7.4 Regression Splines

Zongyi Liu

2023-06-04

7.4 Regression Splines

7.4.1 Piecewise Polynomials

When fitting a high-degree polynomial, we do the fitting over the entire range of X, however, when doing the piecewise polynomial regression, we fit separate low-degree polynomials over different regions of X. For example, we can fit a piecewise cubic regression model as below:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \epsilon_i,$$

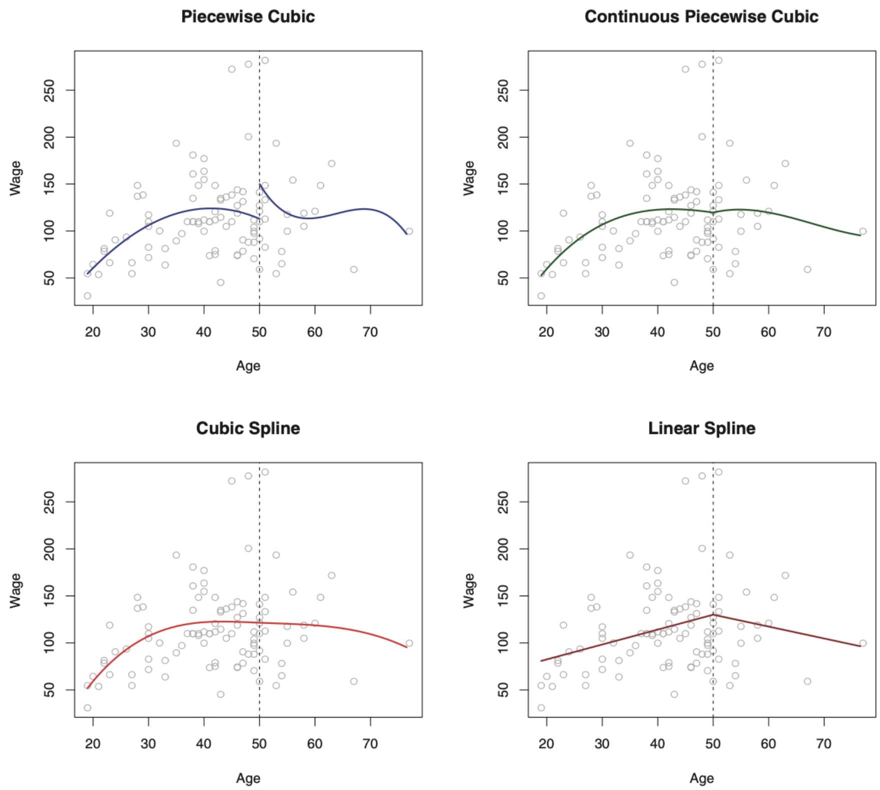
Here the coefficients β_0, β_1, β_2, and β_3 differ in different part of the range of X. The points where the coefficients change are called **knots**.

A piecewise cubic polynomial with a single knot at a point c can be written as:

$$y_i = \begin{cases} \beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2 + \beta_{31}x_i^3 + \epsilon_i & \text{if } x_i < c; \\ \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2 + \beta_{32}x_i^3 + \epsilon_i & \text{if } x_i \geq c. \end{cases}$$

7.4.2 Constraints and Splines

There are many kinds of piecewise polynomials that can be fitted to the `Wage` data, with a knot at `age=50`.



- Top Left: undcontrained
- Top Right: constrained to be continuous at `age=50`
- Bottom Left: constrained to be continuous, and to have continuous first and second derivatives.
- Bottom Right: constrained to be continuous and it's also linear

7.4.3 The Spline Basis Representation

The regression splines that we saw in the previous sections seems to be complex.

A cubic spline with K knots can be modeled as

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_{K+3} b_{K+3}(x_i) + \epsilon_i,$$

for an appropriate choice of basis functions b_1, b_2,..., b_{K+3}. The model can be fit using least squares. Just as there were ways to represent polynomials, there are also many equivalent ways to represent cubic spines using different choices of basis functions.

