University of Basra

College of Engineering

Chemical Engineering Department

Engineering Statistics

The second stage

Dr. Mohammad N. Fares

Curve Fitting and Solution of Equation

FITTING OF STRAIGHT LINE

Let a straight line $y = a + bx$ ...(1)

which is fitted to the given data points (x_1, y_1), (x_2, y_2), (x_3, y_3),......, (x_n, y_n).

Let y_{λ_1} be the theoretical value for x_1 then $e_1 = y_1 - y_{\lambda_1}$

$\Rightarrow \quad e_1 = y_1 - (a + bx_1) \quad \Rightarrow \quad e_1^2 = (y_1 - a - bx_1)^2$

Now we have $S = e_1^2 + e_2^2 + e_3^2 + \ldots\ldots\ldots\ldots + e_n^2$

$$S = \sum_{i=1}^{n} e_i^2 \quad \Rightarrow \quad S = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

By the principle of least squares, the value of S is minimum therefore,

$$\frac{\partial S}{\partial a} = 0 \quad \ldots(2) \qquad \frac{\partial S}{\partial b} = 0 \quad \ldots(3)$$

On solving equations (2) and (3), and dropping the suffix, we have

$$\sum y = na + b\sum x \quad \ldots(4)$$

$$\sum xy = a\sum x + b\sum x^2 \quad \ldots(5)$$

The equation (3) and (4) are known as normal equations.

On solving equations (3) and (4), we get the value of a and b. Putting the value of a and b in equation (1), we get the equation of the line of best fit.

FITTING OF PARABOLA

Let a parabola $y = a + bx + cx^2$ ...(1)

which is fitted to a given data (x_1, y_1), (x_2, y_2), (x_3, y_3) ,......, (x_n, y_n) .

Let y_λ be the theoretical value for x_1 then $e_1 = y_1 - y_\lambda$

$$\Rightarrow \quad e_1 = y_1 - (a + bx_1 + cx_1^2)$$

$$\Rightarrow \quad e_1^2 = (y_1 - a - bx_1 - cx_1^2)^2$$

Now we have
$$S = \sum_{i=1}^{n} e_i^2$$

$$S = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)^2$$

By the principle of least squares, the value of S is minimum, therefore

$$\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0 \text{ and } \frac{\partial S}{\partial c} = 0 \qquad ...(2)$$

Solving equation (2) and dropping suffix, we have

$$\sum y = na + b\sum x + c\sum x^2 \qquad ...(3)$$

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \qquad ...(4)$$

$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \qquad ...(5)$$

The equation (3), (4) and (5) are known as normal equations.

On solving equations (3), (4) and (5), we get the values of a,b and c. Putting the values of a, b and c in equation (1), we get the equation of the parabola of best fit.

Example:

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

Solution

$$y = a_0 + a_1 x$$

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

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

n	x_i	y_i	$x_i y_i$	x_i^2	Y_i	$Y_i - y_i$
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

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

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

we get

$$7a_0 + 21a_1 = 42$$
$$21a_0 + 91a_1 = 164$$

Solving the system[2] we get

$$a_0 = 1.9286$$
$$a_1 = 1.3571$$

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

$$Y_i = 1.9286 + 1.3571x_i$$

Example : Curve fit the data below using a parabola:

$$y = a_0 + a_1 x + a_2 x^2$$

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

Solution

$$y = a_0 + a_1 x + a_2 x^2$$

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

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

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

n	x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4
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

$$9a_0 + 53a_1 + 381a_2 = 76$$
$$53a_0 + 381a_1 + 3017a_2 = 489$$
$$381a_0 + 3017a_1 + 25317a_2 = 3547$$

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

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

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

Example 1: Find the best-fit values of a and b so that $y = a + bx$ fits the data given in the table.

x :	0	1	2	3	4
y :	1	1.8	3.3	4.5	6.3

Sol. Let the straight line is $y = a + bx$ ...(1)

x	y	xy	x^2
0	1	0	0
1	1.8	1.8	1
2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16
$\sum x = 10$	$\sum y = 16.9$	$\sum xy = 47.1$	$\sum x^2 = 30$

Normal equations are, $\sum y = na + b\sum x$ ...(2)

$$\sum xy = a\sum x + b\sum x^2 \quad ...(3)$$

Here $n = 5$, $\sum x = 10$, $\sum y = 16.9$, $\sum xy = 47.1$, $\sum x^2 = 30$

Putting these values in normal equations, we get

$$16.9 = 5a + 10b$$

$$47.1 = 10a + 30b$$

On solving these two equations, we get

$$a = 0.72\ ,\quad b = 1.33.$$

Example 2: Fit a straight line to the given data regarding x as the independent variable.

