# **Model Evaluation Criteria**

There are broadly two categories of model evaluation that are hideout and cross-validation. Set dataset is used in both of them to evaluate models Performance. It is not essentially a good practice to use the dataset that we trained on as a testing dataset because the model could completely remember the data set to perform well in evaluation criteria which could induce overfitting.

## **Holdout**

In hideout evaluation the model is evaluated on a different dataset then it was trained on. Unbiased estimation of learning performance is provided as a result.

Three subset of dataset is created randomly in this method:

1. **Training set** A subset of dataset used for the training
2. **Validation set** A subset of dataset used for the model evaluation of the model during the training phase. Helps in fine tuning the model during the training phase.
3. **Test set** it is a subset of a dataset which is unseen by the model and is used for the evaluation during the testing phase.

## **Cross-Validation**

An independent dataset is used for the evaluation of the model apart from the dataset used during the training phase.

K-fold cross validation is the most common validation technique in which the original data observations are divided into k equal size subsamples. K is chosen by the user and is also can be considered as a hyperparameter. 5 to 10 is the usually the preferred value for k. The process is repeated k times and one of the k subset is used as the test/validation set and the other part that is the other k-1 sets are put together to create a training set.he error estimation is averaged over all k trials to get the total effectiveness of our model.

# **Model Evaluation Metrics**

The quantification of our models performance is represented by models evaluation metrics. The general choice of matrices chosen depends on the different machine learning problem. Some of the supervised learning Model evaluation metrics are as follows

# **Classification Metrics**

## **Classification Accuracy**

The most common evaluation matrix for model evaluation is Accuracy. It is the ratio of the total number of correct predictions made to all the predictions made.

## **Confusion Matrix**

More detailed breakdown of correct and incorrect classification is provided in a confusion matrix. For any classification problem the diagonal elements present in the matrices are the number of points for which the prediction was correct and anything apart from the diagonal element is a bad classification. Thus as a general rule, the higher the diagonal values indicates better model performance.

## **Logarithmic Loss**

The performance of classification models where the prediction input is a probability value between 0 and 1 is done using a logarithmic loss matrix. Smaller value of log loss is better as the function increases the predicted probability diverging from the actual value. A perfect model would have 0 logloss.

## **Area under Curve (AUC)**

A quantified discriminate between positive and negative of a binary classifier is measured using the area under the ROC curve.

## **F-Measure**

Test’s accuracy for both precision and the recall is computed to measure the F-score or F-measure. Precision is the number of correct positives divided by the total predicted data observations. Number of correct positive results divided by the number of all relevant samples is the recall.