7.9 EXERCIESE – CONCEPTUAL

1. It was mentioned in the chapter that a cubic regression spline with one knot at ξ can be obtained using a basis of the form x, x2, x3, (x − ξ)3 +, where (x − ξ)3 + = (x − ξ)3 if x>ξ and equals 0 otherwise. We will now show that a function of the form

$$\mathbf{f(x) = \beta 0 + \beta 1x + \beta 2x^2 + \beta 3x^3 + \beta 4(x - \xi)^3 +}$$

is indeed a cubic regression spline, regardless of the values of β0, β1, β2, β3, β4.

(a) Find a cubic polynomial

$$\mathrm{f1(x) = a1 + b1x + c1x^2 + d1x^3}$$

such that f(x) = f1(x) for all x ≤ ξ. Express a1, b1, c1, d1 in terms of β0, β1, β2, β3, β4.

a.) $f_1(x) = a_1 + b_1x + c_1x^2 + d_1x^3$ —(1)

for $x < \epsilon$,

$(x - \epsilon)^3 = 0$

so, $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ —(2)

from (1) & (2)

$a_1 = \beta_0$, $b_1 = \beta_1$, $c_1 = \beta_2$, $d_1 = \beta_3$

(b) Find a cubic polynomial

$$\mathrm{f2(x) = a2 + b2^x + c2x^2 + d2x^3}$$

such that f(x) = f2(x) for all x>ξ. Express a2, b2, c2, d2 in terms of β0, β1, β2, β3, β4. We have now established that f(x) is a piecewise polynomial.

b.) $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - \epsilon)^3 \ \{\text{for } x > \epsilon\}$

$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x^3 - \epsilon^3 - 3\epsilon x^2 + 3\epsilon^2 . x)$

$f(x) = \beta_0 - \epsilon^3 \beta_4 + (\beta_1 + 3\epsilon^2 \beta_4)x + (\beta_2 \cancel{+} - 3\epsilon\beta_4)x^2 +$

$(\beta_3 + \beta_4)x^3$ - (1)

$f_2(x) = a_2 + b_2 x + c_2 x^2 + d_2 x^3$ - (2)

comparing (1) & (2),

$a_2 = \beta_0 - \epsilon^3 \beta_4$

$b_2 = \beta_1 + 3\epsilon^2 \beta_4$

$c_2 = \beta_2 - 3\epsilon\beta_4$

$d_2 = \beta_3 + \beta_4$

Now, $f(x)$ is,

$$f(x) = \begin{cases} \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, & x \leq \epsilon \\ (\beta_0 - \epsilon^3 \beta_4) + (\beta_1 + 3\epsilon^2 \beta_4)x + \\ \quad (\beta_2 - 3\epsilon\beta_4)x^2 + (\beta_3 + \beta_4)x^3, & x > \epsilon \end{cases}$$

(c) Show that f1(ξ) = f2(ξ). That is, f(x) is continuous at ξ c.) $f_1(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

$$f_1(\epsilon) = \beta_0 + \beta_1 \epsilon + \beta_2 \epsilon^2 + \beta_3 \epsilon^3 \quad -(1)$$

$$f_2(x) = (\beta_0 - \epsilon^3 \beta_4) + (\beta_1 + 3\epsilon^2 \beta_4) x + (\beta_2 - 3\epsilon \beta_4) x^2 + (\beta_3 + \beta_4) x^3$$

$$f_2(\epsilon) = \beta_0 - \cancel{\epsilon^3 \beta_4} + \beta_1 \epsilon + \cancel{3\epsilon^3 \beta_4} + \epsilon^2 \beta_2 - \cancel{3\epsilon^3 \beta_4} + \beta_3 \epsilon^3 + \cancel{\beta_4 \epsilon^3}$$

$$f_2(\epsilon) = \beta_0 + \beta_1 \epsilon + \beta_2 \epsilon^2 + \beta_3 \epsilon^3 \quad -(2)$$

from (1) & (2),

$$f_1(\epsilon) = f_2(\epsilon)$$

(d) Show that f1(ξ) = f2(ξ). That is, f(x) is continuous at ξ

d.) $f_1'(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$

$$f_1'(\epsilon) = \beta_1 + 2\beta_2 \epsilon + 3\beta_3 \epsilon^2 \quad -(1)$$

