# Curve Fitting

## Lesson 1: Straight Line and Quadratic Curve

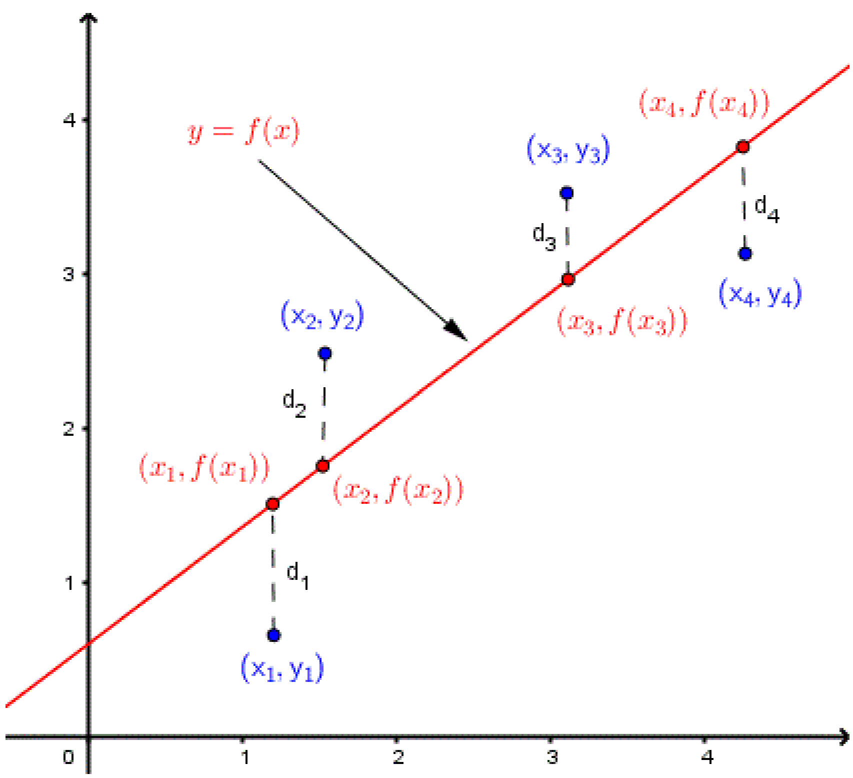
### Scatter Diagram:

From the scatter diagram, it is expected to obtain a functional relationship in the form of y = f (x) between two sets of variables x and y, giving an approximate curve which fit the data $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$. This method is called curve fitting.

### Method of Least Squares:

This method assumes that the best fit curve of a given type is the curve that has the minimum least squares error (sum of the squares of the deviations) from a given data set.

Let the data set be $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ where y is dependent variable and x is independent variable.



Then deviations or errors of estimate or residuals of $y_i$ from each data point is given by

$$d_1 = y_1 - f(x_1)$$

$$d_2 = y_2 - f(x_2)$$

$$d_3 = y_3 - f(x_3)$$

$$\dots\dots\dots\dots\dots$$

$$\dots\dots\dots\dots\dots$$

$$d_n = y_n - f(x_n)$$

According to the principle of least squares, the summation of squares of these residuals should be minimum, i.e,

$$D = {d_1}^2 + {d_2}^2 + {d_3}^2 + \cdots + {d_n}^2$$

$$= \sum_{i=1}^{n} {d_i}^2$$

$$= \sum_{i=1}^{n} [y_i - f(x_i)]^2 \text{ -------------- (1)}$$

should be minimum

## Straight Line:

Let the equation of the best fit line be y = a + bx --------- (2)

Applying the least squares method, we need to find values of a and b so that from equation (1), D is minimum. We can write D as

$$D = \sum_{i=1}^{n} [y_i - a - bx_i]^2 - - - - -(3)$$

In order that D is minimum, we must have $\frac{\partial D}{\partial a} = 0$ and $\frac{\partial D}{\partial b} = 0$ from equation (3). These two conditions give the following two restrictions involving the unknowns a and b as

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

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

These two equations are called normal equations or least square equations. Solving them, we get the values of a and b and with these values of a and b, equation (2) gives the line of best fit to the given data set of points $(x_i, y_i)$, i = 1, 2, …., n.

