

# cross validation

The machine learning culture revolves around effective evaluation of predictive models. **Out-of-sample** testing in the form of **Cross Validation** is heavily relied on when training and testing a model.

**Cross Validation** is the most widely used method of machine learning algorithm evaluation on data.

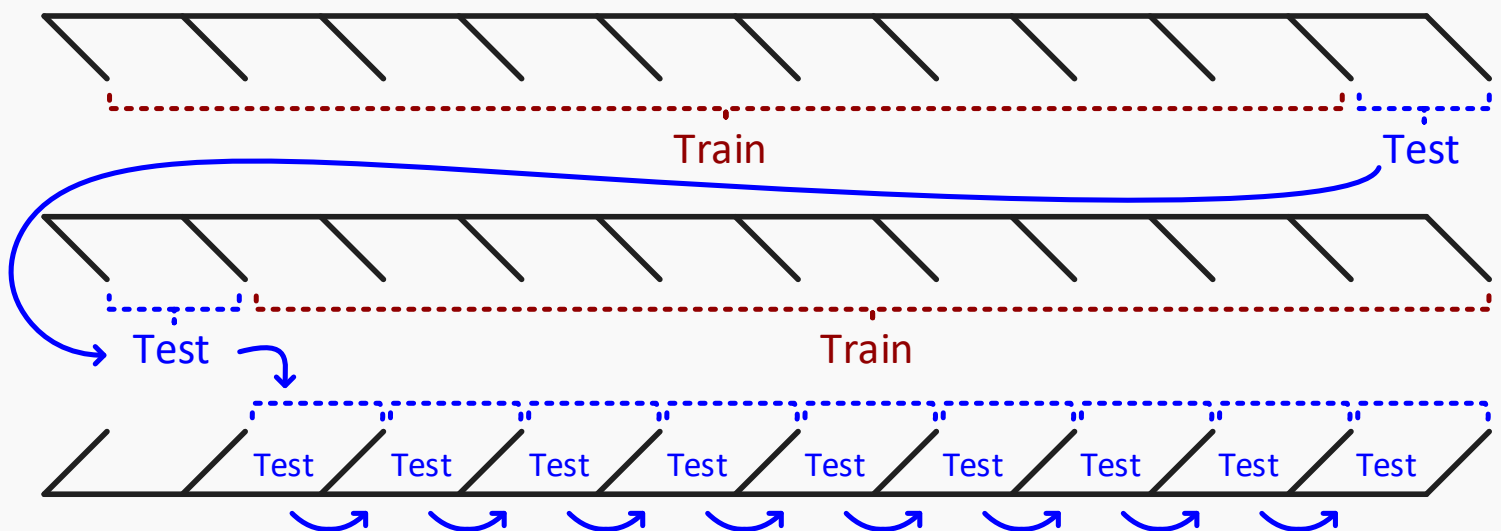
Performing **Cross Validation** requires the follows:

- A dataset
- An algorithm
- An evaluation metric for the quality of result
- Often metrics are squared error between the predicted and actual values

**Cross Validation** is thus executed as follows:

- Dataset is partitioned into approximately 10 equally-sized "**folds**".
- The algorithm is trained on 9 folds, with the evaluation metric computed on the 10<sup>th</sup> fold.
- The training sequence reiterates on the proceeding fold until all folds are evaluated.
  - The process is ultimately repeated 10 times, with each fold in turn as the test fold.
- The mean and standard deviation of the evaluation metric are reported over the 10 folds.

## Cross Validation



The algorithm with the highest measured performance (**average out-of-sample performance across the 10 test folds**) is then applied to the model. Additionally, significance tests can be computed against the performance across the folds.