$$f_2'(x) = \beta_1 + 3\epsilon^2 \beta_4 + 2(\beta_2 - 3\epsilon \beta_4) x + 3(\beta_3 + \beta_4) x^2$$

$$f_2'(\epsilon) = \beta_1 + \cancel{3\beta_4 \epsilon^2} + 2\beta_2 \epsilon - \cancel{6\beta_4 \epsilon^2} + 3\beta_3 \epsilon^2 + \cancel{3\beta_4 \epsilon^2}$$

$$f_2'(\epsilon) = \beta_1 + 2\beta_2 \epsilon + 3\beta_3 \epsilon^2 \quad -(2)$$

from (1) & (2),

$$f_1'(\epsilon) = f_2'(\epsilon)$$

(e) Show that f1 (ξ) = f2 (ξ). That is, f(x) is continuous at ξ e.) $f_1''(x) = 2\beta_2 + 6\beta_3 x \Rightarrow f_1''(\epsilon) = 2\beta_2 + 6\beta_3\epsilon$ ①

$f_2''(x) = 2(\beta_2 - 3\epsilon\beta_4) + 6(\beta_3 + \beta_4)x$

$f_2''(\epsilon) = 2\beta_2 - 6\epsilon\beta_4 + 6\beta_3\epsilon + 6\epsilon\beta_4$

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

from ① & ②

$f_1''(\epsilon) = f_2''(\epsilon)$

2. Suppose that a curve ˆg is computed to smoothly fit a set of n points using the following formula

:gˆ = arg ming n i = 1 (yi − g(xi))2 + λ − g(m)(x)2 dx ,

where g(m) represents the mth derivative of g (and g(0) = g). Provide example sketches of ˆg in each of the following scenarios

$n.x^2 + \beta_3 x$

2.
$$\hat{g} = \arg\min_g \left(\underbrace{\sum_{i=1}^{n}(y_i - g(x_i))^2}_{\text{This term forces the curve to be flexible}} + \lambda \underbrace{\int [g^{(m)}(x)]^2 dx}_{\text{This term forces the curve to be smooth.}} \right)$$

a.) $\lambda = \infty$, $m = 0$

- first term has no effect. ($\lambda \to \infty$)
  ~~mini~~ $g^0(x)$ will be minimized to 0
  hence, $\hat{g} = 0$

b.) $\lambda = \infty$, $m = 1$

- $g'(x) \to 0$
  integrating the above eq.
  $g(x) = c$ (where c is constant)

c.) $\lambda = \infty$, $m = 2$

- $g^2(x) \to 0$

integrating $\Rightarrow$ $g'(x) = c$

integrating again $\Rightarrow$ $g(x) = cx + d$

d.) $\lambda = \infty$, $m = 3$

- $g^3(x) \to 0$

integrating $g^2(x) = c$

〃 again, $g'(x) = cx + d$

〃 〃, $g(x) = cx^2 + dx + e$

〃 〃, $g(x) = cx^2 + dx + e$

e.) $\lambda = 0$, $m = 3$

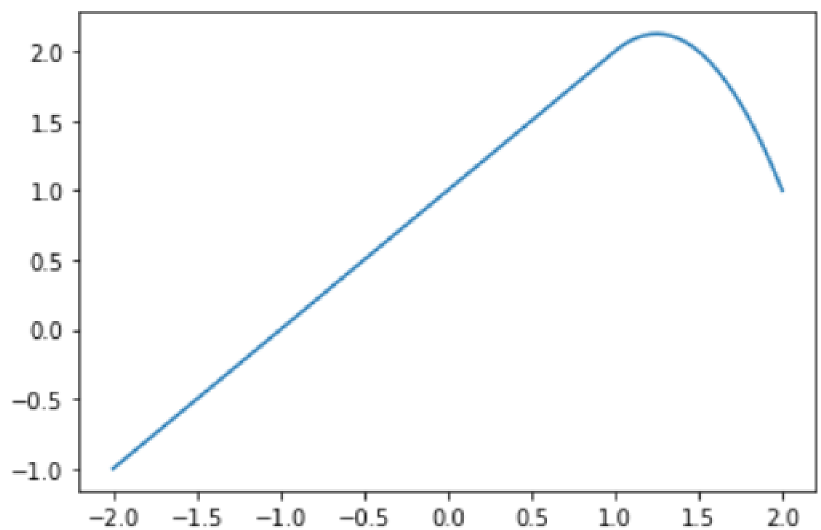
Now, curve will only be determined by first term. Hence $\hat{g}(x)$ will be the one which minimize RSS

