# Ch 7.4: Cubic splines

## Lecture 23 - CMSE 381

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

Mon, Mar 23, 2026

# Announcements

**Last time:**

- 7.2 Step functions
- 7.3 Basis functions

**This lecture:**

- 7.4 Cubic splines

**Announcements:**

- Homework # 5 is due Sunday.

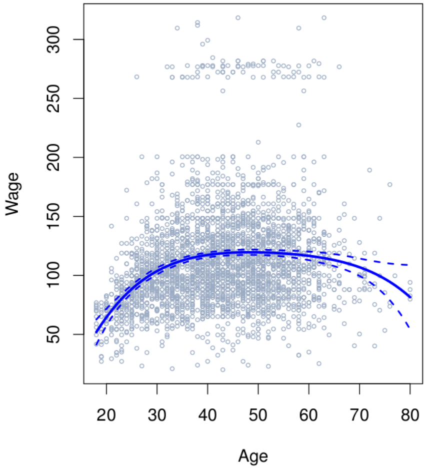
| | | | | | | |
|---|---|---|---|---|---|---|
| 21 | W | 3/18 | Polynomial & Step Functions | 7.1-7.2 | | |
| 22 | F | 3/20 | Step Functions; Basis functions; Start Splines | 7.2-7.4 | | |
| 23 | M | 3/23 | Regression Splines | 7.4 | | |
| 24 | W | 3/25 | Decision Trees | 8.1 | | Q7 |
| 25 | F | 3/27 | Random Forests | 8.2.1, 8.2.2 | HW #5 Due Sun 3/29 | |
| 26 | M | 3/30 | Maximal Margin Classifier | 9.1 | | |
| 27 | W | 4/1 | SVC | 9.2 | | Q8 |
| 28 | F | 4/3 | SVM | 9.3, 9.4 | | |
| 29 | M | 4/6 | Single Layer NN | 10.1 | | |
| 30 | W | 4/8 | Multi Layer NN | 10.2 | | Q9 |
| 31 | F | 4/10 | CNN | 10.3 | HW #6 Due Sun 4/12 | |
| 32 | M | 4/13 | Unsupervised learning / clustering | 12.1, 12.4 | | |
| 33 | W | 4/15 | Virtual: Project Office Hours | | | Q10 |
| | F | 4/17 | ***Review*** | | | |
| | M | 4/20 | **Midterm #3** | | | |
| | W | 4/22 | | | | |
| | F | 4/24 | | | **Project Due** | |

Section 1

# Previously

# Polynomial regression

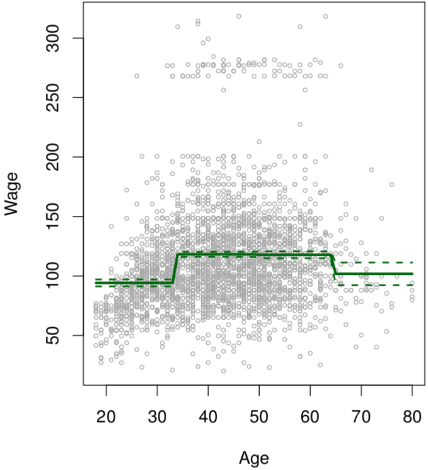
$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2 + \cdots + \beta_d x_i^d + \varepsilon_i$$



# Step function regression

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

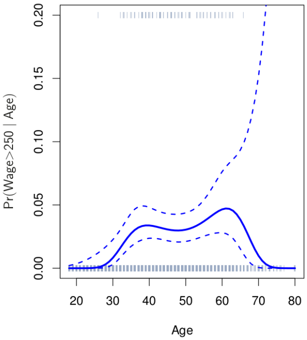
$$y_i = \beta_0 + \beta_1 C_1(x_i) + \beta_2 C_2(x_i) + \cdots + \beta_K C_K(x_i) + \varepsilon_i$$



