Cross Validation

Cynthia Rudin

Machine Learning Course, Duke

Cross-Validation

- I tried ridge regression and I tried least square regression, how do I know which one performed better?

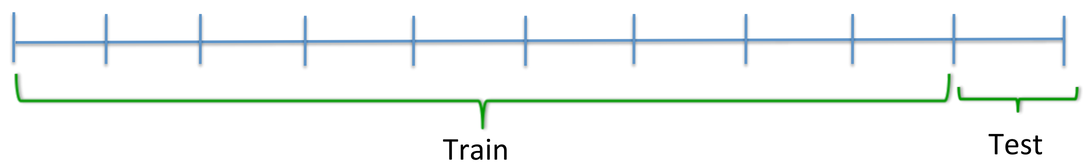
Cross-Validation

- Cross Validation (CV) is the most popular way to evaluate a machine learning algorithm on a dataset.
- You will need a dataset, an algorithm, and an evaluation measure for the quality of the result. The evaluation measure might be the squared error between the predictions and the truth.

Cross-Validation

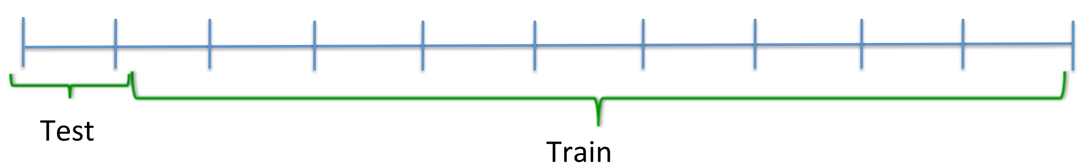
- Cross Validation (CV) is the most popular way to evaluate a machine learning algorithm on a dataset.
- You will need a dataset, an algorithm, and an evaluation measure for the quality of the result. The evaluation measure might be the squared error between the predictions and the truth.
- Divide the data into approximately-equally sized 10 “folds”
- Train the algorithm on 9 folds, compute the evaluation measure on the last fold.
- Repeat this 10 times, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 folds.

Cross-Validation



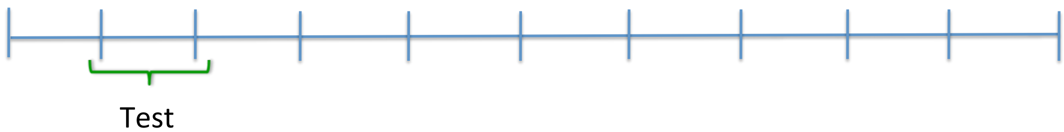
- Divide the data into approximately-equally sized 10 “folds”
- Train the algorithm on 9 folds, compute the evaluation measure on the last fold.
- Repeat this 10 times, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 folds.

Cross-Validation



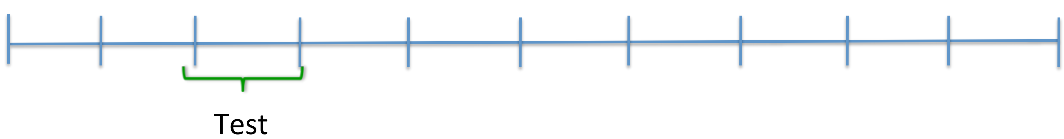
- Divide the data into approximately-equally sized 10 “folds”
- Train the algorithm on 9 folds, compute the evaluation measure on the last fold.
- Repeat this 10 times, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 folds.

Cross-Validation



- Divide the data into approximately-equally sized 10 “folds”
- Train the algorithm on 9 folds, compute the evaluation measure on the last fold.
- Repeat this 10 times, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 folds.

Cross-Validation



- Divide the data into approximately-equally sized 10 “folds”
- Train the algorithm on 9 folds, compute the evaluation measure on the last fold.
- Repeat this 10 times, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 folds.

Cross-Validation

- The algorithm that performed the best was the one with the best average out-of-sample performance across the 10 test folds.
- If desired, compute significance tests on performance across folds.

