23 January 2018
17:01

Leave-One-Out Cross-Validation

LOOCV involves splitting the set of observations into two parts.

1. Create two subsets of comparable size, a single observation (x1, y1) is used for the validation set, and the remaining observations {(x2, y2), . . . , (xn, yn)} are used for the training set.

2. The statistical learning method is fit on the n − 1 training observations, and a prediction y_1 is made for the excluded observation, using its value x1.

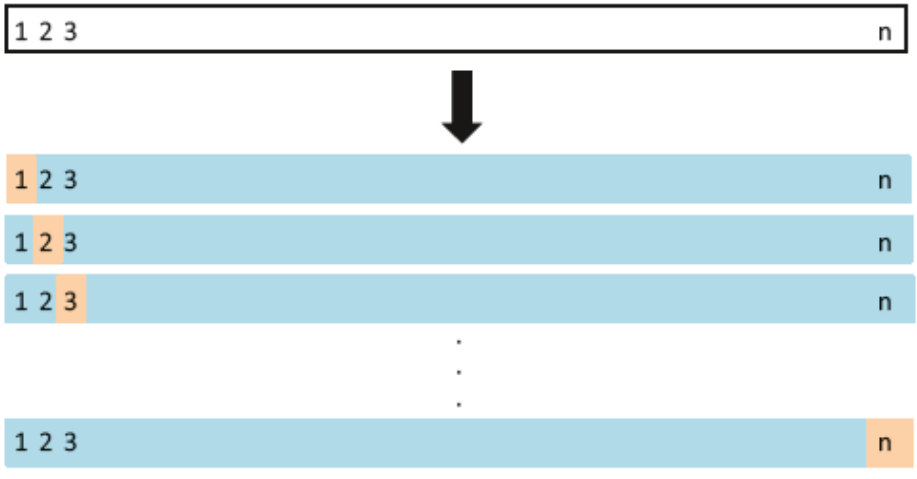
3. Calculate MSE1 =(y1 − y_1)2 provides an approximately unbiased estimate for the test error.

4. We can repeat the procedure by selecting (x2, y2) for the validation data, training the statistical learning procedure on the n − 1 observations {(x1, y1), (x3, y3), . . . , (xn, yn)}, and computing MSE2 = (y2−ˆy2)2.

5. Repeating this approach n times produces n squared errors, MSE1, . . . , MSEn.

6. The LOOCV estimate for the test MSE is the average of these n test error estimates:

$$\mathrm{CV}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{MSE}_i.$$



Draw backs:

- If the data set is very huge we cannot remove each observation and do the process.
- Time consuming.

Advantages:

- It is less bias and high variance.
- Building the model using (n-1) observations.
- When compare with validation set approach, LOOCV give same kind of models since every time removing only 1 observation from data set.
- LOOCV is a general method can be used for any model/classifier.