**What is Cross Validation?**

Cross Validation is a technique which involves reserving a sample of a dataset on which you do not train the model. Later, you test your model on this sample before finalizing it.

Here are the steps involved in cross validation:

* You reserve a sample data set
* Train the model using the remaining part of the dataset
* Use the reserve sample of the test (validation) set. This will help you in gauging the effectiveness of your model’s performance.

**Methods used in Cross Validation**

1. **Leave one out cross validation (LOOCV)**

In this approach, we reserve only one data point from the available dataset and train the model on the rest of the data. This process iterates for each data point. This also has its own advantages and disadvantages.

* We make use of all data points hence the bias will be low.
* We repeat the cross-validation process n times (where n is number of data points) which results in a higher execution time

This approach leads to higher variation in testing model effectiveness because we test against one data point. So, our estimation gets highly influenced by the data point. If the data point turns out to be an outlier, it can lead to a higher variation.

1. **k-fold cross validation**

* Randomly split your entire dataset into k”folds”.
* For each k-fold in your dataset, build your model on k – 1 folds of the dataset. Then, test the model to check the effectiveness for kth fold.
* Record the error you see on each of the predictions.
* Repeat this until each of the k-folds has served as the test set.
* The average of your k recorded errors is called the cross-validation error and will serve as your performance metric for the model.

A screen shot of a computer

Description automatically generated

1. **Stratified k-fold cross validation**

Stratification is the process of rearranging the data to ensure that each fold is a good representative of the whole. For example, in a binary classification problem where each class comprises of 50% of the data, it is best to arrange the data such that in every fold, each class comprises of about half the instances.

