ME 704

Computational Methods in Thermal and Fluids Engineering

(Linear Regression)

Kannan Iyer
Kiyer@iitb.ac.in

Department of Mechanical Engineering
Indian Institute of Technology, Bombay

Review

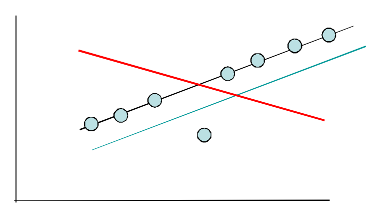
- Introduce the concept of spline interpolation, wherein a higher order curve is fitted between two points
- Cubic Spline passes third order equation through each pair of points
- This was accomplished by not only matching the functional values, but also the continuity of first and second derivatives
- The algebra and the algorithm was fully laid out

General

- When imprecise data is handled, regression is recommended
- Linear regression is most common
- Among the functions used polynomial is most preferred
- Piecewise lower order is always better that a single higher order polynomial
- In regression lower order polynomial is passed through a large number of points

Criterion for Regression

- Maximum error minimised?
- Sum of error minimised?
- Square of the sum of error minimised?



- $e_i = y_i - y_{fit}$
- Least squares criterion implies

$$\sum e_i^2 = minimum$$

Linear Regression-I

- If the function to be fitted is

$$f(x) = a_1 f_1(x) + a_2 f_2(x)$$

- Then sum of squares of error implies

$$\sum E_i^2 = \sum_{i=1}^{N} \left(y_i - (a_1 f_1(x_i) + a_2 f_2(x_i)) \right)^2$$

- a_1 and a_2 are determined by using

$$\frac{\partial \sum E_i^2}{\partial a_1} = \frac{\partial \sum E_i^2}{\partial a_2} = 0$$

5/15

Linear Regression-II

$$\frac{\partial \sum E_i^2}{\partial a_1} \Rightarrow 2\sum_{i=1}^{N} \left(y_i - (a_1 f_1(x_i) + a_2 f_2(x_i)) \right) f_1(x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N} (y_i f_1(x_i)) - a_1 \sum_{i=1}^{N} (f_1(x_i))^2 - a_2 \sum_{i=1}^{N} f_2(x_i) f_1(x_i) = 0$$

Similarly $\frac{\partial \sum E_i^2}{\partial a_2}$ will lead to

$$\Rightarrow \sum_{i=1}^{N} (y_i f_2(x_i)) - a_1 \sum_{i=1}^{N} f_1(x_i) f_2(x_i) - a_2 \sum_{i=1}^{N} (f_2(x_i))^2 = 0$$

6/15

Linear Regression-III

$$\begin{bmatrix} \sum_{i=1}^{N} (f_1(x_i))^2 & \sum_{i=1}^{N} f_1(x_i) f_2(x_i) \\ \sum_{i=1}^{N} f_1(x_i) f_2(x_i) & \sum_{i=1}^{N} (f_2(x_i))^2 \end{bmatrix} \begin{Bmatrix} a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{N} y(x_i) f_1(x_i) \\ \sum_{i=1}^{N} y(x_i) f_2(x_i) \end{Bmatrix}$$

- If $f_1(x) = 1$ and $f_2(x) = x$, then we have a linear fit
- Note that the coefficient matrix is symmetric
- a_1 and a_2 can be obtained by Gauss elimination

7/15

- The above procedure can be **generalised** for any number of functions. This is called **generalised least squares**
- Let us denote the F matrix as

$$F = \begin{bmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ ..... & ..... \\ ..... & ..... \\ f_1(x_n) & f_2(x_n) \end{bmatrix}$$

n x 2

$$\Rightarrow F^T = \begin{bmatrix} f_1(x_1) & f_1(x_2) & .... & .... & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & .... & .... & f_2(x_n) \end{bmatrix}$$

2 x n

8/15

Generalized least Squares-II

$$F^T F = \begin{bmatrix} \sum_{i=1}^{N} (f_1(x_i))^2 & \sum_{i=1}^{N} f_1(x_i) f_2(x_i) \\ \sum_{i=1}^{N} f_1(x_i) f_2(x_i) & \sum_{i=1}^{N} (f_2(x_i))^2 \end{bmatrix}$$

$$F^T \begin{Bmatrix} y(x_1) \\ y(x_2) \\ .. \\ .. \\ y(x_n) \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{N} y(x_i) f_1(x_i) \\ \sum_{i=1}^{N} y(x_i) f_2(x_i) \end{Bmatrix}$$

Generalized least Squares-III

- The linear least square can therefore be generalised as

$$F^T F \begin{Bmatrix} a_1 \\ a_2 \end{Bmatrix} = F^T \begin{Bmatrix} y(x_1) \\ y(x_2) \\ .. \\ .. \\ y(x_n) \end{Bmatrix}$$

Goodness of Fit

- The quality of fit is normally given by coefficient of regression, which is an obscure quantity
- In engineering parlance, the more relevant parameter that can be easily connected is the RMS error

$$\frac{\sum_{i=1}^{n} e_i^{\,2}}{n}$$

- Often the error is normalised with the true value to express the RMS error as a fraction or as percentage

Power Law Fit-I

- Functions of the form f = c Re^n can be fitted with linear regression
- Taking log of both sides, we get
  ln(f) = ln(c) + n ln(Re)
- This is of the form y = a + bx, where,
  y = ln(f), x = ln(Re)

Power Law Fit-II

- Multiple functions of the form Nu = c Re^n Pr^m can be fitted with linear regression

  ln (Nu) = ln(c) + n ln(Re) + m ln(Pr)

$$\Rightarrow z = a_0 f_0(x, y) + a_1 f_1(x, y) + a_2 f_2(x, y)$$

- In the above equation

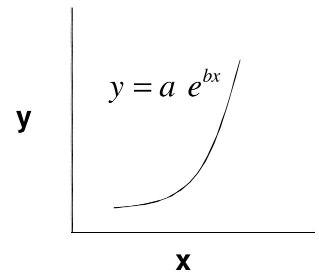
  z = ln (Nu), $f_0(x,y) = 1$, $f_1(x,y) = \ln(Re)$

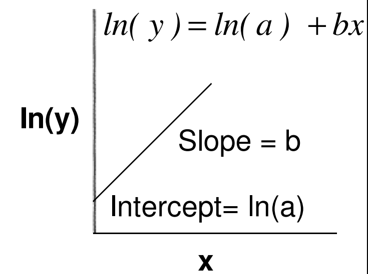
  $f_2(x,y) = \ln(Pr)$

13/15

Other Functions-I

- Other than power laws, exponential and saturation functions are also popular, which can be transformed into linear regression





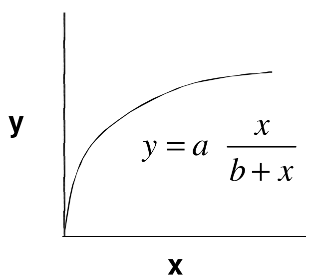
14/15

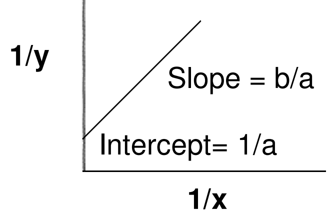
Other Functions-II

- Saturation Function

$$\frac{1}{y} = \frac{1}{a}\ \frac{b+x}{x}$$

$$\frac{1}{y} = \frac{b}{a}\frac{1}{x} + \frac{1}{a}$$





15/15