Que 2)

(a) Given,

kernal is $k(x_i, x) = \frac{\exp \|x_i - x\|_2^2}{\sigma^2}$ — (1)

In kernal regression, given a kernal k the output $\hat{y}$ is approximated as :-

$$\hat{y} = \frac{\sum_{i=1}^{n} k(x_i - x) \cdot y_i}{\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)} \quad \text{— (2)}$$

as the value of kernal is given in the question, let's substitute that value in the equation (2)

then, we have

$$\hat{y} = \frac{\sum_{i=1}^{n} \frac{\exp \|x_i - x\|_2^2}{\sigma^2 n} \cdot y_i}{\sum_{i=1}^{n} \frac{\exp \|x_i - x\|_2^2}{\sigma^2}}$$

do in
d like

of class
ecision
H.

(1)

according to the question

$$\hat{y} = l(x)^T y$$

where $l(x) = (l_1(x), \ldots, l_n(x))^T$

and also y is a $n \times 1$ vector

→ let $w_i = \exp \frac{\|x_i - x\|_2^2}{\sigma^2}$

So,

$$\hat{y} = \frac{\sum_{i=1}^{n} \exp \frac{\|x_i - x\|_2^2}{\sigma^2} \, y_i}{\sum_{i=1}^{n} \exp \frac{\|x_i - x\|_2^2}{\sigma^2}}$$

$l_i(x)$

$$l_i(x) = \frac{w_i}{\sum_{i=1}^{n} w_i}$$

So, $$\hat{y} = \sum_{i=1}^{n} l_i(x) \, y_i$$

Thus, kernal regression is linear smoother. because it is in the form of $l(x)^T y$.

Hence proved

P.T.O

Que 2) b)

To prove :-

If we fit a linear regression model by minimizing the sum of absolute values of resaduals i.e $\|Hw - Y\|_1$ instead of $\|Hw - Y\|_2^2$ then, it is not a linear smoother -

Proof - Minimizing L_1 norm is less stable than minimizing L_2 norm.

Example :-

Let, in a model if one point is having a error of 0 and other point is having a error of 10, if we take L_1 norm but if we take L_2 norm it will give an error of 100

→ If other model has 2 points and both having a error of 5 and 5 So, total error given by L_1 norm is 10 and total error by L_2 norm is $25+25 = 50$

So, if we take L_2 norm then it will suggest second model to be best, which is true because it is considering both the points and also penalising the the points having larger error.

→ Minimizing sum of absolute value of errors can be seen as finding median,

by
values of
of
linear

than

is
point is
L1 norm
give an

both
al

Suggest
e
and
rger

dian,

where an optimal median, makes same number of +ve errors and -ve errors and median minimizes the L_1 norm. But, median is a non-linear function because, it violates additivity.

Ques 2) c)

Given,

$$\hat{y} = \frac{1}{|B_x|} \sum_{i : x_i \in B_x} y_i \quad \text{— (1)}$$

if we write the above eqn in the form of Indicator random variable.

→ Indicator random variable is

$$\begin{cases} 1, & \text{if } x_i \in B_x \\ 0, & \text{if } x_i \notin B_x \end{cases}$$

→ Also $B_1, \ldots B_k$ are the bins

$$\Rightarrow \hat{y} = \frac{\sum_{i=1}^{n} I(x_i \in B_x)\, y_i}{\sum_{i=1}^{n} I(x_i \in B_x)} \quad \text{— (2)}$$

eqn (1) in terms of Indicator random variable.

Here, $l_i(x) = \dfrac{I(x_i \in B_x)}{\sum_{j=1}^{n} I(x_j \in B_x)}$

as, in (a) part of this question we got to know that linear smoother is defined as $\hat{y} = \sum_{i=1}^{n} l_i(x)\, y_i$

$$= l(x)^T Y$$

as eqn (2) can be written in the form

$$\sum_{i=1}^{n} y_i \, l(x_i)$$

$$\hat{y} = \frac{\sum_{i=1}^{n} y_i \, I(x_i \in B_x) \, l(x_i)}{\sum_{i=1}^{n} I(x_i \in B_x)}$$

Hence it is a linear smoother