

36-402 Homework 5

Eu Jing Chua

eujingc

February 19, 2019

Question 1

Q1 a)

$$\frac{1}{2h} \int_{x_0-h}^{x_0+h} (m + t(x - x_0) + c(x - x_0)^2) dx = \frac{1}{2h} \int_{-h}^h (m + tu + cu^2) du, \quad u = x - x_0 \quad (1)$$

$$= \frac{1}{2h} \left[mu + \frac{t}{2}u^2 + \frac{c}{3}u^3 \right]_{-h}^h \quad (2)$$

$$= \frac{1}{2h} (2mh + \frac{2}{3}ch^3) \quad (3)$$

$$= m + \frac{1}{3}ch^2 \quad (4)$$

$$\implies k = \frac{1}{3} \quad (5)$$

Q1 b)

$$\mathbb{E}[m(X)] = \int_{x_0-h}^{x_0+h} m(x) \frac{1}{2h} dx \quad (6)$$

$$= \frac{1}{2h} \int_{x_0-h}^{x_0+h} m(x) dx \quad (7)$$

$$\approx \frac{1}{2h} \int_{x_0-h}^{x_0+h} m(x_0) + m'(x_0)(x - x_0) + \frac{1}{2}m''(x_0)(x - x_0)^2 dx \quad (8)$$

$$= m(x_0) + \frac{m''(x_0)}{3}h^2 \quad (9)$$

Q1 c) Let $A_{x_0} = \{i : |X_i - x_0| \leq h\}$, so $\hat{\mu}(x_0) = \frac{1}{|A_{x_0}|} \sum_{i \in A_{x_0}} Y_i$

$$\mathbb{E}[\hat{\mu}(x_0)] = \mathbb{E}\left[\frac{1}{|A_{x_0}|} \sum_{i \in A_{x_0}} Y_i\right] \quad (10)$$

$$= \mathbb{E}\left[\frac{1}{|A_{x_0}|} \sum_{i \in A_{x_0}} \mu(X_i) + \epsilon_i\right] \quad (11)$$

$$= \frac{1}{|A_{x_0}|} \sum_{i \in A_{x_0}} \mathbb{E}[\mu(X_i)] + \mathbb{E}[\epsilon_i] \quad (12)$$

$$= \mathbb{E}[\mu(X) \mid |X - x_0| \leq h] + \mathbb{E}[\epsilon \mid |X - x_0| \leq h] \quad (13)$$

$$\approx \mu(x_0) + \frac{\mu''(x_0)}{3} h^2 + 0 \quad (14)$$

$$= \mu(x_0) + O(h^2) \quad (15)$$

$$\text{Bias}[\hat{\mu}(x_0)] = \mathbb{E}[\hat{\mu}(x_0)] - \mu(x_0) \quad (16)$$

$$\approx O(h^2) \quad (17)$$

Q1 d)

$$\mathbb{E}[m(X)] = \int_{x_0-h}^{x_0+h} m(x) f(x) dx \quad (18)$$

$$\approx \int_{x_0-h}^{x_0+h} \left(m(x_0) + m'(x_0)(x - x_0) + \frac{1}{2} m''(x_0)(x - x_0)^2 \right) \left(f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 \right) dx \quad (19)$$

$$\approx \int_{x_0-h}^{x_0+h} \left(m(x_0) f(x_0) + \left(\frac{1}{2} m(x_0) f''(x_0) + m'(x_0) f'(x_0) + \frac{1}{2} m''(x_0) f(x_0) \right) (x - x_0)^2 \right) dx \quad (20)$$

$$= 2hm(x_0)f(x_0) + \frac{m(x_0)f''(x_0) + 2m'(x_0)f'(x_0) + m''(x_0)f(x_0)}{3} h^3 \quad (21)$$

$$\text{Let } B \text{ be the event that } |X - x_0| \leq h \quad (22)$$

$$f(x \mid B) \approx \frac{f(x)}{2hf(x_0)} \quad (23)$$

$$f'(x \mid B) \approx \frac{f'(x)}{2hf(x_0)} \quad (24)$$

$$f''(x \mid B) \approx \frac{f''(x)}{2hf(x_0)} \quad (25)$$

$$\mathbb{E}[\hat{\mu}(x_0)] = \mathbb{E}[\mu(X) \mid B] \text{ as shown in c)} \quad (26)$$