## Problems:

Ex.1. Using the method of least squares, find the best fitting straight line to the given data:

| | | | | | |
|---|---|---|---|---|---|
| x | 1 | 2 | 3 | 4 | 5 |
| y | 1 | 3 | 5 | 6 | 5 |

Solution: Let us consider that y = a + bx is the required straight line of the best fit. Let us now make the table for calculation:

| $x_i$ | $y_i$ | $x_i y_i$ | ${x_i}^2$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 3 | 6 | 4 |
| 3 | 5 | 15 | 9 |
| 4 | 6 | 24 | 16 |
| 5 | 5 | 25 | 25 |
| **15** | **20** | **71** | **55** |

We have the normal equations as:

$$20 = 5a + 15b$$

$$\text{and} \quad 71 = 15a + 55b$$

Solving we get, a = 0.7 and b = 1.1

Hence the line of best fit is given by: y = 0.7 + 1.1x.

Ans.

Ex.2. Fit a straight line to the following data:

| x | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| y | 3 | 4 | 5 | 6 | 8 |

Solution: Let y = a + bx is the line of the best fit. Let us make the table for calculation:

| $x_i$ | $y_i$ | $x_i y_i$ | ${x_i}^2$ |
|---|---|---|---|
| 1 | 3 | 3 | 1 |
| 2 | 4 | 8 | 4 |
| 3 | 5 | 15 | 9 |
| 4 | 6 | 24 | 16 |
| 5 | 8 | 40 | 25 |
| **15** | **26** | **90** | **55** |

Thus the normal equations are:

$$26 = 5a + 15b$$

$$\text{and} \quad 90 = 15a + 55b$$

Solving we get, a = 1.6 and b = 1.2

Hence the line of best fit is given by: y = 1.6 + 1.2x.

Ans.

## Quadratic Curve:

Let the equation of the second degree curve be given as

$$y = a + bx + cx^2 - - - (4)$$

which is the curve of best fit. Applying the principle of least squares, we get the normal equations in this case as

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

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

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

Solving these equations, we get the values of a, b and c and with these values of a, b and c, equation (4) gives the quadratic curve of best fit to the given data set of points $(x_i, y_i)$, i = 1, 2, …., n.

## Problems:

Ex.1. Find the quadratic curve of best fit by the method of least squares from the following data:

| x | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| y | 6 | 11 | 18 | 27 |

Solution: Let $y = a + bx + cx^2$ be the required quadratic curve of best fit. Let us make the table for calculation:

| $x_i$ | $y_i$ | ${x_i}^2$ | ${x_i}^3$ | ${x_i}^4$ | $x_i y_i$ | ${x_i}^2 y_i$ |
|---|---|---|---|---|---|---|
| 1 | 6 | 1 | 1 | 1 | 6 | 6 |
| 2 | 11 | 4 | 8 | 16 | 22 | 44 |
| 3 | 18 | 9 | 27 | 81 | 54 | 162 |
| 4 | 27 | 16 | 64 | 256 | 108 | 432 |
| **10** | **62** | **30** | **100** | **354** | **190** | **644** |

The normal equations are:

62 = 4a + 10b + 30c

190 = 10a + 30b + 100c

and 644 = 30a + 100b + 354c

Solving these equations, we get a = 3, b = 2 and c = 1. Therefore the required quadratic curve of best fit is given by: $y = 3 + 2x + x^2$

Ans.

Ex.2. Find a second degree curve to fit the following data:

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

Solution: Let $y = a + bx + cx^2$ be the required quadratic curve of best fit. Let us make the table for calculation:

| $x_i$ | $y_i$ | ${x_i}^2$ | ${x_i}^3$ | ${x_i}^4$ | $x_i y_i$ | ${x_i}^2 y_i$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 1 | 1 | 2 | 2 |
| 2 | 6 | 4 | 8 | 16 | 12 | 24 |
| 3 | 7 | 9 | 27 | 81 | 21 | 63 |
| 4 | 8 | 16 | 64 | 256 | 32 | 128 |
| 5 | 10 | 25 | 125 | 625 | 50 | 250 |
| 6 | 11 | 36 | 216 | 1296 | 66 | 396 |
| 7 | 11 | 49 | 343 | 2401 | 77 | 539 |
| 8 | 10 | 64 | 512 | 4096 | 80 | 640 |
| 9 | 9 | 81 | 729 | 6561 | 81 | 729 |
| **45** | **74** | **285** | **2025** | **15333** | **421** | **2771** |

The normal equations are:

$$74 = 9a + 45b + 285c$$

$$421 = 45a + 285b + 2025c$$

and $$2771 = 285a + 2025b + 15333c$$

Solving these equations, we get $a = -0.9283$,

$b = 3.523$

and $c = -0.2673$

Therefore the required quadratic curve of best fit is given by:

$$y = -0.9283 + 3.523x - 0.2673x^2$$

Ans.