Nested Cross Validation

Cynthia Rudin

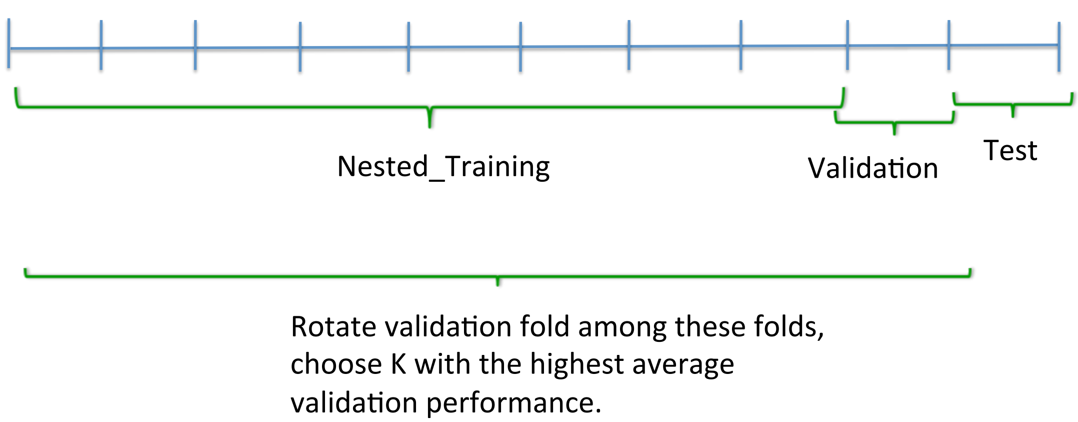
Machine Learning Course, Duke

Nested Cross-Validation

- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.
- You will need a dataset, an algorithm, an evaluation measure for the quality of the result, and a parameter that needs tuning. The parameter is called K and for this example is either 1, 10, 100, 1000, or 10000.

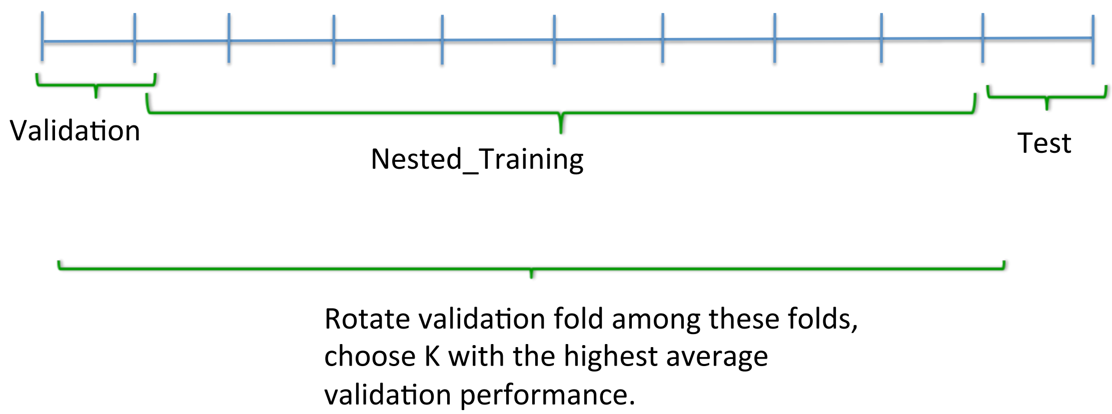
Nested Cross-Validation

- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.