$$= 2h\mu(x_0)f(x_0 \mid B) + \frac{\mu(x_0)f''(x_0 \mid B) + 2\mu'(x_0)f'(x_0 \mid B) + \mu''(x_0)f(x_0 \mid B)}{3} h^3 \quad (27)$$

$$\approx \mu(x_0) + \frac{\mu(x_0)f''(x_0) + 2\mu'(x_0)f'(x_0) + \mu''(x_0)f(x_0)}{6f(x_0)} h^2 \quad (28)$$

$$= \mu(x_0) + O(h^2) \quad (29)$$

$$\text{Bias}[\hat{\mu}(x_0)] = \mathbb{E}[\hat{\mu}(x_0)] - \mu(x_0) \quad (30)$$

$$\approx O(h^2) \quad (31)$$

Question 2

Q2 a)

Let $u = x_1 - x_{01}$ and $v = x_2 - x_{02}$.

$$\frac{1}{(2h)^2} \int_{B_h} u(\vec{x}) d\vec{x} \quad (32)$$

$$= \frac{1}{(2h)^2} \int_{x_{01}-h}^{x_{01}+h} \int_{x_{02}-h}^{x_{02}+h} u(\vec{x}) d\vec{x} \quad (33)$$

$$= \frac{1}{(2h)^2} \int_{-h}^h \int_{-h}^h (m + t_1 u + t_2 v + c_1 u^2 + c_2 v^2 + c_3 uv) du dv \quad (34)$$

$$= \frac{1}{(2h)^2} \int_{-h}^h (2hm + 2ht_2 v + \frac{2c_1}{3} h^3 + 2hc_2 v^2) dv \quad (35)$$

$$= \frac{1}{2h} \int_{-h}^h (m + t_2 v + \frac{c_1}{3} h^2 + c_2 v^2) dv \quad (36)$$

$$= \frac{1}{2h} (2hm + \frac{2c_1}{3} h^3 + \frac{2c_2}{3} h^3) \quad (37)$$

$$= m + \left(\frac{c_1}{3} + \frac{c_2}{3} \right) h^2 \quad (38)$$

Hence $k_1 = \frac{1}{3}$ and $k_2 = \frac{1}{3}$ and $k_3 = 0$.

Q2 b)

$$m(\vec{x}) \approx m(\vec{x}_0) + (x_1 - x_{01})m_{x_1}(\vec{x}_0) + (x_2 - x_{02})m_{x_2}(\vec{x}_0) + \frac{1}{2} ((x_1 - x_{01})^2 m_{x_1 x_1}(\vec{x}_0) + (x_2 - x_{02})^2 m_{x_2 x_2}(\vec{x}_0)) \quad (39)$$

$$+ (x_1 - x_{01})(x_2 - x_{02})m_{x_1 x_2}(\vec{x}_0) \quad (40)$$

$$\mathbb{E} [m(\vec{X})] = \int_{B_h} m(\vec{x}) f(\vec{x}) d\vec{x} \quad (41)$$

$$= \frac{1}{(2h)^2} \int_{B_h} m(\vec{x}) d\vec{x} \quad (42)$$

$$\approx m(\vec{x}_0) + \frac{1}{3} \left(\frac{m_{x_1 x_1}(\vec{x}_0)}{2} + \frac{m_{x_2 x_2}(\vec{x}_0)}{2} \right) h^2 \quad (43)$$

$$= m(\vec{x}_0) + O(h^2) \quad (44)$$

Q2 c)

