Least-Squares Curve Fitting Procedures

P. Sam Johnson

October 14, 2014

Overview

The study of approximation theory involves two general types of problems. **One problem** arises when a function is given explicitly, but we wish to find a "simpler" type of function, such as a polynomial, that can be used to determine approximate values of the given function.

The Taylor polynomial of degree n about the number x_0 is an excellent approximation to an $(n+1)$-times differentiable function f in a small neighbourhood of x_0.

The **second problem** in approximation theory is concerned with fitting functions to given data and find the "best" function in a certain class to represent the data.

The Lagrange interpolating polynomials are approximating polynomials and they fit certain data.

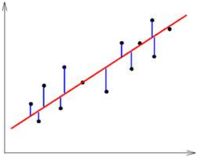
Limitations of these techniques are considerd, and other avenues of approach are disussed.

Introduction

Usually a mathematical equation is fitted to experimental data by plotting the data on a graph paper and then passing a straight line through the data points.

The method has the obvious drawback in that the straight line drawn may not be unique. The method of least squares is probably the most systematic procedure to fit a unique curve using given data points and is widely used in practical computations. It can also be easily implemented on a digital computer.

Fitting a Staight Line



Let $y = a_0 + a_1 x$ be the straight line to be fitted to the given data.

The problem of finding the equation of the best linear approximation requires that values of a_0 and a_1 be found to minimize

$$S(a_0, a_1) = \sum_{i=1}^{m} |y_i - (a_0 + a_1 x_i)|.$$

This quantity is called the **absolute deviation.**

To minimize a function of two variables, we need to set its partial derivatives to zero and simultaneously solve the resulting equations.

In the case of the absolute deviation, we need to find a_0 and a_1 with

$$0 = \frac{\partial S}{\partial a_0} = 0 = \sum_{i=1}^{m} |y_i - (a_0 + a_1 x_i)|.$$

and

$$0 = \frac{\partial S}{\partial a_1} = 0 = \sum_{i=1}^{m} |y_i - (a_0 + a_1 x_i)|.$$

The difficulty is that the absolute-value funciton is not differentiable at zero, and we may not be able to find solutions to this pair of equations.

The **least squares** approach to this problem involves determining the best approximating line when the error involved is the sum of the squares of the differences between the y-values on the approximating line and the given y-values.

Hence, the sum of the squares of the errors,

$$S = \sum_{i=1}^{m} [y_i - (a_0 + a_1 x_i)]^2.$$

For S to be minimum, we have

$$\frac{\partial S}{\partial a_0} = 0 = -2 \sum_{i=1}^{m} [y_i - (a_0 + a_1 x_i)].$$

and

$$\frac{\partial S}{\partial a_1} = 0 = -2 \sum_{i=1}^{m} x_i [y_i - (a_0 + a_1 x_i)].$$

The above equations are simplified to

$$ma_0 + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i$$

and

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

Since the x_i and y_i are known quantities, the above two equations (called the **normal equations**), can be solved for the two unknown a_0 and a_1.

Differentiating $\frac{\partial S}{\partial a_0}$ and $\frac{\partial S}{\partial a_1}$ with respect to a_0 to a_1 respectively, we find that

$$\frac{\partial^2 S}{\partial a_0^2} \text{ and } \frac{\partial^2 S}{\partial a_1^2}$$

will both be positive at the points a_0 and a_1. Hence these values provide a minimum of S.

Another approach of finding the equation of the best linear approximation requires that values of a_0 and a_1 be found to minimize

$$S(a_0, a_1) = \max_{1 \le i \le m} \left\{ \, |y_i - (a_0 + a_1 x_i)| \, \right\}.$$

This is commonly called a **minimax** problem and cannot be handled by elementary techniques.

The **minimax approach** generally assigns too much weight to a bit of data that is badly in error, whereas the **absolute deviation method** does not give sufficient weight to a point that is considerably out of line with the approximation.

The **least squares approach** puts substantially more weight on a point that is out of line with the rest of the data but will not allow that point to completely dominate the approximation.

