CISE301: Numerical Methods

Topic 4:

Least Squares Curve Fitting

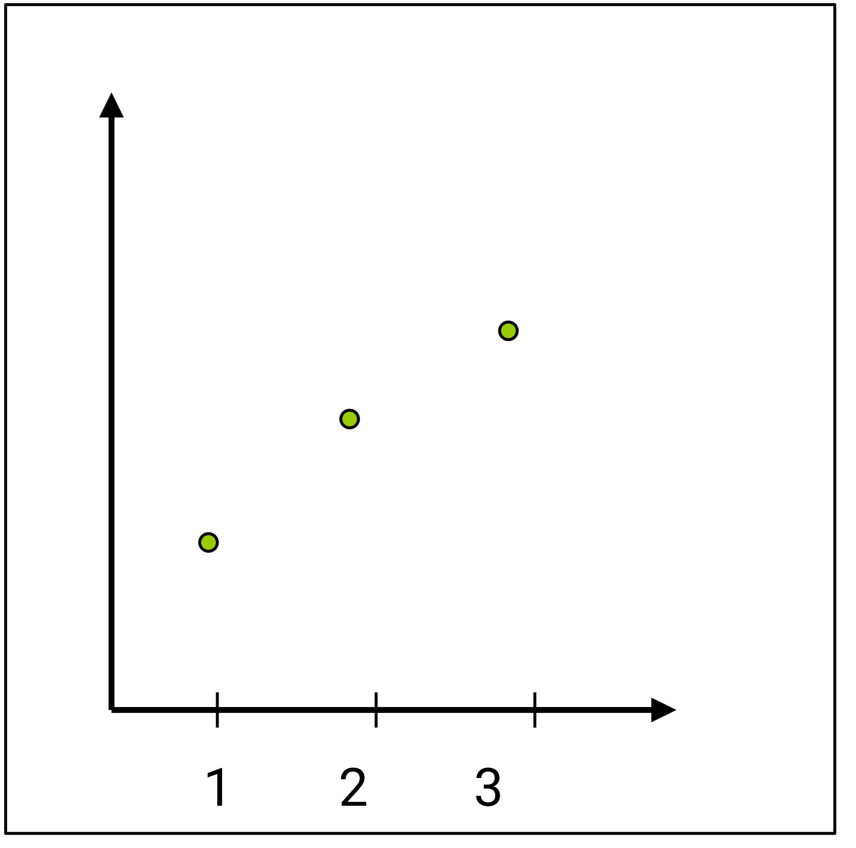
Lectures 18-19:

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.
- Find a function ***f(x)*** that best fit the data

Motivation

- 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 function f that represents the dataset?
2. What is the best criterion to assess the fitting of the function f to the data?

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	$f(x) = a + bx$
Quadratic	$f(x) = a + bx + cx^2$
Polynomial	$f(x) = \sum_{k=0}^{n} a_k \ x^k$
General	$f(x) = \sum_{k=0}^{m} a_k \ g_k(x)$

$g_k(x)$ are known.

Decide on the Criterion

1. Least Squares Regression:

$$\text{minimize} \quad \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Chapter 17

2. Exact Matching (Interpolation):

$$y_i = f(x_i)$$

Chapter 18

Least Squares Regression

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.

$$e_i^{\,2} = (y_i - f(x_i))^2 = (f(x_i) - y_i)^2$$

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

Determine the Unknowns

We want to find a and b to minimize :

$$\Phi(a, b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2$$

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

Determine the Unknowns

Necessary condition for the minimum :

$$\boxed{\begin{gathered}\frac{\partial\ \Phi(a,b)}{\partial\ a}=0\\[1em]\frac{\partial\ \Phi(a,b)}{\partial\ b}=0\end{gathered}}$$

