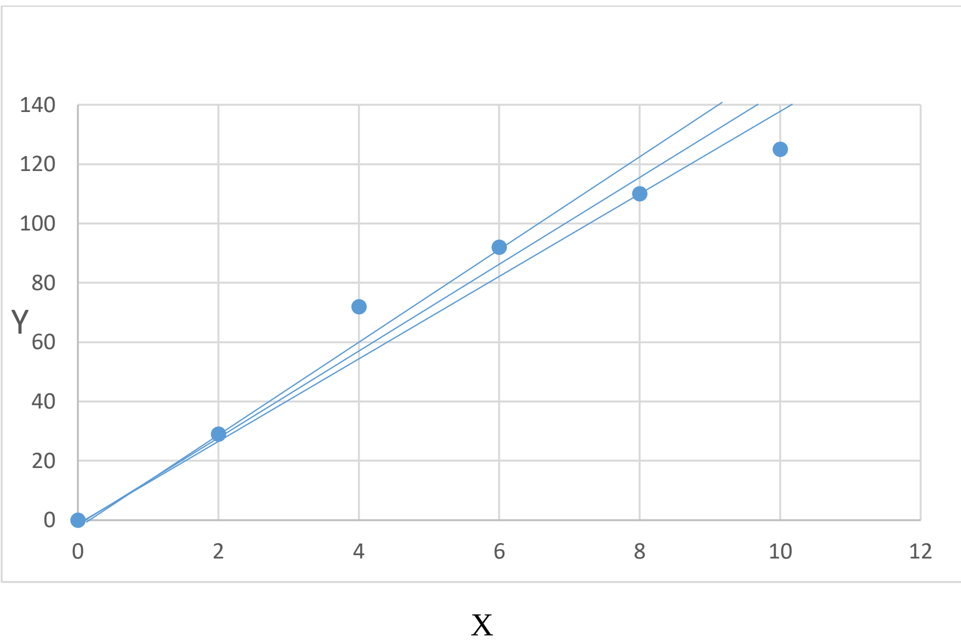
Lecture

So far we have considered there is no discrepancy between observed data which is represented by a function f(x) and the fitting function g(x). However, in the case of experimental data we come across situations where there is difference between experimental and fitting curve. In this lecture, we shall address this issue.

Fitting of curve:

Let us take an example. We consider below is the graph of verification of Ohm's law. X represents V (Voltage) and Y represents I (Current). We know the equation for Ohm's law is given by V = IR, R represents resistance. Therefore, it should be a straight line passing through origin. The solid circles are data obtained from an experiment. Now, what should be the way to fit this curve by straight line? We have drawn number of ways in which these data could be fit and more are possible. It is to be mentioned here that experimental data and fitting data must possess minimum discrepancy. How to do it in unique way? This is what precisely we are going to study here.



Least square curve fitting:

Let us assume set of data points (x_i,y_i), i = 1,2,3,…,n and the fitting curve for set of data is given by y = f(x). If we consider a point $x = x_i$, at which the observed value of y is y_i and the value of fitting curve is $f(x_i)$. So the difference between observed and fitting value is

$E_i = y_i - f(x_i)$ (1)

i = 0, 1, 2, ……, n.

And so E_i is the error.

Now, we write

$$S = \sum_{i=1}^{n}[y_i - f(x_i)]^2 = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots \dots \dots \dots + [y_n - f(x_n)]^2 \quad (2)$$

Or $S = \sum_{i=1}^{n} E_i^2 = E_1^2 + E_2^2 + \cdots \dots \dots .. + E_n^2$ (3)

i.e. S is the sum of squares of errors in each data point.

In this method S is to be minimized. Let us see different cases.

