CURVE FITTING

Fitting Equations to Data

This module introduces Curve Fitting[1] and in particular the Least Squares method. Curve fitting, like Interpolation, is a collection of methods used to represent a set of data by an equation. Unlike Interpolation, curve fitting methods do not require the data points to be taken at equidistant values of the independent variable.

Learning Objectives

Upon completion of this module the student should :

- Understand the Least Squares Method
- Be able to curve fit data using several types of curves.

Topics

The following topics will be covered in this module:

- Linear Curve Fitting
- Polynomial Curve Fitting
- Exponential Curve Fitting
- Logarithmic Curve Fitting
- Power Function Curve Fitting
- Examples

[1] The basic material discussed in this section was first presented by Karl Friedrich Gauss to the Royal Society of Gottingen during the period 1821-1826, in a series of three papers. For example, the French translation by J. Bertrand (authorized by Gauss) and published in 1855.

Least-Squares Method:

The method of least-squares is based on the least square criterion that states: The sum of the squares of the differences between the *curve-fitted function values* and the given *tabular values* of the function must be minimum.

If we assume that the given tabular data belong to the function y=f(x) and that the curve fitting curve is Y=f(x), then the mathematical interpretation of the least squares criterion is:

$$Q = \sum_{i=1}^{n} (Y_i - y_i)^2 \quad \Rightarrow \quad \min imum$$

Linear Curve Fitting:

Linear curve fitting implies that a set of *n* data points { (x_0, y_0), (x_1, y_1), …(x_n, y_n)}, can be represented by a linear equation of the form:

$$Y_i = a_0 + a_1 x_i \qquad (1)$$

provided that Q, the least-squares criterion, conforms to the following requirement:

$$Q = \sum_{i=1}^{n} (Y_i - y_i)^2 \quad = \quad \min imum \qquad (2)$$

Therefore, in order to curve-fit a set of data with a linear curve we only need to find values for a_0 and a_1 and substitute them into eq. (1).

To do this we recall from calculus that the *maximum* or the *minimum* of a function can be determined by setting the first derivative of the function to zero and solving for the variable(s).

Therefore, values for a_0, and a_1 can be found by setting the derivative of eq. (2) equal to zero:

$$dQ = \sum_{i=1}^{n} (Y_i - y_i)^2 \quad = \quad 0$$

However, *Q* is a function of two independent variables, a_0 and a_1 . Therefore, we have to use partial differentiation. By substituting equation (1) into (2) equation (2) becomes

$$Q = \sum_{i=1}^{n} (a_0 + a_1 x_i - y_i)^2 \quad = \quad \min imum \qquad (3)$$

Therefore setting the derivative of (3) equal to zero, we get:

$$\frac{\partial Q}{\partial a_0} = \sum_{i=1}^{n} 2(a_0 + a_1 x_i - y_i) \quad = \quad 0 \qquad (4)$$

$$\frac{\partial Q}{\partial a_1} = \sum_{i=1}^{n} 2(a_0 + a_1 x_i - y_i) x_i = \quad 0 \qquad (5)$$

dividing both sides of (4) and (5) by two, and expanding, we form a system of two equations in two unknowns

$$na_0 + a_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \quad = \quad 0$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \quad = \quad 0$$

or

$$na_0 + a_1 \sum_{i=1}^{n} x_i \quad = \quad \sum_{i=1}^{n} y_i$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 \quad = \quad \sum_{i=1}^{n} x_i y_i$$

Solve the above system we get values for a_0 and a_1 . Substituting a_0 and a_1 in (1) with the found values for a_0 and a_1 we obtain the function for fitting the data with a straight line (linear curve fitting).

