Subject Name: Numerical analysis

3rd Class, Second Semester

Academic Year: 2024-2025

Lecturer: Dr. Amir N.Saud

Email: amir-najah@uomus.edu.iq

Lecture No. 4

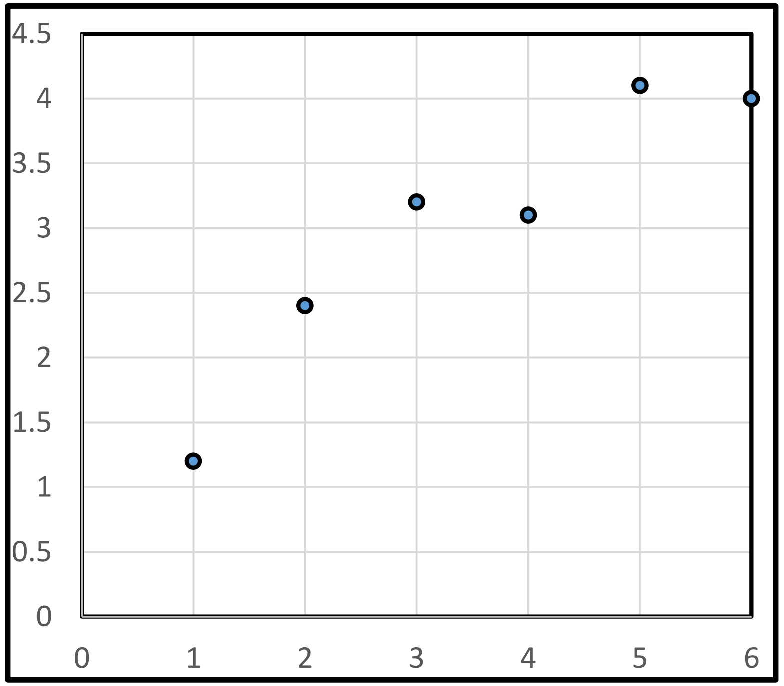
Lecture Title: Least square fitting: the best-fitted line; parabolic least square fitting