(i) The curve is straight line:

Suppose, the data points are to be fitted with a straight line y = mx+c; m and c are slope and intercept, respectively.

$$S = \sum_{i=1}^{n}[y_i - (mx_i + c)]^2 = [y_1 - (mx_1 + c)]^2 + [y_2 - (mx_2 + c)]^2 + \cdots \dots \dots \dots \dots \dots + [y_n - (mx_n + c)]^2 \quad (4)$$

To minimize,

$$\frac{\partial y}{\partial c} = 0 = -2[y_1 - (mx_1 + c)] - 2[y_2 - (mx_2 + c)] - \cdots \dots \dots \dots .. -2[y_n - (mx_n + c)] \quad (5)$$

and

$$\frac{\partial y}{\partial m} = 0 = -2x_1[y_1 - (mx_1 + c)] - 2x_2[y_2 - (mx_2 + c)] - \cdots . -2x_n[y_n - (mx_n + c)] \quad (6)$$

Equation (5) can be written as

$$nc + m(x_1 + x_2 + \cdots .. + x_n) = y_1 + y_2 + \cdots \dots . + y_n$$

And equation (6),

$$c(x_1 + x_2 + \cdots .. + x_n) + m(x_1^2 + x_2^2 + \cdots ... + x_n^2) = x_1y_1 + x_2y_2 + \cdots .. + x_ny_n$$

These equations can also be written as

$$nc + m\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad (7)$$

$$c\sum_{i=1}^{n} x_i + m\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_iy_i \quad (8)$$

Solving equations (7) and (8) give m and c. Thus, we get a unique straight line with minimum error.

Example: Fit the given data with straight line y = mx+c.

x	2	4	6	8	10
y	4	6	7	8	10

Solution:

x		y	x^2	xy
2		4	4	8
4		6	16	24
6		7	36	42
8		8	64	64
10		10	100	100
Total	30	35	220	238

Here n = 5.

Hence, using equations (7) and (8),

5c + 30 m = 35 (9)

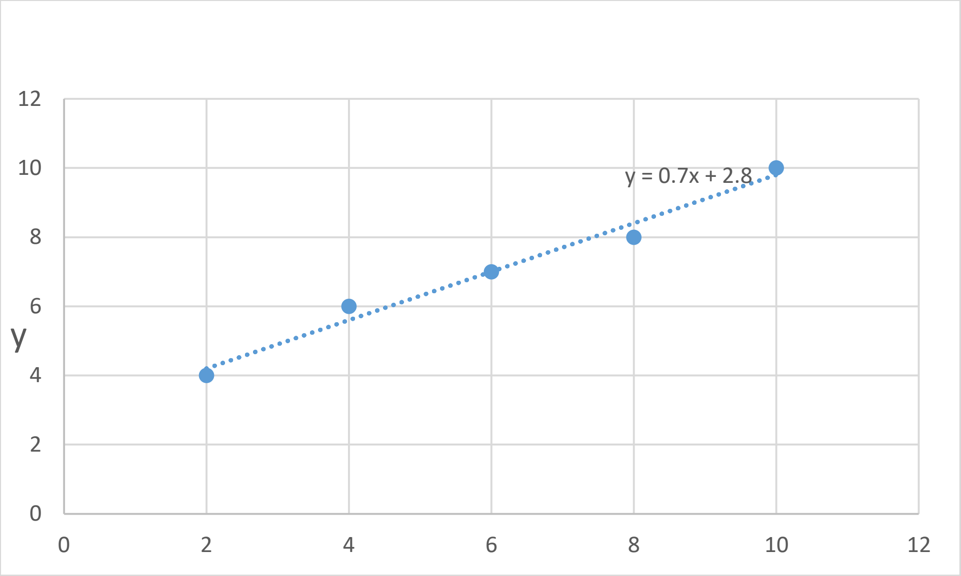
30c+ 220 m = 238 (10)

Solving equations (9) and (10), we get

m = 0.7 and c = 2.8

Hence, equation of fitting curve becomes

y = 0.7 x + 2.8, the result is shown in graph below. Dotted line is the fit to data points.



x

(ii) Nonlinear curve fitting

(A) Power functions:

Let $y = ax^b$

Taking log, we get

$\log y = \log a + b \log x$

$y = mx+c$

$y = \log y$, $x = \log x$, $c = \log a$, $m = b$

(B) Other curves:

$y = cx+d/x$

$xy = cx^2+d$

Now, $xy = Y$ and $x^2 = X$, converts into linear equation.