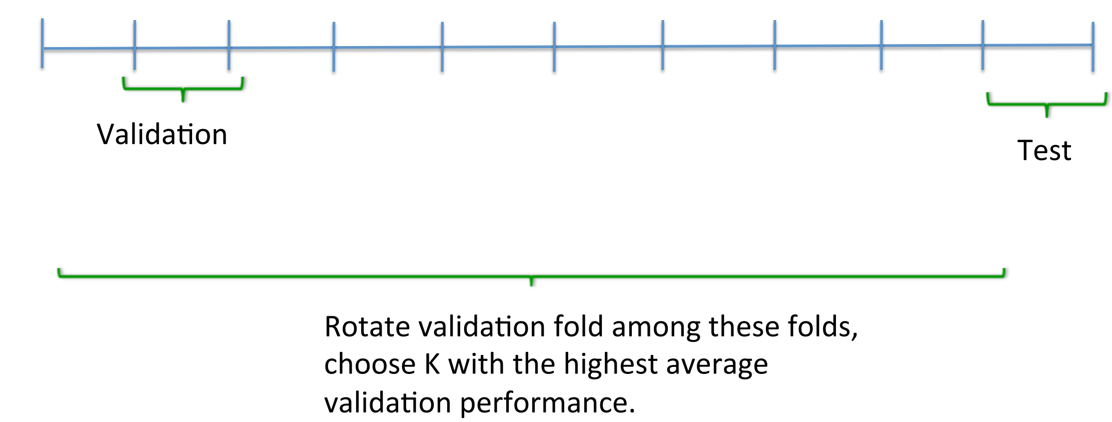
Nested Cross-Validation

- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.



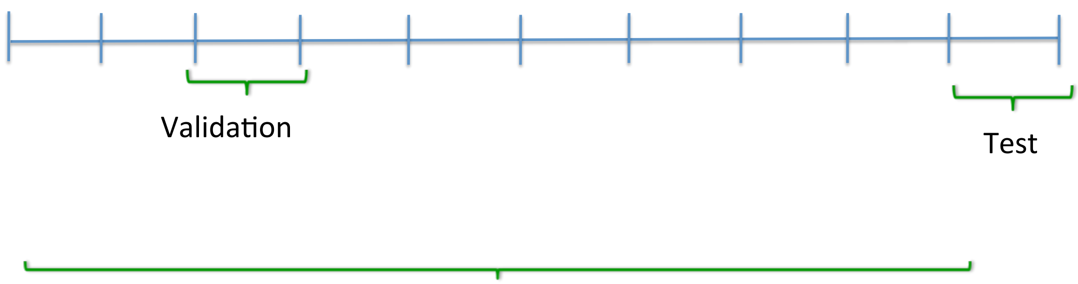
Nested Cross-Validation

- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.



Nested Cross-Validation

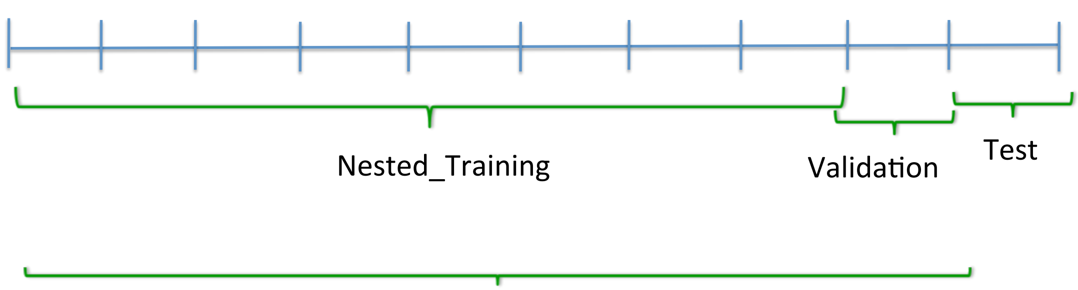
- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.



Rotate validation fold among these folds, choose K with the highest average validation performance.

Nested Cross-Validation

- Nested Cross Validation (CV) is the most popular way to tune parameters of an algorithm.



Rotate validation fold among these folds, choose K with the highest average validation performance.

Nested Cross-Validation

- Divide the data into 10 folds, reserve one for test.
- Reserve one of the 9 training folds for validation.
- For K=1,10,100,1000,10000, train on the 8 remaining training folds and evaluate on the validation fold. We now have 5 measurements.
- Repeat this procedure 9 times. Rotate which training fold is the validation fold. We now have 9*5 (that is, 9 folds*5 K's) measurements.
- Choose K that minimizes the average training error over the 9 folds. Use that K to evaluate on the test set.
- Repeat 10 times from step 2, using each fold in turn as the test fold.
- Report the mean and standard deviation of the evaluation measure over the 10 test folds.

Nice Happy Evaluation Procedure

- The algorithm that performed the best was the one with the best average out-of-sample performance across the 10 test folds, where nested CV was performed 10 times (once for each training set)
- If desired, compute significance tests on performance, where we measure performance of each algorithm across the 10 folds.