Determining the Unknowns

$$\frac{\partial \Phi(a,b)}{\partial a} = \sum_{i=1}^{n} 2(a + bx_i - y_i) = 0$$

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

Normal Equations

$$n\,a + \left(\sum_{i=1}^{n} x_i\right) b = \left(\sum_{i=1}^{n} y_i\right)$$

$$\left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b = \left(\sum_{i=1}^{n} x_i\, y_i\right)$$

Solving the Normal Equations

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

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

Example 1: Linear Regression

Assume :

$f(x) = a + bx$

x	1	2	3
y	5.1	5.9	6.3

Equations :

$$n\,a + \left(\sum_{i=1}^{n} x_i\right) b = \left(\sum_{i=1}^{n} y_i\right)$$

$$\left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b = \left(\sum_{i=1}^{n} x_i\, y_i\right)$$

Example 1: Linear Regression

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

Equations :

$$3a + 6b = 17.3$$

$$6a + 14b = 35.8$$

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

Multiple Linear Regression

Example:

Given the following data:

t	0	1	2	3
x	0.1	0.4	0.2	0.2
y	3	2	1	2

Determine a function of two variables:

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

That best fits the data with the least sum of the square of errors.

Solution of Multiple Linear 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
y	3	2	1	2

Solution of Multiple Linear Regression

$$f(x,t) = a + bx + ct, \quad \Phi(a,b,c) = \sum_{i=1}^{n} (a + bx_i + ct_i - y_i)^2$$

Necessary conditions :

$$\frac{\partial \Phi(a,b,c)}{\partial a} = 2\sum_{i=1}^{n} (a + bx_i + ct_i - y_i) = 0$$

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

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

System of Equations

$$a\,n + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} (x_i)^2 + c\sum_{i=1}^{n} (x_i\ t_i) = \sum_{i=1}^{n} (x_i\ y_i)$$

$$a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} (x_i\ t_i) + c\sum_{i=1}^{n} (t_i)^2 = \sum_{i=1}^{n} (t_i\ y_i)$$

Example 2: Multiple Linear Regression

i	1	2	3	4	Sum
t_i	0	1	2	3	6
x_i	0.1	0.4	0.2	0.2	0.9
y_i	3	2	1	2	8
x_i^2	0.01	0.16	0.04	0.04	0.25
$x_i t_i$	0	0.4	0.4	0.6	1.4
$x_i y_i$	0.3	0.8	0.2	0.4	1.7
t_i^2	0	1	4	9	14
$t_i y_i$	0	2	2	6	10

Example 2: System of Equations

$$4a + 0.9b + 6c = 8$$

$$0.9a + 0.25b + 1.4c = 1.7$$

$$6a + 1.4b + 14c = 10$$

$$\textit{Solving} \quad :$$

$$a = 2.9574, \quad b = -1.7021, \quad c = -0.38298$$

$$f(x,t) = a + bx + ct = 2.9574 - 1.7021x - 0.38298t$$

Lecture 19
Nonlinear Least Squares Problems

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

Polynomial Regression

- The least squares method can be extended to fit the data to a higher-order polynomial

$$f(x) = a + bx + cx^2, \quad e_i^2 = (f(x) - y_i)^2$$

$$\text{Minimize} \quad \Phi(a, b, c) = \sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)^2$$

$$\text{Necessary conditions:}$$

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

Equations for Quadratic Regression

Minimize $\Phi(a, b, c) = \sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)^2$

$$\frac{\partial \Phi(a, b, c)}{\partial a} = 2\sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i) = 0$$

$$\frac{\partial \Phi(a, b, c)}{\partial b} = 2\sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)\, x_i = 0$$

$$\frac{\partial \Phi(a, b, c)}{\partial c} = 2\sum_{i=1}^{n} (a + bx_i + cx_i^2 - y_i)\, x_i^2 = 0$$