Let $A_{\vec{x}_0} = \{i : |X_{i1} - x_{01}| \leq h \wedge |X_{i2} - x_{02}| \leq h\}$, so $\hat{\mu}(x_0) = \frac{1}{|A_{\vec{x}_0}|} \sum_{i \in A_{\vec{x}_0}} Y_i$

$$\mathbb{E}[\hat{\mu}(\vec{x}_0)] = \mathbb{E}\left[\frac{1}{|A_{\vec{x}_0}|} \sum_{i \in A_{\vec{x}_0}} Y_i\right] \quad (45)$$

$$= \mathbb{E}\left[\frac{1}{|A_{\vec{x}_0}|} \sum_{i \in A_{\vec{x}_0}} \mu(\vec{X}_i) + \epsilon_i\right] \quad (46)$$

$$= \frac{1}{|A_{\vec{x}_0}|} \sum_{i \in A_{\vec{x}_0}} \mathbb{E}[\mu(\vec{X}_i)] + \mathbb{E}[\epsilon_i] \quad (47)$$

$$= \mathbb{E}\left[\mu(\vec{X}) \mid |X_1 - x_{01}| \leq h \wedge |X_2 - x_{02}| \leq h\right] + \mathbb{E}[\epsilon \mid |X_1 - x_{01}| \leq h \wedge |X_2 - x_{02}| \leq h] \quad (48)$$

$$\approx \mu(\vec{x}_0) + O(h^2) \quad (49)$$

$$\text{Bias}[\hat{\mu}(\vec{x}_0)] = \mathbb{E}[\hat{\mu}(\vec{x}_0)] - \mu(\vec{x}_0) \quad (50)$$

$$\approx O(h^2) \quad (51)$$

Q3 a)

$$P(\vec{X} \in B_h) = \int_{B_h} f(\vec{x}) d\vec{x} \quad (52)$$

$$\approx \int_{B_h} f(\vec{x}_0) d\vec{x} \quad (53)$$

$$= f(\vec{x}_0)(2h)^p \quad (54)$$

This approximation is only valid for small h , and if the pdf was smooth. Then, we can approximate the surface around \vec{x}_0 as “flat”, having the same density as the center, which is $f(\vec{x}_0)$. The integral just becomes this average density multiplied by the p -dimensional volume of B_h .

Q3 b)

By linearity of expectations, if $P(\vec{x} \in B_h) \approx f(\vec{x}_0)(2h)^p$ for one point, then for n samples, the expected number of points N is:

$$N = \sum_{i=1}^n N_i \quad (55)$$

$$\mathbb{E}[N_i] = 1 \times P(\vec{X}_i \in B_h) + 0 \times (1 - P(\vec{X}_i \in B_h)) = P(\vec{X}_i \in B_h) \quad (56)$$

$$\mathbb{E}[N] = \sum_{i=1}^n \mathbb{E}[N_i] \quad (57)$$

$$\approx n f(\vec{x}_0)(2h)^p \quad (58)$$

Q3 c)

Let $A_{\vec{x}_0} = \{i : \vec{X}_i \in B_h\}$, where $N = |A_{\vec{x}_0}|$, so $\hat{\mu}(x_0) = \frac{1}{|A_{\vec{x}_0}|} \sum_{i \in A_{\vec{x}_0}} Y_i$

$$\text{Var} [\hat{\mu}(\vec{x}_0)] = \frac{1}{N^2} \text{Var} \left[\sum_{i \in A_{\vec{x}_0}} \mu(\vec{x}_i) + \epsilon_i \right] \quad (59)$$

$$= \frac{1}{N^2} \sum_{i \in A_{\vec{x}_0}} \text{Var} [\epsilon_i] \quad (60)$$

$$= \frac{\sigma^2}{N}, \text{ assuming that for all } i, \text{Var} [\epsilon_i] = \sigma^2 \quad (61)$$

$$\approx \frac{\sigma^2}{nf(\vec{x}_0)(2h)^p} \quad (62)$$

$$= O(n^{-1}h^{-p}) \quad (63)$$

Question 4

8.3 1)

$$\begin{aligned} f(x_1, x_2, \dots, x_i, \dots, x_p) - f(x_1, x_2, \dots, x_i + c, \dots, x_p) &= \left(a + f_i(x_i) + \sum_{j \neq i}^p f_j(x_j) \right) - \left(a + f_i(x_i + c) + \sum_{j \neq i}^p f_j(x_j) \right) \\ &= f_i(x_i) - f_i(x_i + c) \end{aligned} \quad \begin{matrix} (64) \\ (65) \end{matrix}$$

8.3 2)

Let $g(x) = f(x, x_2)$ and $h(x) = f(x, x'_2)$ for fixed x_2 and x'_2 .

$$g(x) - h(x) = (a + f_1(x) + f_2(x_2)) - (a + f_1(x) + f_2(x'_2)) \quad (66)$$

$$= f_2(x_2) - f_2(x'_2) \quad (67)$$

This difference is a constant, as it does not depend on x .

