A Simple Formula for LOO-CV

The Leave-one-out (LOO) CV is given by:

$$\text{LOO-CV} = \frac{1}{n}\sum_{i=1}^{n} [\ y_i - \hat{g}^{[-i]}(x_i)]^2$$

where $\hat{g}^{[-i]}$ is the smoothing spline model trained on $n-1$ samples, excluding the i-th sample. It may seem as if we need to fit n distinct models to compute this, but surprisingly, the LOO prediction error can be derived from the model trained on all n samples:

$$y_i - \hat{g}^{[-i]}(x_i) = \frac{y_i - \hat{g}(x_i)}{1 - S_{ii}}, \tag{1}$$

where S is the $n \times n$ smoother matrix.

Proof for $i = 1$. The proof for other i follows analogously.

Consider two smoothing spline models:

Model I: This model, denoted by $\hat{g}$, trained on the original training data $(x_i, y_i)_{i=1}^n$. The predicted values at x_1 to x_n are equal to

$$\begin{pmatrix} \hat{g}(x_1) \\ \cdots \\ \cdots \\ \cdots \end{pmatrix} = S_{n\times n}\mathbf{y}_{n\times 1} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}.$$

Focusing on the prediction for the first data point, x_1, we have

$$\hat{g}(x_1) = S_{11} \cdot y_1 + S_{12} \cdot y_2 + \cdots + S_{1n} \cdot y_n. \tag{2}$$

Model II: Define a modified dataset $(x_i, y_i^*)_{i=1}^n$, which mirrors the original data $(x_i, y_i)_{i=1}^n$, with the sole exception being the first data point, for which

$$y_1^* = \hat{g}^{[-1]}(x_1).$$

Our first goal is to prove that the smoothing spline trained on the modified dataset is equivalent to $\hat{g}^{[-1]}$.

Given the smoothing spline objective function for this new dataset:

$$\begin{aligned} & \sum_{i=1}^n (y_i^* - h(x_i))^2 + \lambda \int_a^b [h''] dx \\ = \quad & (y_1^* - h(x_1))^2 + \sum_{i=2}^n (y_i - h(x_i))^2 + \int_a^b [h''] dx, \end{aligned}$$

letting $h(x) = \hat{g}^{[-1]}(x)$, we observe:

- the first term is zero, since by definition $y_1^* = \hat{g}^{[-1]}(x_1)$.
- the other two terms are minimized by $\hat{g}^{[-1]}$ as they comprise the objective function for the $n-1$ observations.

Thus, the optimal smoothing spline based on $(x_i, y_i^*)_{i=1}^n$ is $\hat{g}^{[-1]}$.

Further, the predicted values from this model are $S\mathbf{y}^*$ Importantly, both Model I and II utilize the same smoother matrix, S, because it solely depends on the x_i values, which remain unchanged.

$$\begin{pmatrix} \hat{g}^{[-1]}(x_1) \\ \cdots \\ \cdots \\ \cdots \end{pmatrix} = S\mathbf{y}^* = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} \hat{g}^{[-1]}(x_1) \\ y_2 \\ \cdots \\ y_n \end{pmatrix}.$$

Focusing on the prediction for the first data point, x_1, we have

$$\begin{aligned} \hat{g}^{[-1]}(x_1) &= S_{11} \cdot \hat{g}^{[-1]}(x_1) + S_{12} \cdot y_2 + \cdots + S_{1n} \cdot y_n \\ &= S_{11} \cdot \big(\hat{g}^{[-1]}(x_1) - y_1\big) + S_{11} \cdot y_1 + S_{12} \cdot y_2 + \cdots + S_{1n} \cdot y_n \\ &= S_{11} \cdot \big(\hat{g}^{[-1]}(x_1) - y_1\big) + \hat{g}(x_1) \end{aligned}$$

where we use (2) at the last equality.

Thus

$$\begin{aligned} y_1 - \hat{g}^{[-1]}(x_1) &= y_1 - S_{11} \cdot \big(\hat{g}^{[-1]}(x_1) - y_1\big) - \hat{g}(x_1) \\ &= S_{11} \cdot \big(y_1 - \hat{g}^{[-1]}(x_1)\big) + y_1 - \hat{g}(x_1) \end{aligned}$$

which implies the Simple Formula for LOO-CV (1).