Example:

Curve-fit the data with a straight line: (0,2), (1,3), (2,5), (3,5), (4,9), (5,8), (6,10).

n	x_i	y_i	$x_i y_i$	x_i^2	**Y_i**	Y_I-y_i	$(Y_I-y_I)^2$
1	0	2	0	0	**1.9287**	-0.0714	
2	1	3	3	1	**3.2857**	0.2857	
3	2	5	10	4	**4.6428**	-0.3571	
4	3	5	15	9	**6.0000**	1.0000	
5	4	9	36	16	**7.3571**	-1.6428	
6	5	8	40	25	**8.7143**	0.71428	
7	6	10	60	36	**10.0714**	0.0714	
Σ	21	42	164	91		0.00008	

Substituting the values from the table into the system below

$$
\begin{aligned}
na_0 + a_1\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} y_i \\
a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i
\end{aligned}
$$

we get

$$
\begin{aligned}
7a_0 + 21a_1 &= 42 \\
21a_0 + 91a_1 &= 164
\end{aligned}
$$

Solving the system[2] we get

$$
\begin{aligned}
a_0 &= 1.9286 \\
a_1 &= 1.3571
\end{aligned}
$$

Therefore, the equation to curve-fit the data with a straight line is:

$$\mathbf{Y}_i = 1.9286 + 1.3571x_i$$

[2] Using Cramer's Rule

Curve Fitting with an n-degree polynomial:

For an n-degree polynomial fit the data can be fitted by

$$Y_i = a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_m x_i^m$$

Applying the least square criterion:

$$\frac{\partial Q}{\partial a_0} = \sum_{i=1}^{n} 2\,(a_0 + a_1 x_i \quad a_2 x_i^2 + \dots + a_m x_i^m - y_i) \quad = \quad 0$$

$$\frac{\partial Q}{\partial a_1} = \sum_{i=1}^{n} 2\,(a_0 + a_1 x_i \quad a_2 x_i^2 + \dots + a_m x_i^m - y_i)\, x_i \quad = \quad 0$$

$$\vdots$$

$$\frac{\partial Q}{\partial a_m} = \sum_{i=1}^{n} 2\,(a_0 + a_1 x_i \quad a_2 x_i^2 + \dots + a_m x_i^m - y_i)\, x_i^m \quad = \quad 0$$

we obtain the following system of m equations in m unknowns:

$$na_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{1}^{n} x_2^2 + \dots + a_m \sum_{1}^{n} x_i^m = \sum_{i=1}^{n} y_i$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{1}^{n} x_i^2 + a_2 \sum_{1}^{n} x_i^3 + \dots + a_m \sum_{1}^{n} x_i^{m+1} = \sum_{i=1}^{n} x_i y_i$$

$$\dots$$

$$a_0 \sum_{i=1}^{n} x_i^m + a_1 \sum_{1}^{n} x_i^{m+1} + a_2 \sum_{1}^{n} x_i^{m+2} + \dots + a_m \sum_{1}^{n} x_i^{m+m} = \sum_{i=1}^{n} x_{i_i}^m y_i$$

CURVE FITTING

Example

Curve fit the data below using a parabola:

x :	1	3	4	5	6	7	8	9	10
y:	2	7	8	10	11	11	10	9	8

n	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4	$\mathbf{Y}_i$	$\mathbf{Y}_i - y_i$	$(\mathbf{Y}_I - y_I)^2$
1	1	2	2	1	2	1	1			
2	3	7	21	9	63	23	81			
3	4	8	32	16	128	64	256			
4	5	10	50	25	250	125	625			
5	6	11	66	36	396	216	1296			
6	7	11	77	49	539	343	2401			
7	8	10	80	64	640	512	4096			
8	9	9	81	81	729	729	6561			
9	10	8	80	100	800	1000	10000			
Σ	53	76	489	381	3547	3013	25317			

$$na_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i-1}^{n} x_i^2 = \sum_{i=1}^{n} y_i$$

$$a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i-1}^{n} x_i^3 = \sum_{i=1}^{n} xy_i$$

$$a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i-1}^{n} x_i^4 = \sum_{i=1}^{n} x^2 y_i$$

$$\begin{aligned} 9a_0 + 53a_1 + 381a_2 &= 76 \\ 53a_0 + 381a_1 + 3017a_2 &= 489 \\ 381a_0 + 3017a_1 + 25317a_2 &= 3547 \end{aligned}$$

Solve the system to find a_0 , a_1 , and a_2.

