Machine Learning

Cubic Splines

Irving Gómez Méndez

August-December, 2021

UNIVERSIDAD IBEROAMERICANA
CIUDAD DE MÉXICO ®

Introduction

It is difficult that a phenomenon can be modeled by a polynomial inn the whole range of its domain. In several cases, it is more reasonable to think that in some range of values of the independent variable the behavior follows a polynomial relation while in other range it can be modeled by another polynomial. For example, suppose that we have the data $\mathcal{D}_n = (X_1, Y_1), \dots (X_{30}, Y_{30})$ where

$$\begin{aligned} 0 &\leq X_1, \dots, X_{10} \leq 1, \\ 1 &< X_{11}, \dots, X_{20} \leq 2, \\ 2 &< X_{21}, \dots, X_{30} \leq 3. \end{aligned}$$

We can adjust the piecewise linear model:

$$\mathbb{E}[Y|X] = \begin{cases} a + bX & \text{if } 0 \leq X \leq 1 \\ c + dX & \text{if } 1 < X \leq 2 \\ e + fX & \text{if } 2 < X \leq 3 \end{cases}$$

This model can be written in the form $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, with

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & X_{11} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & X_{20} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & X_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & X_{30} \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}$$

The previous model is unacceptable because if we approach $X = 1$ from the left the model predicts some value, while if we approach $X = 1$ from the right, the model predicts a different valur for Y.
This model has a lack of continuity. To glue the lines in $X = 1$ and $X = 2$ (these points are called knots), the next equations must be satisfied:

$$a + k_1 b = c + k_1 d$$
$$c + k_2 d = e + k_2 f$$

where k_1 and k_2 are the knots.

In other words, we need that the parameters of the model satisfy a system of linear equations (2 equations and 6 variables):

$$\begin{bmatrix} 1 & k_1 & -1 & -k_1 & 0 & 0 \\ 0 & 0 & 1 & k_2 & -1 & -k_2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \equiv \quad \mathbf{K}\beta = \mathbf{0}$$

Thus, to estimate the parameters of the model we need to solve the next optimization problem with restrictions:

$$\min_{\beta} (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) \quad \text{subject to} \quad \mathbf{K}\beta = \mathbf{0}$$

The solution to this problem can be obtained through the method of Lagrange multipliers, given by:

$$\tilde{\beta} = \hat{\beta} - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{K}^T\left[\mathbf{K}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{K}^T\right]^{-1}\mathbf{K}\hat{\beta},$$

where $\hat{\beta}$ is OLS estimator.

We can consider a piecewise cubic model satisfying the continuity conditions, that is, we want to adjust the model:

$$\mathbb{E}(Y|X) = \begin{cases} a + bX + cX^2 + dX^3 & \text{if } 0 \leq X \leq 1 \\ e + fX + gX^2 + hX^3 & \text{if } 1 < X \leq 2 \\ i + jX + kX^2 + lX^3 & \text{if } 2 < X \leq 3 \end{cases}$$

This model can be written in the form $\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$, with

$$
\mathbf{X} = \begin{bmatrix}
1 & X_1 & X_1^2 & X_1^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & X_{10} & X_{10}^2 & X_{10}^3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & X_{11} & X_{11}^2 & X_{11}^3 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 1 & X_{20} & X_{20}^2 & X_{20}^3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{21} & X_{21}^2 & X_{21}^3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{30} & X_{30}^2 & X_{30}^3
\end{bmatrix}
$$

And

$$\beta = \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \\ k \\ l \end{bmatrix}$$

Since we want a continuous model, the equations to satisfied are:

$$a + bk_1 + ck_1^2 + dk_1^3 = e + fk_1 + gk_1^2 + hk_1^3,$$
$$e + fk_2 + gk_2^2 + hk_2^3 = i + jk_2 + kk_2^2 + lk_2^3,$$

which, can be written as

$$\begin{bmatrix} 1 & k_1 & k_1^2 & k_1^3 & -1 & -k_1 & -k_1^2 & -k_1^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & k_2 & k_2^2 & k_2^3 & -1 & -k_2 & -k_2^2 & -k_3^3 \end{bmatrix} \beta_{12\times 1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\equiv \mathbf{K}\beta = \mathbf{0}$$

where k_1 and k_2 are the knots.
We want to estimate β as the value that minimizes the sum of squared residuals, $SSR(\beta) = (\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta)$ subject to $\mathbf{K}\beta = \mathbf{0}$.

Even while we have continuity, the model has a lack of smoothness in the knots, hence we can ask the first derivatives to be continuous in the knows.

The derivative of the first cubic polynomial is $b + 2cx + 3dx^2$, which must be equal to the derivative of the second cubic polynomial, $f + 2gx + 3hx^2$ in $x = k_1$. Similarly, in the second knot, we want the derivatives to be equal.

These conditions are translated into the equations:

$$b + 2ck_1 + 3dk_1^2 = f + 2gk_1 + 3hk_1^2$$
$$f + 2gk_2 + 3hk_2^2 = j + 2kk_2 + 3lk_2^2$$

Thus, the total of restrictions are:

$$\begin{bmatrix} 1 & k_1 & k_1^2 & k_1^3 & -1 & -k_1 & -k_1^2 & -k_1^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & k_2 & k_2^2 & k_2^3 & -1 & -k_2 & -k_2^2 & -k_3^3 \\ 0 & 1 & 2k_1 & 3k_1^3 & 0 & -1 & -2k_1 & -3k_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2k_2 & 3k_2^3 & 0 & -1 & -2k_2 & -3k_2^2 \end{bmatrix} \beta_{12\times 1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{K}\beta = \mathbf{0}$$

One more condition. We add continuity in the second derivative, which would make the model less "curve". That is, we want to add the conditions:

$$2c + 6dk_1 = 2g + 6hk_1$$
$$2g + 6hk_2 = 2k + 6lk_2$$

Adding this conditions we the so-called **Cubic Spline** model.