Normal Equations

$$a\,n + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i$$

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i$$

Example 3: Polynomial Regression

Fit a second-order polynomial to the following data

x_i	0	1	2	3	4	5	$\sum=15$
y_i	2.1	7.7	13.6	27.2	40.9	61.1	$\sum=152.6$
x_i^2	0	1	4	9	16	25	$\sum=55$
x_i^3	0	1	8	27	64	125	225
x_i^4	0	1	16	81	256	625	$\sum=979$
$x_i\, y_i$	0	7.7	27.2	81.6	163.6	305.5	$\sum=585.6$
$x_i^2\, y_i$	0	7.7	54.4	244.8	654.4	1527.5	$\sum=2488.8$

Example 3: Equations and Solution

$$6a + 15b + 55c = 152.6$$

$$15a + 55b + 225c = 585.6$$

$$55a + 225b + 979c = 2488.8$$

Solving . . .

$$a = 2.4786, \quad b = 2.3593, \quad c = 1.8607$$

$$f(x) = 2.4786 + 2.3593x + 1.8607x^2$$

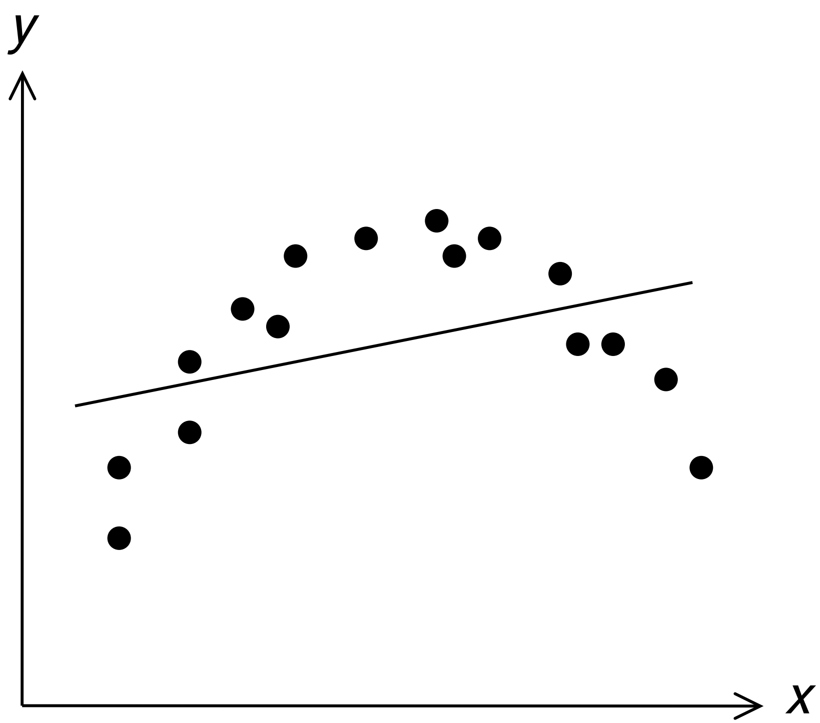
How Do You Judge Functions?