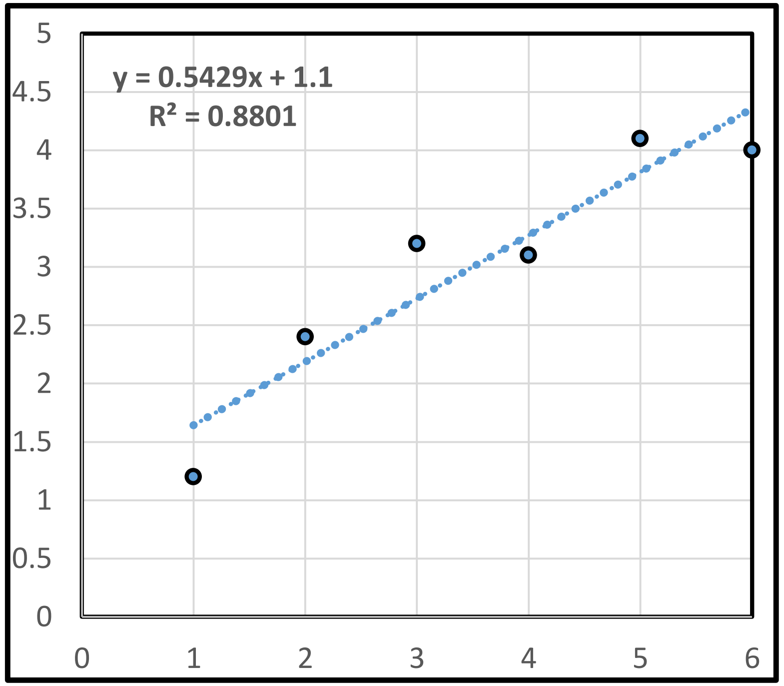
ROOT FINDING

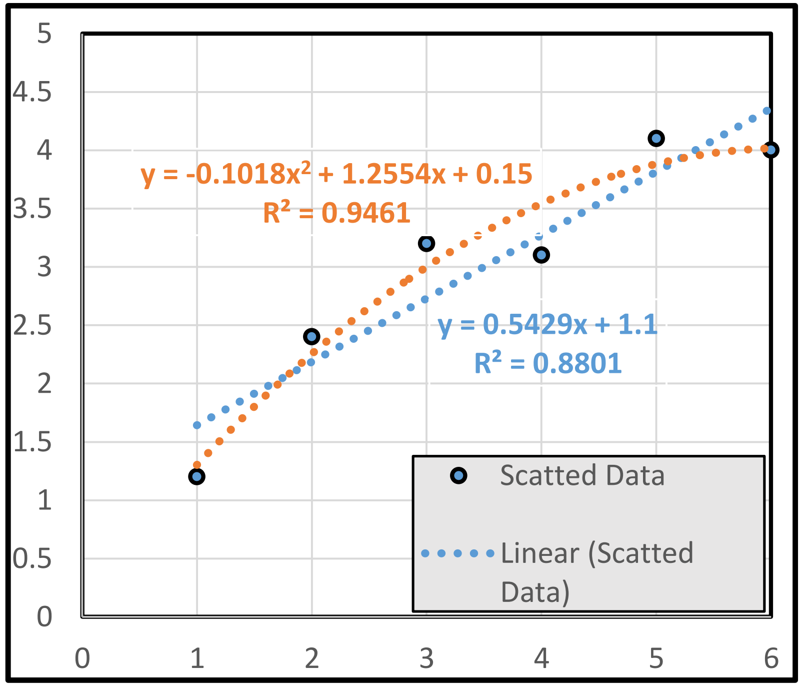
CURVE FITTING

Curve fitting: is a process of finding a curve (or mathematical function) that best represents a set of data points. This is especially useful when the relationship between variables is not perfectly linear or when there are uncertainties or errors in the data. By using curve fitting, we can create a model that helps predict or describe the behavior of the data.

In engineering, data is often collected from experiments or field measurements, and curve fitting is used to approximate the relationship between the variables. For example, if you measure the deflection of a beam under varying loads, curve fitting can be used to model the relationship between load and deflection.







Mathematically, the problem of curve fitting or function approximation may be stated as follows:

To find a functional relationship y = g(x), that relates the set of observed data values P_i (x_i , y_i), i = 1, 2,..., n as closely as possible, so that the graph of y = g(x) goes near the data points P_i's though not necessarily through all of them.

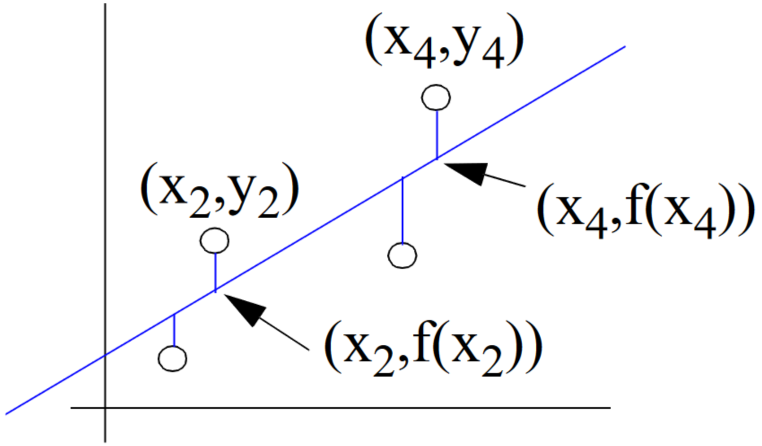
Thus g (x) may be any of the following:

$$g(x) = \alpha + \beta x$$ — **Linear Interpolation Function**

$$g(x) = \alpha + \beta x + \gamma x^2$$ — **2nd Order Interpolation Function**

$$g(x) = \alpha e^{\pm \beta x}$$ — **Exponent Interpolation Function**

$$g(x) = \alpha \log(\beta x)$$ — **Log Interpolation Function**



Here α, β *and* γ are parameters which are to be evaluated so that the curve y = g(x), fits the data well. A measure of how well the curve fits is called the goodness of fit.

