

Ch 7.1-7.2: Polynomial regression and Step Functions

Lecture 21 - CMSE 381

Prof. Guanqun Cao

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Friday, Oct 24, 2025

Announcements

Last time:

- Test 2

This lecture:

- 7.1 Polynomial regression
- 7.2 Step functions

Announcements:

- HW #6 due *Sunday* 11/2
- Project stuff

	F	10/17	Review		
	M	10/20	Fall Break		
	W	10/22	Midterm #2		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2
25	M	11/3	Random Forests	8.2.1, 8.2.2	
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16
31	M	11/17	CNN	10.3	
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	Virtual: Project Office Hours		HW #9 Due Sun 11/23
	M	11/24	Review		
	W	11/26	Midterm #3		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		
	F	12/5			Project Due

Section 1

Polynomial Regression

Polynomial regression

Replace linear model

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

with

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

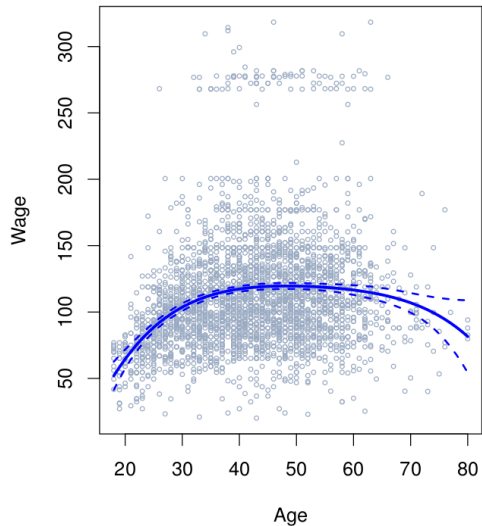
Example: Wage Data

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

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

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

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	



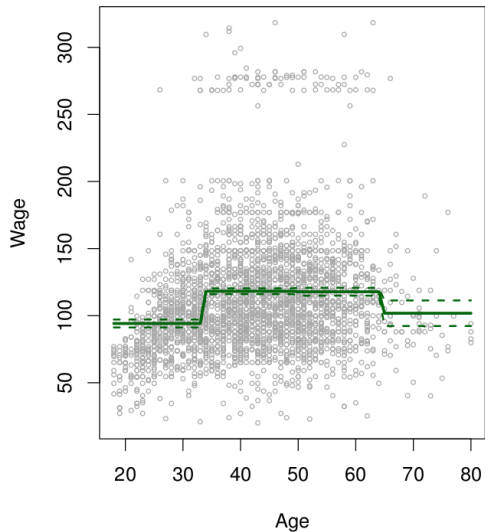
Step function: Learned model

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

Coding bit

Back to the wage data set

Step function example



Next time

	F	10/17	Review		
	M	10/20	Fall Break		
	W	10/22	Midterm #2		
21	F	10/24	Polynomial & Step Functions	7.1-7.2	HW #5 Due Sun 10/28
22	M	10/27	Step Functions; Basis functions; Start Splines	7.2-7.4	
23	W	10/29	Regression Splines	7.4	
24	F	10/31	Decision Trees	8.1	HW #6 Due Sun 11/2
25	M	11/3	Random Forests	8.2.1, 8.2.2	
26	W	11/5	Maximal Margin Classifier	9.1	
27	F	11/7	SVC	9.2	HW #7 Due Sun 11/9
28	M	11/10	SVM	9.3, 9.4	
29	W	11/12	Single Layer NN	10.1	
30	F	11/13	Multi Layer NN	10.2	HW #8 Due Sun 11/16
31	M	11/17	CNN	10.3	
32	W	11/19	Unsupervised learning / clustering	12.1, 12.4	
33	F	11/21	Virtual: Project Office Hours		HW #9 Due Sun 11/23
	M	11/24	Review		
	W	11/26	Midterm #3		
	F	11/28	Thanksgiving		
	M	12/1	Virtual: Project Office Hours		
	W	12/3	Virtual: Project Office Hours		