.

$$\Phi = \sum_{i=1}^{3} \left(ae^{bx_i} - y_i \right)^2$$

Normal Equations are obtained using:

$$\frac{\partial \Phi}{\partial a} = 0 = \sum_{i=1}^{3} \left(ae^{bx_i} - y_i \right) e^{bx_i}$$

$$\frac{\partial \Phi}{\partial b} = 0 = \sum_{i=1}^{3} \left(ae^{bx_i} - y_i \right) a\, x_i\, e^{bx_i}$$

Alternative Solution
(Linearization Method)

Given:

x	1	2	3
y	2.4	5	9

Find a function of the form ae^{bx} that best fits the data.

$$\text{Define: } z = \ln(y) = \ln(a) + bx$$

$$Let: \alpha = \ln(a) \ and \ z_i = \ln(y_i)$$

$$\text{Instead of using: } \Phi = \sum_{i=1}^{3}\left(ae^{bx_i} - y_i\right)^2$$

$$\text{We will use: } \Phi = \sum_{i=1}^{3}\left(\alpha + bx_i - z_i\right)^2 \quad \text{(Easier to solve)}$$

Example
(Linearization Method)

Given:

x	1	2	3
y	0.23	.2	.14

Find a function of the form $1/(ax+b)$ that best fits the data.

$$\mathrm{f(x)} = \frac{1}{\mathrm{ax+b}}$$

$$\text{Define: } \mathrm{z} = \frac{1}{y} = \mathrm{ax+b}$$

$$Let:\ z_i = \frac{1}{y_i}$$

$$\text{Instead of using: } \Phi = \sum_{i=1}^{3}\left(\frac{1}{\mathrm{ax+b}} - y_i\right)^2$$

$$\text{We will use: } \qquad \Phi = \sum_{i=1}^{3}\left(ax_i + b - z_i\right)^2$$

Inconsistent System of Equations

Problem:

Solve the following system of equations:

$$\begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 4.1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 10 \end{bmatrix}$$

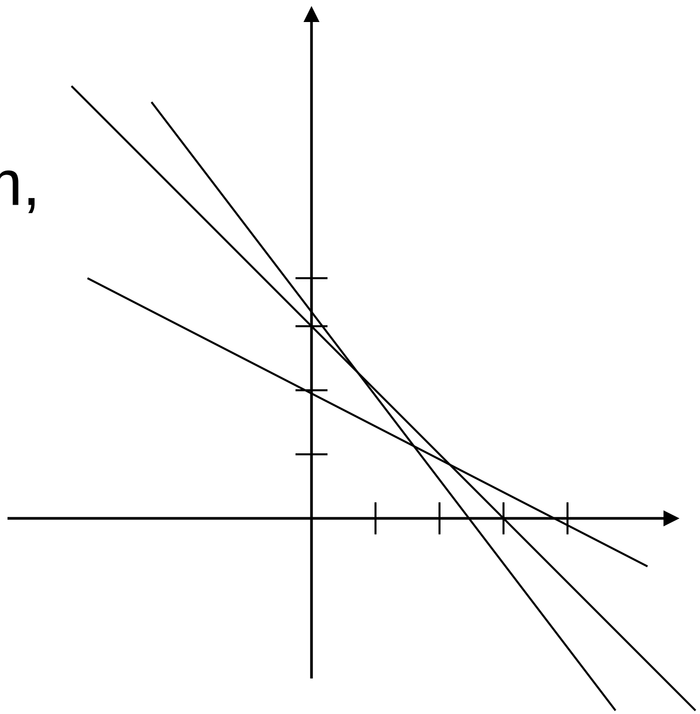
This is an inconsistent system of Equations.

No solution exists.

Inconsistent System of Equations

- Reasons -

- Inconsistent equations may occur because of:
  - Errors in formulating the problem,
  - Errors in collecting the data, or
  - Computational errors.



Solution exists if all lines intersect at one point.

Inconsistent System of Equations

- Formulation as a Least Squares Problem -

We can view the equations as:

$$\begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 4.1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 10 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

Find x_1 and x_2 to minimize the least squares error.

Solution

$$\Phi = (x_1 + 2x_2 - 4)^2 + (2x_1 + 2x_2 - 6)^2 + (3x_1 + 4.1x_2 - 10)^2$$

Find x_1 and x_2 to minimize Φ.

$$\frac{\partial \Phi}{\partial x_1} = 0 = 2(x_1 + 2x_2 - 4) + 4(2x_1 + 2x_2 - 6) + 6(3x_1 + 4.1x_2 - 10)$$

$$0 = (2 + 8 + 18)x_1 + (4 + 8 + 24.6)x_2 + (-8 - 24 - 60)$$

$$28x_1 + 36.6x_2 = 92$$

$$\frac{\partial \Phi}{\partial x_2} = 0 = 4(x_1 + 2x_2 - 4) + 4(2x_1 + 2x_2 - 6) + 8.2(3x_1 + 4.1x_2 - 10)$$

$$0 = (4 + 8 + 24.6)x_1 + (8 + 8 + 33.62)x_2 + (-16 - 24 - 82)$$

$$36.6x_1 + 49.62x_2 = 122$$

Solution

$$\text{Normal equations}:$$

$$28x_1 + 36.6x_2 = 92$$

$$36.6x_1 + 49.62x_2 = 122$$

$$\textit{Solution}:$$

$$\Rightarrow x_1 = 2.0048, x_2 = 0.9799$$