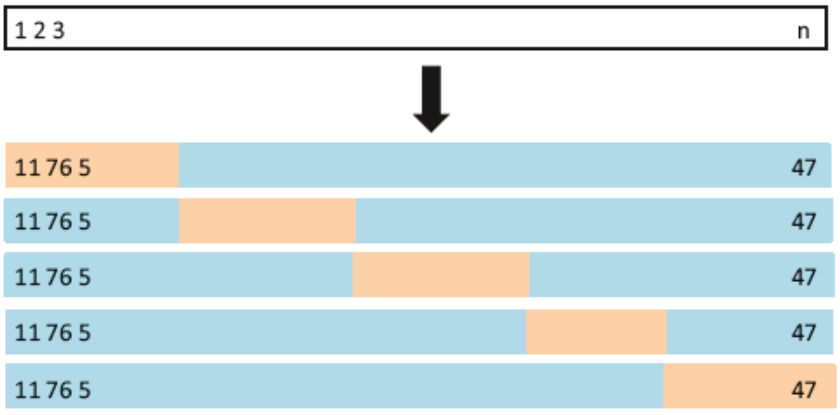
k-Fold Cross-Validation

An alternative to LOOCV is k-fold CV. We can say that LOOCV is a special case of k-fold CV.

1. Randomly k-fold CV dividing the set of observations into k groups, or folds, of approximately equal size.
2. The first fold is treated as a validation set, and the method is fit on the remaining k − 1 folds.
3. Calculate the mean squared error and treated as MSE1.
4. Repeat the procedure for k times, each time your validation set is going to be changed and the MSE is calculated on the remaining groups.
5. This process results in k estimates of the test error, MSE1,MSE2, . . . ,MSEk.
6. The k-fold CV estimate is computed by averaging these values,

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



Advantages:

1. Comparing with LOOCV, it is less time consuming and most feasible.
2. It takes care of both drawbacks of validation-set methods as well as LOOCV.
3. In Practice, we can take k = 5 to 10.
4. It is not having too much of variance or too much of Bias. It will be balanced.
5. It's a General method applied to any statistical learning like Regression/classification/ any machine learning methods.

? Which is going to be good approach?

- Test Error of K-fold is lies between validation set approach and LOOCV
- If we want to reduce Bias you need to go for LOOCV.
- If we want to reduce variance you need to go for Validation set.

NOTE:

★ When K is the number of observations (K = N) leave-one-out cross-validation is used and all the possible splits of the data are used.

★ When K is less than the number of observations (K < N) K splits to be used are found by randomly partitioning the data into K groups of approximately equal size.

In this latter case a certain amount of bias is introduced. This can be reduced by using a simple adjustment. The second value returned in delta is the estimate adjusted by this method.