A piecewise cubic polynomial with a single knot at x_0:

(1) $$y = \begin{cases} a + bx + cx^2 + dx^3 & x < x_0 \\ e + fx + gx^2 + hx^3 & x \geq x_0 \end{cases}$$

8 unknown parameters
or 8 degrees of freedom

In order for (1) to be a cubic spline, we need to have

1. continuity at x_0:

$$a + bx_0 + cx_0^2 + dx_0^3 = e + fx_0 + gx_0^2 + hx_0^3$$

2. continuity of the first derivative at x_0:

$$b + 2cx_0 + 3dx_0^2 = f + 2gx_0 + 3hx_0^2$$

3. continuity of the second derivative at x_0:

$$2c + 6dx_0 = 2g + 6hx_0$$

For this example we have 5 degrees of freedom.

- The general definition of a degree-d spline is that it is a piecewise degree-d polynomial, with continuity in derivatives up to degree $d-1$ at each knot.
- Therefore, a linear spline is obtained by fitting a line in each region of the predictor space defined by the knots, requiring continuity at each knot.
- How can we fit a piecewise degree-d polynomial under the constraint that it (and possibly its first $d-1$ derivatives) be continuous?

we can use the basis model to represent a regression spline.

- A cubic spline with K knots can be modeled as

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

for an appropriate choice of basis functions $b_1, b_2, \ldots, b_{K+3}$. The model can then be fit using least squares.

In general, a cubic spline with K knots uses a total of 4+K degrees of freedom.

- We can represent this model with truncated power basis functions

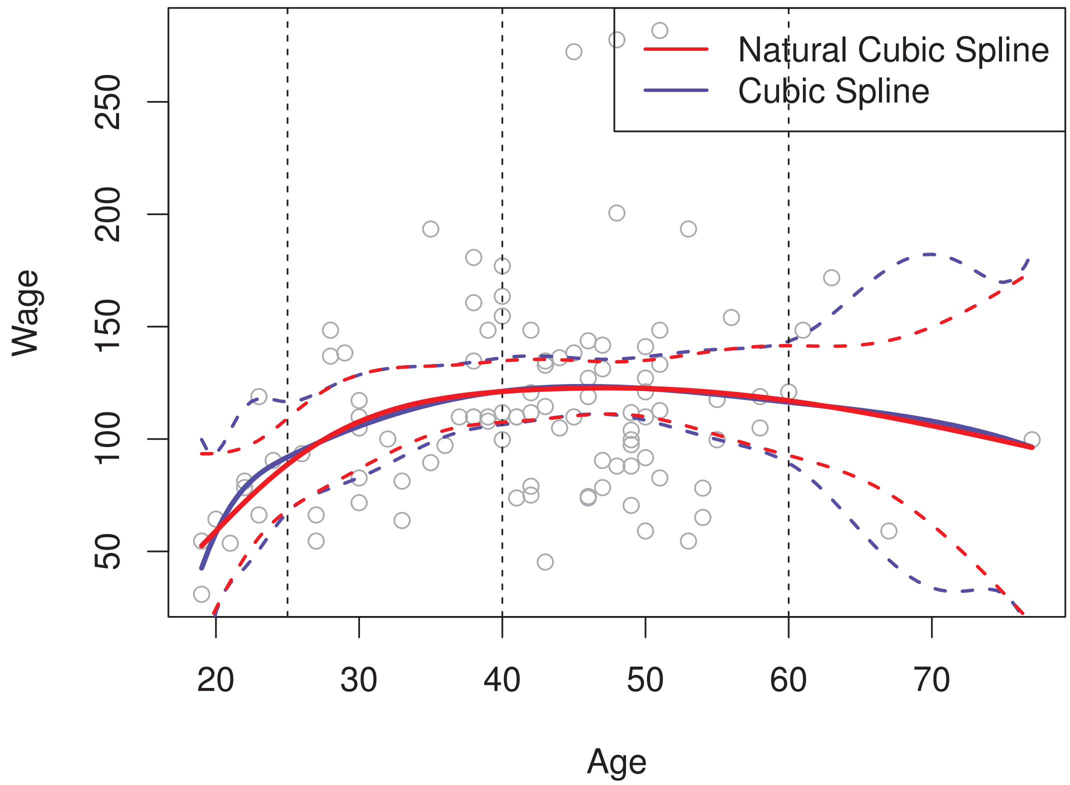
$$\begin{aligned} b_1(x_i) &= x_i \\ b_2(x_i) &= x_i^2 \\ b_3(x_i) &= x_i^3 \\ b_{k+3}(x_i) &= (x_i - \xi_k)_+^3,\ k = 1, \ldots, K \end{aligned}$$

