Data Analysis-I

Interpolation

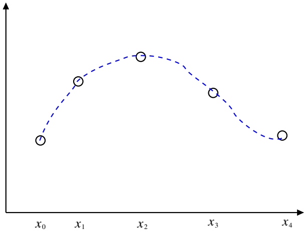
Soon-Hyung Yook

December 4, 2015

Table of Contents

Interplation

Estimate the value between data



- Polynomial interpolation
  - for a set of data $\{(x_i, y_i)\}$
    - no knowledge of the relationship between y and x
    - e.g.: estimate $(x_{i+\delta}, y_{i+\delta})$ from the given dataset.
  - Weierstrass theorem: In general a continuous function $f(x)$ in a finite interval $x \in [a, b]$ can be fitted by a polynomial $P(x)$.
  - Find a polynomial approximation for $f(x)$ from N pairs of numbers $\{(x_i, f(x_i))\}$, for $i = 0, 1, \cdots, N-1$.

Lagrange Interpolation

- Construct a polynomial of the form

$$P(x) = \sum_{k=0}^{N-1} p_k(x) f(x_k) \tag{1}$$

Lagrange Polynomial

$$p_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_{N-1})}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_{N-1})} \tag{2}$$

- If we use the Lagrange polynomial then it is clear that

$$P(x_i) = f(x_i) \tag{3}$$

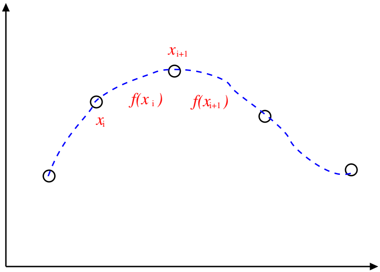
Cubic-Spline

- spline (in dictionary): a piece of "flexible" wood or plastic that can be bent into arbitrary smooth shapes
  - in the days before computers, it was used to trace a smooth curve between points on a sheet of graph paper.
- Idea: use a simple function to approximate the relation between the dependent and independent variables
  - use the third-degree polynomial : "cubic"-spline

$$f(x) = a + bx + cx^2 + dx^3 \tag{4}$$

- To find the coefficients in each interval
  - use the condition for the continuity across each boundary

Cubic-Spline



continuity condition:

$$f_i^{(n)}(x_i) = f_{i+1}^{(n)}(x_{i+1}) \qquad \text{for } n\text{th derivative}$$

Cubic-Spline

Conditions for cubic spline

Given a dataset $\{(x_i, y_i)\}$ in the interval $[a, b]$ ($a = x_0 < x_1 < \cdots < x_n = b$), a cubic spline interpolant, f, for $\{y_i\}$ is a function that satisfies the following condition:

(a) f is a cubic polynomial, denoted by f_i, on the subinterval $[x_i, x_{i+1}]$.

(b) $f(x_i) = y_i$ for each i.

(c) $f_{i+1}(x_{i+1}) = f_i(x_{i+1})$ for each i.

(d) $f'_{i+1}(x_{i+1}) = f'_i(x_{i+1})$ for each i.

(e) $f''_{i+1}(x_{i+1}) = f''_i(x_{i+1})$ for each i.

(f) One of the following sets of boundary condition is frequently used:

- (i) $f''(x_0) = f''(x_n)$ for natural or free boundary
- (ii) $f'(x_0) = y'(x_0)$ and $f'(x_n) = y'(x_n)$ for clamped boundary
- * Other boundary conditions are also possible.

Cubic-Spline

- To construct the cubic-spline interpolant for a ginven set of data $\{y_i\}$, we assume a third-order polynomial

$$f_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \tag{5}$$

for all i.

- from condition (b) and Eq. (5):

$$f_i(x_i) = a_i = y_i \tag{6}$$

- from condition (c): $a_{i+1} = f_{i+1}(x_{i+1}) = f_i(x_{i+1})$.
  - and if we let $h_i \equiv x_{i+1} - x_i$ then we obtain

$$a_{i+1} = a_i + b_i h_i + c_i h_i^2 + d_i h_i^3 \tag{7}$$

- From Eq. (5)

$$b_i = f_i'(x_i) \tag{8}$$

Cubic-Spline

- From condition (d) and $f_i'(x) = b_i + 2c_i(x - x_i) + 3d_i(x - x_i)^2$:

$$b_{i+1} = b_i + 2c_i h_i + 3d_i h_i^2 \tag{9}$$

- From condition (e):

$$c_i = f''(x_i)/2 \tag{10}$$

and

$$c_{i+1} = c_i + 3d_j h_j \tag{11}$$

Cubic-Spline

Solving for d_i in Eq. (11) and substituting d_i into Eq. (7) and Eq. (9)

$$a_{i+i} = a_i + b_i h_i + \frac{h_i^3}{3}(2c_i + c_{j+1}) \tag{12}$$

and

$$b_{i+i} = b_i + h_i(c_i + c_{i+1}) \tag{13}$$

- Note: Eq. (8) and Eq. (10) do not give us any numerical information to determine $\{b_i\}$ and $\{c_i\}$.
- We cannot determine $\{b_i\}$ and $\{c_i\}$ from Eq. (8) and Eq. (10).
- Therefore, we have to do more things!

Cubic-Spline

The final relationship between the coefficients is obtained by first solving the appropriate equation in the form of Eq. (12) for b_i. From Eq. (12)

$$b_i = \frac{1}{h}(a_{i+1} - a_i) - \frac{h_i}{3}(2c_i + c_{i+1}) \tag{14}$$

and then, with reduction of the index ($i \to i-1$):

$$b_{i-1} = \frac{1}{h}(a_i - a_{i-1}) - \frac{h_{i-1}}{3}(2c_{i-1} + c_i) \tag{15}$$

Substituting Eq. (14) and Eq. (15) into Eq. (13), when the index is reduced by 1 ($i \to i-1$), we obtain the linear system of equations:

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = \frac{3}{h_i}(a_{i+1} - a_i) - \frac{3}{h_{i-1}}(a_i - a_{i-1}), \tag{16}$$

for each $i = 1, 2, \cdots, n-1$.

Cubic-Spline

- In Eq. (16), only $\{c_i\}$ is a set of unknown parameters.
- In Eq. (16), $\{h_i\}$ and $\{a_i\}$ are given by the spacing between each data point and the value of $\{y_i\}$.
- Once the values of $\{c_i\}$ is obtained, we can easily find $\{b_i\}$ from Eq. (14).
- Then we can obtain $\{d_i\}$ from Eq. (11).
- Eq. (16) is an n-coupled linear equation.
    - solve Eq. (16) using matrix.