Method of Least Squares

Let (x_1, f_1), (x_2, f_2),..., (x_n, f_n) be a set of observed values and g(x) be the approximating function. We form the sums of the squares of the deviations of the observed values f_i from the estimated values $g(x_i)$,

$$S = \sum_{i=1}^{n} \{f_i - g(x_i)\}^2$$

The function g(x) may have parameters, α, β *and* γ, to determine these parameters we have to form the necessary conditions for S to be **minimum**, which are:

$$\frac{\partial S}{\partial \alpha} = 0, \frac{\partial S}{\partial \beta} = 0, \frac{\partial S}{\partial \gamma} = 0,$$

1. Curve Fitting by a Straight Line:

- Let $g(x) = \alpha + \beta x$ be the straight line which fits a set of observed data points (x_i, y_i), i = 1, 2, ..., n.
- Let S be the sum of the squares of the deviations $g(x_i)-y_i$ for i=1,2,3,..,n, given by:

$$S = \sum_{i=1}^{n} \{\alpha + \beta x_i - y_i\}^2$$

We now employ the method of least squares to determine α and β so that S will be minimum. The normal equations are,

$$\frac{\partial S}{\partial \alpha} = 0 \quad then \quad \sum_{i=1}^{n} (\alpha + \beta x_i - y_i) = 0$$

$$\frac{\partial S}{\partial \beta} = 0 \quad then \quad \sum_{i=1}^{n} x_i(\alpha + \beta x_i - y_i) = 0$$

$$\frac{\partial S}{\partial \alpha} = 0 \quad then \quad \sum_{i=1}^{n} (\alpha + \beta x_i - y_i) = 0$$

$$\frac{\partial S}{\partial \beta} = 0 \quad then \quad \sum_{i=1}^{n} x_i(\alpha + \beta x_i - y_i) = 0$$

$$n\alpha + \sum_{i=1}^{n} x_i \beta - \sum_{i=1}^{n} y_i = 0$$

$$\sum_{i=1}^{n} x_i \alpha + \sum_{i=1}^{n} x_i^2 \beta - \sum_{i=1}^{n} x_i y_i = 0$$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$\boldsymbol{\alpha} = \frac{(\sum \boldsymbol{y})(\sum \boldsymbol{x}^2) - (\sum \boldsymbol{x})(\sum \boldsymbol{xy})}{\boldsymbol{n} \sum \boldsymbol{x}^2 - (\sum \boldsymbol{x})^2}$$

$$\boldsymbol{\beta} = \frac{\boldsymbol{n}(\sum \boldsymbol{xy}) - (\sum \boldsymbol{x})(\sum \boldsymbol{y})}{\boldsymbol{n} \sum \boldsymbol{x}^2 - (\sum \boldsymbol{x})^2}$$

Example – The below table is a scattered data, use the curve fitting method to find the best linear behavior for it:

