Ch 7.4: Cubic splines

Lecture 15 - CMSE 381

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Mon, March 18, 2024

Announcements

Last time:

- 7.2 Step functions
- 7.3 Basis functions

This lecture:

- 7.4 Cubic splines

Announcements:

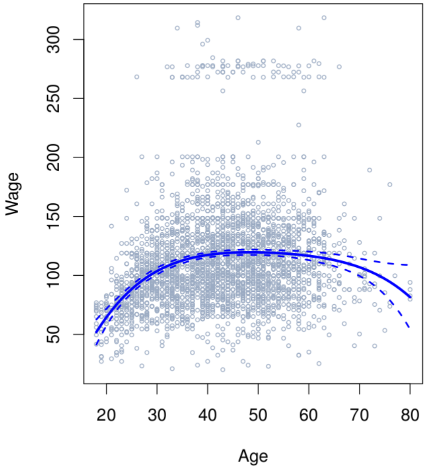
- Next lecture: decision tree

Section 1

Last time

Polynomial regression

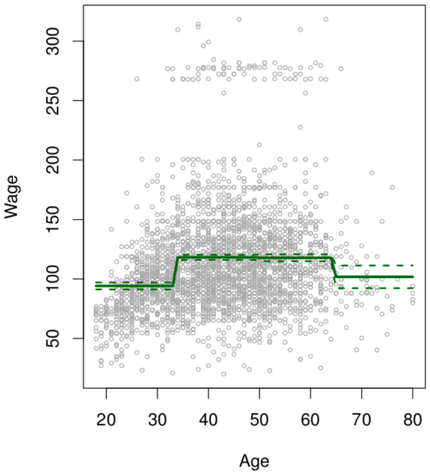
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \varepsilon_i$$



Step function regression

$$
\begin{aligned}
C_0(X) &= I(X < c_1),\\
C_1(X) &= I(c_1 \le X < c_2),\\
C_2(X) &= I(c_2 \le X < c_3),\\
&\vdots\\
C_{K-1}(X) &= I(c_{K-1} \le X < c_K),\\
C_K(X) &= I(c_K \le X),
\end{aligned}
$$

$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \varepsilon_i$$

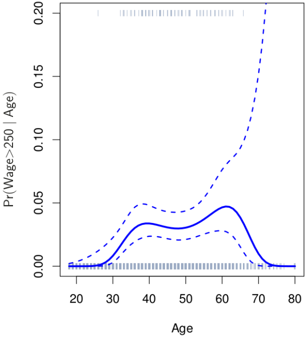


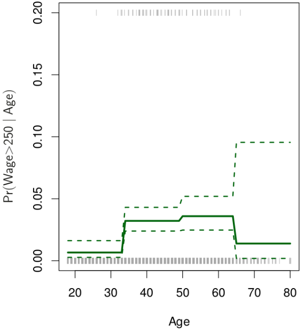
Classification version

$$\Pr(y_i > 250 \mid x_i) =$$

$$\frac{\exp(\beta_0 + \beta_1 x_i + \cdots + \beta_d x_i^d)}{1 + \exp(\beta_0 + \beta_1 x_i + \cdots + \beta_d x_i^d)}$$

$$\frac{\exp(\beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i))}{1 + \exp(\beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i))}$$





Basis Functions Setup

Polynomial and piecewise-constant regression models are special cases of a *basis function* approach.

$$y_i = \beta_0 + \beta_1 b_1(x_i) + \beta_2 b_2(x_i) + \cdots + \beta_K b_K(x_i) + \varepsilon_i$$

- Pick a collection of basis functions $b_1(X), \cdots, b_K(X)$
- Use least squares to figure out the constants
- Lots of possible options for these
  - Some examples are wavelets or Fourier series
  - Next section is we'll talk about regression splines

Section 2

Regression Splines

Piecewise polynomials

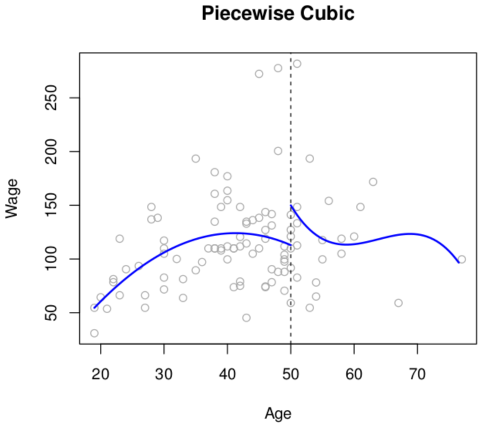
- Fit a polynomial regression

  $$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \varepsilon_i$$

- Let the β_i's be different at different locations of the range.