x	1	2	3	4	5	6
y	1200	900	600	200	110	50

Sol. Let the straight line obtained from the given data by $y = a + bx$ ...(1)

Then the normal equations are $\sum y = na + b\sum x$ ...(2)

$\sum xy = a\sum x + b\sum x^2$ ...(3)

x	y	x^2	xy
1	1200	1	1200
2	900	4	1800
3	600	9	1800
4	200	16	800
5	110	25	550
6	50	36	300
$\sum x = 21$	$\sum y = 3060$	$\sum x^2 = 91$	$\sum xy = 6450$

Putting all values in the equations (2) and (3), we get

$$3060 = 6a + 21b$$

$$6450 = 21a + 91b$$

Solving these equations, we get

$$a = 1361.97 \quad \text{and} \quad b = -243.42$$

Hence the fitted equation is $y = 1361.97 - 243.42x$. **Ans.**

Example 4: Find the least square polynomial approximation of degree two to the data.

x	0	1	2	3	4
y	-4	-1	4	11	20

Sol. Let the equation of the polynomial be $y = a + bx + cx^2$ ...(1)

x	y	xy	x^2	x^2y	x^3	x^4
0	-4	0	0	0	0	0
1	-1	-1	1	-1	1	1
2	4	8	4	16	8	16
3	11	33	9	99	27	81
4	20	80	16	320	64	256
$\sum x = 10$	$\sum y = 30$	$\sum xy = 120$	$\sum x^2 = 30$	$\sum x^2y = 434$	$\sum x^3 = 100$	$\sum x^4 = 354$

The normal equations are, $\sum y = na + b\sum x + c\sum x^2$ ...(2)

$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3 \quad ...(3)$$

$$\sum x^2y = a\sum x^2 + b\sum x^3 + c\sum x^4 \quad ...(4)$$

Here $n = 5, \sum x = 10, \sum y = 30, \sum xy = 120, \sum x^2 = 30, \sum x^2y = 434, \sum x^3 = 100,$ $\sum x^4 = 354.$

Putting all these values in (2), (3) and (4), we get $30 = 5a + 10b + 30c$ ...(5)

$$120 = 10a + 30b + 100c \quad ...(6)$$

$$434 = 30a + 100b + 354c \quad ...(7)$$

On solving these equations, we get $a = -4$, $b = 2$, $c = 1$. Therefore required polynomial is

$y = -4 + 2x + x^2$ **Ans.**

FITTING OF AN EXPONENTIAL CURVE

Suppose an exponential curve of the form

$$y = ae^{bx}$$

Taking logarithm on both the sides, we get

$$\log_{10} y = \log_{10} a + bx \log_{10} e$$

i.e., $$Y = A + Bx$$

where $$Y = \log_{10} y,\ A = \log_{10} a \text{ and } B = b \log_{10} e.$$

The normal equations for (1) are,

$$\sum Y = nA + B\sum x$$

$$\sum xY = A\sum x + B\sum x^2$$

On solving the above two equations, we get A and B

then $$a = \text{antilog } A,\ b = \frac{B}{\log_{10} e}$$

FITTING OF THE CURVE $y = ax + bx^2$

Error of estimate for ith point (x_i, y_i) is $\quad e_i = (y_i - ax_i - bx_i^2)$

We have,
$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - ax_i - bx_i^2)^2$$

By the principle of least square, the value of S is minimum

$$\therefore \quad \frac{\partial S}{\partial a} = 0 \text{ and } \frac{\partial S}{\partial b} = 0$$

Now
$$\frac{\partial S}{\partial a} = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} 2(y_i - ax_i - bx_i^2)(-x_i) = 0$$

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

and
$$\frac{\partial S}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - ax_i - bx_i^2)(-x_i^2) = 0 \qquad \text{...(1)}$$

or
$$\sum_{i=1}^{n} x_i^2 y_i = a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^4 \qquad \text{...(2)}$$

Dropping the suffix i from (1) and (2), then the normal equations are,

$$\sum xy = a\sum x^2 + b\sum x^3$$
$$\sum x^2 y = a\sum x^3 + b\sum x^4$$

FITTING OF THE CURVE $y = ax + \frac{b}{x}$

Error of estimate for ith point (x_i, y_i) is

$$e_i = (y_i - ax_i - \frac{b}{x_i})$$

We have, $S = \sum_{i=1}^{n} e_i^2 \quad = \sum_{i=1}^{n} (y_i - ax_i - \frac{b}{x_i})^2$

