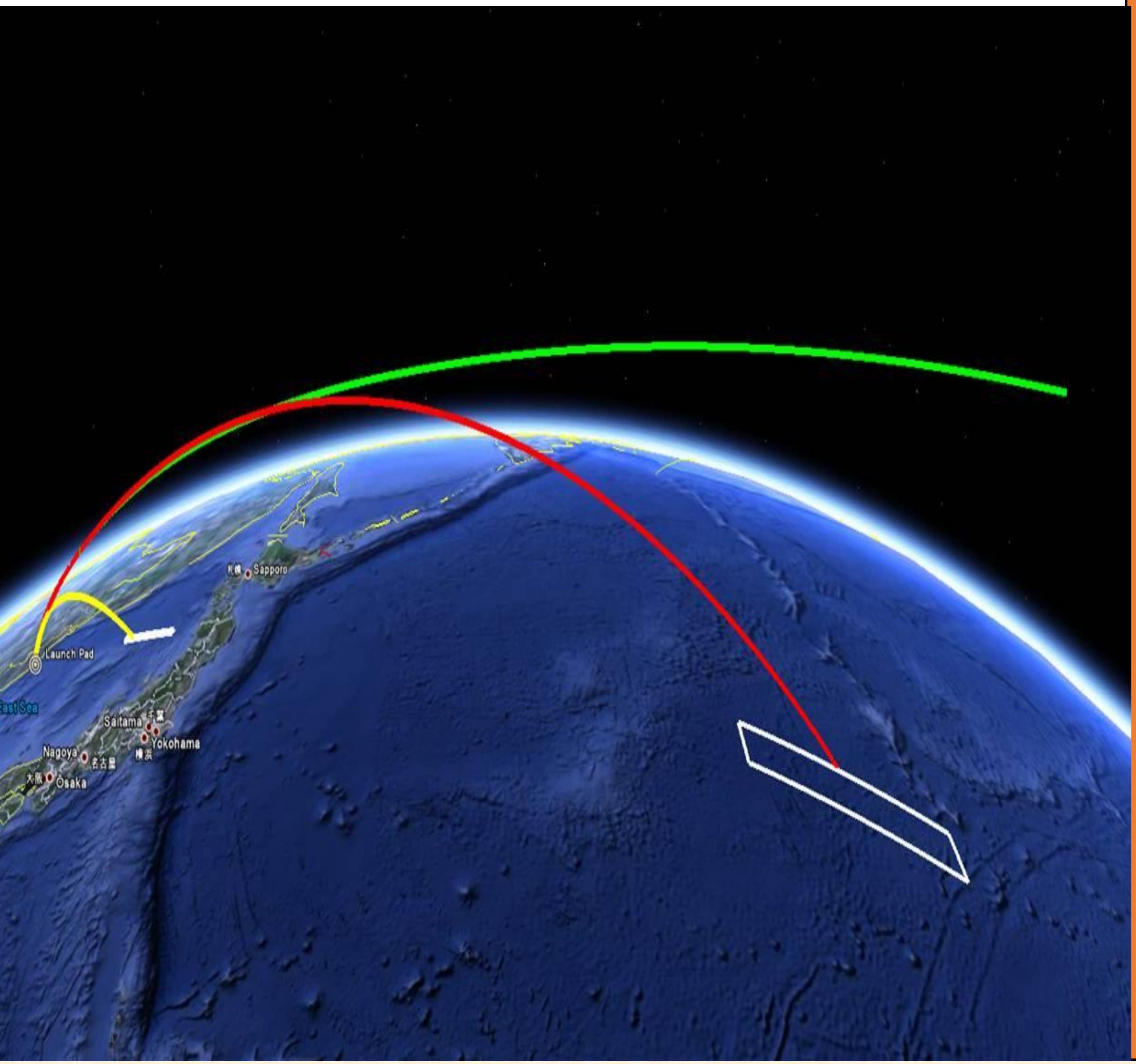


CURVE FITTING

DEPARTMENT OF APPLIED SCIENCES
PARUL UNIVERSITY

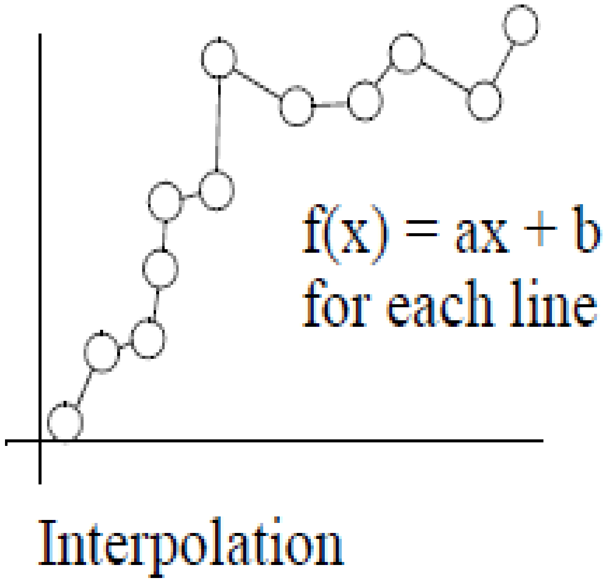
Introduction

