Homework 8

1) a) $H = X(X^TX)^{-1}X^T$

$\hat{\beta}_{(-i)} = (X^T_{(-i)}X_{(-i)})^{-1}X^T_{(-i)}Y_{(-i)}$

We also have $X^T_{(-i)}X_{(-i)} = (X^TX - x_ix_i^T)$

$h_{ii} = x_i^T(X^TX)^{-1}x_i$ is the i^{th} diagonal entry of H

$\Rightarrow (1-h_{ii})^{-1} = \dfrac{1}{1-x_i^T(X^TX)^{-1}x_i}$

We want to prove $(X^T_{(-i)}X_{(-i)})^{-1} = (X^TX)^{-1} + (1-h_{ii})^{-1}(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}$

Let ~~left hand side~~ $(X^T_{(-i)}X_{(-i)}) = X^TX - x_ix_i^T = A$

and right hand size $= B$

$\Rightarrow$ We need to prove $AB = I$ and $BA = I$ to get to $(A)^{-1} = B$

- Proving $AB = I$

$\Rightarrow AB = (X^TX - x_ix_i^T)\left((X^TX)^{-1} + \dfrac{(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}}{1-x_i^T(X^TX)^{-1}x_i}\right) = I$

$\Rightarrow I - x_ix_i^T(X^TX)^{-1} + \dfrac{\overbrace{X^TX(X^TX)^{-1}}^{I}x_ix_i^T(X^TX)^{-1} - x_ix_i^T(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}}{1-x_i^T(X^TX)^{-1}x_i} = I$

$\Rightarrow I - x_ix_i^T(X^TX)^{-1} + \dfrac{x_i(1-x_i^T(X^TX)^{-1}x_i)x_i^T(X^TX)^{-1}}{1-x_i^T(X^TX)^{-1}x_i} = I$

$\Rightarrow I - x_ix_i^T(X^TX)^{-1} + x_ix_i^T(X^TX)^{-1} = I$

$\Rightarrow I = I$ (proved)

- Proving $BA = I$

$\Rightarrow BA = \left((X^TX)^{-1} + \dfrac{(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}}{1-x_i^T(X^TX)^{-1}x_i}\right)(X^TX - x_ix_i^T) = I$

$\Rightarrow I - (X^TX)^{-1}x_ix_i^T + \dfrac{(X^TX)^{-1}x_ix_i^T - (X^TX)^{-1}x_ix_i^T(X^TX)^{-1}x_ix_i^T}{1-x_i^T(X^TX)^{-1}x_i} = I$

$\Rightarrow I - (X^TX)^{-1}x_ix_i^T + \dfrac{(X^TX)^{-1}x_i(1-x_i^T(X^TX)^{-1}x_i)x_i^T}{1-x_i^T(X^TX)^{-1}x_i} = I$

$\Rightarrow I - (X^TX)^{-1}x_ix_i^T + (X^TX)^{-1}x_ix_i^T = I$

$\Rightarrow I = I$ (proved)

$\Rightarrow AB = BA = I \Rightarrow A^{-1} = B$

$\Rightarrow (X^T_{(-i)}X_{(-i)})^{-1} = (X^TX)^{-1} + (1-h_{ii})^{-1}(X^TX)^{-1}x_ix_i^T(X^TX)^{-1}$

b) $\hat{\beta}_{(-i)} = (X_{(-i)}^T X_{(-i)})^{-1} X_{(-i)}^T y_{(-i)}$

$$= \left((X^TX)^{-1} + \frac{(X^TX)^{-1} x_i x_i^T (X^TX)^{-1}}{1-h_{ii}} \right) (X^T y - x_i y_i)$$

$$= \underbrace{(X^TX)^{-1} X^T y}_{\hat{\beta}_{(OLS)}} + (X^TX)^{-1} x_i \left(-y_i + \frac{x_i^T \overbrace{(X^TX)^{-1} X^T y}^{\hat{\beta}_{(OLS)}}}{1-h_{ii}} - \frac{\overbrace{x_i^T (X^TX)^{-1} x_i}^{h_{ii}} y_i}{1-h_{ii}} \right)$$

