

Ch 7.1-7.2: Polynomial regression and Step Functions

Lecture 21 - CMSE 381

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Wed, Mar 18, 2026

Announcements

Last time:

- Exam 2

This lecture:

- 7.1 Polynomial regression
- 7.2 Step functions

Announcements:

- HW #5 due Sun 3/29
- Project stuff

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		
32	M	4/13	Unsupervised learning / clustering	12.1, 12.4	HW #6 Due Sun 4/12	
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

What should you learn from today's lecture?

- Why do people want to use polynomial regression rather than simple linear regression
- How to fit polynomial regression models?
- Why do people want to use step functions rather than simple linear regression?
- How to fit step function models?
- What is common between polynomial regression and step functions?
- What is different between polynomial regression and step functions?

Section 1

Polynomial Regression

Polynomial regression

Replace linear model

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

with

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

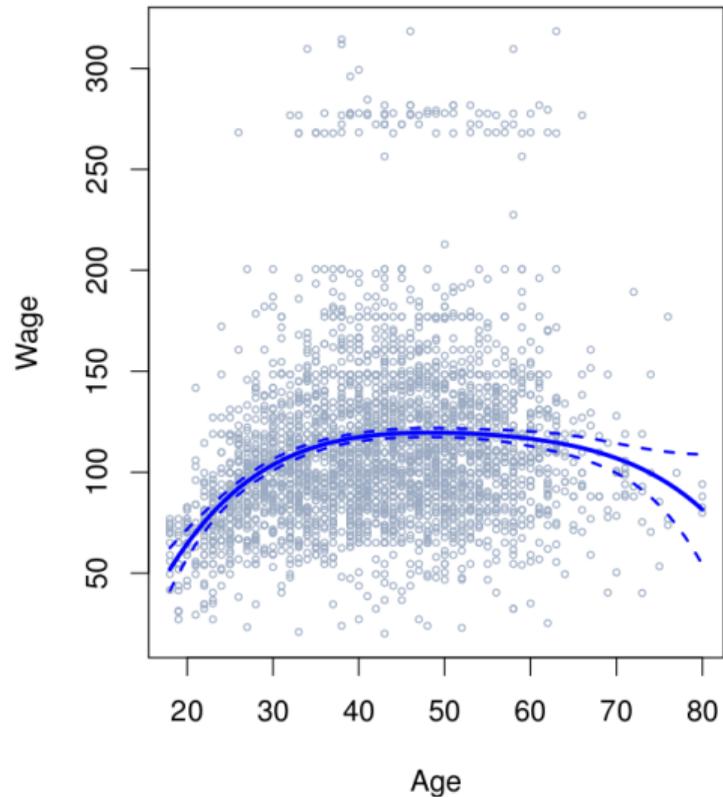
Example: Wage Data

$$\text{wage} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \dots + \beta_p \text{age}^p + \varepsilon.$$

My code learned:

$$-184.1542 + 21.24552 \cdot \text{age} + -0.56386 \cdot \text{age}^2 + 0.00681 \cdot \text{age}^3 + (-3 \cdot 10^{-5}) \cdot \text{age}^4$$

Example with wage data



Section 2

Step function

Step functions

$$I(X < c) \quad I(c_1 \leq X < c_2) \quad I(c \leq X)$$

More on step function setup

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

Test your understanding: [PollEv](#)

Example

Given knots $c_1 = 3$, $c_2 = 5$, $c_3 = 7$, determine the entries in the columns for $C_i(X)$ in the below matrix.

X	$C_0(X)$	$C_1(X)$	$C_2(X)$	$C_3(X)$
1				
2				
3				
4				
5				

X	$C_0(X)$	$C_1(X)$	$C_2(X)$	$C_3(X)$
6				
7				
8				
9				
10				

Draw the function

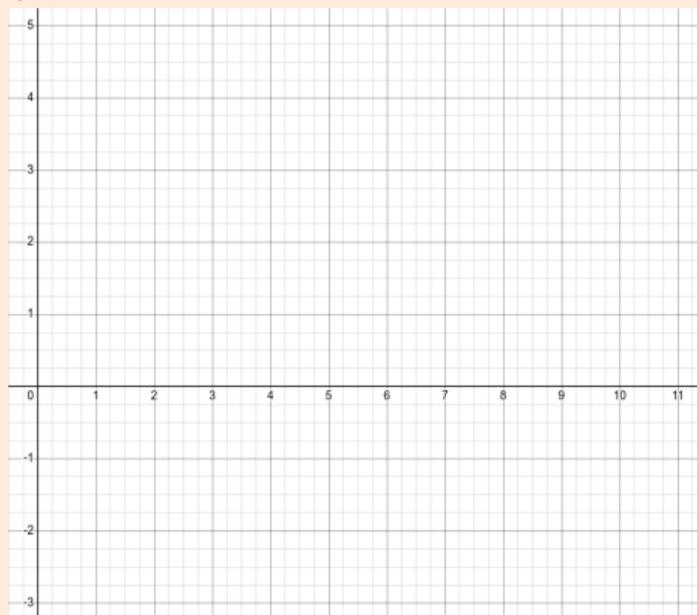
My code doing regression on the step function input returned the function.

$$f(X) = -1 + 3C_1(X) + 4C_2(X) - 2C_3(X).$$

Fill in the table of values, then draw this function below.

X	F(X)
1	
2	
3	
4	
5	

X	F(X)
6	
7	
8	
9	
10	



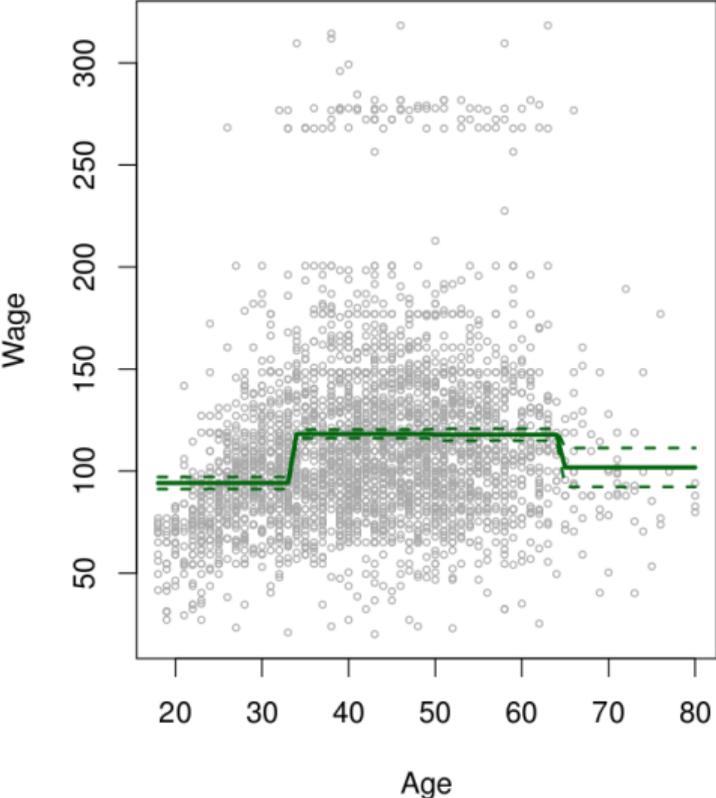
Step function: Learned model

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

Coding bit

Back to the wage data set

Step function example



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		
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		
32	M	4/13	Unsupervised learning / clustering	12.1, 12.4	HW #6 Due Sun 4/12	
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	