CS 5565, HW6(Moving Beyond Linearity) 85pts (65pts Undergrad)

Name Anusha Muppalla

16286311

1. (15 points total) Suppose that a curve $\hat{g}$ is computed to smoothly fit a set of n points using the following formula:

$$\hat{g} = \arg\min_{g} \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(m)}(x) \right]^2 dx \right)$$

where $g^{(m)}$ represents the mth derivative of g (and $g^{(0)} = g$). Provide example sketches of $\hat{g}$ in each of the following scenarios.

   (a) (3 points) $\lambda = \infty,\ m = 0$

   (b) (3 points) $\lambda = \infty,\ m = 1$

   (c) (3 points) $\lambda = \infty,\ m = 2$

   (d) (3 points) $\lambda = \infty,\ m = 3$

   (e) (3 points) $\lambda = 0,\ m = 3$

2. (10 points total) Suppose we fit a curve with basis functions $b_1(X) = X$, $b_2(X) = (X-1)^2 I(X \geq 1)$. (Note that $I(X \geq 1)$ equals 1 for $X \geq 1$ and 0 otherwise.) We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \epsilon,$$

and obtain coefficient estimates $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -2$.

Sketch the estimated curve between $X = -2$ and $X = 2$. Note the intercepts, slopes, and other relevant information.

3. (10 points total) Suppose we fit a curve with basis functions $b_1(X) = I(0 \leq X \leq 2) - (X-1)I(1 \leq X \leq 2)$, $b_2(X) = (X-3)I(3 \leq X \leq 4) + I(4 < X \leq 5)$. We fit the linear regression model

$$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \epsilon,$$

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.

4. (15 points total) Consider two curves, $\hat{g}_1$ and $\hat{g}_2$ defined by

$$\begin{aligned} \hat{g}_1 &= \arg\min_{g} \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(3)}(x) \right]^2 dx \right), \\ \hat{g}_2 &= \arg\min_{g} \left(\sum_{i=1}^{n} (y_i - g(x_i))^2 + \lambda \int \left[g^{(4)}(x) \right]^2 dx \right), \end{aligned}$$

where $g^{(m)}$ represents the mth derivative of g.

(a) (5 points) As $\lambda \longrightarrow \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training RSS?

(b) (5 points) As $\lambda \longrightarrow \infty$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller test RSS?

(c) (5 points) For $\lambda = 0$, will $\hat{g}_1$ or $\hat{g}_2$ have the smaller training and test RSS?

5. (20 points total (extra credit undergraduate)) 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, x^2, x^3, $(x-\xi)^3_+$, where $(x-\xi)^3_+ = (x-\xi)^3$ if $x > \xi$ and equals 0 otherwise. We will now show that a function of the form

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

is indeed a cubic regression spline, regardless of the values of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

(a) (5 points) Find a cubic polynomial

$$f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$$

such that $f(x) = f_1(x)$ for all $x \leq \xi$. Express a_1, b_1, c_1, d_1 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

(b) (5 points) Find a cubic polynomial

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

such that $f(x) = f_2(x)$ for all $x > \xi$. Express a_2, b_2, c_2, d_2 in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$.

(c) (5 points) Show that $f_1(\xi) = f_2(\xi)$. That is, $f(x)$ is continuous at ξ.

(d) (5 points) Show that $f_1'(\xi) = f_2'(\xi)$. That is, $f'(x)$ is continuous at ξ.

(e) (5 points) Show that $f_1''(\xi) = f_2''(\xi)$ i.e. $f(x)$ is indeed a cubic spline.

6. (15 points) Construct a natural cubic spline that passes through the points (1,2),(2,4), and (3,3).

