Contents

Statistics

G.N.Dar

May 10, 2020

1 Curve Fitting to the Data

The given data set is fitted with the aim to define a formula or polynomial. The formula is used to estimate or pridict the missing points in the data or out of the data set.We will be knowing here a technique called *Curve Fitting* and in the next unit we will go through the technique called *Interpolation*.

1. *Interpolation*: This is a piece-wise connection of data points. Together two point ,three point or four point or more connection is done and accordingly a linear formula , parabolic,cubic or degree of 4 is defined. More about this technique in the next chapter on *Numerical Analysis*.

2. *Curve Fitting*: Looking at the data trend,a formula is assigned across the entire range, a best possible fit. Depending on the trend a stright line, parabolic ,cubic, exponential etc equation is fitted well and errors are estimated or measured say by least squre method. Therefore the best fit is declared the one where error so quntified is minimum.

The job in both the methods is to find the coefficients of the polynomial such that the function $f(x)$ fits the data well.

1.1 Linear Curve Fitting

If we have data setx, y, then finding or fitting this data set by a linear function $f(x) = ax + b$ if the data is linear. we can have many choices of fitting the given data but we choose the coefficients a and b so that the line fits the data well. That can be acheived by measuring the erros or deviations between the fitted line and the data point itself.Mathematically we add up all the deviations and the line is considered the best fit having minimum error.

Quantifying the Error:

Let $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ are the given data points.
Assume a best steright line fit as

$$f(x) = a * x + b$$

then

$$(x_1, f(x_1), (x_2, f(x_2), (x_3, f(x_3), (x_4, f(x_4)$$

are the correspounding points on the fitted line.

The difference between y_1 and $f(x_1,\ y_2$ and $f(x_2)$ etc measures the error. Therefore sum of the squre of the errors at these points are

$$err = \sum_{i=1}^{4} d_i^2 = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + (y_3 - f(x_3))^2 + (x_4 - f(x_4))^2 = \sum_{i=1}^{4} (y_i - f(x_i))^2$$

To have minimum error we must have

$$\frac{\partial err}{\partial a} = 0 = \frac{\partial err}{\partial b}$$

Therefore we have for n point data set

$$\sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0$$

and

$$\sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

solving amd we get

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

and

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

solving these algebric equation and evaluting coefficients a and b and the job is done

we can write them in matrix form as

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} * \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$
$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

using given data and coefficients can be evaluted and hence the fitted line is the best fit line .

1.2 Second order Curve Fitting

On the similar pattern we assume second order equation as

$$f(x) = a_0 + a_1 x + a_2 x^2$$

$$err = \sum_{i=1}^{n} [y_i - f(x_i)]^2 = \sum_{i=1}^{n} \left[y_i - a_0 - a_1 x_i - a_2 x_i^2\right]^2$$

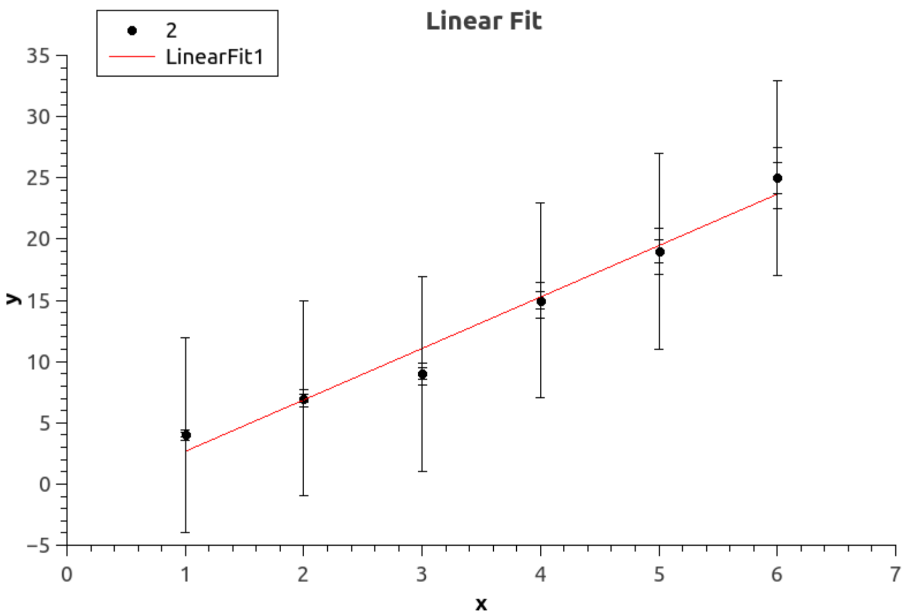

Figure 1: Linear Fitting

for the minimum error we must have

$$\frac{\partial err}{\partial a_0} = 0 = \frac{\partial err}{\partial a_1} = \frac{\partial err}{\partial a_2}$$

