

Ch 7.4: Cubic splines

Lecture 23 - CMSE 381

Prof. Lianzhang Bao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Mon, Mar 24, 2025

Announcements

Last time:

- 7.2 Step functions
- 7.3 Basis functions

This lecture:

- 7.4 Cubic splines

Announcements:

- Homework # 6 is due Wednesday
- Projects

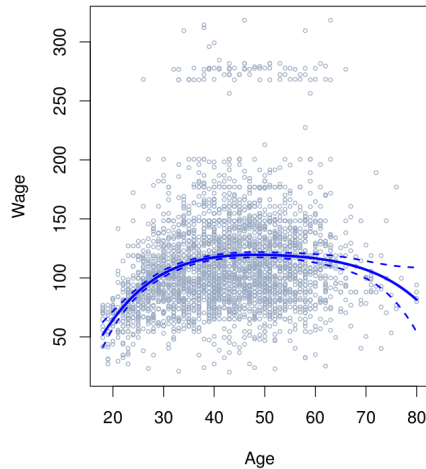
21	W	3/19	Polynomial & Step Functions	7.1-7.2		
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4		
23	M	3/24	Regression Splines	7.4		
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26	Q7
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30	
26	M	3/31	Maximal Margin Classifier	9.1		
27	W	4/2	SVC	9.2		Q8
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6	
29	M	4/7	Single Layer NN	10.1		
30	W	4/9	Multi Layer NN	10.2		Q9
31	F	4/11	CNN	10.3		
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4	HW #9 Due Sun 4/13	Q10
33	W	4/16	Virtual: Project Office Hours			
	F	4/18	Review			
	M	4/21	Midterm #3			
	W	4/23				
	F	4/25			Project Due	
			No final exam			

Section 1

Last time

Polynomial regression

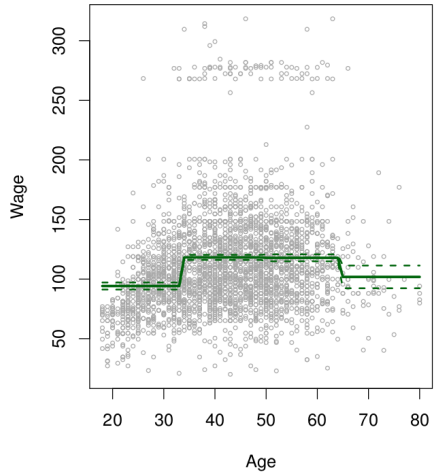
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \varepsilon_i$$



Step function regression

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

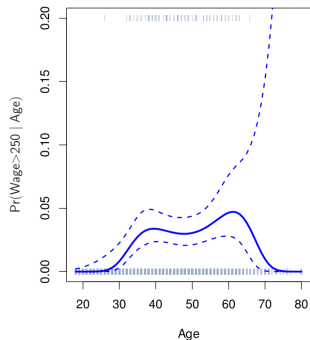
$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \varepsilon_i$$



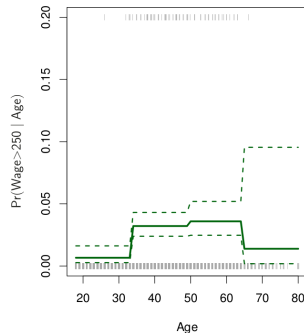
Classification version

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

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



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



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

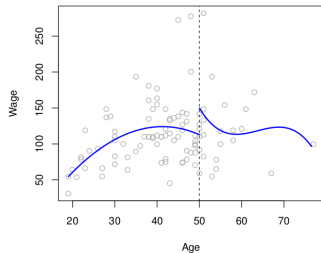
- Fit a polynomial regression

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

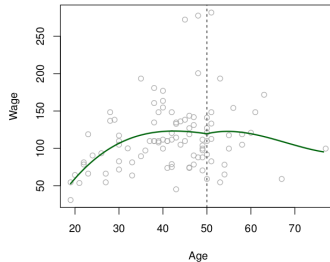
- Let the β_i 's be different at different locations of the range.