HOMEWORK ASSIGNMENT 6

Anusha muppalla
16286311

1)

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

In this case $\hat{g}=0$ because a large smoothing parameter forces $g^{(0)}(x)\to 0$

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

Here, $\hat{g}=c$ because a large smoothing parameter forces $g^{(1)}(x)\to 0$

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

in this case, $\hat{g}=cx+d$ because a large smoothing parameter forces $g^{(2)}(x)\to 0$

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

Here $\hat{g}=cx^2+dx+e$ because a large smoothing parameter forces $g^{(3)}(x)\to 0$

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

The penalty term doesn't play any role, so in this case g in the interpolating spline

4)

a) The smoothing spline $\hat{g}_2$ will probably have the smaller training RSS because it will be a higher order polynomial due to order of the penalty term

b) $\lambda \to \infty$ $\hat{g}_1$ & $\hat{g}_2$ have the smaller test RSS
As mentioned above we expect $\hat{g}_2$ to be more flexible, so it may overfit the data. it will probably be $\hat{g}_1$ that have the smaller test RSS

c) For $\lambda = 0$, $\hat{g}_1$ & $\hat{g}_2$ have smaller training & test RSS
if $\lambda = 0$, we have $\hat{g}_1 = \hat{g}_2$, so they will have the same training & test RSS.

5)

a) $f_1(x) = a_1 + b_1 x + c_1 x^2 + d_1 x^3$
By comparing $f_1(x)$ and $f(x)$
$a_1 = \beta_0$, $b_1 = \beta_1$, $c_1 = \beta_2$, $d_1 = \beta_3$

b) by comparing
$a_2 = \beta_0 - \beta_4 \xi^3$, $b_2 = \beta_1 + 3\xi^2 \beta_4$, $c_2 = \beta_2 - 3\beta_4 \xi$, $d_2 = \beta_3 + \beta_4$

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

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

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

$$f_2''(\varepsilon) = 2(\beta_2 - 3\beta_4\varepsilon) + 6(\beta_3 + \beta_4)\varepsilon$$
$$= 2\beta_2 + 6\beta_3\varepsilon.$$

6) given points $(1,2)$ $(2,4)$ $(3,3)$

$[1,2]$ $s_0(x) = a_0 + b_0(x-1) + c_0(x-1)^2 + d_0(x-1)^3$

$[2,4]$ $s_1(x) = a_1 + b_1(x-2) + c_1(x-2)^2 + d_1(x-2)^3$

$s_0(1) = 2$ $a_0 = 2$

$s_0(2) = 4$ $a_0 + b_0 + c_0 + d_0$

$2 + b_0 + c_0 + d_0 \Rightarrow b_0 + c_0 + d_0 = 2$ — ①

$s_1(2)$. $a_1 = 4$

$s_1(3) = 3$ $a_1 + b_1 + c_1 + d_1 = 3$

$b_1 + c_1 + d_1 = 3 - 4 = -1$

$b_1 + c_1 + d_1 = 1$ — ②

$s_0'(x) = b_0 + 2c_0(x-1) + 3d_0(x-1)^2$

$s_1'(x) = b_1 + 2c_1(x-2) + 3d_1(x-2)^2$

$s_0'(2) = s_1'(2) \Rightarrow [b_0 + 2c_0 + 3d_0 = b_1]$ — ③

$s_0''(x) = 2c_0 + 6d_0(x-1)$

$s_1''(x) = 2c_1 + 6d_1(x-2)$

$\therefore s_0''(2) = s_1''(2)$ [$\because$ Boundary condition]

$\Rightarrow 2c_0 + 6d_0 = 2c_1$

$\boxed{c_0 + d_0 3 = c_1}$ — ④

another $\boxed{\text{Boundary condition}}$ $s_i''(1)=0$; $s_i''(3)=0$

$c_0=0$ $\quad$ $2c_1+6d_1=0$

$c_0=0$ $\quad$ $[c_1+3d_1=0]$ - (5)

$[b_0+3d_0=b_1]$ - (6)

$b_0+d_0=2$;

$b_1+c_1+d_1=-1$

$c_1+3d_1=0$

$3d_0=c_1$

$b_0+3d_0=b_1$

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 3 & 0 & -1 & 0 \\ 1 & 3 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ d_0 \\ b_1 \\ c_1 \\ d_1 \end{bmatrix} = \begin{bmatrix} +2 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$AX=B \Rightarrow X=A^{-1}B$

By calculating inverse and multiplying the fetched matrix with "B", we got below value.