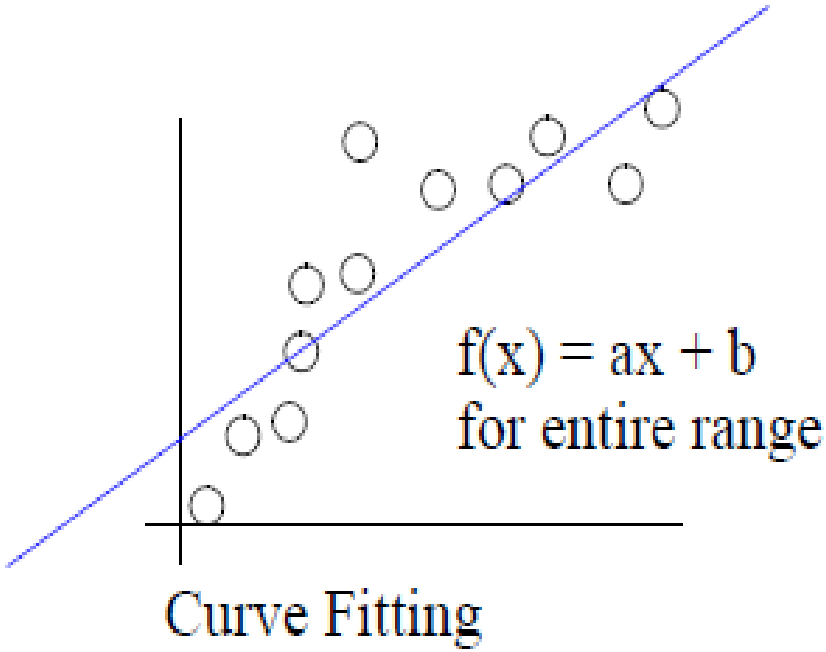
- Motivation
- Method of Least Squares
  - Linear equation $y=ax+b$
  - Quadratic equation $y=a+bx+cx^2$
  - Exponential $y=ax^b$, $y=Ce^{ax}$

