BEKG 2452
NUMERICAL METHODS
Curve Fitting

Ser Lee Loh[a], Wei Sen Loi[a]

[a]Fakulti Kejuruteraan Elektrik
Universiti Teknikal Malaysia Melaka

Lesson Outcome

Upon completion of this lesson, the student should be able to:

1. Determine polynomial regression function from a given set of paired observations.
2. Estimate the value of a polynomial regression function for any intermediate value of the independent variable.

Curve Fitting

A method to construct a curve that best fits to a set of data points.

Linear Regression

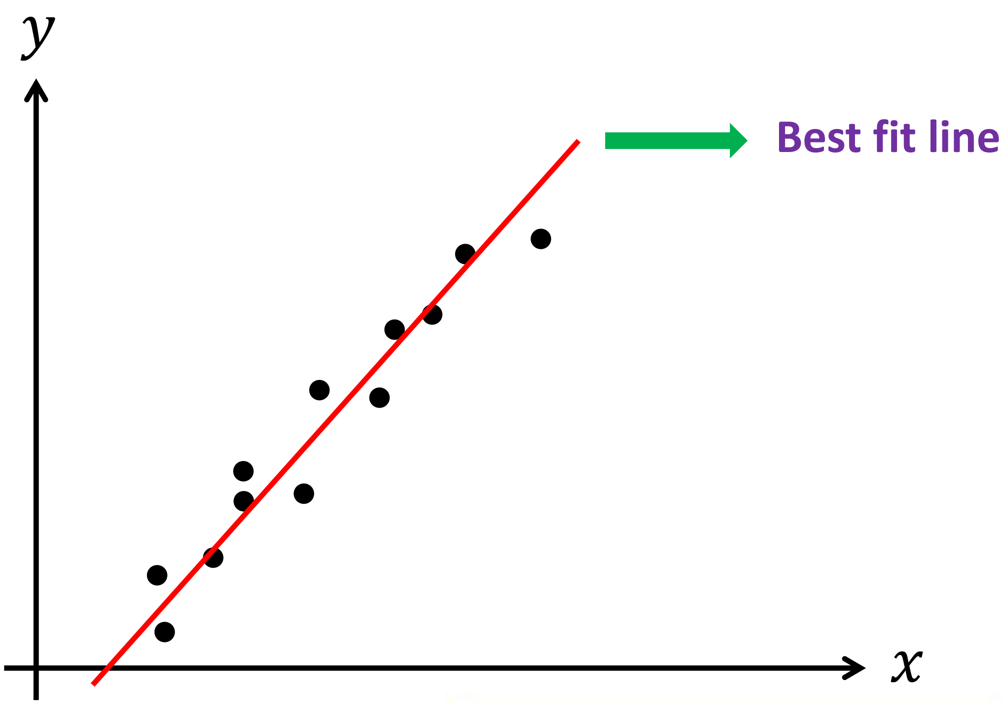
Polynomial Regression

Multiple Linear Regression

- Find the **polynomial function** that best fits a given set of points
- Estimate intermediate value

4.3 Extrapolation by Curve Fitting
4.3.1 Linear Regression

- Fitting a straight line to a set of given observation points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.



Equation of a straight line:

$$\boxed{y = a + bx}$$

4.3.1.1 Illustration of Linear Regression

Example 4.7:

Find the linear regression line that best fits the following data.

x_i	2	3	4	6
y_i	8	10	11	12

4.3.1.1 Illustration of Linear Regression

Solution:

Step 1: Find the values of $\sum x$, $\sum x^2$, $\sum y$ and $\sum xy$

x_i	y_i	x_i^2	$x_i y_i$
2	8	4	16
3	10	9	30
4	11	16	44
6	12	36	72
$\sum x_i = 15$	$\sum y_i = 41$	$\sum x_i^2 = 65$	$\sum x_i y_i = 162$

Note: These values may be obtained from calculator

4.3.1.1 Illustration of Linear Regression

Step 2: Form the linear system as follows:

$$\sum y = an + b\sum x$$

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

where n is the number of points and hence,

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

4.3.1.1 Illustration of Linear Regression

With $\sum x_i = 15, \sum y_i = 41, \sum x_i^2 = 65$ and $\sum x_i y_i = 162$,

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{4(162) - (15)(41)}{4(65) - (15)^2}$$

$$= 0.9429$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{15}{4} = 3.75; \quad \bar{y} = \frac{\sum y_i}{n} = \frac{41}{4} = 10.25,$$

$$a = \bar{y} - b\bar{x} = 10.25 - 0.9429(3.75) = 6.7141$$

Linear regression line:

$$y = a + bx \quad \Rightarrow \quad y = 6.7141 + 0.9429x$$

4.3.1.1 Illustration of Linear Regression

Alternative Step 2: Form the linear system and solve it by Gauss Elimination, Gauss Seidel or LU decomposition.

