In linear regression (ridge), we can compute the loo error without actually doing leave-one-out computation.

Because

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i^{loo}\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i-\hat{Y}_i}{1-H_{ii}}\right)^2$$

Y_i : the i^{th} response in the training data $\qquad H = X(X^TX+\lambda I)^{-1}X^T$,

$\hat{Y}_i$: fitted value of Y_i $= x_i^T\hat{\beta}$ $\qquad H_{ii}$: the (i,i)th entry of H.

$\hat{Y}_i^{loo}$: fitted value of Y_i by loo

Proof:

We know

$$\hat{\beta}=(X^TX+\lambda I)^{-1}X^Ty$$

is the solution to ridge regression

so

$$\hat{Y}=X\hat{\beta}=X(X^TX+\lambda I)^{-1}X^Ty,\quad X=\begin{bmatrix}x_1^T\\ x_2^T\\ \vdots\\ x_n^T\end{bmatrix}\in\mathbb{R}^{n\times p}$$

$$\hat{Y}_i^{loo}=x_i^T\hat{\beta}^{loo}=x_i^T\left(X_{-i}^TX_{-i}+\lambda I\right)^{-1}X_{-i}^Ty_{-i},\quad X_{-i}=\begin{bmatrix}x_1^T\\ x_2^T\\ \vdots\\ x_{i-1}^T\\ x_{i+1}^T\\ \vdots\\ x_n^T\end{bmatrix},\ y_{-i}=\begin{bmatrix}Y_1\\ Y_2\\ \vdots\\ Y_{i-1}\\ Y_{i+1}\\ \vdots\\ Y_n\end{bmatrix}$$

$$=x_i^T\left[\lambda I+X^TX-x_ix_i^T\right]^{-1}\left[X^Ty-x_iY_i\right]$$

By Matrix Inversion Lemma: $(A+UCV)^{-1}=A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$

we have

$$(\underbrace{\lambda I+X^TX}_{A}-\underbrace{x_i}_{U}\underbrace{x_i^T}_{V})^{-1}=(\lambda I+X^TX)^{-1}-\frac{(\lambda I+X^TX)^{-1}x_ix_i^T(\lambda I+X^TX)^{-1}}{-1+x_i^T(\lambda I+X^TX)^{-1}x_i}$$

$C=-1$

Note that $x_i^T(\lambda I+X^TX)^{-1}x_i=H_{ii}$

Using this, we have

$$\hat{Y}_i^{loo} = x_i^T \left[(\lambda I + X^T X)^{-1} - \frac{(\lambda I + X^T X)^{-1} x_i x_i^T (\lambda I + X^T X)^{-1}}{-1 + H_{ii}} \right] \left[X^T y - x_i Y_i \right]$$

$$= \left[x_i^T (\lambda I + X^T X)^{-1} - \frac{H_{ii}\, x_i^T (\lambda I + X^T X)^{-1}}{-1 + H_{ii}} \right] \left[X^T y - x_i Y_i \right]$$

$$= \left[x_i^T \hat{\beta} - \frac{H_{ii}\, x_i^T \hat{\beta}}{-1 + H_{ii}} - H_{ii} Y_i + \frac{H_{ii}^2}{H_{ii} - 1} Y_i \right]$$

$$= \left(1 - \frac{H_{ii}}{-1 + H_{ii}} \right) \hat{Y}_i + \left(-H_{ii} + \frac{H_{ii}^2}{H_{ii} - 1} \right) Y_i$$

$$= \frac{-1}{-1 + H_{ii}} \hat{Y}_i + \frac{H_{ii}}{H_{ii} - 1} Y_i .$$

So $$Y_i - \hat{Y}_i^{loo} = \frac{\hat{Y}_i - Y_i}{-1 + H_{ii}}$$

$$\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i^{loo} \right)^2 = \sum_{i=1}^{n} \frac{(\hat{Y}_i - Y_i)^2}{(1 - H_{ii})^2}$$ ※.

