Marco Reina 7066486
Samuele Serri 7069839

1

a)

$f_1(x) = a_1 + b_1x + c_1x^2 + d_1x^3 \quad x \leq \xi$
if $x \leq \xi$ then $(x-\xi)^3_+ = 0$ by definition, so we can simply set:
$a_1 = \beta_0$
$b_1 = \beta_1$
$c_1 = \beta_2$
$d_1 = \beta_3$

b)

$f_2(x) = a_2 + b_2x + c_2x^2 + d_2x^3 \quad x > \xi$
if $x > \xi$ then $(x-\xi)^3_+ = (x-\xi)^3$ by definition.
Thus $f(x) = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4(x-\xi)^3 =$
$= \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4(x^3 - \xi^3 - 3x^2\xi + 3x\xi^2) =$
$= \beta_0 - \beta_4\xi^3 + x(\beta_1 + 3\beta_4\xi^2) + x^2(\beta_2 - 3\beta_4\xi) + x^3(\beta_3 + \beta_4)$
And we set:
$a_2 = \beta_0 - \beta_4\xi^3$
$b_2 = \beta_1 + 3\beta_4\xi^2$
$c_2 = \beta_2 - 3\beta_4\xi$
$d_2 = \beta_3 + \beta_4$

c)

$\lim_{x \longrightarrow \xi} f_2(x) = \lim_{x \longrightarrow \xi} \beta_0 - \beta_4\xi^3 + x(\beta_1 + 3\beta_4\xi^2) + x^2(\beta_2 - 3\beta_4\xi) + x^3(\beta_3 + \beta_4) =$
$\beta_0 - \beta_4\xi^3 + \xi(\beta_1 + 3\beta_4\xi^2) + \xi^2(\beta_2 - 3\beta_4\xi) + \xi^3(\beta_3 + \beta_4) = \beta_0 + \beta_1\xi + \beta_2\xi^2 + \beta_3\xi^3 =$

$= \lim_{x \longrightarrow \xi} f_1(x).$
Hence f is continuous in ξ.

d)

$f_1^{'}(x) = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$ and
$f_2^{'}(x) = \beta_1 + 3\beta_4 \xi^2 + 2x(\beta_2 - 3\beta_4 \xi) + 3x^2(\beta_3 + \beta_4)$

$\lim_{x \longrightarrow \xi} f_2^{'}(x) = \beta_1 + 3\beta_4\xi^2 + 2\xi(\beta_2 - 3\beta_4\xi) + 3\xi^2(\beta_3 + \beta_4) = \beta_1 + 2\beta_2\xi + 3\beta_3\xi^2 =$
$= \lim_{x \longrightarrow \xi} f_1^{'}(x)$

e)

$f_1^{''}(x) = 2\beta_2 + 6\beta_3 x$ and
$f_2^{''}(x) = 2(\beta_2 - 3\beta_4\xi) + 6x(\beta_3 + \beta_4)$

$\lim_{x \longrightarrow \xi} f_2^{''}(x) = 2(\beta_2 - 3\beta_4\xi) + 6\xi(\beta_3 + \beta_4) = 2\beta_2 + 6\beta_3\xi = \lim_{x \longrightarrow \xi} f_1^{''}(x)$

2

a)

$\hat{g}_1$ will have smaller training RSS because is less regular than $\hat{g}_1$ and will fit the training data better (i.e. small bias).

b)

$\hat{g}_2$ will have smaller test RSS because is more regular and less sensitive to over-fitting.

c)

For $\lambda = 0$ $\hat{g}_1 = \hat{g}_2$ and RSS on train is equal to zero because the function g will be an interpolation function of the given points.