$b_0=2.75$

$d_0=-0.75$

$b_1=0.25$

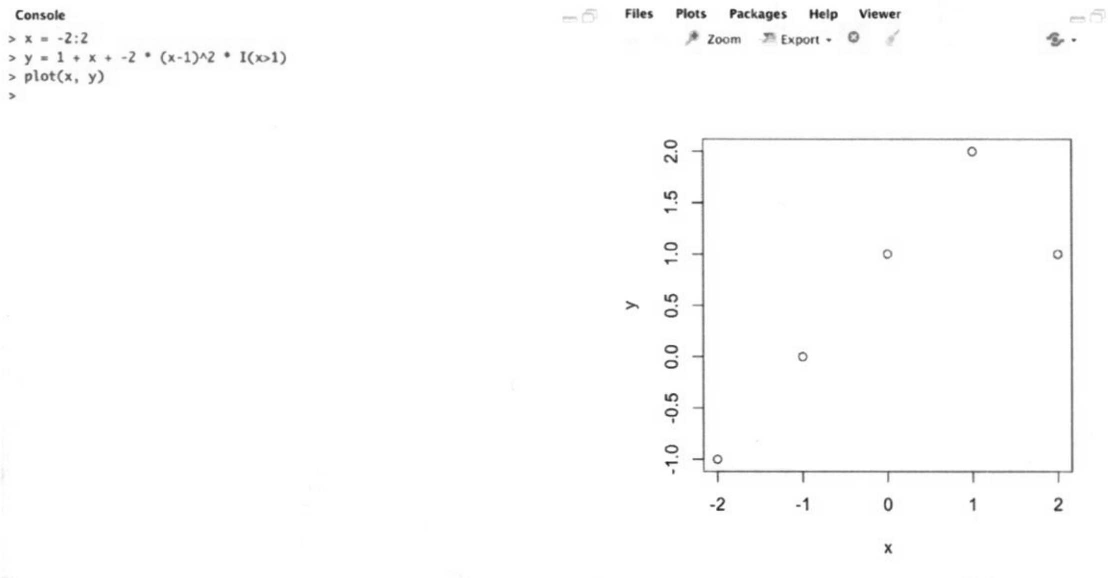
$c_1=0.50$

$d_1=-0.75$.

2. (10 points total) Suppose we fit a curve with basis functions $b_1(X) = X$, $b_2(X) = (X-1)^2 I(X \geq 1)$. (Note that $I(X \geq 1)$ equals 1 for $X \geq 1$ and 0 otherwise.) We fit the linear regression model

$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \varepsilon$, and obtain coefficient estimates $\hat{\beta}_0 = 1$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -2$.

Sketch the estimated curve between $X = -2$ and $X = 2$. Note the intercepts, slopes, and other relevant information.



The curve is linear between -2 and 1 : $y=1+x$ and quadratic between 1, and 2 : $y=1+x-2(x-1)2$.

3.(10 points total) Suppose we fit a curve with basis functions $b_1(X) = I(0 \leq X \leq 2)-(X-1)I(1 \leq X \leq 2), b_2(X)=(X-3)I(3 \leq X \leq 4)+I(4<X \leq 5)$. Wefitthe linear regression model

$Y = \beta_0 + \beta_1 b_1(X) + \beta_2 b_2(X) + \varepsilon$, 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.

Console

```
> x = -2:2
> y = 1 + x + -2 * (x-1)^2 * I(x>1)
> plot(x, y)
> x = -2:2
> y = c(1 + 0 + 0, # x = -2
+       1 + 0 + 0, # x = -1
+       1 + 1 + 0, # x = 0
+       1 + (1-0) + 0, # x = 1
+       1 + (1-1) + 0 # x =2
+ )
> plot(x,y)
>
```