49. Cross-Validation in Machine Learning

- It gives the information about how long your model can give you highest accuracy on particular data
- Cross-validation is a technique for validating the model efficiency by training it on the subset of input data and testing on previously unseen subset of the input data
- It will give the range which will tell that your data has how much min accuracy and max accuracy it can attain

![No description has been provided for this image]

49.1 Methods used for cross-validation:

- Leave p out cross-validation
- Leave one out cross-validation
- Holdout cross-validation
- Repeated random subsampling validation
- k-fold cross-validation
- Stratified k-fold cross-validation
- Time Series cross-validation
- Nested cross-validation

49.1.1 K-Fold Cross-Validation:

- The original dataset is equally partitioned into k subparts or folds.
- Out of the k-folds or groups, for each iteration, one group is selected as validation data,
- and the remaining (k-1) groups are selected as training data.
- Not suitable for an imbalanced data

49.1.2 Stratified Cross-Validation:

- It works when the data is in classification nature
- It works on unbalanced data
- The original dataset is equally partitioned into k subparts or folds.
- Out of the k-folds or groups, for each iteration, one group is selected as validation data,
- and the remaining (k-1) groups are selected as training data.
- Stratified k-fold cross-validation solved the problem of imbalanced data

49.1.3 Leave-One-Out Cross-Validation:

- It gets trained on whole data
- It is an exhaustive cross-validation technique
- it is a category of Lp OCV with the case of p=1.
- It is slower in case of large data b/c of this issue it is used less
- The model trained by this method is very accurate

49.1.4 Leave-P-Out Cross-Validation:

- It is an exhaustive cross-validation technique, that involves using p-obervation as validation data
- It is slower in case of large data b/c of this issue it is used less

In [ ]: