2) a) Kernel regression should satisfy

$$\int K(x)\,dx = 1 \quad \text{and} \quad K(x) = K(-x)$$

$$K(x_i, x^*) = \exp\left(\frac{\|x - x^*\|_2^2}{\sigma^2}\right)$$

Any smoother satisfies

$$\hat{y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

For kernel smoother the $\sum_{i=1}^{n} w_i$ is replaced by $\sum_{i=1}^{n} K(x_i, x_i)$

$$\Rightarrow \hat{y} = \frac{\sum_{i=1}^{n} K(x_i, x) \cdot y}{\sum_{i=1}^{n} K(x_i, x)}$$

Here

$$l^T(x_i) = \frac{K(x_i, x)}{\sum_{i=1}^{n} K(x_i, x)}$$

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

$\Rightarrow$ kernel regression is a linear smoother

2) (b) An optimal w should make the same no. of (+)ve and (-)ve error. ~~A~~ for a linear smoother, the weights 'w' must be a linear function 'y'. ~~which~~ when we try to minimise the absolute difference, the minimum error is obtained when 'w' is taken ~~as~~ the median of 'y' values (Given in hint).

Since median cannot be written as a function of 'y', it cannot be a linear smoother.

eg: For ~~any~~ value $x_i = 3$ when we have multiple y values like 1, 2, 3, 4, 5. The best split is obtain by drawing line through 3 ($y = 3$). This ~~'y'~~ ~~'w'~~ changes when the 'y' values are changed.

∴ No it is not a linear smoother

classmate

(c) Yes it is a linear smoother. The vector

$$l(x) = \frac{I(x_j \in B_i)}{|B_i|}$$

as $\hat{y} = \frac{1}{|B_k|} \sum_{i : x \in B_k} y_i$

$\Downarrow$

$l(x)$