

Design and Analysis of Machine Learning Experiments

Ethem Alpaydın
Boğaziçi University, Istanbul

alpaydin@boun.edu.tr
<http://www.cmpe.boun.edu.tr/~ethem/>

It has become more and more apparent in recent years that machine learning systems need to be tested better. We have gone a long way from using a single test set to doing cross-validation and then to paired t tests. Still, unlike most practice, testing should not be a separate step done after all runs are completed; the whole process of experimentation should be designed beforehand, relevant factors defined, strategy of experimentation and the proper experimental procedure should be decided upon and then, only then, the runs should be done and the results analyzed properly (while being aware of all assumptions made) so that any conclusion we arrive at is minimally affected by external sources of variability and is *significant*.

This is called the Design and Analysis of Experiments, and the aim of this tutorial is to show how it can be used in the particular context of machine learning experiments.

We will cover the following topics:

- Basics of Experiment Design: Factors, Response, and the Strategy of Experimentation
- Factorial Design and Response Surfaces
- Randomization, Replication, and Blocking (Pairing)
- Cross-Validation and Resampling Methods
- Measures for Assessing Performance: Accuracy, Precision-Recall, ROC, AUC
- Interval Estimation, Confidence Intervals, and Hypothesis Testing
- Tests for the Expected Error of a Single Classification Algorithm
- Comparing the Expected Errors of Two Classification Algorithms
- Comparing Multiple Algorithms: Analysis of Variance
- Nonparametric Rank Tests for Comparison over Multiple Datasets
- Conclusions: Some General Guidelines for ML Experiments