Curve fitting – Motivation

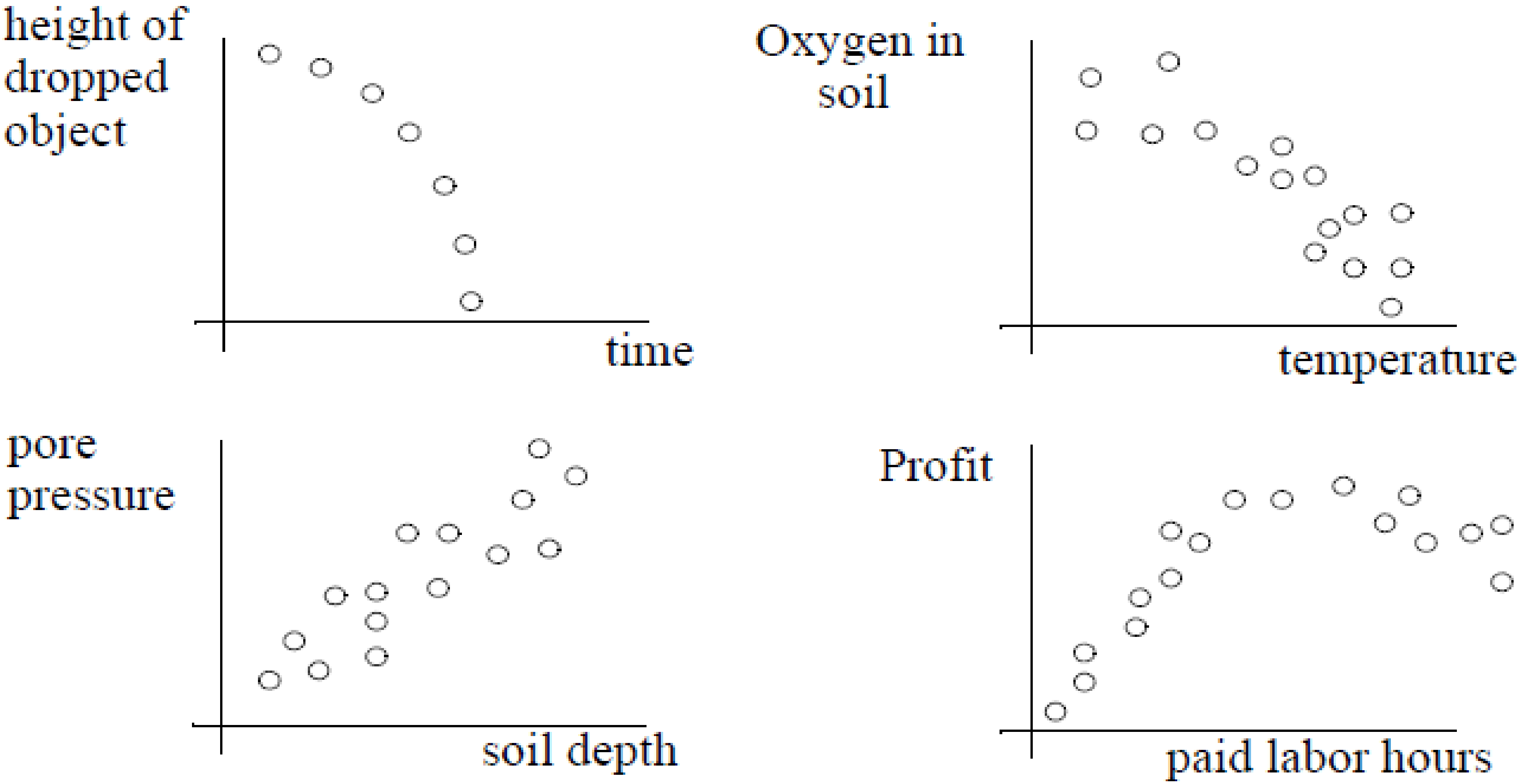
For root finding, we used a given function to identify where it crossed zero where does ?

Curve fitting - capturing the trend in the data by assigning a single function across the entire range. The example below uses a straight line function





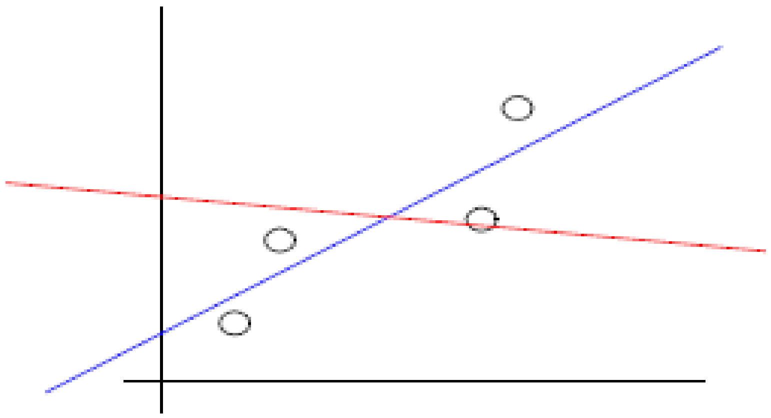
examples of data sets that we can fit a function to

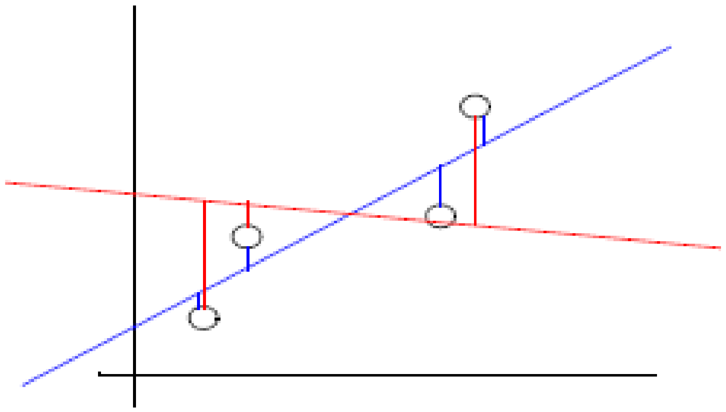


Linear curve fitting (linear regression)

