Assignment #3

Elements of Machine Learning

Saarland University – Winter Semester 2024/25

Rabin Adhikari
7072310
raad00002@stud.uni-saarland.de

Dhimitrios Duka
7059153
dhdu00001@stud.uni-saarland.de

3 Problem 3 (Beyond linearity: Polynomial and Splines)

3.1 Cubic Regression Spline with One Knot

Cubic regression spline with one knot at ξ can be obtained using a basis of the form $x, x^2, x^3, (x-\xi)_+^3$, where $(x-\xi)_+^3 = (x-\xi)^3$ if $x > \xi$ and equals 0 otherwise. We can show that a function of the following form is indeed a cubic regression spline, regardless of the values of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x-\xi)_+^3$$

3.1.1 Find a cubic polynomial $f_1(x)$

Answer: Find a cubic polynomial

$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$

such that $f(x) = f_1(x)$ for all $x \le \xi$. Express a_1, b_1, c_1, d_1 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

Answer: When $x \le \xi$, $(x-\xi)_+^3 = 0$, so $f(x)$ can be written as follows.

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Comparing the above coefficients with that of $f_1(x)$, we can write the following.

$$\begin{aligned} a_1 &= \beta_0 \\ b_1 &= \beta_1 \\ c_1 &= \beta_2 \\ d_1 &= \beta_3 \end{aligned}$$

3.1.2 Find a cubic polynomial $f_2(x)$

Question: Find a cubic polynomial

$$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$$

such that $f(x) = f_2(x)$ for all $x > \xi$. Express a_2, b_2, c_2, d_2 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$. We have now established that $f(x)$ is a piecewise polynomial.

Answer: When $x > \xi$, $(x-\xi)_+^3 = (x-\xi)^3$, so $f(x)$ can be written as follows.

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x-\xi)^3$$

Also, we know $(x-\xi)^3 = x^3 - 3x^2\xi + 3x\xi^2 - \xi^3$. So, the above equation can be expanded as follows.

$$\begin{aligned} f(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 \left(x^3 - 3x^2\xi + 3x\xi^2 - \xi^3\right) \\ &= \beta_0 - \beta_4\xi^3 + \left(\beta_1 + 3\beta_4\xi^2\right) x + \left(\beta_2 - 3\beta_4\xi\right) x^2 + \left(\beta_3 + \beta_4\right) x^3 \end{aligned}$$

Comparing the above coefficients with that of $f_2(x)$, we can write the following.

$$\begin{aligned} a_2 &= \beta_0 - \beta_4\xi^3 \\ b_2 &= \beta_1 + 3\beta_4\xi^2 \\ c_2 &= \beta_2 - 3\beta_4\xi \\ d_2 &= \beta_3 + \beta_4 \end{aligned}$$

3.1.3 Continuity at ξ

Question: Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at ξ.

Answer: First, $f_1(\xi)$ can be written as follows.

$$f_1(x) = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3$$

Now, $f_2(\xi)$ can be written as follows.

$$\begin{aligned} f_2(x) &= \beta_0 - \beta_4\xi^3 + \left(\beta_1 + 3\beta_4\xi^2\right)\xi + \left(\beta_2 - 3\beta_4\xi\right)\xi^2 + \left(\beta_3 + \beta_4\right)\xi^3 \\ &= \beta_0 - \beta_4\xi^3 + \beta_1\xi + 3\beta_4\xi^3 + \beta_2\xi^2 - 3\beta_4\xi^3 + \beta_3\xi^3 + \beta_4\xi^3 \\ &= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 \\ &= f_1(x) \end{aligned}$$

3.1.4 First Derivative Continuity at ξ

Question: Show that $f_1'(\xi) = f_2'(\xi)$. That is, $f'(x)$ is continuous at ξ.

Answer: First, $f_1'(x)$ can be written as follows.

$$f_1'(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$$

Therefore, $f_1'(\xi)$ is,

$$f_1'(x) = \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2$$

Also, $f_2'(x)$ can be written as follows.

$$f_2'(x) = \beta_1 + 3\beta_4\xi^2 + 2\left(\beta_2 - 3\beta_4\xi\right)x + 3\left(\beta_3 + \beta_4\right)x^2$$

Therefore, $f_2'(\xi)$ is,

$$\begin{aligned} f_2'(x) &= \beta_1 + 3\beta_4\xi^2 + 2\left(\beta_2 - 3\beta_4\xi\right)\xi + 3\left(\beta_3 + \beta_4\right)\xi^2 \\ &= \beta_1 + 3\beta_4\xi^2 + 2\beta_2\xi - 6\beta_4\xi^2 + 3\beta_3\xi^2 + 3\beta_4\xi^2 \\ &= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 \\ &= f_1'(x) \end{aligned}$$

3.1.5 Second Derivative Continuity at ξ

Question: Show that $f_1''(\xi) = f_2''(\xi)$. That is, $f''(x)$ is continuous at ξ. Therefore, $f(x)$ is indeed a cubic spline.

Answer: First, $f_1''(x)$ can be written as follows.

$$f_1''(x) = 2\beta_2 + 6\beta_3 x$$

Therefore, $f_1'(\xi)$ is,

$$f_1''(x) = 2\beta_2 + 6\beta_3 \xi$$

Also, $f_2''(x)$ can be written as follows.

$$f_2''(x) = 2(\beta_2 - 3\beta_4\xi) + 6(\beta_3 + \beta_4)x$$

Therefore, $f_2''(\xi)$ is,

$$\begin{aligned} f_2''(x) &= 2(\beta_2 - 3\beta_4\xi) + 6(\beta_3 + \beta_4)\xi \\ &= 2\beta_2 - 6\beta_4\xi + 6\beta_3\xi + 6\beta_4\xi \\ &= 2\beta_2 + 6\beta_3\xi \\ &= f_1''(x) \end{aligned}$$

3.2 Comparing Smoothing Splines

Consider two curves, $\hat{g}_1$ and $\hat{g}_2$, defined by

$$\begin{aligned} \hat{g}_1 &= \arg\min_g \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(3)}(x) \right]^2 dx \right) \\ \hat{g}_2 &= \arg\min_g \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(4)}(x) \right]^2 dx \right) \end{aligned}$$

where $g^{(m)}$ represents the m-th derivative of g.

3.2.1 Training RSS as $\lambda \to \infty$

Question: As $\lambda \to \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training RSS?

Answer: When $\lambda \to \infty$, the minimization would want to make the second term minimum. The minimum value attainable by the term is zero and the value of $g^{(m)}$ would be zero for $(m-1)^{th}$ polynomial. So, when $\lambda \to \infty$, $\hat{g}_1$ would be a quadratic polynomial and $\hat{g}_2$ would be the cubic one. Since $\hat{g}_2$ has more capacity than $\hat{g}_1$, there is a high chance that $\hat{g}_2$ would have a smaller training RSS.

3.2.2 Test RSS as $\lambda \to \infty$

Question: As $\lambda \to \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller test RSS?

Answer: Since $\hat{g}_2$ is a quadratic polynomial and $\hat{g}_1$ is a cubic polynomial, there may be some cases where $\hat{g}_2$ is overfitting the data and $\hat{g}_1$ is not. However, we can't be sure until we have some knowledge about that underlying function. In a nutshell, the RSS in the test set depends on the underlying function.

3.2.3 RSS for $\lambda = 0$

Question: For $\lambda = 0$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training RSS and test RSS?

Answer: When $\lambda = 0$, both the equations just minimize the RSS on the training set resulting in the same model since the nature of the curves is identical. Hence, the test RSS would also be the same.