with this we have with respect a_0

$$2\sum_{i=1}^{n}\left[y_i - a_0 - a_1x_i - a_2x_i^2\right] = 0$$

which give

$$n.a_0 + a_1\sum_{i=1}^{n}x_i + a_2.\sum_{i=1}^{n}x_i^2 = \sum_{i=1}^{n}y_i$$

Similarly with respect a_1

$$\sum_{i=1}^{n}\left[y_i - a_0 - a_1x_i - a_2x_i^2\right](x_i) = 0$$

which gives

$$a_0\sum_{i=1}^{n}x_i + a_1\sum_{i=1}^{n}x_i^2 + a_2\sum_{i=1}^{n}x_i^3 = \sum_{i=1}^{n}x_iy_i$$

and with respect a_2

$$a_0\sum_{i=1}^{n}x_i^2 + a_1\sum_{i=1}^{n}x_i^3 + a_2\sum_{i=1}^{n}x_i^4 = \sum_{i=1}^{n}x_i^2y_i$$

Again we write in matrix form and solve for coefficient as

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} * \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix}$$

$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

using given data and coefficients can be evaluted and hence the fitted line is the best fit line .

similarly one can proceed for third, fourth degree curve fitting

1.3 polynomial Curve Fitting

let the polynomial of j^{th} order fitted to given data set is

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_j x^j = a_0 + \sum_{k=1}^{j} a_k x^k$$

Error will be

$$err = \sum_{i=1}^{n} d_i^2 = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + (y_3 - f(x_3))^2 + \cdots + (x_n - f(x_n))^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

with the given form of $f(x)$ we have

$$err = \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right]^2$$

for minimum with respect various coefficients, solve the above equation on the previous pattern we have

with respect a_0

$$-2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right] = 0$$

With respect a_1

$$-2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right] x_i = 0$$

$$-2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right] x_i^2 = 0$$

.... with respect a_j

$$-2 \sum_{i=1}^{n} \left[y_i - \left(a_0 + \sum_{k=1}^{j} a_k x_i^k \right) \right] x_i^j = 0$$

there are $j+1$ equations (recall for linear we have two equations, for parabolic three and now $j+1$ for j^{th} order .)

we can write in matrix form and solve for coefficients as

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 & ..... & \sum x_i^J \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & ..... & \sum x_i^{J+1} \\ . & . & . & ..... & . \\ . & . & . & ..... & . \\ . & . & . & ..... & . \\ \sum x_i^j & \sum x_i^{j+1} & \sum x_i^{j+2} & ..... & \sum x_i^{j+j} \end{pmatrix} * \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ . \\ . \\ a_j \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ . \\ . \\ \sum x_i^j y_i \end{pmatrix}$$

$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

using given data and coefficients can be evaluted and hence the fitted line is the best fit line .

1.4 Some Problems on curve fitting

a) for the given table , fit a steright line.

i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	0	1.5	3.0	4.5	6.0	7.5

Solution:

We see from the given table

$$\sum x_i = 7.5, \sum x_i^2 = 13.75, \sum y_i = 22.5, \sum x_i y_i = 41.25$$

we know for linear fit..

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} * \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

$$\begin{pmatrix} 6 & 7.5 \\ 7.5 & 13.75 \end{pmatrix} * \begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 22.5 \\ 41.25 \end{pmatrix}$$

$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

we see

$$\begin{pmatrix} b \\ a \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

b) for the given table , fit a steright line.

i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	2	-0.43261	-0.1656	3.1253	4.7877	8.6909

Solution: Home Work
c) Fit a second order polynomial to the give data set

i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	0	0.25	1.0	2.25	4.0	6.25

Solution:
for second order $j = 2$ that is $f(x) = a_0 + a_1 x + a_2 x^2$

$$\begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} * \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix}$$

$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

using given data and coefficients can be evaluted and hence the fitted line is the best fit line .

$$\begin{pmatrix} 6 & 7.5 & 13.75 \\ 7.5 & 13.75 & 8.125 \\ 13.75 & 28.125 & 61.1875 \end{pmatrix} * \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 13.75 \\ 28.125 \\ 61.1875 \end{pmatrix}$$

$$A * X = B$$

we can write it like a matrix equation as above and solve for coefficient matrix

$$AX = B \rightarrow X = A^{-1} * B$$

using given data and coefficients can be evaluted and hence the fitted line is the best fit line .

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Therefore the fitted line is

$$f(x) = a_0 + a_1 x + a_2 x^2 = 0 + 0 + x^2 = x^2$$

behaving just a parabolic