Given the general form of a straight line

$$f(x) = ax + b$$



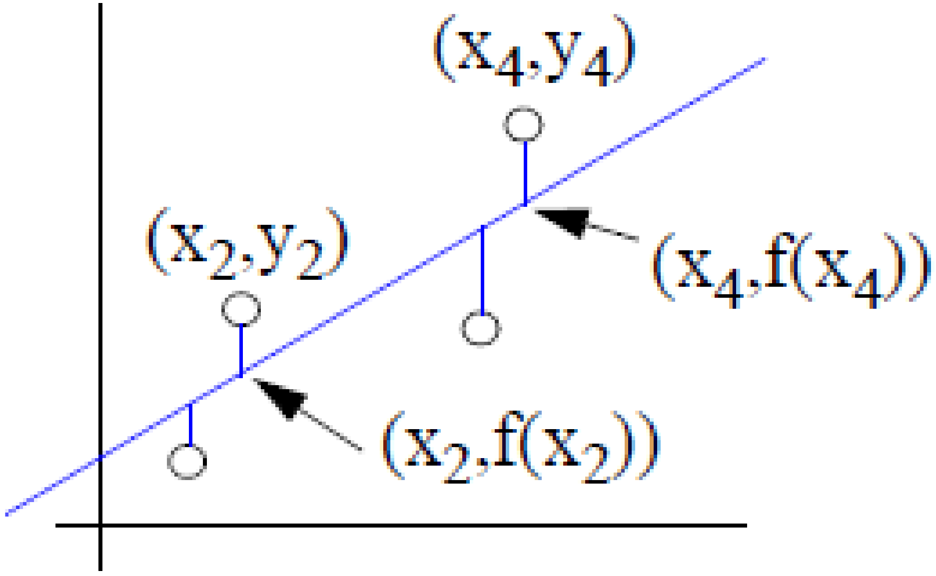


Quantifying errors in a curve fit

Assumption:

(1) positive or negative error have the same value
(data point is above or below the line)

(2) Weight greater errors more heavily
we can do both of these things by squaring the distancedenote data values as
(x, y) ====>> denote points on the fitted line as (x, f(x)) sum the error at the four data points



$$err = \sum_{i=1}^{n} d_i^{\,2} = \left(y_1 - f\left(x_1\right)\right)^2 + \left(y_2 - f\left(x_2\right)\right)^2 + \ldots\ldots\ldots\left(y_n - f\left(x_n\right)\right)^2$$

$$= \left(y_1 - \left(ax_1 + b\right)\right)^2 + \left(y_2 - \left(ax_2 + b\right)\right)^2 + \ldots\ldots\ldots + \left(y_n - \left(ax_n + b\right)\right)^2$$

$$= \sum_{i=1}^{n} \left(y_i - \left(ax_i + b\right)\right)^2$$

Error is minimum

$$\frac{\partial(err)}{\partial a} = \sum_{i=1}^{n} -2x_i\left(y_i - \left(ax_i + b\right)\right) = 0$$

$$\therefore \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i^{\,2} - b\sum_{i=1}^{n} x_i = 0$$

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

$$\frac{\partial(err)}{\partial b} = \sum_{i=1}^{n} -2\left(y_i - \left(ax_i + b\right)\right) = 0$$