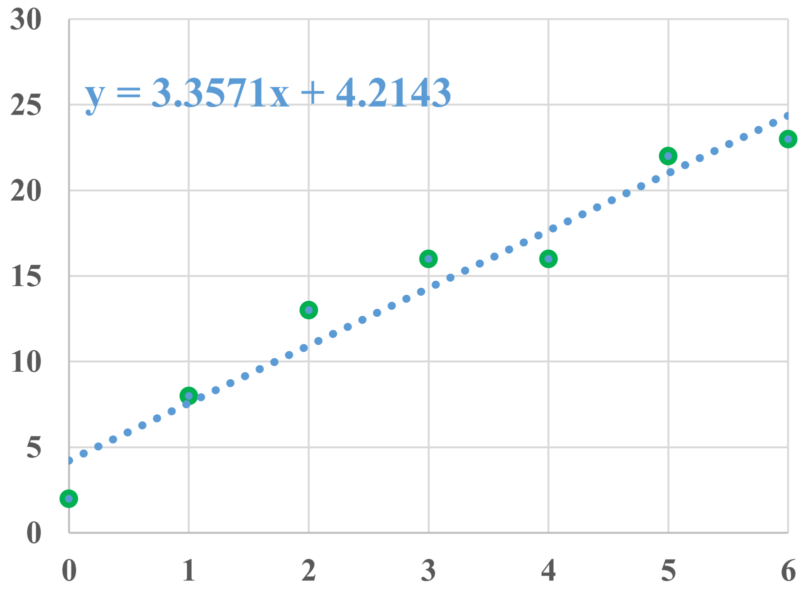
x	0	1	2	3	4	5	6
y	2	8	13	16	16	22	23

	x	y	x^2	xy	
	0	2	0	0	
	1	8	1	8	
	2	13	4	26	
	3	16	9	48	
	4	16	16	64	
	5	22	25	110	
	6	23	36	138	
Sum	**21**	**100**	**91**	**394**	
		alpha	**4.21429**		826
					196
		beta	**3.35714**		658
					196

x	Y_table	Y_fit	Error	E_Abs
0	2	4.21429	111%	111%
1	8	7.57143	-5%	5%
2	13	11.78571	-9%	9%
3	16	16.00000	0%	0%
4	16	20.21429	26%	26%
5	22	24.42857	11%	11%
6	23	28.64286	25%	25%
		average%	**23%**	**27%**

Then Find the Overall Average Error percentage?

Y-fit(1) = 3.3571 (0)+4.2143= 4.2143



$$Percentage\ Error = \frac{|Predicated\ - True|}{|True|} \times 100$$

Example – Find the straight line fitting the following data:

x	4	6	8	10	12
y	13.72	12.9	12.01	11.14	10.31

x	y	x^2	xy
4	13.72	16	54.88
6	12.9	36	77.4
8	12.01	64	96.08
10	11.14	100	111.4
12	10.31	144	123.72
40	60.08	360	463.48
	a	**15.44800**	3089.6
			200
	b	**-0.42900**	-85.8
			200

x	y	y_pred	Error%
4	13.72	13.732	0.09%
6	12.9	12.874	0.20%
8	12.01	12.016	0.05%
10	11.14	11.158	0.16%
12	10.31	10.3	0.10%
			0.12%

Linear Correlation Coefficient

In curve fitting and regression analysis, the linear correlation coefficient is crucial for determining how well a linear model fits the data. A higher correlation coefficient indicates that the linear model is a good fit for the data, whereas a low correlation coefficient suggests that a linear model may not be appropriate.
The linear correlation coefficient R is calculated using the following formula:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{\left(n \sum x^2 - \left(\sum x\right)^2\right)\left(n \sum y^2 - \left(\sum y\right)^2\right)}}$$

Interpreting the Correlation Coefficient

- $0.7 \leq r \leq 1$: Strong positive correlation — a linear relationship where y increases as x increases.
- $0.3 \leq r < 0.7$: Moderate positive correlation — a positive trend exists, but the data points are more spread out.
- $0 \leq r < 0.3$: Weak positive correlation — a weak relationship where the data points do not form a clear linear pattern.
- $-0.3 \leq r < 0$: Weak negative correlation — a weak negative relationship where y decreases slightly as x increases.
- $-0.7 \leq r < -0.3$: Moderate negative correlation — a clear, but not perfect, negative relationship.
- $-1 \leq r < -0.7$: Strong negative correlation — a strong linear relationship where y decreases as x increases.

EX. 2: Specify the correlation type and the degree for the data in example 1.

Solution:

x	y	x^2	xy	y^2
50	37	2500	1850	1369
100	48	10000	4800	2304
150	60	22500	9000	3600
200	71	40000	14200	5041
250	80	62500	20000	6400
300	90	90000	27000	8100
350	102	1222500	35700	10404
400	109	160000	43600	11881
Σx =1800	**Σy = 597**	**$\Sigma x^2 = 510000$**	**Σxy = 156150**	**$\Sigma y^2 = 49099$**

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2)\,(n\sum y^2 - (\sum y)^2)}} = \frac{8(156150) - (1800)(597)}{\sqrt{[8(510000)-(1800)^2][8(49099)-(597)^2]}} = 0.999$$

$\because$ $r \approx 1 \longrightarrow$ Perfect positive linear correlation.

