Exploratory and Inferential Analysis of Benchmark Experiments

Manuel J. A. Eugster and Friedrich Leisch

Department of Statistics, Ludwig-Maximilians-Universität München,
Ludwigstrasse 33, 80539 München, Germany,
firstname.lastname@stat.uni-muenchen.de

Abstract. Benchmark experiments produce data in a very specific format. The observations are drawn from the performance distributions of the candidate algorithms on resampled data sets. `benchmark` is the comprehensive `R` toolbox for the setup, execution and exploratory and inferential analysis of these experiments. The package introduces an additional layer of abstraction (using S4 mechanisms) representing the elements of benchmark experiments. This allows the integration of all statistical learning algorithms available in the `R` system and a consistent way for developing new ones. The consequence of this slight extra work is a standardized setup and analysis of benchmark experiments. The package provides wrapper methods for common learning algorithms available in `R`.

In this presentation we introduce the elements of benchmark experiments and show how to combine them into a flexible framework. The usage is illustrated with exemplary benchmark studies based on common learning algorithms on one or several popular data sets, respectively. We present new visualisation techniques, show how formal test procedures can be used to evaluate the results, and, finally, how to sum up to an overall ranking.