Linear system formed:

$$41 = 4a + 15b$$

$$162 = 15a + 65b$$

By Gauss elimination:

$$\left[\begin{array}{cc|c} 4 & 15 & 41 \\ 15 & 65 & 162 \end{array}\right] \xrightarrow{-\frac{15}{4}r_1 + r_2} \left[\begin{array}{cc|c} 4 & 15 & 41 \\ 0 & 8.75 & 8.25 \end{array}\right]$$

By backward substitution: $b = 0.9429$, $a = 6.7141$

Linear regression line:

$$y = a + bx \quad \Rightarrow \quad y = 6.7141 + 0.9429x$$

Exercise:

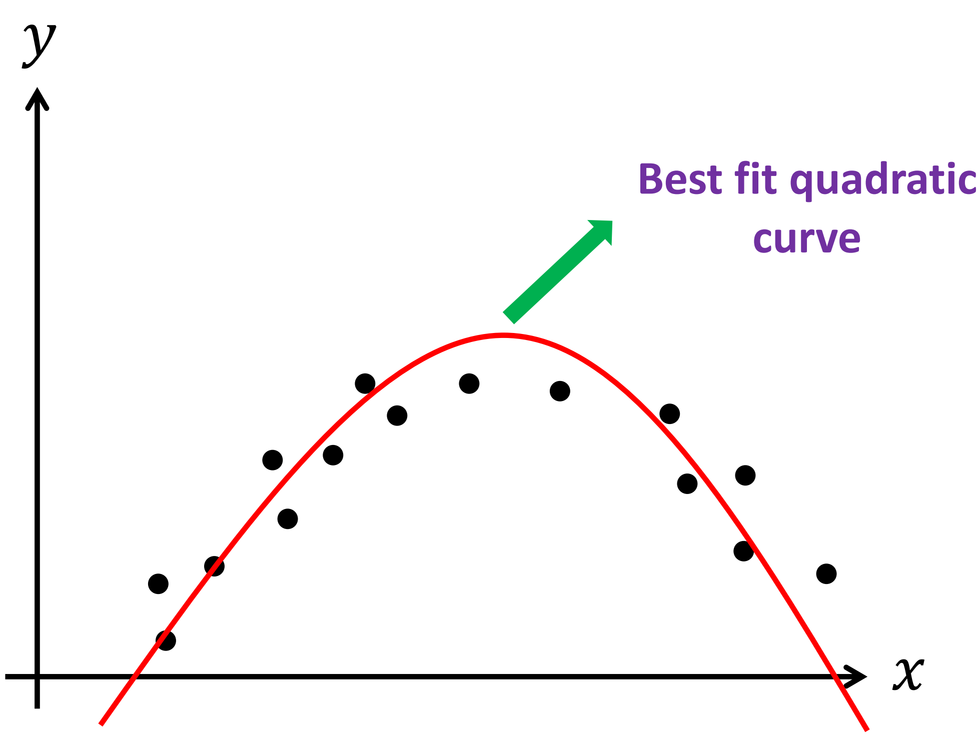
Find the least-square regression line for the following data.

x_i	0	2	5	6	8	9	13	15
$f(x_i)$	4	9	12	16	22	23	36	40

Then evaluate $f(10)$.

4.3 Extrapolation by Curve Fitting
4.3.2 Polynomial Regression

- Fitting a polynomial curve to a set of given observation points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.



Equation of a quadratic polynomial:

$$\boxed{y = a + bx + cx^2}$$

4.3.2.1 Illustration of Polynomial Regression (quadratic)

Example 4.8:

Find a second-order polynomial that best fits the following data.

x_i	0	1	3	5	6
y_i	11	13	16.5	18	20

4.3.2.1 Illustration of Polynomial Regression (quadratic)

Solution:

Step 1: Find the values of $\sum x$, $\sum x^2$, $\sum x^3$, $\sum x^4$, $\sum y$, $\sum xy$ and $\sum x^2 y$

x_i	y_i	$x_i y_i$	x_i^2	$x_i^2 y_i$	x_i^3	x_i^4
0	11	0	0	0	0	0
1	13	13	1	13	1	1
3	16.5	49.5	9	148.5	27	81
5	18	90	25	450	125	625
6	20	120	36	720	216	1296
$\sum x_i = 15$	$\sum y_i = 78.5$	$\sum x_i y_i = 272.5$	$\sum x_i^2 = 71$	$\sum x_i^2 y_i = 1331.5$	$\sum x_i^3 = 369$	$\sum x_i^4 = 2003$

Solution:

Step 2: Form the linear system as follows:

$$\sum y = an + b\sum x + c\sum x^2$$

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

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

where n is the number of points.

The linear system formed:

$$78.5 = 5a + 15b + 71c$$

$$272.5 = 15a + 71b + 369c$$

$$1331.5 = 71a + 369b + 2003c$$

Step 3: Solve the linear system by Gauss Elimination, Gauss Seidel or LU decomposition.