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

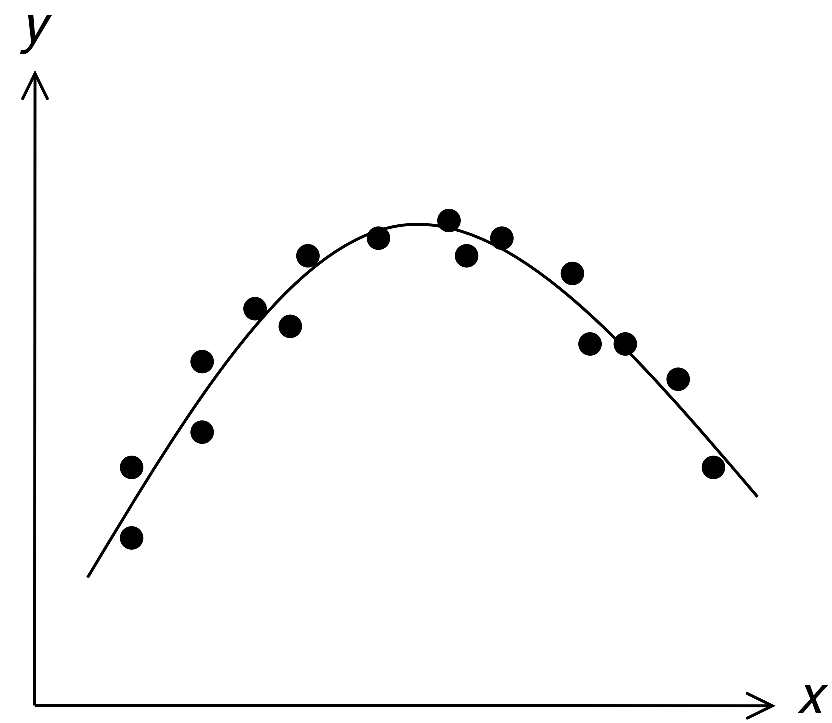
Answer :

Determine the parameters for each function, then compute Φ for each one. The function resulting in smaller Φ (least sum of the squares of the errors) is the best.

is preferable than Linear Regression



Linear Regression



Quadratic Regression

Fitting with Nonlinear Functions

x_i	0.24	0.65	0.95	1.24	1.73	2.01	2.23	2.52
y_i	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.

$$\Phi(a, b, c) = \sum_{i=1}^{n} \left(f(x_i) - y_i \right)^2$$

Fitting with Nonlinear Functions

$$\Phi(a,b,c) = \sum_{i=1}^{n} \left(a \ln(x_i) + b \cos(x_i) + c\, e^{x_i} - y_i \right)^2$$

Necessary condition for the minimum :

$$\left.\begin{array}{l} \dfrac{\partial \Phi(a,b,c)}{\partial a} = 0 \\[2ex] \dfrac{\partial \Phi(a,b,c)}{\partial b} = 0 \\[2ex] \dfrac{\partial \Phi(a,b,c)}{\partial c} = 0 \end{array}\right\} \Rightarrow \textit{Normal Equations}$$

Normal Equations

$$a\sum_{i=1}^{n}(\ln x_i)^2 + b\sum_{i=1}^{n}(\ln x_i)(\cos x_i) + c\sum_{i=1}^{n}(\ln x_i)(e^{x_i}) = \sum_{i=1}^{n} y_i(\ln x_i)$$

$$a\sum_{i=1}^{n}(\ln x_i)(\cos x_i) + b\sum_{i=1}^{n}(\cos x_i)^2 + c\sum_{i=1}^{n}(\cos x_i)(e^{x_i}) = \sum_{i=1}^{n} y_i(\cos x_i)$$

$$a\sum_{i=1}^{n}(\ln x_i)(e^{x_i}) + b\sum_{i=1}^{n}(\cos x_i)(e^{x_i}) + c\sum_{i=1}^{n}(e^{x_i})^2 = \sum_{i=1}^{n} y_i(e^{x_i})$$

Evaluate the sums and solve the normal equations.