8.3 3)

Let $g(x) = f(x_1, \dots, x, \dots, x_p)$ and $h(x) = f(x'_1, \dots, x, \dots, x'_p)$, where both $g(x), h(x)$ only vary by the i^{th} coordinate, and we fix all other coordinates as x_j and x'_j for $g(x)$ and $h(x)$ respectively.

Then the difference along the i^{th} coordinate between the two functions is

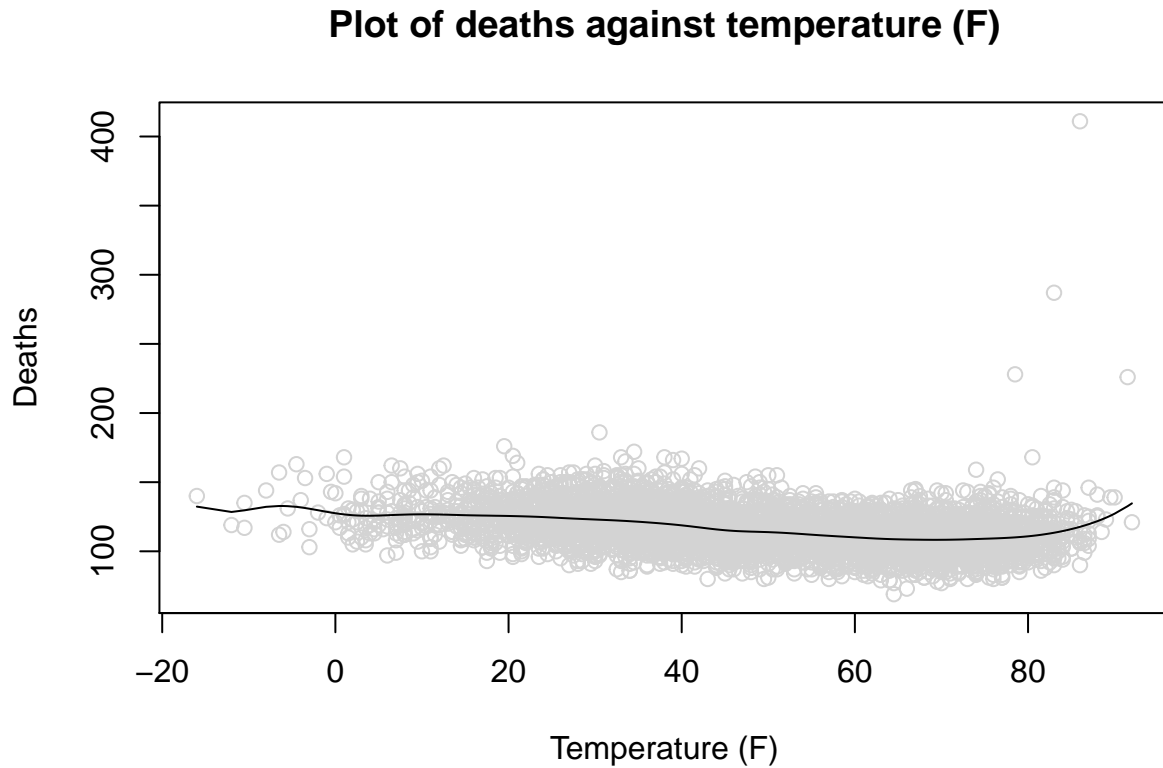
$$g(x) - h(x) = \left(a + f_i(x) + \sum_{j \neq i}^p f_j(x_j) \right) - \left(a + f_i(x) + \sum_{j \neq i}^p f_j(x'_j) \right) \quad (68)$$

$$= \sum_{j \neq i}^p f_j(x_j) - f_j(x'_j) \quad (69)$$

We know that all the curves formed by fixing one coordinate only have a difference that is independent of the coordinate being fixed (constant with respect to x_i), and hence are all parallel.