2. Curve Fitting by a Quadratic (A Parabola):

Let $g(x) = a + bx + cx^2$, be the approximating quadratic to fit a set of data (x_i, y_i), i = 1, 2, ..., n. Here the parameters are to be determined by the method of least squares, i.e., by minimizing the sum of the squares of the deviations given by,

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

$$\frac{\partial S}{\partial a} = 0 \quad then \quad \sum_{i=1}^{n} \left(a + bx_i + cx_i^2 - y_i\right) = 0 \qquad na + \sum_{i=1}^{n} bx_i + \sum_{i=1}^{n} cx_i^2 - \sum_{i=1}^{n} y_i = 0$$

$$\frac{\partial S}{\partial b} = 0 \quad then \quad \sum_{i=1}^{n} x_i\left(a + bx_i + cx_i^2 - y_i\right) = 0 \qquad \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} bx_i^2 + \sum_{i=1}^{n} cx_i^3 - \sum_{i=1}^{n} x_i y_i = 0$$

$$\frac{\partial S}{\partial c} = 0 \quad then \quad \sum_{i=1}^{n} x_i^2\left(a + bx_i + cx_i^2 - y_i\right) = 0 \qquad \sum_{i=1}^{n} ax_i^2 + \sum_{i=1}^{n} bx_i^3 + \sum_{i=1}^{n} cx_i^4 - \sum_{i=1}^{n} x_i^2 y_i = 0$$

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

AX= B
using built in Mathcad matrix inversion, the coefficients >>
X = A-1*B

and are solved

Example – Find the second order polynomial fitting for the following data:

	1	2	3	4	5	6
x	0	0.5	1	1.5	2	2.5
y	0	0.25	1	2.25	4	6.25

$$y = a + bx + cx^2 \rightarrow\rightarrow \boldsymbol{y = x^2}$$

x	y	y_Pred	Error%
0	0	0	
0.5	0.25	0.25	0
1	1	1	0
1.5	2.25	2.25	0
2	4	4	0
2.5	6.25	6.25	0
			0%

x	y	x^2	x^3	x^4	xy	x^2y
0	0	0	0	0	0	0
0.5	0.25	0.25	0.125	0.0625	0.125	0.0625
1	1	1	1	1	1	1
1.5	2.25	2.25	3.375	5.0625	3.375	5.0625
2	4	4	8	16	8	16
2.5	6.25	6.25	15.625	39.0625	15.625	39.0625
7.5	13.75	13.75	28.125	61.1875	28.125	61.1875

6	7.5	13.75
7.5	13.75	28.125
13.75	28.125	61.1875

0.82143	-1.1786	0.35714
-1.1786	2.90714	-1.0714
0.35714	-1.0714	0.42857

13.75
28.125
61.1875

2.24E-13
0
1

https://youtu.be/QtI9ofU6zWg

Example – Find the second order polynomial fitting for the following data:

	1	2	3	4	5	6
x	0	0.5	1	1.5	2	2.5
y	0.0674	-0.9156	1.6253	3.0377	3.3535	7.9409

$$y = -0.18139 - 0.32209x + 1.35369x^2$$

x	y	x^2	x^3	x^4	xy	x^2y
0	0.0674	0	0	0	0	0
0.5	-0.9156	0.25	0.125	0.0625	-0.4578	-0.2289
1	1.6253	1	1	1	1.6253	1.6253
1.5	3.0377	2.25	3.375	5.0625	4.55655	6.834825
2	3.3535	4	8	16	6.707	13.414
2.5	7.9409	6.25	15.625	39.0625	19.85225	49.63063
7.5	15.1092	13.75	28.125	61.1875	32.2833	71.27585

x	y	y_Pred	Error%
0	0.0674	-0.18139	369%
0.5	-0.9156	-0.00401	100%
1	1.6253	0.850211	48%
1.5	3.0377	2.381274	22%
2	3.3535	4.58918	37%
2.5	7.9409	7.473929	6%

6	7.5	13.75
7.5	13.75	28.125
13.75	28.125	61.1875

0.821429	-1.17857	0.357143
-1.17857	2.907143	-1.07143
0.357143	-1.07143	0.428571

15.1092
32.2833
71.27585

-0.18139
-0.32209
1.353686