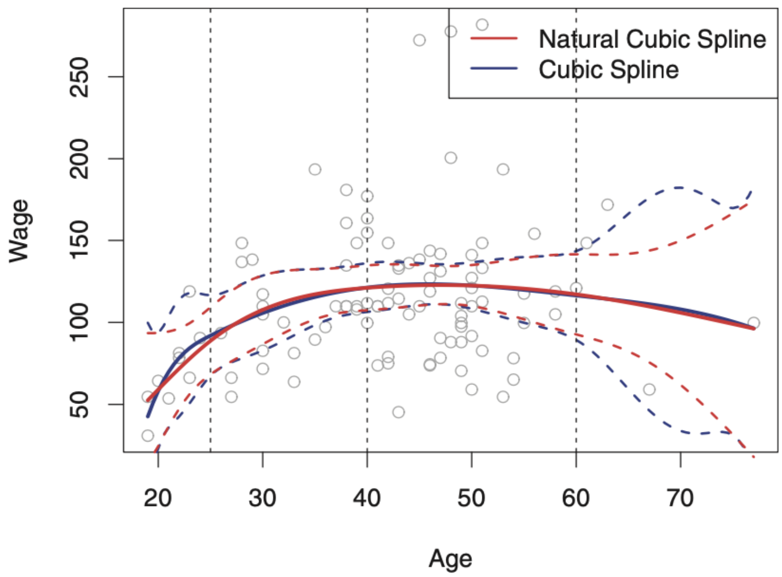
A truncated power basis function is defined as

$$h(x, \xi) = (x - \xi)^3_+ = \begin{cases} (x-\xi)^3 & \text{if } x > \xi \\ 0 & \text{otherwise,} \end{cases}$$

where ϵ is the knot. We can show that adding a term of the form $\beta_4 h(x, \epsilon)$ to the model above for a cubic polynomial will led to a discontinuity in only the third derivative ar ϵ; the function will remain continuous with continuous first and second derivatives at each of the knots.

Thus, in order to fit a cubic splines to a data set with K knots, we perform least squares squares regression with an intercept and $3+K$ predictors, of the form X, X^2, $X3$, $h(X, \epsilon_1), h(X, \epsilon_2), ..., h(X, \epsilon_3)$ where $\epsilon_1, ..., \epsilon_K$

are the knots. This amounts to estimating a total of $K+4$ regression coefficients; for this reason, fitting a cubic spline with K knots uses $K+4$ dof.

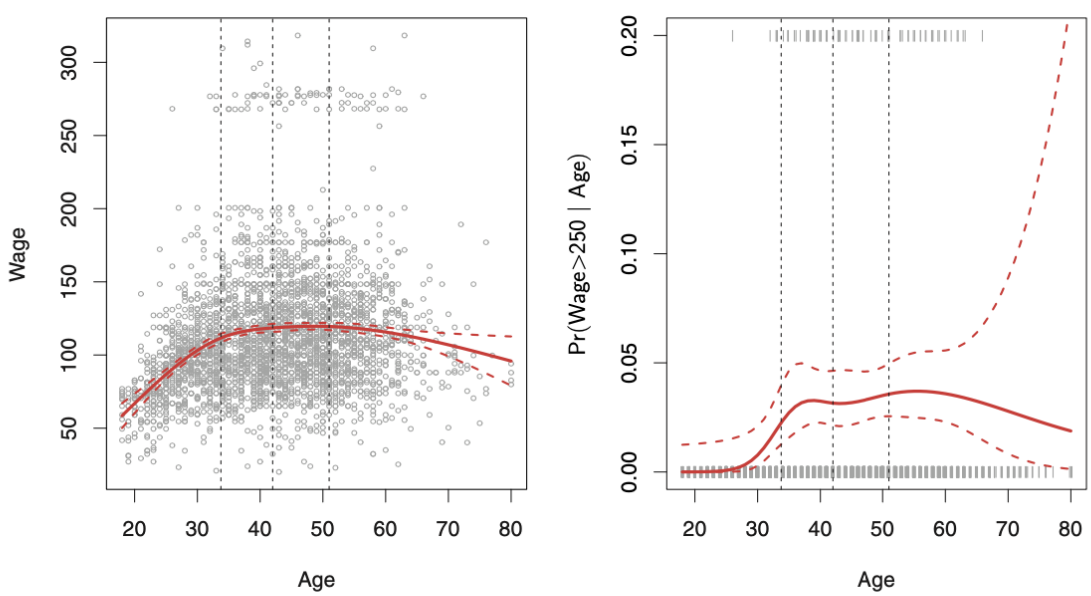


However, splines can have high variance at the outer range of the predictors, when X takes on either a very small or large value. In the figure above, we can see how to fit to the `Wage` data with three knots. We can see that the confidence bands in the boundary region appear fairly wild.

