ISLR | Chapter 7 Exercises

Marshall McQuillen

9/12/2018

Conceptual

1

- **A**. The cubic piecewise polynomial:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \xi)_+^3 \quad where \quad (x - \xi)_+^3 = \begin{cases} 0, & x \leq \xi \\ (x - \xi)^3, & otherwise \end{cases}$$

...can be broken up and rewritten to be:

$$f(x) = \begin{cases} f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3, & x \leq \xi \\ f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3, & otherwise \end{cases}$$

In $f_1(x)$, since $(x - \xi)_+^3 = 0$ (because $x \leq \xi$), the fifth term (of $f(x)$) zeroes out and the coefficients can be expresses as $a_1 = \beta_0,\ b_1 = \beta_1,\ c_1 = \beta_2$ and $d_1 = \beta_3$.

- **B**. Expanding the fifth term in $f(x)$ allows for the various powers of x to be grouped together and then recondensed. $a_2,\ b_2,\ c_2$ and d_2 are expressed in terms of the cofficients below.

$$\begin{aligned} f_2(x) &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \xi)^3 && (1) \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \xi)(x - \xi)(x - \xi) && (2) \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x^2 - 2x\xi + \xi^2)(x - \xi) && (3) \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x^3 - x^2\xi - 2x^2\xi + 2x\xi^2 + \xi^2 x - \xi^3) && (4) \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x^3 - 3x^2\xi + 3x\xi^2 - \xi^3) && (5) \\ &= \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^3 - \beta_4 3x^2\xi + \beta_4 3x\xi^2 - \beta_4 \xi^3 && (6) \\ &= (\beta_0 - \beta_4 \xi^3) + (\beta_1 x + \beta_4 3x\xi^2) + (\beta_2 x^2 - \beta_4 3x^2\xi) + (\beta_3 x^3 + \beta_4 x^3) && (7) \\ &= (\beta_0 - \beta_4 \xi^3) + (\beta_1 + 3\beta_4 \xi^2)x + (\beta_2 - 3\beta_4 \xi)x^2 + (\beta_3 + \beta_4)x^3 && (8) \end{aligned}$$

$$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3 \quad where \quad \begin{cases} a_2 = \beta_0 - \beta_4 \xi^3 \\ b_2 = \beta_1 + 3\beta_4 \xi^2 \\ c_2 = \beta_2 - 3\beta_4 \xi \\ d_2 = \beta_3 + \beta_4 \end{cases} \qquad (9)$$

- **C**. Showing that $f(x)$ is continuous at ξ is illustrated by showing that $f(\xi)_1 = f(\xi)_2$.

$$\begin{aligned}
f_1(\xi) &= a_1 + b_1(\xi) + c_1(\xi)^2 + d_1(\xi)^3 && (10)\\
&= \beta_0 + \beta_1(\xi) + \beta_2(\xi)^2 + \beta_3(\xi)^3 && (11)\\
& && (12)\\
f_2(\xi) &= a_2 + b_2(\xi) + c_2(\xi)^2 + d_2(\xi)^3 && (13)\\
&= (\beta_0 - \beta_4\xi^3) + (\beta_1 + 3\beta_4\xi^2)(\xi) + (\beta_2 - 3\beta_4\xi)(\xi)^2 + (\beta_3 + \beta_4)(\xi)^3 && (14)\\
&= (\beta_0 - \beta_4\xi^3) + (\beta_1\xi + 3\beta_4\xi^3) + (\beta_2\xi^2 - 3\beta_4\xi^3) + (\beta_3\xi^3 + \beta_4\xi^3) && (15)\\
&= \beta_0 - \beta_4\xi^3 + \beta_1\xi + 3\beta_4\xi^3 + \beta_2\xi^2 - 3\beta_4\xi^3 + \beta_3\xi^3 + \beta_4\xi^3 && (16)\\
&= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 + 3\beta_4\xi^3 - 3\beta_4\xi^3 + \beta_4\xi^3 - \beta_4\xi^3 && (17)\\
&= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 + (3\beta_4\xi^3 - 3\beta_4\xi^3) + (\beta_4\xi^3 - \beta_4\xi^3) && (18)\\
f_2(\xi) &= \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 && (19)
\end{aligned}$$

$$f_2(\xi) = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 = f_1(\xi)$$

- **D**. In order to show that $f_1'(\xi) = f_2'(\xi)$, we must first find $f'(x)$ with respect to x and then simplify both $f_1'(\xi)$ and $f_2'(\xi)$.