Question 5

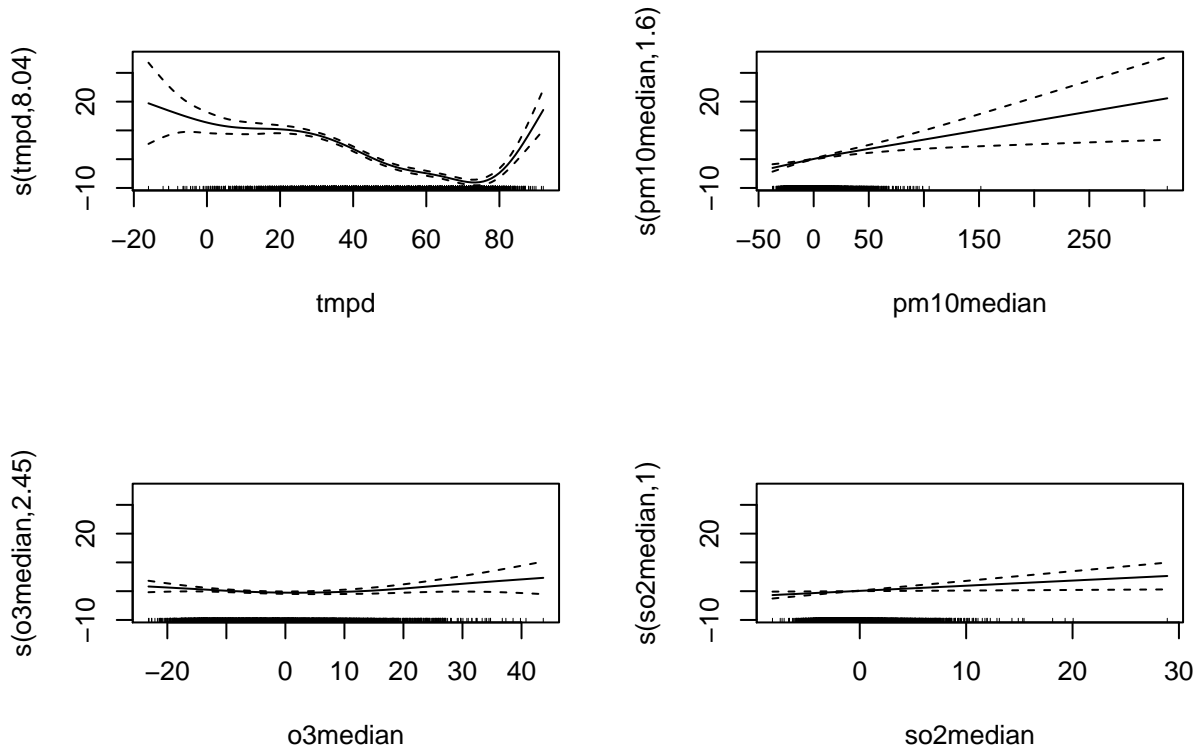
a)



A kernel regression is used to fit a non-parametric regression of deaths on temperature. Least-squares cross-validation was used to determine an optimal bandwidth for kernel regression.

The plot looks relatively smooth, being quite linear with a negative relationship for most of the temperature range up till past around 70F, where the number of deaths with each increase in temperature increases.

Q5 b)



The partial response function for temperature decreases from -20F to around 70F, before it sharply increases again. The partial response functions for pm10median, o3median, and so2median all look relatively linear and smooth, with all three having positive slopes.

Q5 c)

The shape of the partial response function for temperature roughly matches the shape of the curve, as both curves are decreasing up till around 70F before increasing again. Both curves should have roughly the same shape, as both methods of fitting a curve non-parametrically via kernel regression and splines do not make strong assumptions about the true $\mu(x)$ but instead fit by finding a good hyperparameter via cross-validation.

Q5 d)

Table 1: Cross-validated MSE

	x
Kernel Reg.	195.0461
GAM	191.8537

Since the additive model has a lower MSE from cross-validation, we conclude that it predicts better.