A natural spline is a regression spline with additional boundary constraints: the function is required to be linear at the boundary. This additional constraint means that natural splines generally produce more stable estimates at the boundaries.

7.4.4 Choosing the Number and Locations of the Knots

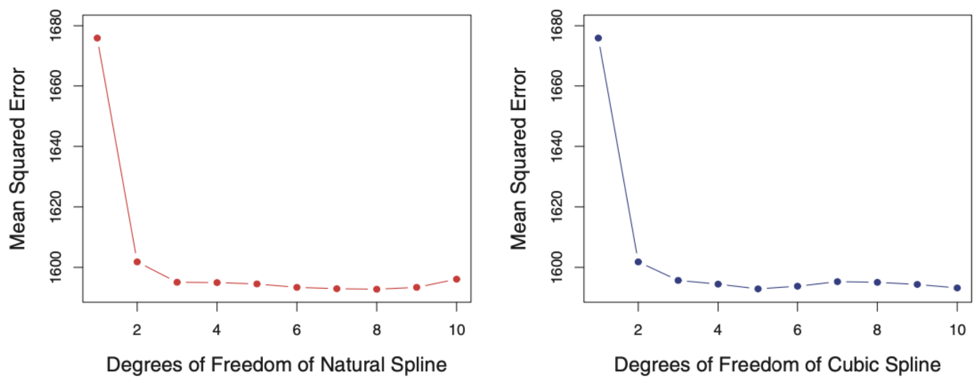
When we fit a spline, we should consider where to place the knots.



A natural cubic spline function with four degrees of freedom is fit to the `Wage` data.

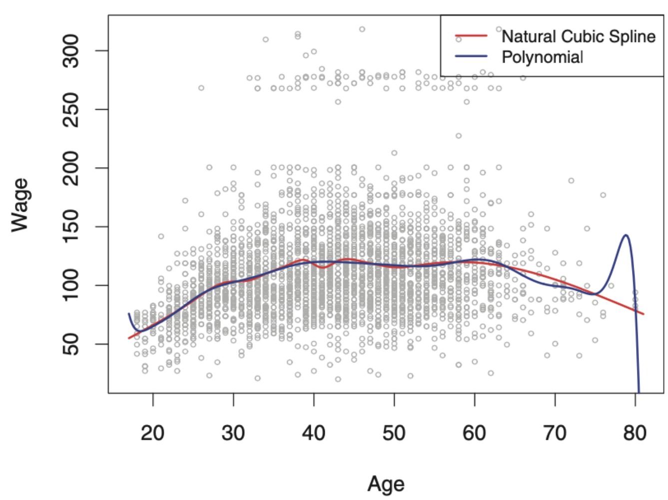
- Left: A spline is fit to `wage` (in thousands of dollars) as a function of `age`.
- Right: Logistic regression is used to model the binary event `wage>250` as a function of `age`. The fitted posterior probability of `wage` exceeding $250,000 is shown.

To decide how many knots should we use: try out different numbers of knots and see which produces the best looking curve. A somewhat more objective approach is to use cross-validation, as discussed in before. With this method, we remove a portion of the data (say 10 %), fit a spline with a certain number of knots to the remaining data, and then use the spline to make predictions for the held-out portion. We repeat this process multiple times until each observation has been left out once, and then compute the overall cross-validated RSS. This procedure can be repeated for different numbers of knots K. Then the value of K giving the smallest RSS is chosen.



7.4.5 Comparison to Polynomial Regression

Regression splines often give superior results to polynomial regression. This is because unlike polynomials, which must use a high degree to produce flexible fits, splines introduce flexibility by increasing the number of knots but keeping the degree fixed.



On the `Wage` data set, a natural cubic spline with 15 degrees of freedom is compared to a degree-15 polynomial. Polynomials can show wild behavior, especially near the tails.