Ch 7.4: Cubic splines

Lecture 24 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Nov 6, 2023

Announcements

Last time:

- 7.2 Step functions
- 7.3 Basis functions

This lecture:

- 7.4 Cubic splines

Announcements:

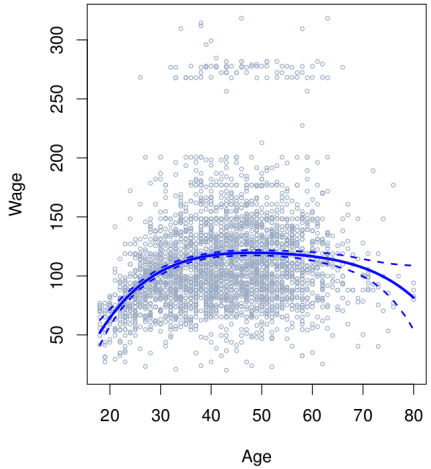
- Homework # 6 is now due Wednesday

Section 1

Last time

Polynomial regression

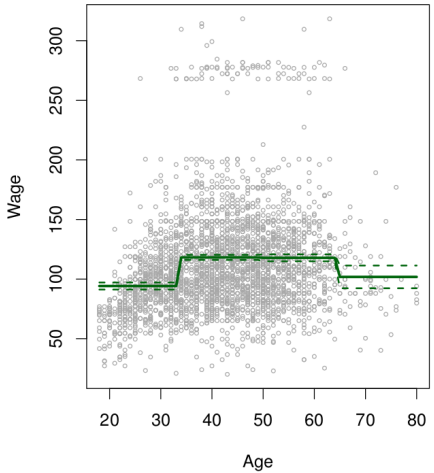
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \varepsilon_i$$



Step function regression

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

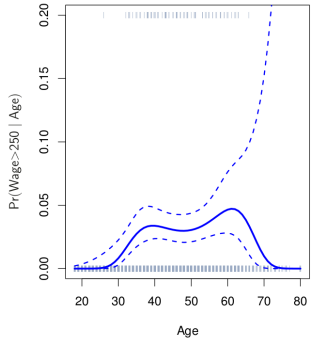
$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \varepsilon_i$$



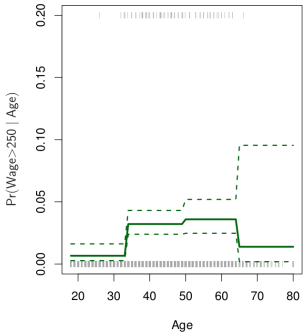
Classification version

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

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



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



Basis Functions Setup

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

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

Section 2

Regression Splines

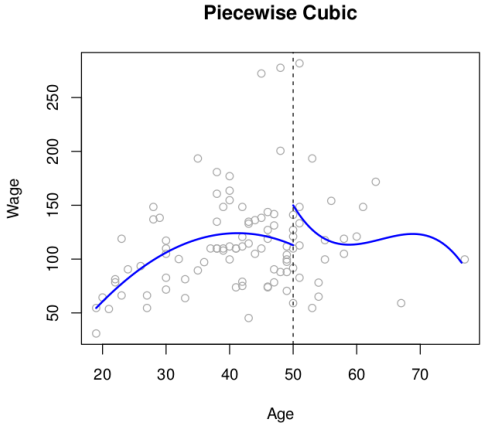
Piecewise polynomials

- 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.

Example of piecewise polynomial

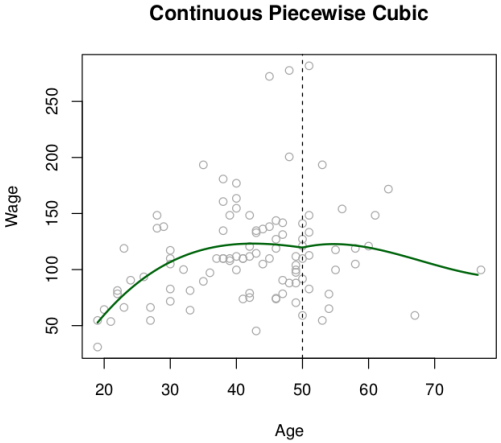


Example:

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

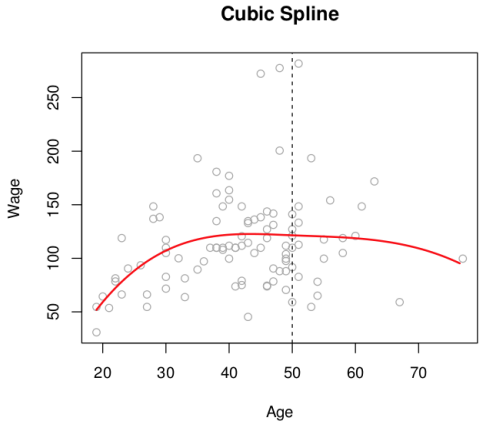
The fix

- Fit piecewise polynomial
- Require continuity at knots



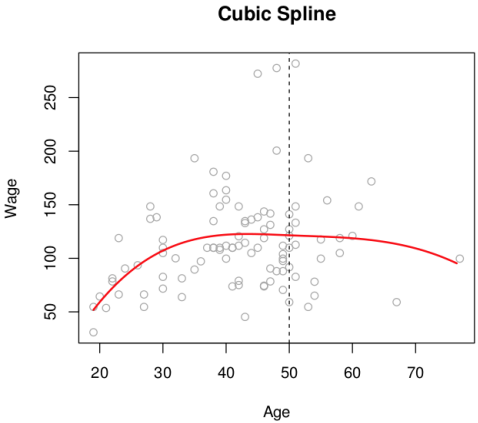
The better fix: Cubic splines

- Fit piecewise polynomial
- Require continuity at knots
- Require the first and second derivatives to be continuous at knots



