Ex. 6.7

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{f}^{(-i)}(x_i)\right)^2$$

for $\hat{f}(x_0) = b(x_0)^T\left(B^T W(x_0) B\right)^{-1} B^T W(x_0) Y$,

We will show that:

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\left(y_i - \cancel{h_i}\hat{f}(x_i)\right)^2}{(1-h_i)^2}$$; where h_i is a function of x_i.

Notation: for a matrix or vector A, $A^{(-i)}$ is the corresponding matrix or vector with the i^{th} row/element removed. Additionally, $A^{(i)}$ refers to that i^{th} row/element.

The proof is split into 2 parts:

Part 1/2

Wish to show that:

$$(*) \quad \hat{f}^{(-i)}(x_i) = \frac{\hat{f}(x_i) - h_i y_i}{1-h_i}$$

first some results we will use later:

The Sherman-Morrison formula states that for V an invertible $P\times P$ matrix; a P-vector q; and λ constant:

$$① \quad (V + \lambda q q^T)^{-1} = V^{-1} - \frac{\lambda V^{-1} q q^T V^{-1}}{1 + \lambda q^T V^{-1} q}$$

for B an $N\times P$ matrix and W a diagonal $N\times N$ matrix with i^{th} diagonal element W_i:

$$② \quad \sum_{i=1}^{N} W_i B^{(i)T} B^{(i)} = B^T W B$$

Since $B^{(i)}$ is simply the basis expansion of x_i we can follow the books notation in referring to it as $b(x_i)$ from here on.

We will need to prove one more identity:
Now, using ② we can see that:

$$\left(B^{(-i)T} W^{(-i)} B^{(-i)}\right)^{-1} = \left(B^T W B - b(x_i) W(x_i) b(x_i)^T\right)^{-1}$$

$$= (B^T W B)^{-1} + \frac{W(x_i)(B^T W B)^{-1} b(x_i) b(x_i)^T (B^T W B)^{-1}}{1 - W(x_i) b(x_i)^T (B^T W B)^{-1} b(x_i)} \quad \text{(using ①)}$$

$$\Rightarrow \left(B^{(-i)T} W^{(-i)} B^{(-i)}\right)^{-1} b(x_i) W(x_i) = (B^T W B)^{-1} b(x_i) W(x_i)$$

$$+ (B^T W B)^{-1} b(x_i) W(x_i) \left(\frac{b(x_i)^T (B^T W B)^{-1} b(x_i) W(x_i)}{1 - W(x_i) b(x_i)^T (B^T W B)^{-1} b(x_i)} \right)$$

Giving:

③ $$\left(B^{(-i)T} W^{(-i)} B^{(-i)}\right)^{-1} b(x_i) W(x_i) = (B^T W B)^{-1} b(x_i) W(x_i) \left(\frac{1}{1 - h_i}\right)$$

Where $h_i = b(x_i)^T (B^T W B)^{-1} b(x_i) W(x_i)$

Finally, we can make some progress on the original problem. From the definition of $\hat{\beta}$ in 6.8 we have:

$$\hat{\beta} = (B^T W B)^{-1} B^T W y$$

$$\Rightarrow (B^T W B)\hat{\beta} = B^T W y$$

$$\Rightarrow \left(B^{(-i)T} W^{(-i)} B^{(-i)} + b(x_i) W(x_i) b(x_i)^T\right)\hat{\beta} = B^{(-i)T} W^{(-i)} y^{(-i)} + b(x_i) W(x_i) y_i \quad \text{(similar logic to ②)}$$

$$\Rightarrow \left(\mathbb{I} + \left(B^{(-i)T} W^{(-i)} B^{(-i)}\right)^{-1} b(x_i) W(x_i) b(x_i)^T\right)\hat{\beta}$$

$$= \hat{\beta}^{(-i)} + \left(B^{(-i)T} W^{(-i)} B^{(-i)}\right)^{-1} b(x_i) W(x_i)\, y_i$$

Now the residual of observation i is defined as:

$r_i = y_i - b(x_i)^T\hat{\beta}$ implying $y_i = b(x_i)^T\hat{\beta} + r_i$ (4)

Now when we expand y_i in the previous equation some convenient cancelation occours and we are left with:

$$\hat{\beta} = \hat{\beta}^{(-i)} + \left(B^{(-i)T}W^{(-i)}B^{(-i)}\right)^{-1} b(x_i)W(x_i)r_i$$

$$\Rightarrow b(x_i)^T\hat{\beta} = b(x_i)^T\hat{\beta}^{(-i)} + b(x_i)^T\left(B^{(-i)T}W^{(-i)}B^{(-i)}\right)^{-1} b(x_i)W(x_i)r_i$$

(using (3))

$$\Rightarrow \hat{f}(x_i) = \hat{f}^{(-i)}(x_i) + b(x_i)^T\left(B^TWB\right)^{-1}b(x_i)W(x_i)\left(\frac{1}{1-h_i}\right)r_i$$

$$\Rightarrow \hat{f}(x_i) = \hat{f}^{(-i)}(x_i) + \left(\frac{h_i}{1-h_i}\right)r_i$$ (From definition of h_i)

$$\Rightarrow \hat{f}^{(-i)}(x_i) = \hat{f}(x_i) - \left(\frac{h_i}{1-h_i}\right)(y_i - \hat{f}(x_i))$$ (using (4) again)

$$= \hat{f}(x_i)\left(1 - \frac{h_i}{1-h_i}\right) - \frac{h_i y_i}{1-h_i}$$

$$\hat{f}^{(-i)}(x_i) = \frac{\hat{f}(x_i) - h_i y_i}{1-h_i}$$ Proving (*) as required

Part 2//

The next step is to sub (*) into the orignal cross validation expression in particular:

$$y_i - \hat{f}^{(-i)}(x_i) = y_i - \frac{\hat{f}(x_i) - h_i y_i}{1-h_i}$$

$$= \frac{y_i - h_i y_i - \hat{f}(x_i) + h_i y_i}{1 - h_i}$$

$$= \frac{y_i - \hat{f}(x_i)}{1 - h_i}$$

Proving that the leave-one-out cross-validated residual sum-of-squares for local Polynomial regression is given by:

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f^{(-i)}(x_i)\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{f}(x_i)}{1 - h_i}\right)$$

where $h_i = b(x_i)^T (B^T W B)^{-1} b(x_i) W(x_i)$