Building up to cubic splines

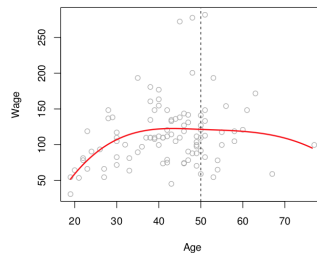
Piecewise Cubic



Continuous Piecewise Cubic



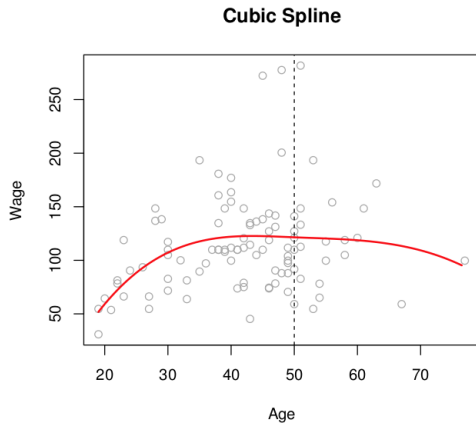
Cubic Spline



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

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

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

Version 1: Truncated power basis function

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

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

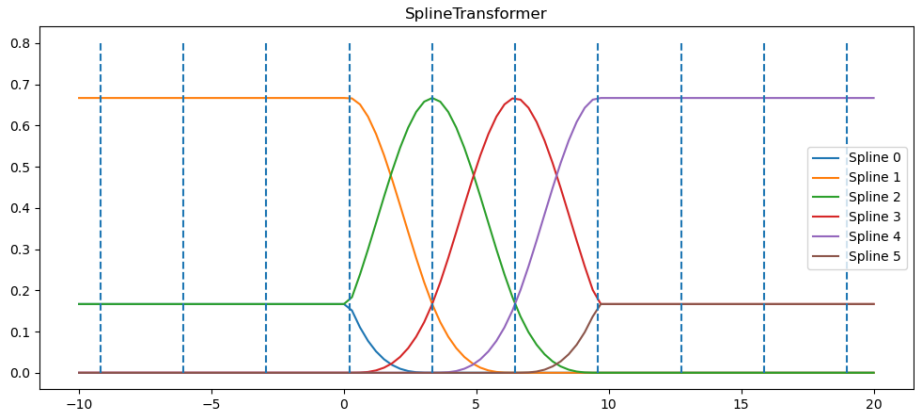
The (book) basis for cubic splines

Given knots at $z_1, \dots, z_K$

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

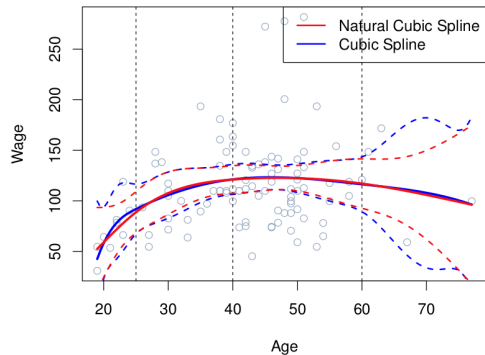
$$f(X) = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 h(X, z_1) + \beta_5 h(X, z_2) + \dots + \beta_{k+3} h(X, z_K)$$

Version 2: B-spline basis function

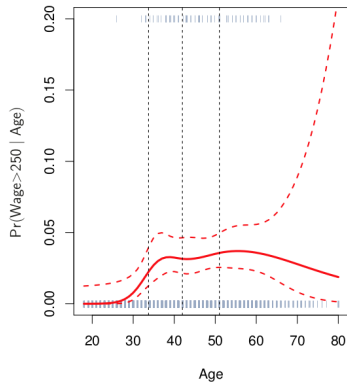
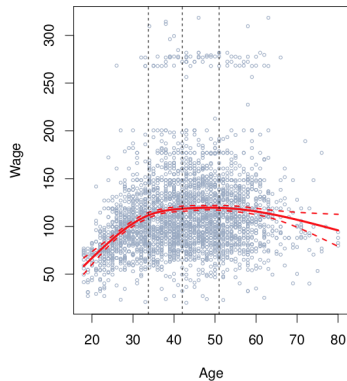


Coding example

Notes on cubic splines

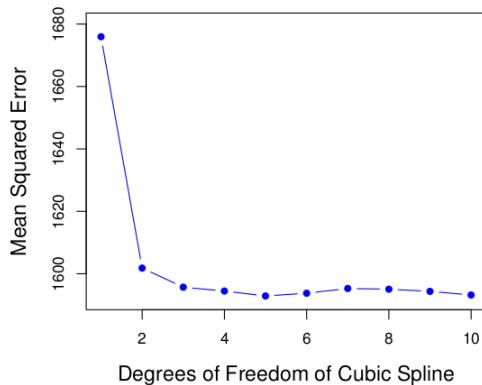


Where to put the knots?

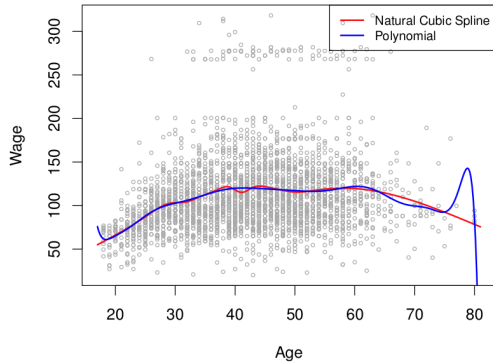


How many knots to use?

When in doubt, Cross-Validate.



Cubic splines vs Polynomial Regression



Next time

21	W	3/19	Polynomial & Step Functions	7.1-7.2		
22	F	3/21	Step Functions; Basis functions; Start Splines	7.2-7.4		
23	M	3/24	Regression Splines	7.4		
24	W	3/26	Decision Trees	8.1	HW #6 Due Wed 3/26	Q7
25	F	3/28	Random Forests	8.2.1, 8.2.2	HW #7 Due Sun 3/30	
26	M	3/31	Maximal Margin Classifier	9.1		
27	W	4/2	SVC	9.2		Q8
28	F	4/4	SVM	9.3, 9.4	HW #8 Due Sun 4/6	
29	M	4/7	Single Layer NN	10.1		
30	W	4/9	Multi Layer NN	10.2		Q9
31	F	4/11	CNN	10.3	HW #9 Due Sun 4/13	
32	M	4/14	Unsupervised learning / clustering	12.1, 12.4		
33	W	4/16	Virtual: Project Office Hours			Q10
	F	4/18	Review			
	M	4/21	Midterm #3			
	W	4/23				
	F	4/25			Project Due	
			No final exam			