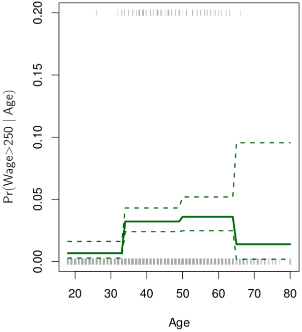
# Classification version

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

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




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



# Basis Functions Setup

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

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

# What will you learn today?

- How to define cubic splines mathematically in terms of piecewise polynomials? (review)
  - Write the equations for the model on each interval between knots.
  - Write the continuity constraints (function, first and second derivatives) at each knot.
  - Compute the total degrees of freedom.
- How to define cubic splines in terms of basis functions?
  - Describe them mathematically using the truncated power basis (textbook version).
  - Visualize them as B-spline basis functions (e.g., in Python / scikit-learn).
- How to fit a cubic spline model to data in Python?
- How to change the model flexibility?
  - What are the relevant metaparameters?
  - How to choose appropriate flexibility?
- What precautions must you take at the outer boundary? Why?

Section 2

# Regression Splines

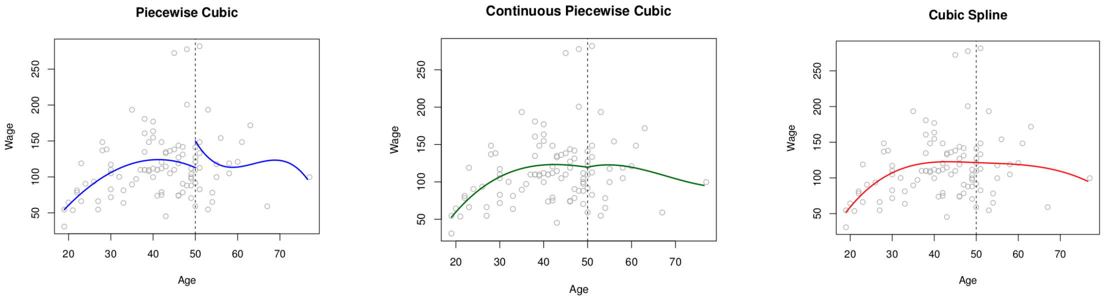
# Piecewise polynomials

- 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 $\beta_i$'s be different at different locations of the range.