- Points where the coefficients change are called "knots"
- More knots gives more flexible polynomial
- Piecewise constant of previous section is degree 0 version of this
- the best trade-off between variance (stability) and bias (flexibility) is $d = 3$!

Example of piecewise polynomials of third order



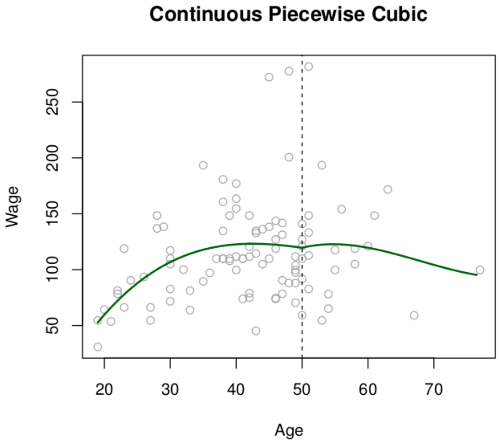
Example:

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

- Uses 8 degrees of freedom to fit the model
- problem: we have a discontinuous function.

The fix

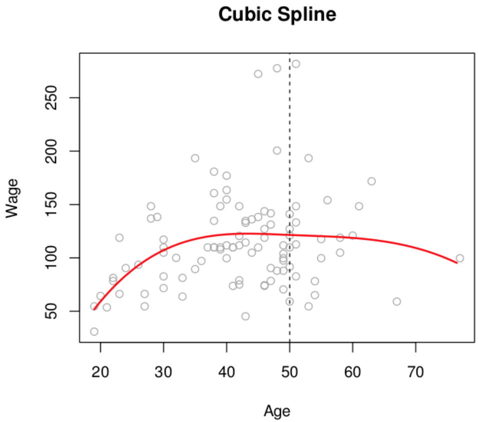
- Fit piecewise polynomial
- Require continuity at knots



Looks better, but pointy join is odd

The better fix: Cubic splines

- Fit piecewise polynomial
- Require continuity at knots
- Require the first and second derivatives to be continuous at knots
- Now we have smooth joins
- This is called a "cubic spline"



An example

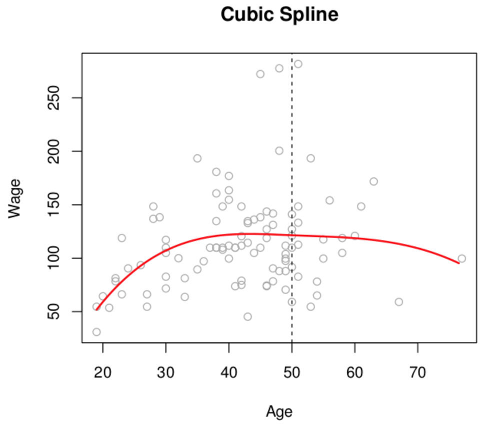
What would the values of $\beta_1, \beta_2, \beta_3$ be for $f(x)$ to a cubic spline?

$$f(x) = \begin{cases} 1 - \beta_1 x + \beta_2 x^3 & \text{if } x < 1 \\ -1 - x + \beta_3 x^2 - x^3 & \text{if } x \geq 1 \end{cases}$$

Cubic splines: degrees of freedom

$$f(x) = \begin{cases} \beta_0^1 + \beta_1^1 x + \beta_2^1 x^2 + \beta_3^1 x^3 & x < c \\ \beta_0^2 + \beta_1^2 x + \beta_2^2 x^2 + \beta_3^2 x^3 & x \geq c \end{cases}$$

- 8 original degrees of freedom, minus three constraints (continuous, 1st deriv continuous, 2nd deriv continuous) leaves 5 degrees of freedom
- cubic spline with K knots uses $4 + K$ degrees of freedom