Cubic splines: degrees of freedom

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



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

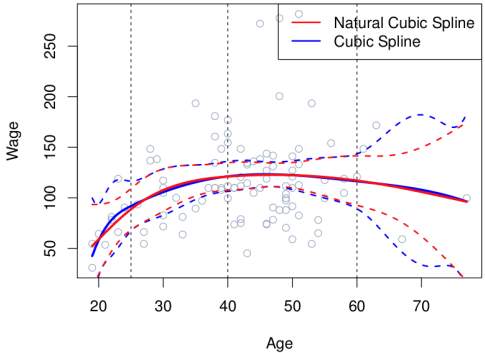
Given knots at $z_1, \cdots, z_K$

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

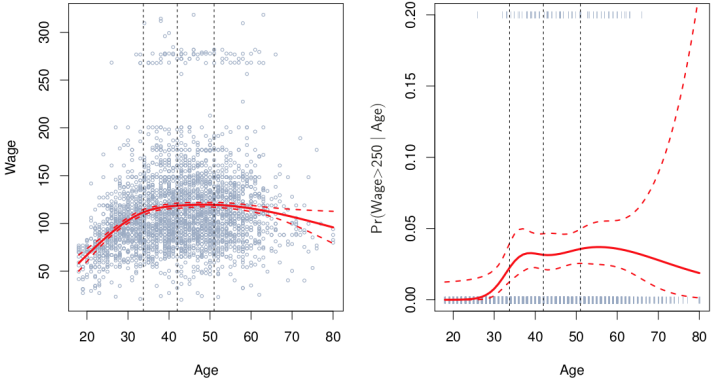
$$f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 h(X, z_1) + \beta_5 h(X, z_2) + \cdots + \beta_{k+3} h(X, z_K)$$

Coding example

Notes on cubic splines

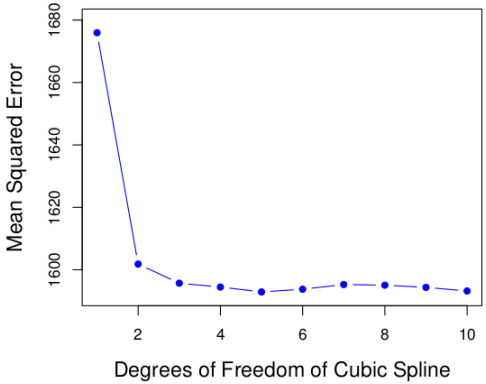


Where to put the knots?

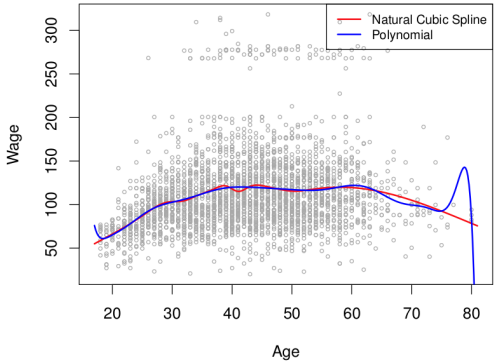


How many knots to use?

When in doubt, Cross-Validate.



Cubic splines vs Polynomial Regression



Next time

Status	Lec #	Date			Reading	Homeworks
		Mon	Oct 23	No class - Fall break		
		Wed	Oct 25	**Midterm #2**		
Done	20	Fri	Oct 27	Dimension Reduction	6.3	
Done	21	Mon	Oct 30	More dimension reduction; High dimensions	6.4	
Done	22	Wed	Nov 1	Polynomial & Step Functions	7.1,7.2	
Pushed	23	Fri	Nov 3	Step Functions; Basis functions; Start Splines	7.2 - 7.4	
	24	Mon	Nov 6	Regression Splines	7.4	~~HW #6 Due~~
	25	Wed	Nov 8	Decision Trees	8.1	HW #6 Due
	26	Fri	Nov 10	Random Forests	8.2.1, 8.2.2	
	27	Mon	Nov 13	Maximal Margin Classifier	9.1	
	28	Wed	Nov 15	SVC	9.2	
	29	Fri	Nov 17	SVM	9.3, 9.4	
	30	Mon	Nov 20	Single layer NN	10.1	
	31	Wed	Nov 22	Virtual: Project office hours		
		Fri	Nov 24	No class - Thanksgiving		
		Mon	Nov 27	**Review**		
		Wed	Nov 29	**Midterm #3**		
	32	Fri	Dec 1	Multi Layer NN	10.2	
	33	Mon	Dec 4	CNN	10.3	
	34	Wed	Dec 6	Unsupervised Learning & Clustering	12.1, 12.4	
	35	Fri	Dec 8	Virtual: Project office hours		Project due