where ξ_k, $k = 1, \ldots, K$, are the knots.

$$(x_i - \xi_k)_+^3 = \begin{cases} (x_i - \xi_k)^3 & x_i > \xi_k \\ 0 & \text{otherwise} \end{cases}$$

is the truncated power basis function.

- Unfortunately, splines can have high variance at the outer range of the predictors—that is, when X takes on either a very small or very large value.



A cubic spline and a natural cubic spline, with three knots, fit to a subset of the Wage data. The dashed lines denote the knot locations.

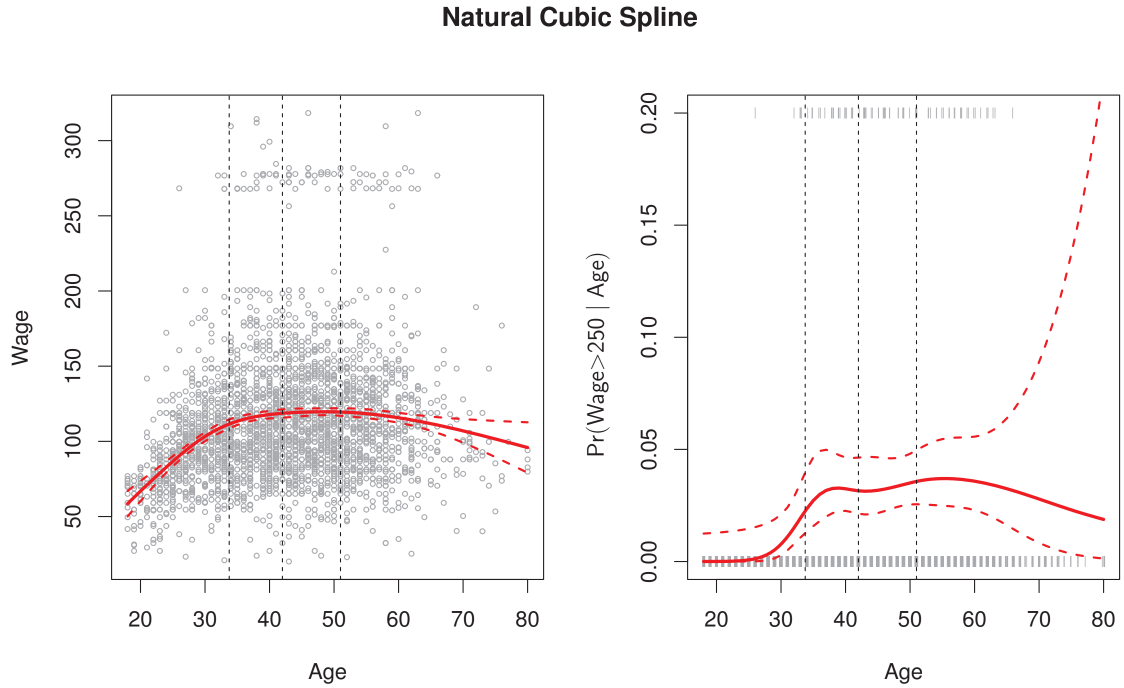
- A natural spline is a regression spline with additional boundary constraints: the function is required to be linear at the boundary (in the region where X is smaller than the smallest knot, or larger than the largest knot).

- This additional constraint means that natural splines generally produce more stable estimates at the boundaries.