Example 4: Evaluating Sums

xi	0.24	0.65	0.95	1.24	1.73	2.01	2.23	2.52	∑=11.57
yi	0.23	-0.23	-1.1	-0.45	0.27	0.1	-0.29	0.24	∑=-1.23
$(\ln xi)^2$	2.036	0.1856	0.0026	0.0463	0.3004	0.4874	0.6432	0.8543	∑=4.556
ln(xi) cos(xi)	-1.386	-0.3429	-0.0298	0.0699	-0.0869	-0.2969	-0.4912	-0.7514	∑=-3.316
$\ln(xi) * e^{xi}$	-1.814	-0.8252	-0.1326	0.7433	3.0918	5.2104	7.4585	11.487	∑=25.219
yi * ln(xi)	-0.328	0.0991	0.0564	-0.0968	0.1480	0.0698	-0.2326	0.2218	∑=-0.0625
$\cos(xi)^2$	0.943	0.6337	0.3384	0.1055	0.0251	0.1808	0.3751	0.6609	∑=3.26307
$\cos(xi) * e^{xi}$	1.235	1.5249	1.5041	1.1224	-0.8942	-3.1735	-5.696	-10.104	∑=-14.481
yi*cos(xi)	0.223	-0.1831	-0.6399	-0.1462	-0.0428	-0.0425	0.1776	-0.1951	∑=-0.8485
$(e^{xi})^2$	1.616	3.6693	6.6859	11.941	31.817	55.701	86.488	154.47	∑=352.39
$yi * e^{xi}$	0.2924	-0.4406	-2.844	-1.555	1.523	0.7463	-2.697	2.9829	∑=-1.9923

Example 4: Equations & Solution

$$4.55643\,a - 3.31547\,b + 25.2192\,c = -0.062486$$

$$-3.31547\,a + 3.26307\,b - 14.4815\,c = -0.848514$$

$$25.2192\,a - 14.4815\,b + 352.388\,c = -1.992283$$

Solving the above equations :

$$a = -0.88815, \; b = -1.1074, \; c = 0.012398$$

Therefore,

$$f(x) = -0.88815\ln(x) - 1.1074\cos(x) + 0.012398\,e^{x}$$

Example 5

Given:

x_i	1	2	3
y_i	2.4	5	9

Find a function $f(x) = ae^{bx}$ that best fits the data.

$$\Phi(a,b) = \sum_{i=1}^{n} (ae^{bx_i} - y_i)^2$$

Normal Equations are obtained using :

$$\frac{\partial \Phi}{\partial a} = 2\sum_{i=1}^{n} (ae^{bx_i} - y_i)e^{bx_i} = 0$$

Difficult to Solve

$$\frac{\partial \Phi}{\partial b} = 2\sum_{i=1}^{n} (ae^{bx_i} - y_i)a\,x_i\,e^{bx_i} = 0$$

Linearization Method

Find a function $f(x) = ae^{bx}$ that best fits the data.

Define $g(x) = \ln(f(x)) = \ln(a) + bx$

Define $z_i = \ln(y_i) = \ln(a) + bx_i$

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

Instead of minimizing : $\Phi(a, b) = \sum_{i=1}^{n} (ae^{bx_i} - y_i)^2$

Minimize : $\Phi(\alpha, b) = \sum_{i=1}^{n} (\alpha + bx_i - z_i)^2$ (Easier to solve)

Example 5: Equations

$$\Phi(\alpha, b) = \sum_{i=1}^{n} (\alpha + b x_i - z_i)^2$$

Normal Equation s are obtained using :

$$\frac{\partial \Phi}{\partial \alpha} = 2\sum_{i=1}^{n} (\alpha + b x_i - z_i) = 0$$

$$\frac{\partial \Phi}{\partial b} = 2\sum_{i=1}^{n} (\alpha + b x_i - z_i) x_i = 0$$

$$\alpha n + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} z_i \quad \text{and} \quad \alpha\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i z_i)$$

Evaluating Sums and Solving

x_i	1	2	3	$\sum=6$
y_i	2.4	5	9	
$z_i=\ln(y_i)$	0.875469	1.609438	2.197225	$\sum=4.68213$
x_i^2	1	4	9	$\sum=14$
$x_i\, z_i$	0.875469	3.218876	6.591674	$\sum=10.6860$

Equations :

$$3\alpha + 6b = 4.68213$$

$$6\alpha + 14b = 10.686$$

Solving Equations :

$$\alpha = 0.23897, \quad b = 0.66087$$

$$\alpha = \ln(a), \quad a = e^{\alpha}$$

$$a = e^{0.23897} = 1.26994$$

$$f(x) = ae^{bx} = 1.26994\, e^{0.66087x}$$