$y = pe^{nx}$

$\log y = \log p + nx$

$y = mx + c$

$Y = \log y$, $x = X$ and $\log p = c$, $m = n$

We can solve.

(C) Polynomial of nth degree:

Let the polynomial of nth degree is

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots \ldots \ldots \ldots \ldots . + a_n x^n \qquad (1)$$

This polynomial is to fit data points (x_i, y_i), $i = 0, 1, 2, \ldots, m$.

So, we have

$$S = \sum_{i=1}^{m} [y_i - (mx_i + c)]^2 = [y_1 - (a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 + \cdots \ldots \ldots \ldots \ldots . + a_n x_1^n)]^2 + [y_2 - (a_0 + a_1 x_2 + a_2 x_2^2 + a_3 x_2^3 + \cdots \ldots \ldots \ldots \ldots . + a_n x_2^n)]^2 + \cdots \ldots \ldots \ldots \ldots \ldots + [y_n - (a_0 + a_1 x_m + a_2 x_m^2 + a_3 x_m^3 + \cdots \ldots \ldots \ldots \ldots . + a_n x_m^n)]^2 \qquad (2)$$

Minimizing the same way and solving, we get

$$m a_0 + a_1 \sum_{i=1}^{m} x_i + a_2 \sum_{i=1}^{m} x_i^2 + \cdots \ldots \ldots . . + a_n \sum_{i=1}^{m} x_i^n = \sum_{i=1}^{m} y_i \qquad (3)$$

$$a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 + \cdots \ldots \ldots . . + a_n \sum_{i=1}^{m} x_i^{n+1} = \sum_{i=1}^{m} x_i y_i \qquad (4)$$

$$a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + \cdots \ldots \ldots . . + a_n \sum_{i=1}^{m} x_i^{2n} = \sum_{i=1}^{m} x_i^n y_i \qquad (5)$$

We can put in matrix form

$$\begin{bmatrix} m & \cdots & \sum_{i=1}^{m} x_i^n \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} x_i^n & \cdots & \sum_{i=1}^{m} x_i^{2n} \end{bmatrix} [a] = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i^2 y_i \\ \vdots \end{bmatrix} \quad (6)$$

where

$$[a] = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \end{bmatrix}$$

We have seen methods to solve such equations (Gauss elimination, etc.). There are (n+1) equations and solving them would get a_0, a_1, etc. One important point to be noted here that equation (6) is an ill-conditioned system. Upto polynomial of degree 3 or 4 problem is not much but if the polynomial is of higher degree then round off errors become large. To treat such problems, special methods are needed. We shall not cover these things here.

Example: Fit the data with $y = a_0+a_1x+a_2x^2$.

x		y	x^2	x^3	x^4	xy	x^2y
1		0.63	1	1	1	0.63	0.63
3		2.05	9	27	81	6.15	18.45
4		4.08	16	64	256	16.32	65.28
6		10.78	36	216	1296	64.68	388.08
Total	14	17.54	62	308	1634	87.78	472.44

n = 4

Equations become

$4\ a_0 + 14\ a_1 + 62\ a_2 = 17.54$

$14\ a_0+62\ a_1+308\ a_2 = 87.78$

$62\ a_0+308\ a_1+1634\ a_2 = 472.44$

Gives, $a_0 = 1.24$, $a_1 = -1.05$, $a_2 = 0.44$

Problem for students.

Plot the graph on graph paper for given data and fitting data of above example and see yourself.