# nested cross validation

**Nested Cross Validation** is specifically the most used method of tuning parameters in an algorithm.

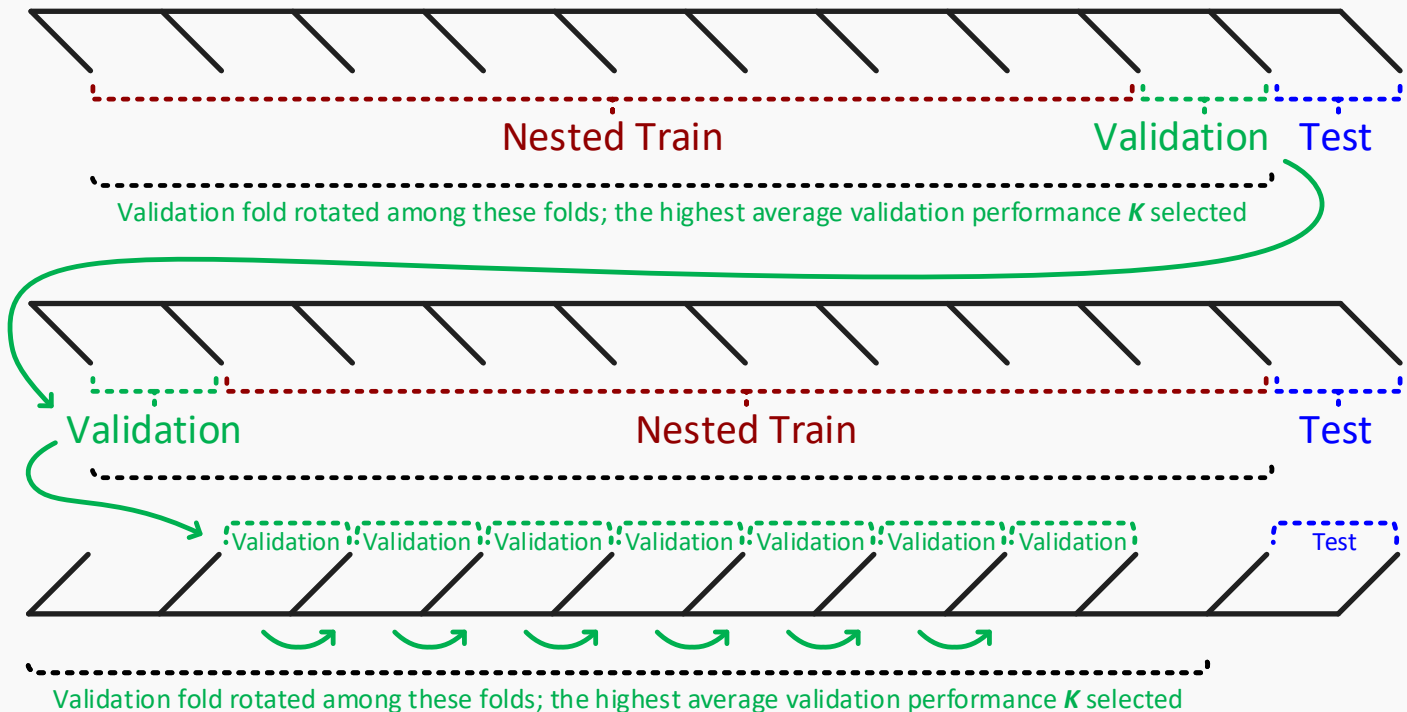
Performing **Nested Cross Validation** requires the follows:

- .. A dataset
- .. An algorithm
- .. An evaluation metric for the quality of result (**additionally a parameter for tuning**)
- .. For illustration, the parameter will be **K** and will be assigned values of **1, 10, 100, 1000, or 10000**.

**Cross Validation** is thus executed as follows:

- .. Dataset is partitioned into approximately 10 equally-sized "**folds**", reserving one for **test**.
- .. Additionally, a fold will be reserved for **validation**.
- .. For **K = 1, 10, 100, 1000, 10000**, the algorithm is trained on the 8 remaining folds, with the evaluation metric computed on the **validation** fold; **5 measurements** are computed.
- .. The training sequence reiterates on the proceeding fold until all folds are evaluated.
  - The process is ultimately repeated 9 times, rotating which training fold is for **validation**.
  - The resulting computations are 9\*5 metrics (**9 folds x 5 K's**).
- .. **K** that minimizes that average training error over the 9 folds is chosen and used to evaluate on the test dataset.
- .. The process is repeated 10 times from the second step, using each fold in turn as the test fold.
- .. The mean and standard deviation of the evaluation metric are reported over the 10 test folds.

## Nested Cross Validation



The algorithm with the highest measured performance (**average out-of-sample performance across the 10 test folds**) where **Nested Cross Validation** was applied is then assigned to the model.

Additionally, significance tests can be computed against the performance across the folds.

**Nested Cross Validation** can be highly computationally expensive

(10 test sets\*10 validation sets\*number of parameter settings being considered).