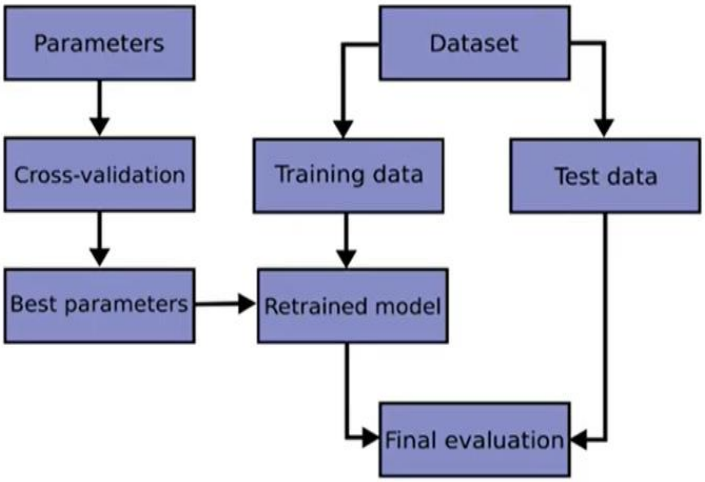
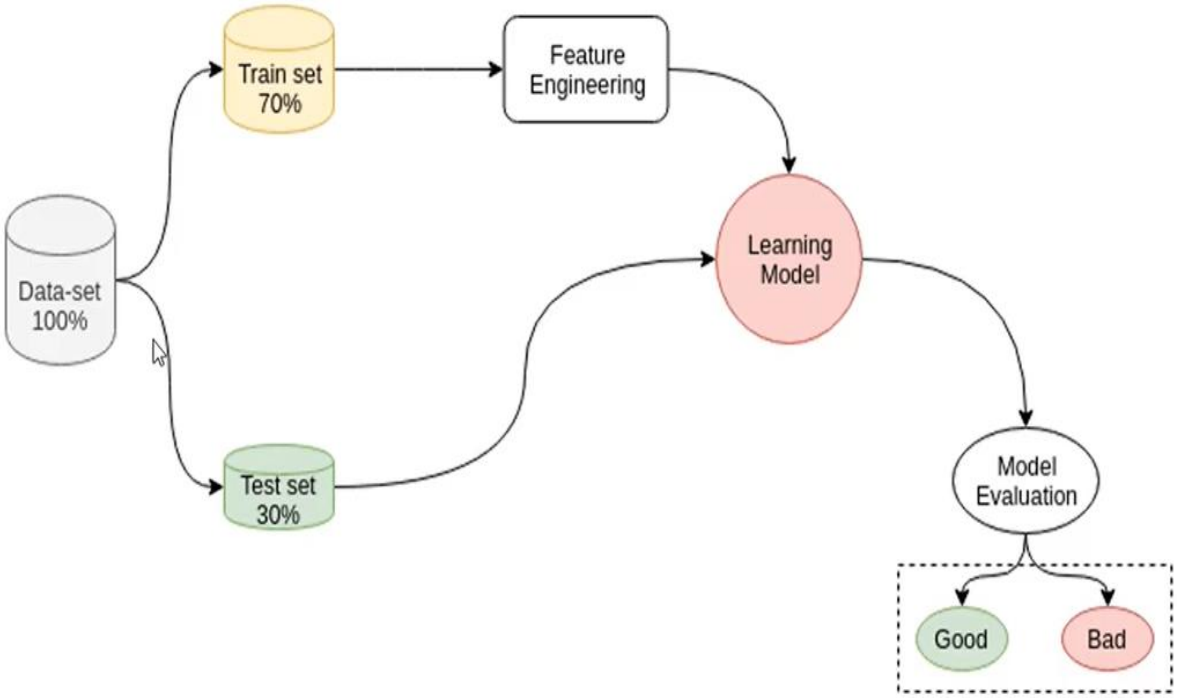
Cross Validation

STUDY MART

When you build your model, you need to evaluate its performance. Cross-validation is a statistical method that can help you with that. It is a technique for evaluating machine learning models by training several models on subsets of the available input data and evaluating them on the complementary subset of the data.