In Locally - Polynomial Regression :

$$\text{Let} \quad M_i = \begin{bmatrix} 1 & X_1 - X_i & \frac{1}{2}(X_1 - X_i)^2 & \cdots & \frac{1}{p!}(X_1 - X_i)^p \\ 1 & X_2 - X_i & \frac{1}{2}(X_2 - X_i)^2 & \cdots & \frac{1}{p!}(X_2 - X_i)^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 1 & X_n - X_i & \cdots & \cdots & \frac{1}{p!}(X_n - X_i)^p \end{bmatrix} \begin{matrix} \\ \\ \\ \rightarrow i^{th} \text{ row} \\ \\ \\ \end{matrix}$$

$$W_i = \begin{bmatrix} K\left(\frac{\|X_1 - X_i\|}{h}\right) & & & & & 0 \\ & \ddots & & & & \\ & & K\left(\frac{\|X_{i-1} - X_i\|}{h}\right) & & & \\ & & & 0 & & \\ & & & & K\left(\frac{\|X_{i+1} - X_i\|}{h}\right) & \\ & 0 & & & & \ddots \\ & & & & & K\left(\frac{\|X_n - X_i\|}{h}\right) \end{bmatrix}$$

$M_{i,-i}$: M_i without the i^{th} row, so $M_{i,-i} \in \mathbb{R}^{(n-1)\times(p+1)}$

$W_{i,-i}$: W_i without the i^{th} row and column, $W_{i,-i} \in \mathbb{R}^{(n-1)\times(n-1)}$

From Eq. (27.34) and Eq (27.35) on P.539 of the notes :

$$\hat{Y}_i^{loo} = e_1^T \left[\left(M_{i,-i}^T W_{i,-i} M_{i,-i} \right)^{-1} M_{i,-i}^T W_{i,-i} Y_{-i} \right] \qquad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \in R^{n\times 1}$$

$$\hat{Y}_i = \underbrace{e_1^T \left[\left(M_i^T W_i M_i \right)^{-1} M_i^T W_i \right.}_{i^{th} \text{ row of the linear smoothing matrix } L} \left. Y \right] \qquad Y_{-i} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{i-1} \\ Y_{i+1} \\ \vdots \\ Y_n \end{bmatrix} \in R^{(n-1)\times 1}$$

$$L_{ii} = \left(e_1^T \left(M_i^T W_i M_i \right)^{-1} M_i^T W_i \right)_i \qquad e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= e_1^T \left(M_i^T W_i M_i \right)^{-1} e_1 W_{ii} .$$

Note that.

$$M_{i,-i}^T W_{i,-i} M_{i,-i} = M_i^T W_i M_i - W_{ii} e_1 e_1^T$$

So by Matrix Inversion lemma

$$\left(M_{i,-i}^T W_{i,-i} M_{i,-i} \right)^{-1} = \Big(\underbrace{M_i^T W_i M_i}_{A} - \underbrace{W_{ii}}_{C} \underbrace{e_1}_{U} \underbrace{e_1^T}_{V} \Big)^{-1}$$

$$= \left(M_i^T W_i M_i\right)^{-1} - \frac{\left(M_i^T W_i M_i\right)^{-1} e_1 e_1^T \left(M_i^T W_i M_i\right)^{-1}}{-W_{ii}^{-1} + e_1^T \left(M_i^T W_i M_i\right)^{-1} e_1}$$

$$= \left(M_i^T W_i M_i\right)^{-1} - \frac{W_{ii} \left(M_i^T W_i M_i\right)^{-1} e_1 e_1^T \left(M_i^T W_i M_i\right)^{-1}}{-1 + L_{ii}} \quad \text{—— (A)}$$

We also have

$$M_{i,-i}^T W_{i,-i} Y_{-i} = M_i^T W_i Y - W_{ii} e_1 Y_i \quad \text{———— (B)}$$

Substituting (A) and (B) into the formula for $\hat{Y}_i^{loo}$ leads

to

$$\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i^{loo} \right)^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i - \hat{Y}_i}{1 - L_{ii}} \right)^2$$