Polynomial of the *n*th degree

Let the polynomial of the *n*th degree, $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be fitted to the data points $(x_i, y_i), i = 1, 2, \ldots, m$. We then have

$$S = \sum_{i=1}^{m} [y_i - (a_0 + a_1 x_i + \cdots + a_n x_i^n)]^2.$$

We get the following **normal equations**

$$\begin{aligned}
ma_0 + a_1 \sum_{i=1}^{m} x_i + a_2 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^n &= \sum_{i=1}^{m} y_i \\
a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 + \cdots + a_n \sum_{i=1}^{m} x_i^{n+1} &= \sum_{i=1}^{m} x_i y_i \\
&\vdots \\
a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + \cdots + a_n \sum_{i=1}^{m} x_i^{2n} &= \sum_{i=1}^{m} x_i^n y_i.
\end{aligned}$$

Nonlinear Curve Fitting : Power Function

We consider a **power function**, $y = ax^c$ to fit the given data points

$$(x_i, y_i), \qquad i = 1, \ldots, m.$$

Taking logarithms of both sides, we obtain the relation

$$\log y = \log a + c \log x,$$

which is of the form $Y = a_0 + a_1 X$, where $Y = \log y, a_0 = \log a, a_1 = c$ and $X = \log x$.

Hence the procedure discussed earlier can be followed to evaluate a_0 and a_1.

Then a and c can be calculated from the formulae $a_0 = \log a$ and $c = a_1$.

Nonlinear Curve Fitting : Exponential Function

Let the curve

$$y = a_0 e^{a_1 x*}$$

be fitted to the given data.

Then, as before, taking logarithms of both sides, we get

$$\log y = \log a_0 + a_1 x,$$

which can be written in the form

$$Z = A + Bx,$$

where $Z = \log y, A = \log a_0$ and $B = a_1$.

The problem therefore reduces to finding a least-squares straight line through the given data.

Fitting the data with given curve

Let the set of data points be

$$(x_i, y_i), i = 1, 2, \ldots, m,$$

and let the curve given by

$$y = f(x)$$

be fitted to this data.

At $x = x_i$, the experimental (or observed) value of the ordinate is y_i and the corresponding value on the fitting curve is $f(x_i)$.

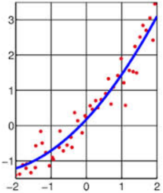
If e_i is the error of approximation at $x = x_i$, then we have

$$e_i = y_i - f(x_i).$$

If we write

$$S = [y_1 - f(x_1)]^2 + [y_2 - f(x_2)]^2 + \cdots + [y_m - f(x_m)]^2$$
$$= e_1^2 + e_2^2 + \cdots + e_m^2$$

then the method of least squares consists in minimizing S, i.e., the sum of the squares of the errors.



Nodes (x_i, y_i) are in red coloured points. The curve $y = f(x)$ is to be fitted to this data is shown in blue.

Weighted Least Squares Approximation

We have minimized the sum of squares of the errors. A more general approach is to minimize the weighted sum of the squares of the errors taken over all data points. If this sum is denoted by S, then we have

$$\begin{aligned} S &= \sum_{i=1}^{m} W_i [y_i - f(x_i)]^2 \\ &= \sum_{i=1}^{m} W_i e_i^2. \end{aligned}$$

In the above equation, the W_i are prescribed positive numbers and are called **weights**.

A weight is prescribed according to the relative accuracy of a data point. If all the data points are accurate, we set $W_i = 1$ for all i. We consider again the linear and nonlinear cases below.

Linear Weighted Least Squares Approximation

Let $y = a_0 + a_1 x$ be the staight line to be fitted to the given data points, $(x_1, y_1), \ldots, (x_m, y_m)$. Then

$$S(a_1, a_2) = \sum_{i=1}^{m} W_i [y_i - (a_0 + a_1 x_i)]^2$$

For maxima or minima, we have

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = 0$$

which given

$$\frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^{m} W_i [y_i - (a_0 + a_1 x_i)] = 0$$

and

$$\frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^{m} W_i [y_i - (a_0 + a_1 x_i)] x_i = 0.$$

Simplification yields the system of equations for a_0 and a_1 :

$$a_0 \sum_{i=1}^{m} W_i + a_1 \sum_{i=1}^{m} W_i x_i = \sum_{i=1}^{m} W_i y_i$$

and

$$a_0 \sum_{i=1}^{m} W_i x_i + a_1 \sum_{i=1}^{m} W_i x_i^2 = \sum_{i=1}^{m} W_i x_i y_i$$

which are the **normal equations in this case** and are solved to obtain a_0 and a_1.