By the principle of least square, the value of S is minimum

$$\therefore \qquad \frac{\partial S}{\partial a} = 0 \text{ and } \frac{\partial S}{\partial b} = 0$$

Now $\quad \frac{\partial S}{\partial a} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} 2(y_i - ax_i - \frac{b}{x_i})(-x_i) = 0$

or $\quad \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + nb \qquad \text{...(1)}$

and $\quad \frac{\partial S}{\partial b} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} 2(y_i - ax_i - \frac{b}{x_i})(-\frac{1}{x_i}) = 0$

or $\quad \sum_{i=1}^{n} \frac{y_i}{x_i} = na + b\sum_{i=1}^{n} \frac{1}{x_i^2} \qquad \text{...(2)}$

Dropping the suffix i from (1) and (2), then the normal equations are,

$$\sum xy = nb + a\sum x^2$$

$$\sum \frac{y}{x} = na + b\sum \frac{1}{x^2}$$

where n is the number of pair of values of x and y.

FITTING OF THE CURVE $y = \frac{c_0}{x} + c_1\sqrt{x}$

Error of estimate for ith point (x_i, y_i) is

$$e_i = (y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i})$$

We have, $S = \sum_{i=1}^{n} e_i^2 \quad = \sum_{i=1}^{n} (y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i})^2$

By the principle of least square, the value of S is minimum

$$\therefore \qquad \frac{\partial S}{\partial c_0} = 0 \text{ and } \frac{\partial S}{\partial c_1} = 0$$

Now $\frac{\partial S}{\partial c_0} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} 2(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i})(-\frac{1}{x_i}) = 0$

or

$$\sum_{i=1}^{n} \frac{y_i}{x_i} = c_0 \sum_{i=1}^{n} \frac{1}{x_i^2} + c_1 \sum_{i=1}^{n} \frac{1}{\sqrt{x_i}} \qquad \ldots(1)$$

and $\frac{\partial S}{\partial c_1} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} 2(y_i - \frac{c_0}{x_i} - c_1\sqrt{x_i})(-\sqrt{x_i}) = 0$

or

$$\sum_{i=1}^{n} y_i\sqrt{x_i} = c_0 \sum_{i=1}^{n} \frac{1}{\sqrt{x_i}} + c_1 \sum_{i=1}^{n} x_i \qquad \ldots(2)$$

Dropping the suffix i from (1) and (2), then the normal equations are,

$$\sum \frac{y}{x} = c_0 \sum \frac{1}{x^2} + c_1 \sum \frac{1}{\sqrt{x}}$$

$$\sum y\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x.$$

PROBLEM SET 5.1

1. Fit a straight line to the given data regarding x as the independent variable:

x	1	2	3	4	6	8
y	2.4	3.1	3.5	4.2	5.0	6.0

[**Ans.** $y = 2.0253 + 0.502x$]

2. Fit a straight line $y = a + bx$ to the following data by the method of least square:

x	0	1	3	6	8
y	1	3	2	5	4

[**Ans.** $1.6 + 0.38x$]

3. Find the least square approximation of the form $y = a + bx^2$ for the following data:

x	0	0.1	0.2	0.3	0.4	0.5
y	1	1.01	0.99	0.85	0.81	0.75

[**Ans.** $y = 1.0032 - 1.1081x^2$]

4. Fit a second degree parabola to the following data:

x	0.0	1.0	2.0
y	1.0	6.0	17.0

[**Ans.** $y = 1 + 2x + 3x^2$]

5. Fit a second degree parabola to the following data:

x	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y	1.1	1.3	1.6	2.0	2.7	3.4	4.1

[**Ans.** $y = 1.04 - 0.193x + 0.243x^2$]

6. Fit a second degree parabola to the following data by the least square method:

x	1	2	3	4	5
y	1090	1220	1390	1625	1915

[**Ans.** $y = 27.5x^2 + 40.5x + 1024$]

7. Fit a parabola $y = a + bx + cx^2$ to the following data:

x	2	4	6	8	10
y	3.07	12.85	31.47	57.38	91.29

[**Ans.** $y = 0.34 - 0.78x + 0.99x^2$]

8. Determine the constants a and b by the method of least squares such that $y = ae^{bx}$ fits the following data:

x	2	4	6	8	10
y	4.077	11.084	30.128	81.897	222.62

[**Ans.** $y = 1.49989e^{0.50001x}$]