Model Evaluation

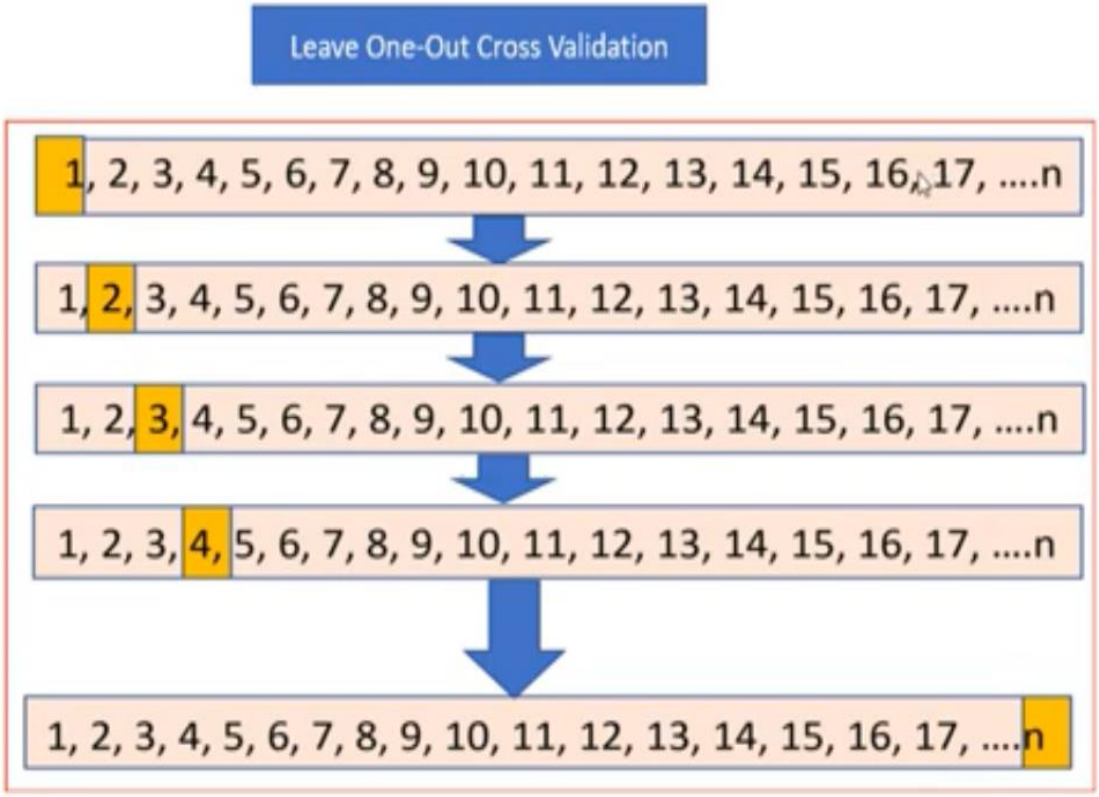


Cross Validation

- K-Fold Cross Validation
- Leave P-Out Cross Validation (LPOCV)
- Leave One-Out Cross Validation (LOOCV)
- Repeated Random Sub-sampling Method
- Holdout Method

- Time Series Split Cross-Validation
- Blocked Cross-Validation

Cross Validation



Cross Validation

STUDY MART

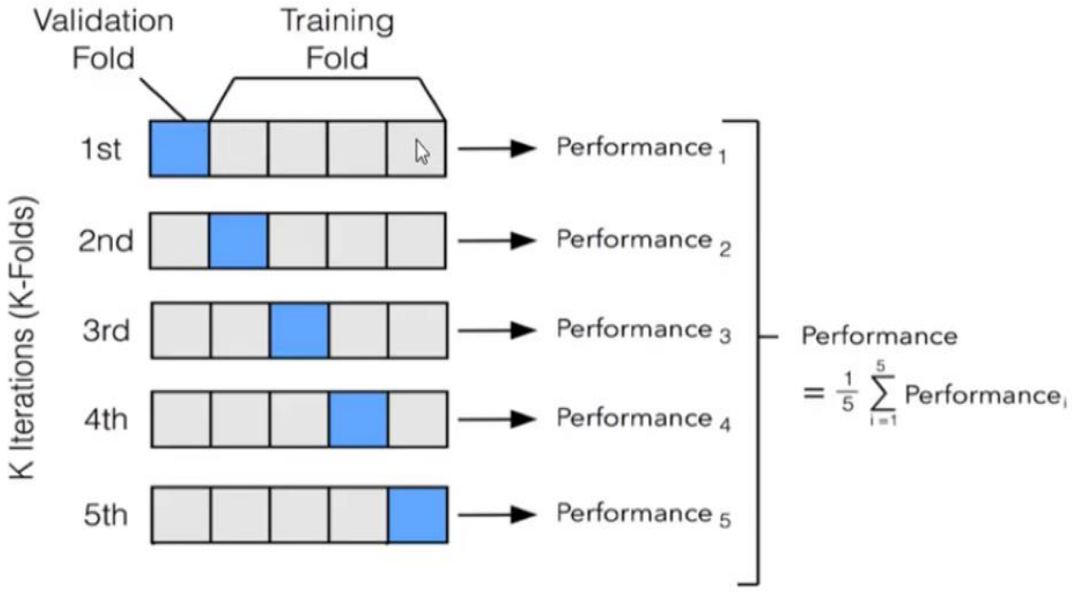
k Fold Cross Validation

The general procedure is as follows:

1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
    1. Take the group as a hold out or test data set
    2. Take the remaining groups as a training data set
    3. Fit a model on the training set and evaluate it on the test set
    4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores

Cross Validation

k Fold Cross Validation



Cross Validation

k Fold Cross Validation

Advantage:

1. Computation time is reduced as we repeated the process only 10 times when the value of k is 10.
2. Reduced bias
3. Every data points get to be tested exactly once and is used in training k-1 times
4. The variance of the resulting estimate is reduced as k increases

Zhom