$$\therefore \sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} 1 = 0$$

$$\therefore \sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + nb$$

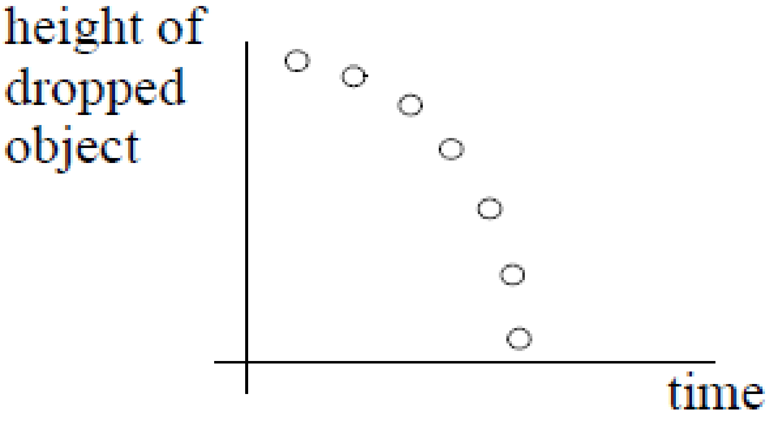
Solve the equations

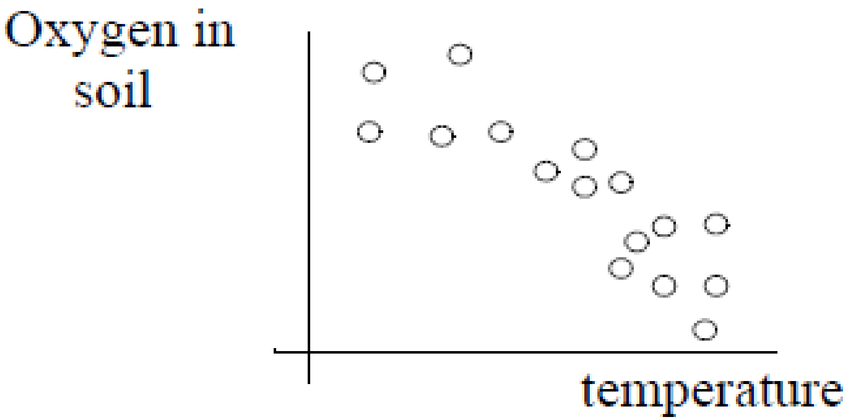
$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + nb \qquad (1)$$

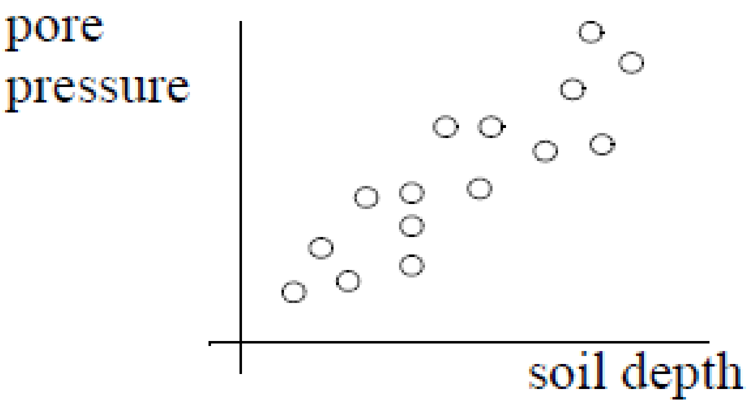
$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^{\,2} + b\sum_{i=1}^{n} x_i \qquad (2)$$

Curve fitting - higher order polynomials

We started the linear curve fit by choosing a generic form of the straight line $f(x) = ax + b$
This is just one kind of function. There are an infinite number of generic forms we could choose from for almost any shape we want.
Let's start with a simple extension to the linear regression concept
recall the examples of sampled data