$$= \hat{\beta} + \frac{(X^TX)^{-1} x_i}{1-h_{ii}} \left(-y_i(1-h_{ii}) + x_i^T \hat{\beta} - h_{ii} y_i \right)$$

$$= \hat{\beta} + \frac{(X^TX)^{-1} x_i}{1-h_{ii}} \left(-y_i + x_i^T \hat{\beta} \right)$$

$$= \hat{\beta} + \frac{(X^TX)^{-1} x_i}{1-h_{ii}} \left(-y_i + \hat{Y}_i \right)$$

$$= \hat{\beta} - \frac{(X^TX)^{-1} x_i}{1-h_{ii}} \left(y_i - \hat{Y}_i \right)$$

$$\Rightarrow y_i - x_i^T \hat{\beta}_{(-i)} = y_i - x_i^T \left(\hat{\beta} - \frac{(X^TX)^{-1} x_i (y_i - \hat{y}_i)}{1-h_{ii}} \right)$$

$$= y_i - x_i^T \hat{\beta} + \frac{x_i^T (X^TX)^{-1} x_i (y_i - \hat{y}_i)}{1-h_{ii}}$$

$$= y_i - \hat{y}_i + \frac{h_{ii}(y_i - \hat{y}_i)}{1-h_{ii}}$$

$$= \frac{(1 - h_{ii} + h_{ii})(y_i - \hat{y}_i)}{1-h_{ii}} = \frac{y_i - \hat{y}_i}{1-h_{ii}}$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - x_i^T \hat{\beta}_{(-i)})^2 = \sum_{i=1}^{n} \left(\frac{y_i - \hat{y}_i}{1-h_{ii}} \right)^2$$

3) a) $\hat{f}^{(h)}(x) = \dfrac{\sum_i K\left(\frac{\|x - x_i\|}{h}\right) Y_i}{\sum_i K\left(\frac{\|x - x_i\|}{h}\right)} = \sum_{i=1}^{n} L_i(x)\, Y_i$

$\Rightarrow L_{ij}(h) = L_i(x_j) = \dfrac{K(\|x_j - x_i\|/h)}{\sum_i K(\|x_j - x_i\|/h)}$

b) i) $L_{ii}(h) = L_i(x_i)$ (using result in a)

$\Rightarrow L_{ii}(h) = \dfrac{K(\|x_i - x_i\|/h)}{\sum_j K(\|x_j - x_i\|/h)} = \dfrac{K(0)}{\sum_j K(\|x_j - x_i\|/h)}$

We have $K(0) > 0$, $\sum_j K(\|x_j - x_i\|/h) > 0$

also $\underset{x \in \mathbb{R}}{\operatorname{argmax}} K(x) = 0$

$\Rightarrow \sum_j K(\|x_j - x_i\| / h)$ is maximized at $K(0)$

$\Rightarrow L_{ii}(h) = \dfrac{K(0)}{\sum_j K(\|x_j - x_i\|/h)} \leq \dfrac{K(0)}{K(0)} = 1$

$\Rightarrow 0 \leq L_{ii}(h) \leq 1$

ii) $df_h = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} Cov(\hat{f}(x_i), y_i) = \dfrac{1}{\sigma^2} \sum_{i=1}^{n} Cov(L_i^h y_i, y_i)$

$= \dfrac{1}{\sigma^2} \cdot \sigma^2 \sum_{i=1}^{n} L_i^h = \sum_{i=1}^{n} L_i^h$

We proved earlier that $0 < L_i^{(h)} \leq 1$

$\Rightarrow$ df_h will be between $0(n)$ and $(1)(n)$

$\Rightarrow$ $0 < df_h \leq n$

$\Rightarrow$ df_h has to be a positive integer

d) $\|x_1\| \leq \|x_2\|$ then $K(\|x_1\|) \geq K(\|x_2\|)$

For $h_2 \leq h_1$

$\Rightarrow \dfrac{\|x_j - x_i\|}{h_2} \geq \dfrac{\|x_j - x_i\|}{h_1}$

$\Rightarrow L_i(h_2) \geq L_i(h_1)$

$\Rightarrow df(h_2) \geq df(h_1)$

This agrees with my intuition, because with smaller bandwidth, we will have larger variance, and therefore higher degrees of freedom