$$\begin{aligned}
f(x) &= a_1 + b_1x + c_1x^2 + d_1x^3 && (20)\\
f'(x) &= b_1 + 2c_1x + 3d_1x^2 && (21)
\end{aligned}$$

Therefore, substituting the necessary coefficients in for $b_1,\ c_1$ and d_1 in both $f_1'(\xi)$ and $f_2'(\xi)$, we get:

$$f'(x) = b_1 + 2c_1x + 3d_1x^2 \quad \textit{then} \quad \begin{cases} f_1'(\xi) = \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 \\ f_2'(\xi) = (\beta_1 + 3\beta_4\xi^2) + 2(\beta_2 - 3\beta_4\xi)\xi + 3(\beta_3 + \beta_4)\xi^2 \end{cases} \qquad (22)$$

$$\begin{aligned}
f_2'(\xi) &= (\beta_1 + 3\beta_4\xi^2) + 2(\beta_2 - 3\beta_4\xi)\xi + 3(\beta_3 + \beta_4)\xi^2 && (23)\\
&= \beta_1 + 3\beta_4\xi^2 + 2\beta_2\xi - 6\beta_4\xi^2 + 3\beta_3\xi^2 + 3\beta_4\xi^2 && (24)\\
&= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 + (3\beta_4\xi^2 + 3\beta_4\xi^2 - 6\beta_4\xi^2) && (25)\\
&= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 + (6\beta_4\xi^2 - 6\beta_4\xi^2) && (26)\\
f_2'(\xi) &= \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 && (27)
\end{aligned}$$

We now see that the derivative $f'(x)$ is continuous at knot ξ, which is to say $f_1'(\xi) = f_2'(\xi)$:

$$f_2'(\xi) = \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 = f_1'(\xi)$$

- **E**. In order to show that $f_1''(\xi) = f_2''(\xi)$, we must first find $f''(x)$ with respect to x and then simplify both $f_1''(\xi)$ and $f_2''(\xi)$.

$$\begin{aligned} f(x) &= a_1 + b_1 x + c_1 x^2 + d_1 x^3 && (28)\\ f'(x) &= b_1 + 2c_1 x + 3d_1 x^2 && (29)\\ f''(x) &= 2c_1 + 6d_1 x && (30) \end{aligned}$$

Therefore, substituting the necessary coefficients in for c_1 and d_1 in both $f_1''(\xi)$ and $f_2''(\xi)$, we come to:

$$f''(x) = 2c_1 + 6d_1 x \quad then \quad \begin{cases} f_1''(\xi) = 2\beta_2 + 6\beta_3\xi \\ f_2''(\xi) = 2(\beta_2 - 3\beta_4\xi) + 6(\beta_3 + \beta_4)\xi \end{cases} \tag{31}$$

$$\begin{aligned} f_2''(\xi) &= 2(\beta_2 - 3\beta_4\xi) + 6(\beta_3 + \beta_4)\xi && (32)\\ &= 2\beta_2 - 6\beta_4\xi + 6\beta_3\xi + 6\beta_4\xi && (33)\\ &= 2\beta_2 + 6\beta_3\xi + (6\beta_4\xi - 6\beta_4\xi) && (34)\\ f_2''(\xi) &= 2\beta_2 + 6\beta_3\xi && (35) \end{aligned}$$

We now see that the second derivative $f''(x)$ is continuous at knot ξ, which is to say $f_1''(\xi) = f_2''(\xi)$:

$$f_2''(\xi) = 2\beta_2 + 6\beta_3\xi = f_1''(\xi)$$

2

(sketches on following page)

- **A**. With $\lambda = \infty$, the second term will dominate the above equation and the RSS will be ignored. Since $g^0 = g$, this comes out to finding $g(x)$ that minimizes the integral of $g(x)$. Therefore, $g(x) = 0$.
- **B**. With $\lambda = \infty$ and $m = 1$, the second term will dominate the above equation and the RSS will be ignored. This then becomes a problem of finding a function $g(x)$ where $\int g'(x)$ is minimized. Therefore, $g(x) = c$ (a flat line) where c is a constant, ensuring that $g'(x) = 0$.
- **C**. With $\lambda = \infty$ and $m = 2$, the second term will dominate the above equation and the RSS will be ignored. This then becomes a problem of finding a function $g(x)$ where $\int g''(x)$ is minimized.