Choosing the Number and Locations of the Knots:

- When we fit a spline, where should we place the knots?
- The regression spline is most flexible in regions that contain a lot of knots, because in those regions the polynomial coefficients can change rapidly.
- Hence, one option is to place more knots in places where we feel the function might vary most rapidly, and to place fewer knots where it seems more stable.
- While this option can work well, in practice it is common to place knots in a uniform fashion.
- One way to do this is to specify the desired degrees of freedom, and then have the software automatically place the corresponding number of knots at uniform quantiles of the data.

A natural cubic spline with K knots has K degrees of freedom.

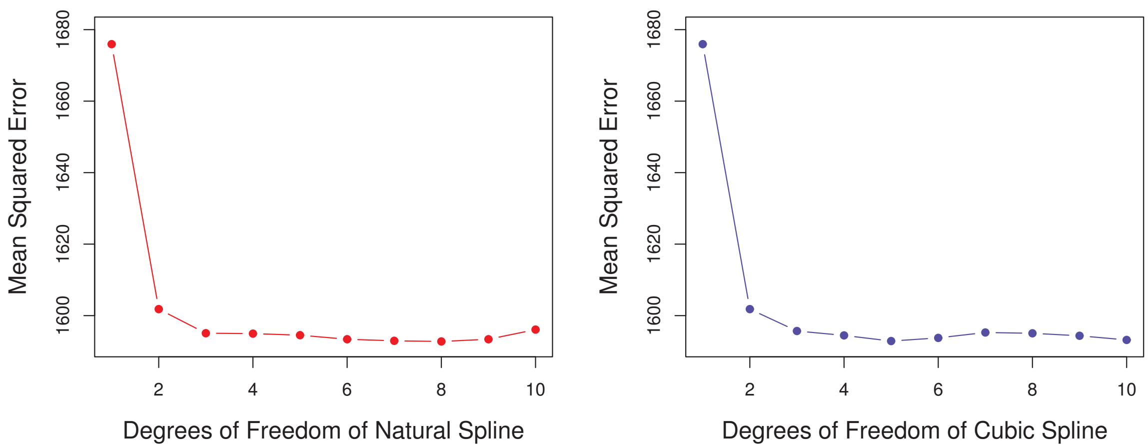


A natural cubic spline function with four degrees of freedom is fit to the Wage data. Left: A spline is fit to wage (in thousands of dollars) as a function of age. Right: Logistic regression is used to model the binary event wage> 250 as a function of age. The fitted posterior probability of wage exceeding $250,000 is shown. The dashed lines denote the knot locations.

- How many knots should we use, or equivalently how many degrees of freedom should our spline contain?

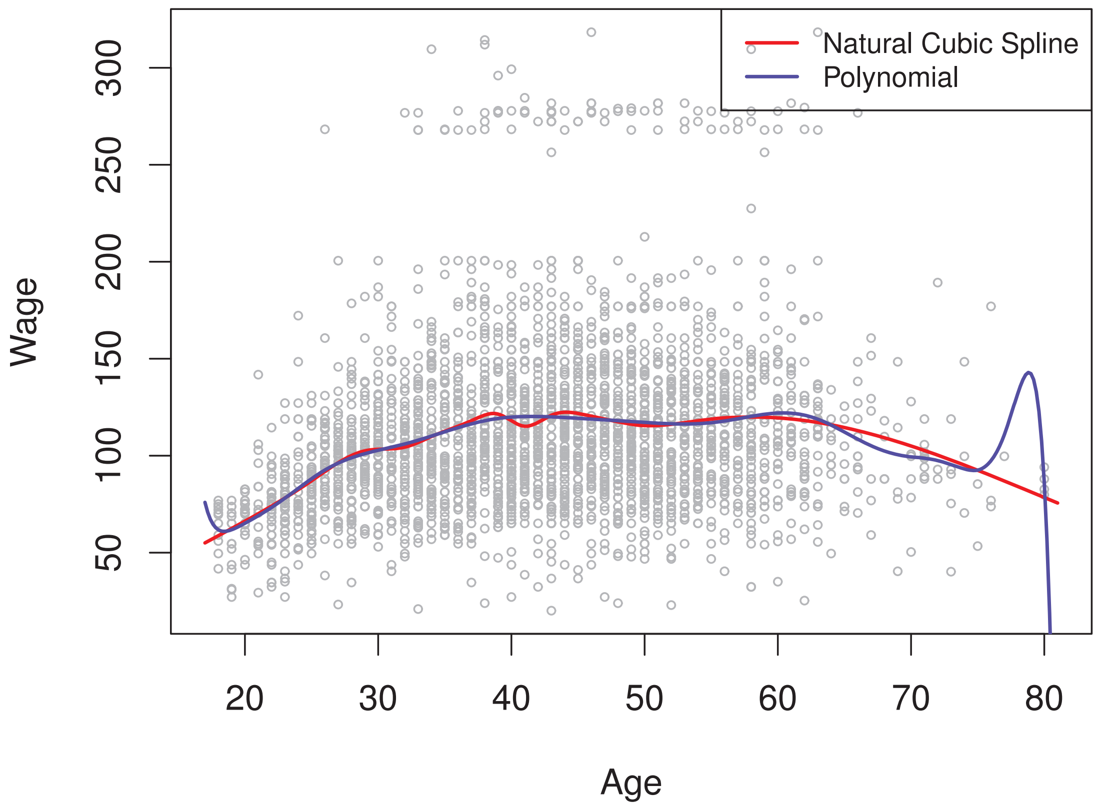
– one option is to try out different numbers of knots and see which produces the best looking curve.