- Suppose that in data, a point (x_0, y_0) is known to be more reliable than the others. Then we prescribe a weight (say, 10) corresponding to this point only and all other weights are taken as unity.
- We consider with an increased weight, say 100, corresponding to (x_0, y_0), then the approximation becomes better when the weight is increased.

Nonlinear Weighted Least Squares Approximation

We now consider the least squares approximation of a set of m data points

$$(x_i, y_i), \quad i = 1, 2, \ldots, m,$$

by a polynomial of degree $n < m$. Let

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

be fitted to the given data points. We then have

$$S(a_0, a_1, \ldots, a_n) = \sum_{i=1}^{m} W_i [y_i - (a_0 + a_1 x_i + \cdots + a_n x_i^n)]^2.$$

If a minimum occurs at $(a_0, a_1, \ldots, a_n)$, then we have

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = \frac{\partial S}{\partial a_2} = \cdots = \frac{\partial S}{\partial a_n} = 0.$$

These conditions yield the normal equations

$$\begin{aligned} a_0 \sum_{i=1}^{m} W_i + a_1 \sum_{i=1}^{m} W_i x_i + \cdots + a_n \sum_{i=1}^{m} W_i x_i^n &= \sum_{i=1}^{m} W_i y_i \\ a_0 \sum_{i=1}^{m} W_i x_i + a_1 \sum_{i=1}^{m} W_i x_i^2 + \cdots + a_n \sum_{i=1}^{m} W_i x_i^{n+1} &= \sum_{i=1}^{m} W_i x_i y_i \\ &\vdots \\ a_0 \sum_{i=1}^{m} W_i x_i^n + a_1 \sum_{i=1}^{m} W_i x_i^{n+1} + \cdots + a_n \sum_{i=1}^{m} W_i x_i^{2n} &= \sum_{i=1}^{m} W_i x_i^n y_i. \end{aligned}$$

There are $(n+1)$ equations in $(n+1)$ unknowns $a_0, a_1, \ldots, a_n$.

If the x_i's are distinct with $n < m$, then the equations possses a 'unique' solution.

Method of Least Squares for Continuous Functions

We considered the least squares approximations of discrete data. We shall discuss the least squares approximation of a continuous function on $[a, b]$.

The summations in the normal equations are now replaced by definite integrals. Let $y(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be chosen to minimize

$$S(a_0, a_1, \ldots, a_n) = \int_a^b W(x)[y(x) - (a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n)]^2 dx.$$

The necessary conditions for minimum are given by

$$\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = \cdots = \frac{\partial S}{\partial a_n} = 0.$$

Hence

$$
\begin{aligned}
-2\int_a^b W(x)[y(x)-(a_0+a_1x+a_2x^2+\cdots+a_nx^n)]dx &= 0\\
-2\int_a^b W(x)[y(x)-(a_0+a_1x+a_2x^2+\cdots+a_nx^n)]xdx &= 0\\
-2\int_a^b W(x)[y(x)-(a_0+a_1x+a_2x^2+\cdots+a_nx^n)]x^2dx &= 0\\
&\vdots\\
-2\int_a^b W(x)[y(x)-(a_0+a_1x+a_2x^2+\cdots+a_nx^n)]x^ndx &= 0
\end{aligned}
$$

Rearrangement of terms gives the system

$$\begin{aligned}
a_0 \int_a^b W(x)dx + a_1 \int_a^b xW(x)dx + \cdots + a_n \int_a^b x^n W(x)dx &= \int_a^b W(x)y(x)dx \\
a_0 \int_a^b xW(x)dx + a_1 \int_a^b x^2 W(x)dx + \cdots + a_n \int_a^b x^{n+1} W(x)dx &= \int_a^b xW(x)y(x)dx \\
&\vdots \\
a_0 \int_a^b x^n W(x)dx + a_1 \int_a^b x^{n+1} W(x)dx + \cdots + a_n \int_a^b x^{2n} W(x)dx &= \int_a^b x^n W(x)y(x)dx
\end{aligned}$$

The system comprises $(n+1)$ normal equations in $(n+1)$ unknowns $a_0, a_1, a_2, \ldots, a_n$ and they always possess a 'unique' solution.

References

- Richard L. Burden and J. Douglas Faires, "*Numerical Analysis – Theory ad Applications*", Cengage Learning, New Delhi, 2005.
- Kendall E. Atkinson, "*An Introduction to Numerical Analysis*", John Wiley & Sons, Delhi, 1989.