

Calculating CV without fitting lots of models

Rob J Hyndman

Wednesday, 5 August 2015

The leave-one-out cross-validation statistic is given by

$$CV = \frac{1}{N} \sum_{i=1}^N e_{[i]}^2,$$

where $e_{[i]} = y_i - \hat{y}_{[i]}$, y_1, \dots, y_N are the observations, and $\hat{y}_{[i]}$ is the predicted value obtained when the model is estimated with the i th case deleted. It turns out that for linear models, we do not actually have to estimate the model N times, once for each omitted case. Instead, CV can be computed after estimating the model once on the complete data set.

Suppose we have a linear regression $Y = X\beta + e$. Then $\hat{\beta} = (X'X)^{-1}X'Y$ and $H = X(X'X)^{-1}X'$ is the “hat-matrix”. It has this name because it is used to compute $\hat{Y} = X\hat{\beta} = HY$. If the diagonal values of H are denoted by h_1, \dots, h_N , then the leave-one-out cross-validation statistic can be computed using

$$CV = \frac{1}{N} \sum_{i=1}^N [e_i / (1 - h_i)]^2,$$

where $e_i = y_i - \hat{y}_i$ and \hat{y}_i is the predicted value obtained when the model is estimated with all data included. This is a remarkable result, and is given without proof in Section 5/5 of my forecasting textbook. What follows is the simplest proof I know (adapted from Seber and Lee, 2003).

Proof

Let $X_{[i]}$ and $Y_{[i]}$ be similar to X and Y but with the i th row deleted in each case. Let x'_i be the i th row of X and let

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

be the estimate of β without the i th case. Then $e_{[i]} = y_i - x'_i\hat{\beta}_{[i]}$.

Now $X'_{[i]}X_{[i]} = (X'X - x_i x'_i)$ and $x_i(X'X)^{-1}x_i = h_i$. So by the Sherman–Morrison–Woodbury formula,

$$(X'_{[i]}X_{[i]})^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x'_i (X'X)^{-1}}{1 - h_i}.$$

Also note that $X'_{[i]}Y_{[i]} = X'Y - x_i y_i$. Therefore

$$\begin{aligned}\hat{\beta}_{[i]} &= \left[(X'X)^{-1} + \frac{(X'X)^{-1}x_i x'_i (X'X)^{-1}}{1 - h_i} \right] (X'Y - x_i y_i) \\ &= \hat{\beta} - \left[\frac{(X'X)^{-1}x_i}{1 - h_i} \right] \left[y_i(1 - h_i) - x'_i\hat{\beta} + h_i y_i \right] \\ &= \hat{\beta} - (X'X)^{-1}x_i e_i / (1 - h_i)\end{aligned}$$

Thus

$$\begin{aligned}e_{[i]} &= y_i - x'_i\hat{\beta}_{[i]} \\ &= y_i - x'_i \left[\hat{\beta} - (X'X)^{-1}x_i e_i / (1 - h_i) \right] \\ &= e_i + h_i e_i / (1 - h_i) \\ &= e_i / (1 - h_i),\end{aligned}$$

and the result follows.

References

Seber, G. A. F. and A. J. Lee (2003). *Linear Regression Analysis*. 2nd. John Wiley & Sons.