  If we work backwards conceptually, we will see that $g(x) = \beta_0 + \beta_1 x$. Since $\int g''(x)$ must be minimized, $g''(x) = 0$. Therefore, $g'(x) = c$ where c is some constant. This implies that $g(x)$ must have a constant slope, c aka β_1. Therefore, $g(x) = \beta_0 + \beta_1 x$
- **D**. With $\lambda = \infty$ and $m = 3$, the second term will dominate the above equation and the RSS will be ignored. This then becomes a problem of finding a function $g(x)$ where $\int g'''(x)$ is minimized. Therefore, $g(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, $g(x)$ will be quadratic in some sense

  Once again, working backwards conceptually, if the goal is to minimize $\int g'''(x)$, then $g'''(x) = 0$. Therefore, $g''(x) = c$, where c is some constant. This implies that $g'(x)$ must have a constant slope, c. if $g'(x)$ has a constant slope, then $g(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. Having a quadratic equation means that the slope of $g(x)$ is changing at a fixed rate, which satisfies our condition that $g'(x) = c$.
- **E**. With $\lambda = 0$ and $m = 3$, the second term in the equation is completely ignored, and $g(x)$ becomes the line that interpolates all data points.

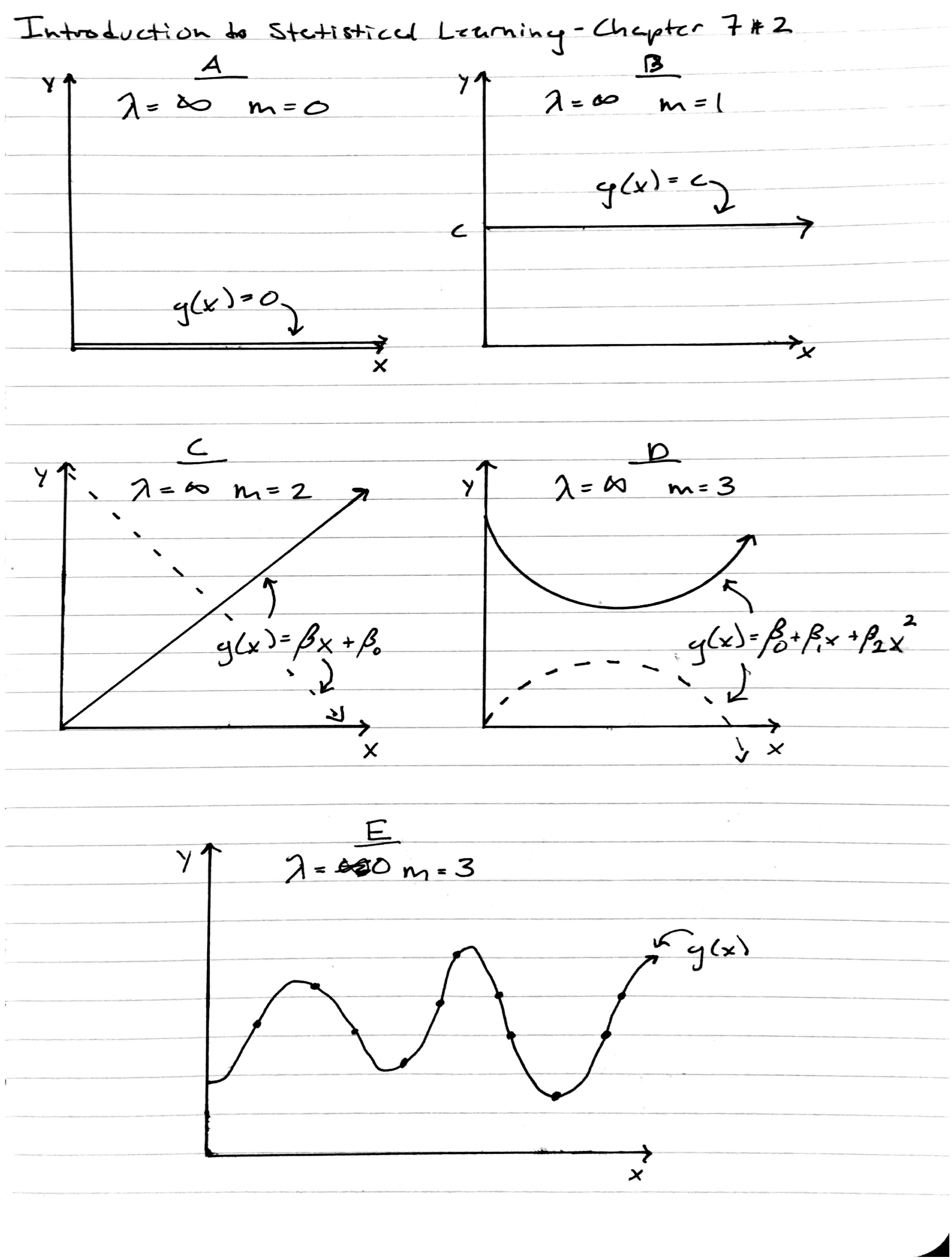
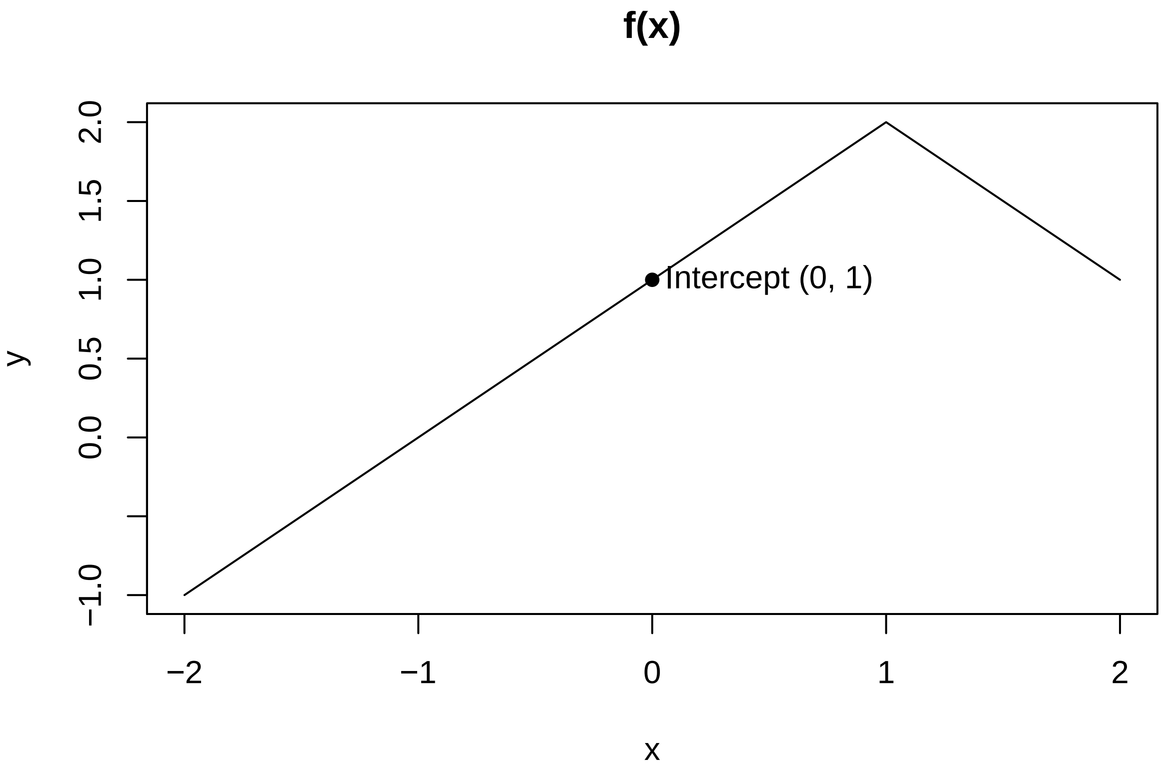


Figure 1: “Conceptual Exercise 2”

3

$$f(x) = 1 + x + \begin{cases} -2(x-1)^2, & x \geq 1 \\ 0, & otherwise \end{cases}$$

The intercept is at $y = 1$, $f(x)$ is linear with a slope equal to 1 up to $x = 1$, after which it becomes quadratic.

