**K-Fold Cross Validation**

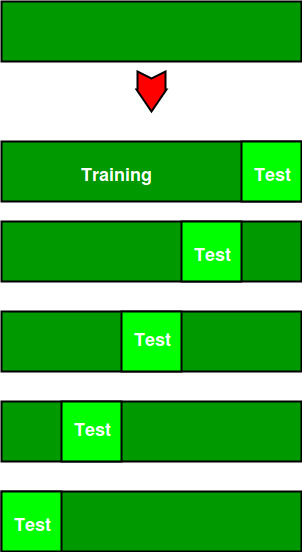
**K-Fold Cross Validation**  
In this method, we split the data-set into k number of subsets(known as folds) then we perform training on the all the subsets but leave one(k-1) subset for the evaluation of the trained model. In this method, we iterate k times with a different subset reserved for testing purpose each time.

***Note:***

It is always suggested that the value of k should be 10 as the lower value

of k is takes towards validation and higher value of k leads to LOOCV method.

**Example**  
The diagram below shows an example of the training subsets and evaluation subsets generated in k-fold cross-validation. Here, we have total 25 instances. In first iteration we use the first 20 percent of data for evaluation, and the remaining 80 percent for training([1-5] testing and [5-25] training) while in the second iteration we use the second subset of 20 percent for evaluation, and the remaining three subsets of the data for training([5-10] testing and [1-5 and 10-25] training), and so on.



Total instances: 25

Value of k : 5

No. Iteration Training set observations Testing set observations

1 [ 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24] [0 1 2 3 4]

2 [ 0 1 2 3 4 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24] [5 6 7 8 9]

3 [ 0 1 2 3 4 5 6 7 8 9 15 16 17 18 19 20 21 22 23 24] [10 11 12 13 14]

4 [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 20 21 22 23 24] [15 16 17 18 19]

5 [ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19]

**Comparison of train/test split to cross-validation**

Advantages of train/test split:

1. This runs K times faster than Leave One Out cross-validation because K-fold cross-validation repeats the train/test split K-times.
2. Simpler to examine the detailed results of the testing process.

Advantages of cross-validation:

1. More accurate estimate of out-of-sample accuracy.
2. More “efficient” use of data as every observation is used for both training and testing.

Python code for k fold cross-validation.

Python3

*# This code may not be run on GFG IDE*

*# as required packages are not found.*

*# importing cross-validation from sklearn package.*

**from** **sklearn** **import** cross\_validation

*# value of K is 10.*

data = cross\_validation.KFold(len(train\_set), n\_folds=10, indices=**False**)