By Gauss elimination:

$$\left[\begin{array}{ccc|c} 5 & 15 & 71 & 78.5 \\ 15 & 71 & 369 & 272.5 \\ 71 & 369 & 2003 & 1331.5 \end{array}\right] \xrightarrow[-14.2r_1+r_3]{-3r_1+r_2} \left[\begin{array}{ccc|c} 5 & 15 & 71 & 78.5 \\ 0 & 26 & 156 & 37 \\ 0 & 156 & 994.8 & 216.8 \end{array}\right]$$

$$\xrightarrow{-6r_2+r_3} \left[\begin{array}{ccc|c} 5 & 15 & 71 & 78.5 \\ 0 & 26 & 156 & 37 \\ 0 & 0 & 58.8 & -5.2 \end{array}\right]$$

By backward substitution:

$$c = -0.0884, b = 1.9535, a = 11.0948$$

Quadratic polynomial:

$$y = a + bx + cx^2 \Rightarrow \quad y = 11.0948 + 1.9535x - 0.0884x^2$$

Exercise:

Find the least-square quadratic polynomial for the following data.

x_i	0	2	3	5	6	8	10	13
$f(x_i)$	2	4	6	10	13	17	22	27

Then evaluate $f(7)$.

4.3 Extrapolation by Curve Fitting
4.3.3 Multiple Linear Regression

- Fitting a multiple linear regression to a set of given observations with several independent variables.

- Equation of a multiple linear regression:

$$\boxed{y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_m}$$

4.3.3 Multiple Linear Regression

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_m$$

where the coefficients $a_0, a_1, \dots, a_m$ are:

$$\sum y = a_0 n + a_1 \sum x_1 + \cdots + a_m \sum x_m$$

$$\sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 + \cdots + a_m \sum x_1 x_m$$

$$\sum x_2 y = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 + \cdots + a_m \sum x_2 x_m \vdots$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\sum x_m y = a_0 \sum x_m + a_1 \sum x_1 x_m + a_2 \sum x_2 x_m + \cdots + a_m \sum x_m^2$$

4.3.3 Multiple Linear Regression

For two independent variables:

$$y = a + bx_1 + cx_2$$

where

$$\sum y = an + b\sum x_1 + c\sum x_2$$

$$\sum x_1 y = a\sum x_1 + b\sum x_1^2 + c\sum x_1 x_2$$

$$\sum x_2 y = a\sum x_2 + b\sum x_1 x_2 + c\sum x_2^2$$

4.3.3.1 Illustration of Multiple Linear Regression

Example 4.9:

Use multiple linear regression to fit these data.

x_1	1	3	6	8	9
x_2	2	5	7	9	10
y	6	9	11	15	20

Solution:

Step 1: Find the values of $\sum y$, $\sum x_1$, $\sum x_2$, $\sum x_1^2$, $\sum x_2^2$, $\sum x_1x_2$, $\sum x_1y$ and $\sum x_2y$

y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
6	1	2	1	4	2	6	12
9	3	5	9	25	15	27	45
11	6	7	36	49	42	66	77
15	8	9	64	81	72	120	135
20	9	10	81	100	90	180	200
$\sum y = 61$	$\sum x_1 = 27$	$\sum x_2 = 33$	$\sum x_1^2 = 191$	$\sum x_2^2 = 259$	$\sum x_1 x_2 = 221$	$\sum x_1 y = 399$	$\sum x_2 y = 469$

Solution:

Step 2: Form the linear system as follows:

$$\sum y = an + b\sum x_1 + c\sum x_2$$

$$\sum x_1 y = a\sum x_1 + b\sum x_1^2 + c\sum x_1 x_2$$

$$\sum x_2 y = a\sum x_2 + b\sum x_1 x_2 + c\sum x_2^2$$

where n is the number of observations.

The linear system formed:

$$61 = 5a + 27b + 33c$$

$$399 = 27a + 191b + 221c$$

$$469 = 33a + 221b + 259c$$

Solution:

Step 3: Solve the linear system by Gauss Elimination, Gauss Seidel or LU decomposition.

By Gauss elimination:

$$\left[\begin{array}{ccc|c} 5 & 27 & 33 & 61 \\ 27 & 191 & 221 & 399 \\ 33 & 221 & 259 & 469 \end{array}\right] \xrightarrow[-6.6r_1+r_3]{-5.4r_1+r_2} \left[\begin{array}{ccc|c} 5 & 27 & 33 & 61 \\ 0 & 45.2 & 42.8 & 69.6 \\ 0 & 42.8 & 41.2 & 66.4 \end{array}\right]$$

$$\xrightarrow{-0.9469r_2+r_3} \left[\begin{array}{ccc|c} 5 & 27 & 33 & 61 \\ 0 & 45.2 & 42.8 & 69.6 \\ 0 & 0 & 0.6727 & 0.4958 \end{array}\right]$$

By backward substitution:

$c = 0.7370, b = 0.8420, a = 2.789$

Multiple linear regression:

$$y = a + bx_1 + cx_2$$

$$\Rightarrow \quad y = 2.789 + 0.842x_1 + 0.737x_2$$

Exercise:

Find the multiple linear regression for the following data.

x_1	2	4	7	9	10	13
x_2	1	3	6	10	12	15
y	3	6	10	12	14	18