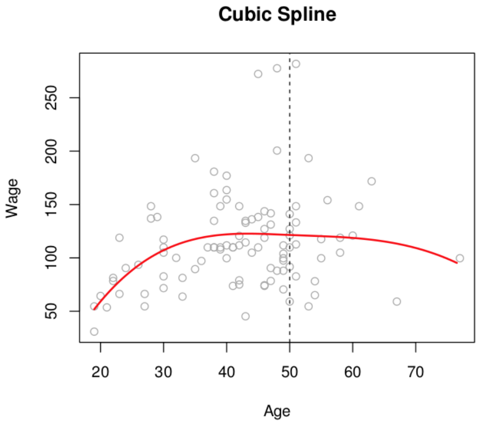
# Building up to cubic splines




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

# 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

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

Desmos link: https://www.desmos.com/calculator/ahllu5glar

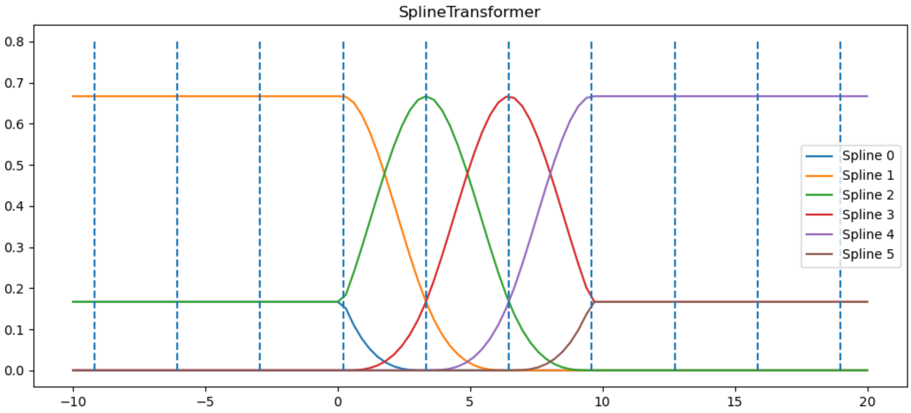
# The (book) basis for cubic splines

Given knots at $z_1, \cdots, z_K$

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

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

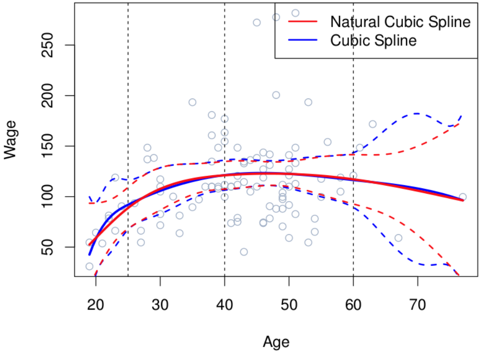
# Version 2: B-spline basis function



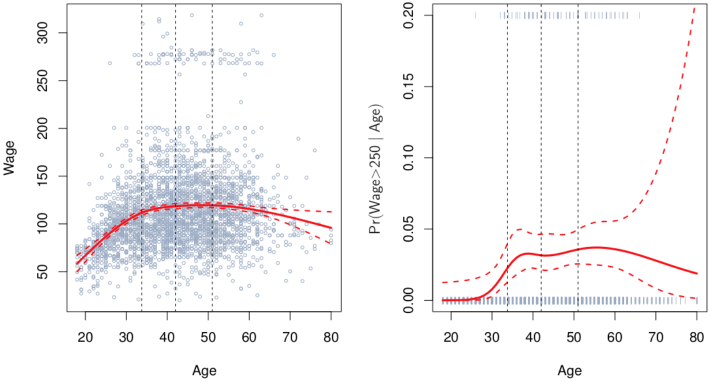

Test your understanding: PollEv

# Coding example

# Notes on cubic splines

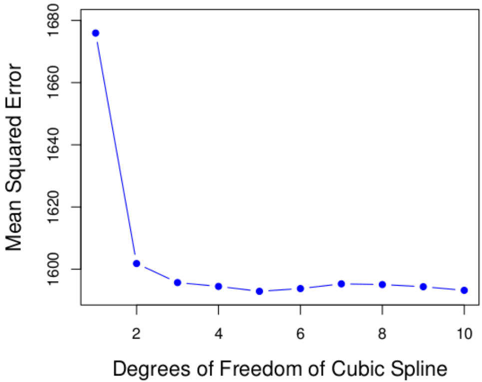


# Where to put the knots?

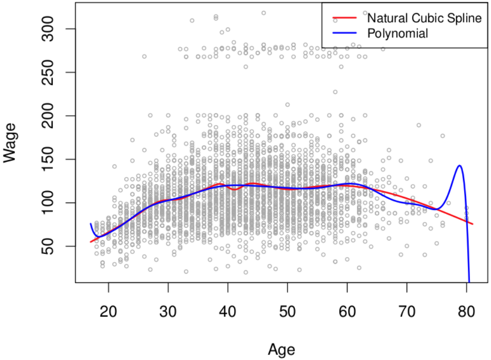


# How many knots to use?

When in doubt, Cross-Validate.



# Cubic splines vs Polynomial Regression



# Next time

| | | | | | | |
|---|---|---|---|---|---|---|
| 21 | W | 3/18 | Polynomial & Step Functions | 7.1-7.2 | | |
| 22 | F | 3/20 | Step Functions; Basis functions; Start Splines | 7.2-7.4 | | |
| 23 | M | 3/23 | Regression Splines | 7.4 | | Q7 |
| 24 | W | 3/25 | Decision Trees | 8.1 | | |
| 25 | F | 3/27 | Random Forests | 8.2.1, 8.2.2 | HW #5 Due Sun 3/29 | |
| 26 | M | 3/30 | Maximal Margin Classifier | 9.1 | | Q8 |
| 27 | W | 4/1 | SVC | 9.2 | | |
| 28 | F | 4/3 | SVM | 9.3, 9.4 | | |
| 29 | M | 4/6 | Single Layer NN | 10.1 | | Q9 |
| 30 | W | 4/8 | Multi Layer NN | 10.2 | | |
| 31 | F | 4/10 | CNN | 10.3 | HW #6 Due Sun 4/12 | |
| 32 | M | 4/13 | Unsupervised learning / clustering | 12.1, 12.4 | | Q10 |
| 33 | W | 4/15 | Virtual: Project Office Hours | | | |
| | F | 4/17 | ***Review*** | | | |
| | M | 4/20 | **Midterm #3** | | | |
| | W | 4/22 | | | | |
| | F | 4/24 | | | **Project Due** | |