height of dropped object
time

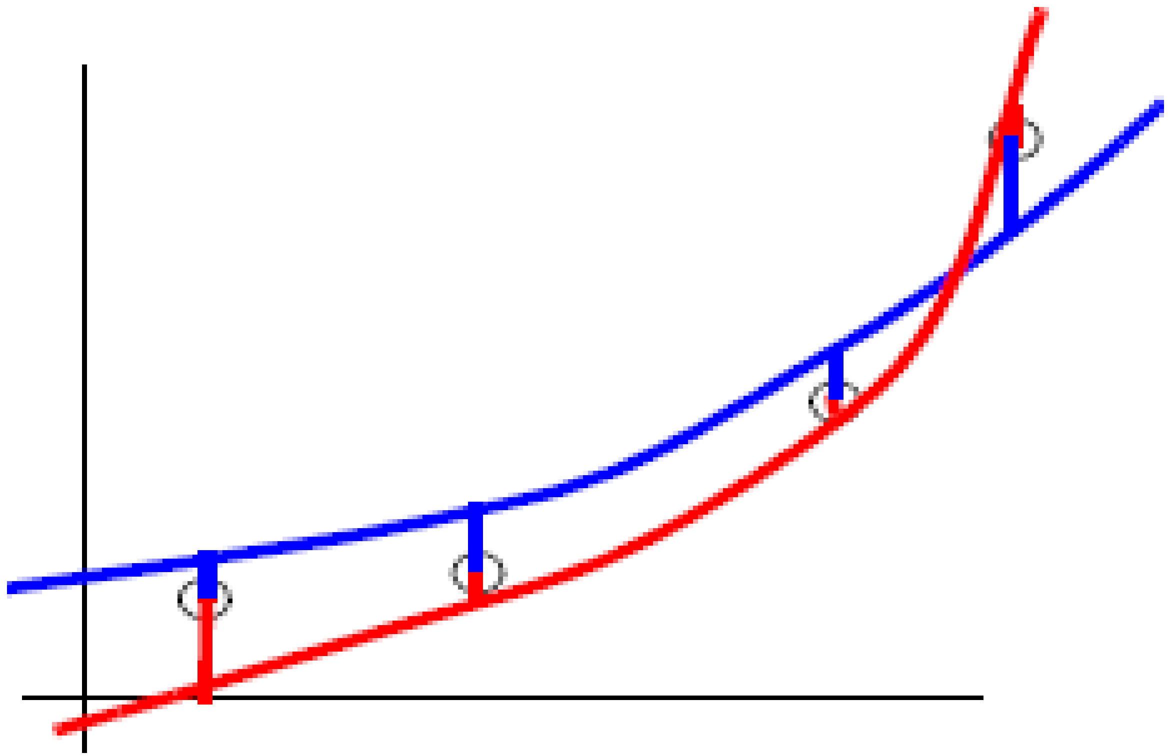

Oxygen in soil
temperature


pore pressure
soil depth


Profit
paid labor hours

Curve fitting – Quadratic polynomial

Let the general form of second order polynomial $f(x) = a + bx + cx^2$



Error - Least squares approach

$$err = \sum_{i=1}^{n} d_i^{\ 2} = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + \ldots\ldots..(y_n - f(x_n))^2$$

$$= \left(y_1 - \left(a + bx_1 + cx_1^{\ 2}\right)\right)^2 + \left(y_2 - \left(a + bx_2 + +cx_2^{\ 2}\right)\right)^2 + \ldots\ldots.. + \left(y_n - \left(a + bx_n + cx_n\right)\right)^2$$

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

To minimize the error

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

$$\frac{\partial(err)}{\partial b} = \sum_{i=1}^{n} -2x_i\left(y_i - \left(a + bx_i + cx_i^{\ 2}\right)\right) = 0$$

$$\frac{\partial(err)}{\partial b} = \sum_{i=1}^{n} -2x_i^{\ 2}\left(y_i - \left(a + bx_i + cx_i^{\ 2}\right)\right) = 0$$

Simplify these equations, We get

$$\sum_{i=1}^{n} y_i = an + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^{\,2}$$

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

$$\sum_{i=1}^{n} x_i^{\,2} y_i = a\sum_{i=1}^{n} x_i^{\,2} + b\sum_{i=1}^{n} x_i^{\,3} + c\sum_{i=1}^{n} x_i^{\,4}$$

Curve fitting - Other nonlinear fits (exponential)

General exponential equation $f(x) = C\,e^{Ax}$

Now, take log on both side, we get

$$\ln y = \ln C + Ax$$

$$Y = b + aX; \qquad \text{where } Y = \ln y,\ X = x, \ln C = b \text{ and } a = \ln A$$

Which is equation of line, the original data in xy- plane mapped into XY-plane. This is called *linearization*. The data (x, y) transformed as $(x, \ln y)$

To find the value of a and b we will use the equations

$$\sum_{i=1}^{n} Y_i = a \sum_{i=1}^{n} X_i + nb \qquad (1)$$

$$\sum_{i=1}^{n} X_i Y_i = a \sum_{i=1}^{n} X_i^{\,2} + b \sum_{i=1}^{n} X_i \qquad (2)$$

$$A = \text{antilog}\, a,\ C = \text{antilog}\, b$$

Example: Fit a straight line using least square method

x_i	0	0.5	1	1.5	2	2.5
y_i	0	1.5	3	4.5	6	7.5

Solution : Solve the equations

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + nb$$

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

Substitute the values from the table, here n=6.

$$7.5 = 2.5a + 6b$$

$$18.75 = 6.25a + 2.5b$$

$$a = 3.561 \text{ and } b = -0.975$$

Hence, the best fit line is $y = 3.561x - 0.975$

Example: Fit a second order polynomial equation to following data

x_i	0	0.5	1.0	1.5	2.0	2.5
y_i	0	0.25	1.0	2.25	4.0	6.25

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

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

$$\sum_{i=1}^{n} x_i^{2} y_i = a\sum_{i=1}^{n} x_i^{2} + b\sum_{i=1}^{n} x_i^{3} + c\sum_{i=1}^{n} x_i^{4}$$

$y = x^2$ is required equation which fits the data.

Example: An experiment gave the following values:

Fit an exponential curve $y = Ce^{Ax}$

x	1	5	7	9
Y	10	15	12	21