Thus the above data can be curve-fitted with the following equation

$$Y_i = -1.4597 + 3.6053x_i - 0.2676x_i^2.$$

CURVE FITTING

Power Function $y = cx^n$

To curve-fit a set of data using a power function we transform $y = cx^n$ to a polynomial. We can accomplish the transformation by taking the log of both side of the equation $y = cx^n$.

$$\log y = \log c + n \log x$$

Letting $\log y = Y$, $\log c = m$, and $\log x = X$ we can re-write the above equation in a linear form as:

$$Y = m + n X$$

To find the values for m and n we solve the system:

$$k\,m + n\sum_{i=1}^{k} X_i = \sum_{i=1}^{k} Y_i$$

$$\sum_{i=1}^{k} X_i \; m + n\sum_{i=1}^{k} X_i^2 = \sum_{i=1}^{k} X_i Y_i$$

Since m=log c, then $c=10^m$. By substituting n and c into the power equation we can use it to curve fit the given data.

Example

Curve-fit the data using a power function

x :	2	4	5	6	8
y:	0.7500	0.1875	0.1200	0.0833	0.0469

k	x_i	y_i	**X** (log x)	**Y** (log y)	**X^2**	**XY_i**
1	2	0.7500	0.3010	-0.1249	0.0906	-0.0376
2	4	0.1875	0.6021	-0.7270	0.3625	-0.4377
3	5	0.1200	0.6988	-0.9208	0.4886	-0.6436
4	6	0.0833	0.7782	-1.0793	0.6055	-0.8399
5	8	0.0469	0.9031	-1.3288	0.8156	-1.2001
Σ			3.2832	-4.1808	2.3628	-3.1589

$$Y = m + n X$$

CURVE FITTING

Since $Y = m + n X$ is a linear function, we form a system of two equations in two unknowns:

$$5 m + 3.2833 n = -4.1809$$
$$3.2833 m + 2.3628 n = -3.1589$$

Solve for m and n:

$$m = 0.4769$$
$$n = -1.9997 \cong -2.0$$

Since $m = \log c$:

$$0.4769 = \log c \quad \text{which implies}$$
$$c = 10^{0.4769} \quad \text{or}$$
$$c = 2.9998 \quad \text{or}$$
$$c \cong 3.0$$

$\therefore$ To curve-fit the given data we use the power function $\mathbf{y = 3\, x^{-2.0}}$

Exponential Function $y = ae^{bx}$

To curve-fit a set of data using an exponential function we use the same procedure as before. That is, we transform $y = ae^{bx}$ into a polynomial. We can accomplish the transformation by taking the log of both side of $y = ae^{bx}$.

$$\ln y = \ln a + b\,x$$

Letting $\ln y = Y$, and $\ln a = c$ we can re-write the above equation in a linear form as:

$$Y = c + b\,x$$

From the system

$$\begin{aligned} k\,c \quad + \quad b\sum_{i=1}^{k} x_i \quad &= \sum_{i=1}^{k} Y_i \\ \sum_{i=1}^{k} x_i\,c \quad + \quad b\sum_{i=1}^{k} x_i^2 \quad &= \sum_{i=1}^{k} x_i Y_i \end{aligned}$$

We find b and c. Since c=ln a, then $e^c=a$. By substituting a and b into the exponential function we can use it to curve fit the given data.

Example

Curve-fit the data using a power function

x :	-4	-2	0	1	2	4
y:	0.57	1.32	4.12	6.65	11.0	30.3

k	x_i	y_i	**Y** (ln y)	x_i^2	$x_i Y_i$
1	-4	0.57	-0.5621	16	2.2485
2	-2	1.32	0.2776	4	-0.5553
3	0	4.12	1.4159	0	0.0000
4	1	6.65	1.8946	1	1.8946
5	2	11.0	2.3980	4	4.7958
6	4	30.3	3.4112	16	13.6446
Σ			8.8352	41	22.0282

$$Y = c + b\,x$$

Since $Y = c + b\,x$ is a linear function, we form a system of two equations in two unknowns:

$$6\,c + 1.0\,b = 8.8350$$
$$1.0\,c + 41\,b = 22.0282$$

Solve for b and c:

$$b = 0.5034$$
$$c = 1.3886$$

Since $c = \ln a$

$1.3886 = \ln a$ which implies

$a = e^{1.3886}$ or

$a = 4.01$

$\therefore$ To curve-fit the given data we use the power function $\mathbf{y = 4.01\, e^{0.5034\, x}}$

CURVE FITTING