3. Suppose we fit a curve with basis functions b1(X) = X, b2(X) = (X − 1)2I(X ≥ 1). (Note that I(X ≥ 1) equals 1 for X ≥ 1 and 0 otherwise.) We fit the linear regression model

$$Y = \beta0 + \beta1b1(X) + \beta2b2(X) + \text{error}$$

and obtain coefficient estimates βˆ0 = 1, βˆ1 = 1, βˆ2 = −2. Sketch the estimated curve between X = −2 and X = 2. Note the intercepts, slopes, and other relevant information.

```
In [17]: X = np.linspace(-2,2,200)

         def b1(x):
             return x

         def b2(x):
             return (x-1)**2 * (x >= 1)

         beta_0 = 1
         beta_1 = 1
         beta_2 = -2

         y = beta_0 + beta_1*b1(X) + beta_2*b2(X)

         plt.plot(X,y)

Out[17]: [<matplotlib.lines.Line2D at 0x1d86e7ee160>]
```



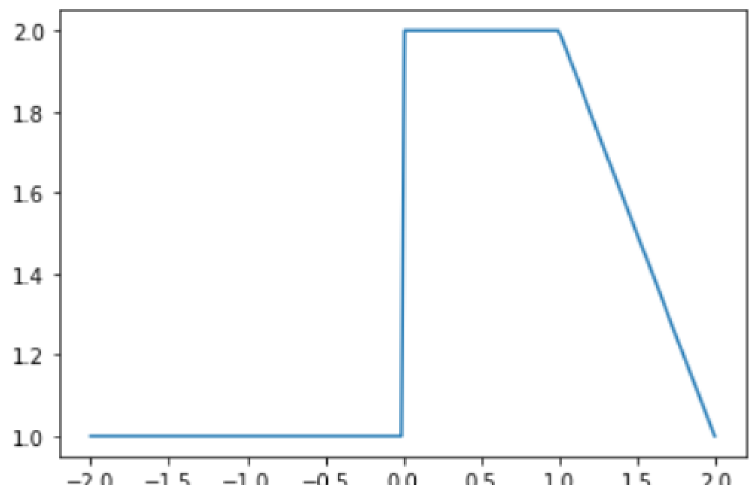
4. Suppose we fit a curve with basis functions $b1(X) = I(0 \le X \le 2) - (X - 1)I(1 \le X \le 2)$, $b2(X)=(X - 3)I(3 \le X \le 4) + I(4 < X \le 5)$. We fit the linear regression model

$$Y = \beta 0 + \beta 1 b1(X) + \beta 2 b2(X) + \text{error}$$

and obtain coefficient estimates $\hat{\beta}0 = 1$, $\hat{\beta}1 = 1$, $\hat{\beta}2 = 3$. Sketch the estimated curve between $X = -2$ and $X = 2$. Note the intercepts, slopes, and other relevant information.

In [43]:
```
X = np.linspace(-2,2,200)

def b1(x):
    return ((x >= 0) & (x <= 2)) - (x-1)*((x >= 1) & (x <= 2))

def b2(x):
    return (x-3)*((x >= 3) & (x <= 4)) + ((x > 4) & (x <= 5))

beta_0 = 1
beta_1 = 1
beta_2 = 3

y = beta_0 + beta_1*b1(X) + beta_2*b2(X)

plt.plot(X,y)
```

Out[43]: [<matplotlib.lines.Line2D at 0x1d86e870f60>]



5. Consider two curves, ˆg1 and ˆg2, defined by

(equations in the book)

where g(m) represents the mth derivative of g.

(a) As λ → ∞, will ˆg1 or ˆg2 have the smaller training RSS?

When lamda will tend to infinity the first term will have no effect to the final curve, which will be totally effected by the second term. In first case, g1 will minimize the third degree derivative of the fitted line. Which will lead to a cubic plot. In second case, it will try to minimize the fourth order derivative to zero, hence the resulting curve will be a cubic curve.

Since, we know as degree inceases, flexibility of the data increases, and line fits the training data better. Therefore, g2 will have a smaller RSS

(b) As $\lambda \rightarrow \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller test RSS?

This cannot be told. The test RSS depends on the true relationship between the predictors and response.

(c) For $\lambda = 0$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training and test RSS?

When lambda is 0, both the functions are similar, and will have value of training and test RSS.