Chapter XXI
Spline Fitting

Michael Wodny
Ernst-Moritz-Arndt-University, Germany

ABSTRACT

Given are the m points (x_i,y_i), i=1,2,…,m. Spline functions are introduced, and it is noticed that the interpolation task in the case of natural splines has a unique solution. The interpolating natural cubic spline is constructed. For the construction of smoothing splines, different optimization problems are formulated. A selected problem is looked at in detail. The construction of the solution is carried out in two steps. In the first step the unknown $D_i=s(x_i)$ are calculated via a linear system of equations. The second step is the construction of the interpolating natural cubic spline with respect to these (x_i,D_i), i=1,2,…,m. Every optimization problem contains a smoothing parameter. A method of estimation of the smoothing parameter from the given data is motivated briefly.

INTRODUCTION

Model fitting in data mining requires the attention of at least three aspects: the data, the model to be fitted, and the optimization criterion. This general situation is specified for simplicity as follows. The data are a two-dimensional set of real numbers (x_i,y_i), i = 1, 2, ..., m, and the model is a class of spline functions. Selected optimization criteria are subsequently explained.

Spline functions are a class of functions that is characterized by general mathematical properties, instead of data-driven or problem-driven properties. Their application is widespread. There are multidimensional splines and many mathematical generalizations, too, especially on Hilbert spaces.

Spline fitting involves the calculation of the parameters of the chosen spline function from the given data. The data, the class of the functions, and the optimization criterion determined the calculation method for the parameters of the spline function.

Let (x_i,y_i), i = 1, 2, …, m be given. We assume that $x_1<x_2<x_3<...<x_m$.

DEFINITION

A spline function s(x) of degree n is a function defined on ○. s(x) is given by some polynomial of degree n or less in each of the intervals $(-\infty, x_1]$, $[x_i, x_{i+1}]$, i = 1, 2, …, m-1, and $[x_m, +\infty)$. All derivatives of s(x) up to order n-1 are supposed to be continuous everywhere. s(x) is called a natural spline if n = 2k-1 is odd and s(x) is given in each of the intervals $(-\infty, x_1]$ and $[x_m, +\infty)$ by polynomials of degree k-1 or less.

For example, the class of all splines of degree n withthe knots x_i includes all polynomials of degree n or less.

INTERPOLATION

A special task of curve fitting is the interpolation problem. One looks for a function f(x) satisfying the strong conditions $f(x_i) = y_i$ for all i from 1 to m.

The interpolation problem has a unique solution for natural splines s(x). The main property of natural splines is proved by deBoor (2001) and Schoenberg and deBoor. In the case n = 3, the result was given already by Holladay (1957).

THEOREM

Let s(x) be the interpolating natural spline of degree n = 2k-1, with respect to (x_i, y_i), i = 1, 2, ..., m, and f(x) any interpolating function with continuous derivatives up to order k. Then

$$\int_a^b s^{(k)}(x)^2\,dx \le \int_a^b f^{(k)}(x)^2\,dx \text{ for all } a \le x_1 \text{ and } b \ge x_m.$$

In the case where k > 1, the strong inequality is valid.

We furthermore refer to natural cubic splines because they are widely used. It is possible to generalize the afterward-derived calculation procedure of the parameters of a natural cubic spline to arbitrary natural splines.

Looking at the definition, a cubic spline can be written in the interval $[x_i, x_{i+1}]$ as:

$$s(x) = A_i(x-x_i)^3 + B_i(x-x_i)^3 + C_i(x-x_i) + D_i.$$

Obviously,

$$s(x_i) = D_i = y_i, \ i=1, 2, ..., m-1. \tag{1}$$

Furthermore, it follows that $s''(x_i) = 2B_i$, and so:

$$B_i = s''(x_i)/2. \tag{2}$$

The second derivation s"(x) is a straight line with the slope $(s''(x_{i+1})-s''(x_i))/(x_{i+1}-x_i)$. Consequently, one obtains:

$$A_i = 1/3(B_{i+1}-B_i))/(x_{i+1}-x_i) \tag{3}$$

for all i = 1, 2, …, m-1.

The C_i can be represented with the data and the B_i as:

$$C_i=(y_{i+1}-y_i)/(x_{i+1}-x_i) - 1/3(x_{i+1}-x_i)(2 B_{i+1}+B_i) \tag{4}$$

as seen in Equations (1), (2), and (3).

From these considerations it follows that an interpolating cubic spline is uniquely determined and completely represented by $(x_i, s(x_i))$ and B_i (especially $B_m = s''(x_m)/2$). One uses the continuity of s'(x) in the x_i, i=1, 2, ..., m to calculate the unknown B_i. Consequently:

$$s'(x_i) = 3A_{i-1}(x_i-x_{i-1})^2 + 2B_{i-1}(x_i-x_{i-1}) + C_{i-1} = C_i, \ i = 2, 3, ..., m-1$$

holds true. Short remodeling together with the specifications $\Delta x_i := x_{i+1}-x_i$ and $\Delta y_i := y_{i+1}-y_i$, i = 1, 2, ..., m-1 leads to:

Related Content