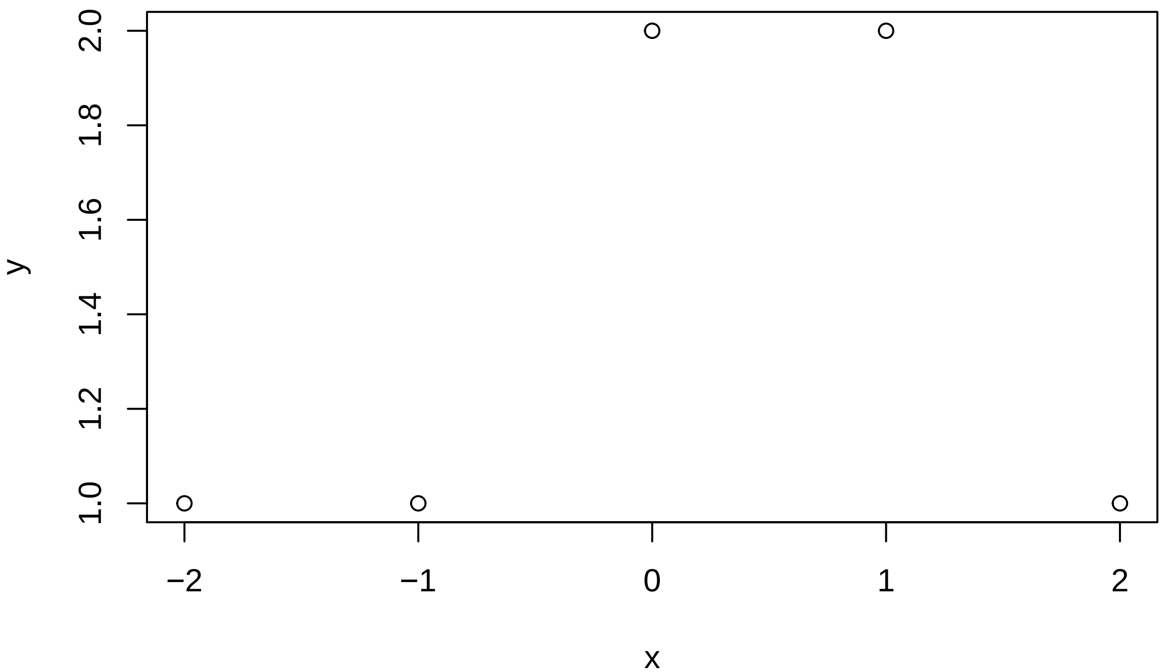


4

$$f(x) = \beta_0 + \beta_1 b_1(x) + \beta_2 b_2(x) \tag{36}$$

$$f(x) = 1 + b_1(x) + 3b_2(x) \quad where \quad \begin{cases} b_1(x) = I(0 \le x \le 2) - (x-1)I(1 \le x \le 2) \\ b_2(x) = (x-3)I(3 \le x \le 4) + I(4 < x \le 5) \end{cases} \tag{37}$$

```r
x <- -2:2
y <- c(1,1,2,2,1)
plot(x, y)
```



5

$$\hat{g_1} = \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \int [g^3(x)]^2 dx \right) \tag{38}$$

$$\hat{g_2} = \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \int [g^4(x)]^2 dx \right) \tag{39}$$

- **A**. As $\lambda \to \infty$, $\hat{g_2}$ will have a smaller training RSS. This is because $\hat{g_2}$ has one more degree of freedom than $\hat{g_1}$; in other words, it is allowed to be more flexible thatn $\hat{g_1}$.
- **B**. As $\lambda \to \infty$, $\hat{g_1}$ will most likely have a lower test RSS, although this is less certain than part **A**. It will most likely have a lower test RSS because we are constraining it more, which is to say there is less of a chance that it incorporates the error term ϵ into the model itself.
- **C**. If $\lambda = 0$, the two equations are the same so they will have the same training and test RSS (one that interpolates all data points).