STATS 7022 – Data Science PG Assignment 2

Dang Thinh Nguyen

2024-08-07

Question 2: Piecewise Cubic Splines

(a).

Each segment between knots can be described by a cubic polynomial:

$$f_i(x_j) = \beta_{0i} + \beta_{1i}x_j + \beta_{2i}x_j^2 + \beta_{3i}x_j^3 + \epsilon_j$$

A cubic spline with k knots creates $k + 1$ polynomial segments, so:

$$i = 1, 2, \dots, k + 1$$

Each segment has 4 parameters ($\beta_{0i}, \beta_{1i}, \beta_{2i}, \beta_{3i}$), so a cubic spline has:

$$4 \times (k + 1) = 4k + 4 \text{ (parameters)}$$

If we denote knots as $x_1, x_2, \dots, x_k$, the cubic polynomial segments for intervals:

$$[x_0, x_1], [x_1, x_2], \dots, [x_{k-1}, x_k]$$

where x_0 and x_k are the endpoints of the entire range.

At each knot x_j:

$$f_i(x_j) = \beta_{0i} + \beta_{1i}x_j + \beta_{2i}x_j^2 + \beta_{3i}x_j^3$$

there is a requirement that the spline function $f(x)$ be continuous. This means the value of the left of the knot must equal the value from the right of the knot.

Since there are k internal knots, this results in k continuity constraints for the spline function itself. Each of these constraints ensures that the value of the spline matches at both sides of each knot.

The spline is continuous up to (and including) the second order derivative:

$$f_i'(x_j) = \beta_{1i} + 2\beta_{2i}x_j + 3\beta_{3i}x_j^2$$

$$f_i''(x_j) = 2\beta_{2i} + 6\beta_{3i}x_j$$

each derivative also has k continuity constraints, similar to the spline function.

Thus, there are $3k$ continuity constraints in total.

After accounting for the constraints, the number of parameters needed to describe this piecewise cubic is:

$$4k + 4 - 3k = k + 4 \text{ (parameters)}$$

(b).

From the linear combination of basis functions:

$$h_1(x) = 1$$

$$h_2(x) = x$$

$$h_3(x) = x^2$$

$$h_4(x) = x^3$$

$$h_5(x) = (x - \xi)_+^3$$

The cubic spline function is given by:

$$h(x) = t_1 h_1(x) + t_2 h_2(x) + t_3 h_3(x) + t_4 h_4(x) + t_5 h_5(x)$$

Let $h_a(x)$ is the spline for $x < \xi$:

$h_5(x) = 0$

$h_a(x) = t_1 h_1(x) + t_2 h_2(x) + t_3 h_3(x) + t_4 h_4(x)$

$\quad = t_1 + t_2 x + t_3 x^2 + t_4 x^3$

which matches the form of $f(x)$ in method 2 where

$$a_1 = t_1$$

$$a_2 = t_2$$

$$a_3 = t_3$$

$$a_4 = t_4$$

So: $h_a'(x) = t_2 + 2t_3 x + 3t_4 x^2$

$\quad h_a''(x) = 2t_3 + 6t_4 x$

match the form of $f'(x)$ and $f''(x)$ in method 2

Let $h_b(x)$ is the spline for $x \geq \xi$:

$h_5(x) = (x - \xi)^3 = x^3 - 3x^2\xi + 3x\xi^2 - \xi^3$

$h_b(x) = t_1 h_1(x) + t_2 h_2(x) + t_3 h_3(x) + t_4 h_4(x) + t_5 h_5(x)$

$\quad = t_1 + t_2 x + t_3 x^2 + t_4 x^3 + t_5(x^3 - 3x^2\xi + 3x\xi^2 - \xi^3)$

$\quad = (t_1 - t_5\xi^3) + (t_2 + 3t_5\xi^2)x + (t_3 - 3t_5\xi)x^2 + (t_4 + t_5)x^3$

which matches the form of $g(x)$ in method 2 where

$$a_1 = t_1 - t_5\xi^3$$

$$a_2 = t_2 + 3t_5\xi^2$$

$$a_3 = t_3 - 3t_5$$

$$a_4 = t_4 + t_5$$

So: $h_b'(x) = (t_2 + 3t_5\xi^2) + 2(t_3 - 3t_5\xi)x + 3(t_4 + t_5)x^2$

$h_b''(x) = 2(t_3 - 3t_5\xi) + 6(t_4 + t_5)x$

match the form of $g'(x)$ and $g''(x)$ in method 2

At $x = \xi$:

$h_5(x) = (x - \xi)^3 = 0$

$h_b(\xi) = t_1h_1(\xi) + t_2h_2(\xi) + t_3h_3(\xi) + t_4h_4(\xi)$

$= t_1 + t_2\xi + t_3\xi^2 + t_4\xi^3$

Because $h_a(x)$ is defined from $x < \xi$, at $x = \xi$:

$$\lim_{x \to \xi} h_a(x) = \lim_{x \to \xi}\big(t_1h_1(x) + t_2h_2(x) + t_3h_3(x) + t_4h_4(x) + t_5h_5(x)\big)$$

$$= \lim_{x \to \xi}(t_1 + t_2x + t_3x^2 + t_4x^3 + t_5(x - \xi)^3)$$

$$= t_1 + t_2\xi + t_3\xi^2 + t_4\xi^3$$

$\Longrightarrow h_a(\xi) = h_b(\xi)$

$\Longrightarrow g(\xi) = f(\xi)$

$h_b'(\xi) = t_2 + 2t_3\xi + 3t_4\xi^2$

$$\lim_{x \to \xi} h_a'(x) = \lim_{x \to \xi}\big(t_1h_1(x) + t_2h_2(x) + t_3h_3(x) + t_4h_4(x) + t_5h_5(x)\big)'$$

$$= \lim_{x \to \xi}(t_1 + t_2x + t_3x^2 + t_4x^3 + t_5(x - \xi)^3)'$$

$$= \lim_{x \to \xi}(t_2 + 2t_3x + 3t_4x^2 + 3t_5(x - \xi)^2)$$

$$= t_2 + 2t_3\xi + 3t_4\xi^2$$

$\Longrightarrow h_a'(\xi) = h_b'(\xi)$

$\Longrightarrow g'(\xi) = f'(\xi)$

$h_b''(\xi) = 2t_3 + 6t_4\xi$

$$\lim_{x \to \xi} h_a''(x) = \lim_{x \to \xi}\big(t_1h_1(x) + t_2h_2(x) + t_3h_3(x) + t_4h_4(x) + t_5h_5(x)\big)''$$

$$= \lim_{x \to \xi}(t_1 + t_2x + t_3x^2 + t_4x^3 + t_5(x - \xi)^3)''$$

$$= \lim_{x \to \xi}\big(2t_3 + 6t_4x + 6t_5(x - \xi)\big)$$

$$= 2t_3 + 6t_4\xi$$

$\Longrightarrow h_a''(\xi) = h_b''(\xi)$

$\Longrightarrow g''(\xi) = f''(\xi)$

Thus, Method 1 implies Method 2.