– Another approach is to use cross-validation.



Ten-fold cross-validated mean squared errors for selecting the degrees of freedom when fitting splines to the Wage data. The response is wage and the predictor age. Left: A natural cubic spline. Right: A cubic spline.

Comparison to Polynomial Regression



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.

- The extra flexibility in the polynomial produces undesirable results at the boundaries, while the natural cubic spline still provides a reasonable fit to the data.

- Regression splines often give superior results to polynomial regression. This is because unlike polynomials, which must use a high degree (exponent in the highest monomial term, e.g. X^{15}) to produce flexible fits, splines introduce flexibility by increasing the number of knots but keeping the degree fixed.

- Generally, this approach produces more stable estimates.

- Splines also allow us to place more knots, and hence flexibility, over regions where the function f seems to be changing rapidly, and fewer knots where f appears more stable.

Smoothing Splines

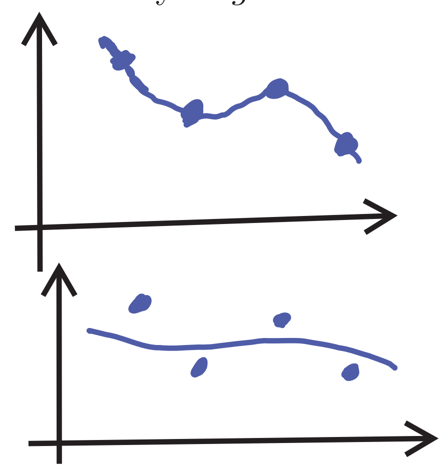
Consider this criterion for fitting a smooth function $g(x)$ to some data:

$$\underset{g \in S}{\text{minimize}} \sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int g''(t)^2 dt$$

- The first term is a loss function that encourages g to fit the data well.
- The second term is a penalty term that penalizes the variability in g.

The smaller λ, the more wiggly the function, eventually interpolating y_i when $\lambda = 0$.

As $\lambda \to \infty$, the function $g(x)$ becomes linear.



- The solution is a natural cubic spline, with a knot at every unique value of x_i. The penalty still controls the roughness via λ.

Choosing the Smoothing Parameter λ

- Although a smoothing spline has n parameters and hence n nominal degrees of freedom, these n parameters are heavily constrained or shrunk down.

- Hence the effective degrees of freedom, df_λ, is a measure of the flexibility of the smoothing spline—the higher it is, the more flexible (and the lower-bias but higher-variance) the smoothing spline.

- The vector of n fitted values can be written as

$$\hat{\mathbf{g}}_\lambda = \mathbf{S}_\lambda \mathbf{y}$$

where $\mathbf{S}_\lambda$ is an $n \times n$ matrix (determined by the x_i and λ).