A better choice of basis: Spline basis representation

Want to pick b_i so that we represent a cubic spline with K knots as

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

Truncated power basis function

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

Desmos link: https://www.desmos.com/calculator/esucuulbgj

The basis for cubic splines

Given knots at $z_1, \cdots, z_K$

- 1
- X
- X^2
- X^3
- $h(X, z_1)$
- $h(X, z_2)$
- $\vdots$
- $h(X, z_K)$

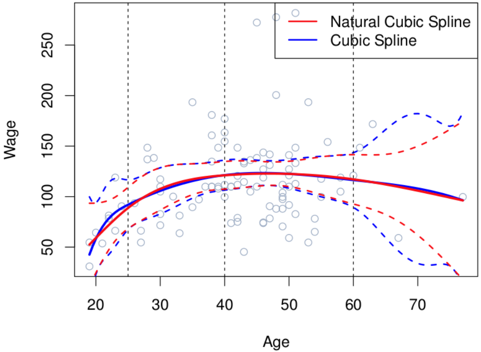
- Adding $h(X, z_i)$ leads to discontinuity at most at the third derivative at z_i
- So use the $4 + K$ functions listed here, result is $K + 4$ coefficients.
- Matches with our $K + 4$ degrees of freedom from earlier

$$f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 h(X, z_1) + \beta_5 h(X, z_2) + \cdots + \beta_{k+3} h(X, z_K)$$

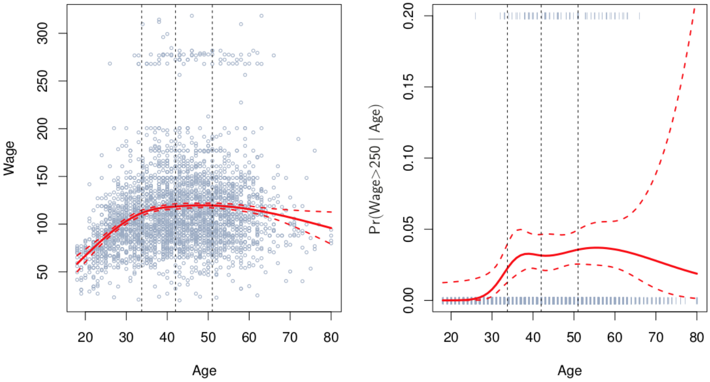
Coding example

Notes on cubic splines

- Still have problem of high variance at the outer range of predictors (big or small X)
- *Natural spline:* regression spline plus boundary constraints where function is required to be linear at the boundary



Where to put the knots?

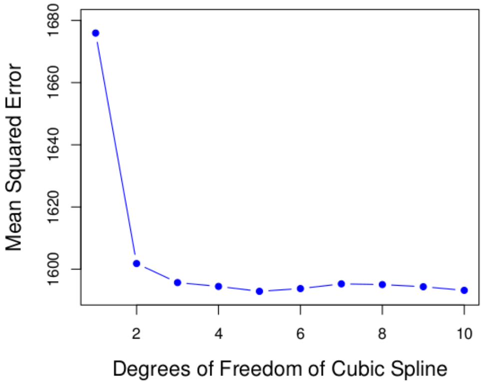


- Plan A: Uniform placement
- Plan B: Data driven. Example here puts it at quantiles of the data

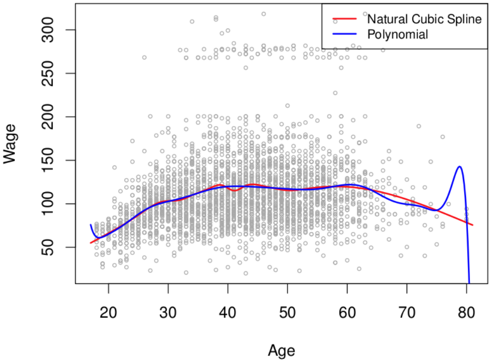
How many knots to use?

When in doubt, Cross-Validate.

- Do CV (k-fold probably) on data for each number of knots
- Figure at right is the MSE resulting from 10-fold CV
- Note that clearly linear is not enough, after that 3-4 degrees of freedom for the cubic spline are enough



Cubic splines vs Polynomial Regression



- Left: natural cubic spline with 15 degrees of freedom vs degree 15 polynomial
- Polynomial flexibility means ugly boundary behavior
- Splines let you add flexibility as more knots rather than more degrees, so allows for flexibility where the function is changing more rapidly