9. Fit a least square geometric curve $y = ax^b$ to the following data:

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

[**Ans.** $y = 0.5012x^{1.9977}$]

10. A person runs the same race track for five consecutive days and is timed as follows:

Day(x)	1	2	3	4	5
Time(y)	15.3	15.1	15	14.5	14

Make a least square fit to the above data using a function $a + \frac{b}{x} + \frac{c}{x^2}$.

$$\left[\textbf{Ans.}\ y = 13.0065 + \frac{6.7512}{x} + \frac{4.4738}{x^2}\right]$$

11. Use the method of least squares to fit the curve $y = \frac{c_0}{x} + c_1\sqrt{x}$ to the following table of values:

x	0.1	0.2	0.4	0.5	1	2
y	21	11	7	6	5	6

$\left[\textbf{Ans. } y = \frac{1.97327}{x} + 3.28182\sqrt{x}\right]$

12. Using the method of least square to fit a parabola $y = a + bx + cx^2$ in the following data:

$(x, y): (-1, 2), (0, 0), (0, 1), (1, 2)$ $\left[\textbf{Ans. } y = \frac{1}{2} + \frac{3}{2}x^2\right]$

13. The pressure of the gas corresponding to various volumes V is measured, given by the following data:

$V(\text{cm}^3)$	50	60	70	90	100
$p(\text{kgcm}^{-2})$	64.7	51.3	40.5	25.9	78

Fit the data to the equation $pV^{\gamma} = c$. [**Ans.** $pV^{0.28997} = 167.78765$]

14. Employ the method of least squares to fit a parabola $y = a + bx + cx^2$ in the following data: (x, y): $(-1, 2)$, $(0, 0)$, $(0, 1)$, $(1, 2)$ [**Ans.** $y = 0.5 + 1.5x^2$]

15. Fit a second degree parabola in the following data: [*U.T.U. 2008*]

x	0.0	1.0	2.0	3.0	4.0
y	1.0	4.0	10.0	17.0	30.0

[**Ans.** $y = 1 + 2x + 3x^2$]

16. Fit at least square quadratic curve to the following data:

x	1	2	3	4
y	1.7	1.8	2.3	3.2

, estimate $y(2.4)$

[**Ans.** $y = 2 - 0.5x + 0.2x^2$ and $y(2.4) = 1.952$]

17. Fit an exponential curve by least squares

x	1	2	5	10	20	30	40	50
y	98.2	91.7	81.3	64.0	36.4	32.6	17.1	11.3

Estimate y when $x = 25$. [**Ans.** $y = 100(0.96)^x$, $y(25) = 33.9$]

18. Fit the curve $y = a + \frac{b}{x}$ to the following data

x	1	2	3	4
y	3	1.5	6	7.5

Estimate y when $x = 2.25$. $\left[\textbf{Ans. } y = 1.3 + \frac{1.7}{x}, y(2.25) = 4.625\right]$

Q1/The relationship between the heat capacity of oil and temperature is given by the following formula:

$$Cp = A + BT + CT^2$$

Using curve fitting equations, find the constants (**a**, **b** and **c**) using data below and calculate amount of heat required to raise **1** kg of oil from **25°C** to **55°C**.

Temperature (C°)	**20**	**40**	**60**
Heat capacity (kJ/kg.°C)	**1.42**	**4.32**	**8.82**

x	y	x2	x3	x4	yx	yx2		
20	1.42E+00	400	8000	160000	28.4	568		
40	4.32E+00	1600	64000	2560000	172.8	6912		
60	8.82E+00	3600	216000	12960000	529.2	31752		
120	14.56	5600	288000	15680000	730.4	39232		
14.56		3		120		5600		
								40
730.4		120		5600		288000		
								46.66667
39232		5600		288000		15680000		
582.4		120		4800		224000		
148		0		800		64000		
34085.33		5600		261333.3		13440000		
								33.33333
5146.667		0		26666.67		2240000		
4933.333				26666.67		2133333		
213.3333				0		106666.7		
c=	0.002							
b=	0.025							
a=	0.12			134.1				
20		28.4						

The relationship between the Nusselt number and the Reynolds number is given by the following formula:

$$Nu = a.Re^{n}$$

Using curve fitting equations, find the constants (**a** and **b**), by using data below,

Nu	27.3	29.1	30.7	33.4
Re	1800	1920	2108	2400