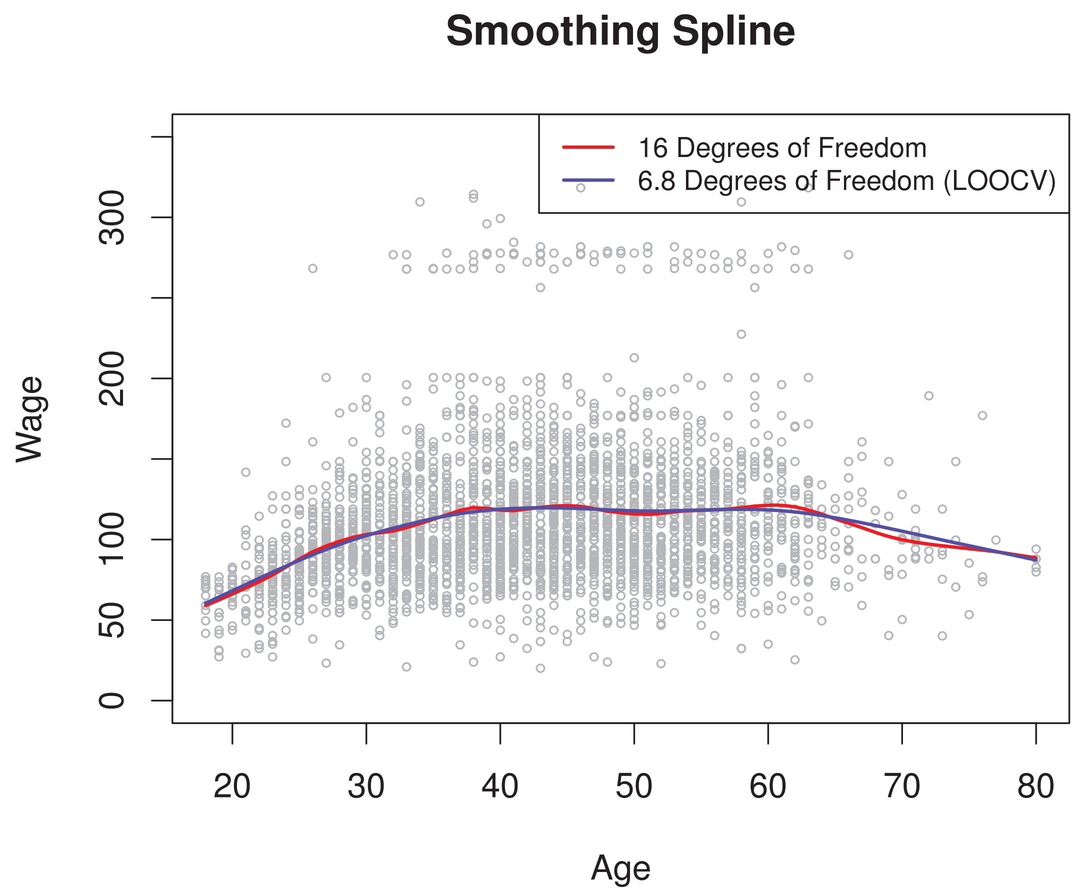
- Then the effective degrees of freedom is defined to be

$$df_\lambda = \sum_{i=1}^{n} \{\mathbf{S}_\lambda\}_{ii}$$: the sum of the diagonal elements of the matrix S_λ

$df_\lambda : n \to 2$ as $\lambda : 0 \to \infty$

- In fitting a smoothing spline, we do not need to select the number or location of the knots—there will be a knot at each training observation, $x_1, \ldots, x_n$. Instead, we need to choose the value of λ.

- One possible solution to this problem is cross-validation. In other words, we can find the value of λ that makes the cross-validated RSS as small as possible.



Smoothing spline fits to the Wage data. The red curve results from specifying 16 effective degrees of freedom. For the blue curve, λ was found automatically by leave-one-out cross-validation, which resulted in 6.8 effective degrees of freedom.

since there is little difference between the two fits, the smoothing spline fit with 6.8 degrees of freedom is preferable, since in general simpler models are better unless the data provides evidence in support of a more complex model.