SSCE 2393 NUMERICAL METHODS

CHAPTER 4

CURVE FITTING

Farhana Johar, Department of Mathematical Sciences, Faculty of Science, UTM.
farhanajohar@utm.my

Overview of Chapter 4

4.1 Introduction to Data Fitting

- An important area in approximation is the problem of fitting a curve to experimental data. Since the data is experimental, we must assume that it is polluted with some degree of error-most commonly measurement error.
- So, we do not necessarily want to construct a curve that goes through every data point.
- Rather, we want to construct a function that represents the 'sense of data' and is, in some sense, a close approximation to the data.

(An Introduction to Numerical Methods and Analysis, J.F. Epperson, 2002, John Wiley & Sons, Inc)

4.2 Lease Square Curve Fiting

The most common approach is known as least squares data fitting.

- Is a method of finding a function which 'best' fit the data. (the line not necessarily pass through all points.)
- Rule: the sum of square error must be minimum.
- This method is called ***Least Squares Method*** and the curve that is obtained is called ***Least Squares Curve***.

Least Squares Method

The easiest method of finding a least squares curve is by using a polynomial function.

- Linear polynomial
- Quadratic/parabolic polynomial

4.2.1 Linear Least Squares

For a given set of data $\{(x_i, y_i)\}_{i=0}^{n}$, we want to find an equation of a line $y = A + Bx$ which minimize the total of squares error,

$$S = \sum_{i=0}^{n}\left(y_i - y(x_i)\right)^2 = \sum_{i=0}^{n}\left(y_i - A - Bx_i\right)^2$$

S is minimum when $\dfrac{\partial S}{\partial A} = 0$ and $\dfrac{\partial S}{\partial B} = 0$.

These yield:

$$\sum_{i=0}^{n} 2\left(y_i - A - Bx_i\right)(-1) = 0$$

$$\sum_{i=0}^{n} 2\left(y_i - A - Bx_i\right)(-x_i) = 0$$

Divide every equation with (-2) and expand it to get *normal equations* as follows:

$$A\sum_{i=0}^{n} 1 + B\sum_{i=0}^{n} x_i = \sum_{i=0}^{n} y_i$$

$$A\sum_{i=0}^{n} x_i + B\sum_{i=0}^{n} x_i^{\,2} = \sum_{i=0}^{n} x_i y_i$$

or in matrix form:

$$\begin{pmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^{\,2} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \end{pmatrix}$$

Solve the system above to find a value of A and B and substitute in $y = A + Bx$ (the least squares line).

Example 1

Given the six data points (0, 0.98), (0.1, 1.01), (0.2, 0.99), (0.3,0.88), (0.4, 0.85), and (0.5, 0.77) which are represent by the function $y(x) = a + bx^2$. Apply the least-squares method to find *a* and *b*. Then, compute the total squares error, *S.*

4.2.2 Parabolic/Quadratic Least Squares

For a given set of data $\{(x_i, y_i)\}_{i=0}^{n}$, we want to find a parabolic equation $y = A + Bx + Cx^2$ which minimize the *total least squares error:*

$$S = \sum_{i=0}^{n} \left(y_i - y(x_i)\right)^2 = \sum_{i=0}^{n} \left(y_i - A - Bx_i - Cx_i^{\,2}\right)^2$$

S is minimum if $\dfrac{\partial S}{\partial A} = 0$, $\dfrac{\partial S}{\partial B} = 0$ and $\dfrac{\partial S}{\partial C} = 0$.

And we get,

$$\sum_{i=0}^{n} 2\left(y_i - A - Bx_i - Cx_i^{\,2}\right)(-1) = 0$$

$$\sum_{i=0}^{n} 2\left(y_i - A - Bx_i - Cx_i^{\,2}\right)(-x_i) = 0$$

$$\sum_{i=0}^{n} 2\left(y_i - A - Bx_i - Cx_i^{\,2}\right)\left(-x_i^{\,2}\right) = 0$$

Divide every equation with (-2) and expand it to get the *normal equations*:

$$A\sum_{i=0}^{n} 1 + B\sum_{i=0}^{n} x_i + C\sum_{i=0}^{n} x_i^{\,2} = \sum_{i=0}^{n} y_i$$

$$A\sum_{i=0}^{n} x_i + B\sum_{i=0}^{n} x_i^{\,2} + C\sum_{i=0}^{n} x_i^{\,3} = \sum_{i=0}^{n} x_i y_i$$

$$A\sum_{i=0}^{n} x_i^{\,2} + B\sum_{i=0}^{n} x_i^{\,3} + C\sum_{i=0}^{n} x_i^{\,4} = \sum_{i=0}^{n} x_i^{\,2} y_i$$

or in the matrix form;

$$\begin{pmatrix} \sum_{i=0}^{n} 1 & \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^{\,2} \\ \sum_{i=0}^{n} x_i & \sum_{i=0}^{n} x_i^{\,2} & \sum_{i=0}^{n} x_i^{\,3} \\ \sum_{i=0}^{n} x_i^{\,2} & \sum_{i=0}^{n} x_i^{\,3} & \sum_{i=0}^{n} x_i^{\,4} \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{n} y_i \\ \sum_{i=0}^{n} x_i y_i \\ \sum_{i=0}^{n} x_i^{\,2} y_i \end{pmatrix}$$

Solve the above system for a value of A, B and C to finally obtain the parabolic equation $y = A + Bx + Cx^2$.

4.3 Least Squares Power Function

Let say a set of data $\{(x_i, y_i)\}_{i=0}^{n}$ is going to be fit into a power function in the form of

$$y = Ax^B$$

where B is a constant.

Therefore, we need to find A so that the total least squares error, *S* is minimum:

$$S = \sum_{i=0}^{n} \left(y_i - y(x_i)\right)^2 = \sum_{i=0}^{n} \left(y_i - Ax_i^{\ B}\right)^2$$

S is minimum when

$$\frac{\partial S}{\partial A} = 0 \longrightarrow \qquad 2\sum_{i=0}^{n} \left(y_i - Ax_i^{\ B}\right)\left(-x_i^{\ B}\right) = 0$$

Divide the above equation with (-2) and factorize it to get:

$$\sum_{i=0}^{n} x_i^{\ B} y_i - A\sum_{i=0}^{n} x_i^{\ 2B} = 0 \quad \text{or} \quad A = \frac{\sum_{i=0}^{n} x_i^{\ B} y_i}{\sum_{i=0}^{n} x_i^{\ 2B}}$$

Example:

Recent world records for men's weight lifting is given as follows:

W (kg)	52	56	60	67.5	75	82.5	90	100	110
L (kg)	252.5	277.5	302.5	345	360	400	415	430	427.5

The higher the weight class, the greater the lift. Accordingly, it is proposed that lift should be proportional to $W^{2/3}$ which arise the equation $L = cW^{2/3}$. Determine the constant c by the least-squares method.

4.4 Data Linearization

Let say a set of data $\{(x_i, y_i)\}_{i=0}^{n}$ is going to be fit into a power function in the form of

$$y = Ce^{Dx}$$

where C and D are constants. Apply *ln* to both sides to yield

$$\underbrace{\ln(y)}_{Y} = \underbrace{\ln(C)}_{A} + \underbrace{Dx}_{BX}$$

- linear form: $Y = A + BX$

Solve for value *A* and *B* to get $y = Ce^{Dx}$.

Example 1:

Given the following data;

V	160	180	200	220	240
T	7	5.5	5.0	3.5	2.0

If $VT^a = b$, (a, b constants), then determine the value of a and b by the least squares method.