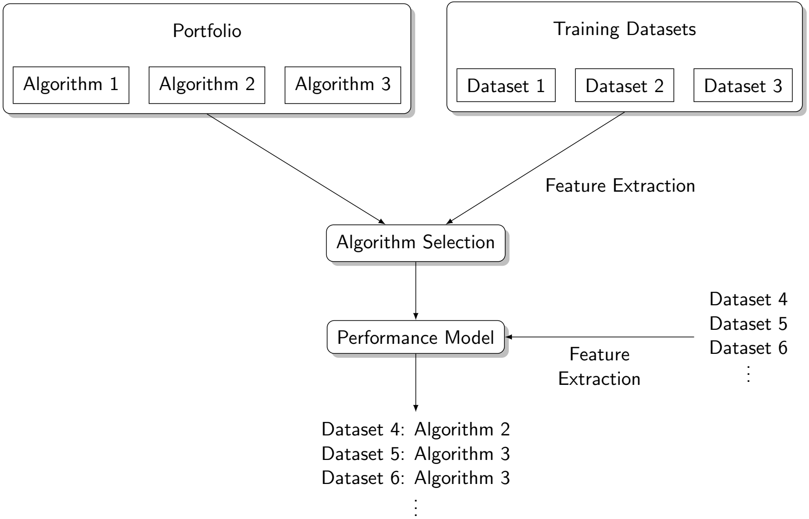
AutoML: Algorithm Selection

Algorithm Selection

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

Algorithm Selection



Algorithm Portfolios

- instead of a single algorithm, use several (hopefully complementary) algorithms
- idea from Economics – minimize risk by spreading it out across several securities
- same here – minimize risk of algorithm performing poorly
- in practice often constructed from algorithms known to perform well
- idea similar to ensembles or boosting – leverage strengths and alleviate weaknesses, but learn which algorithm to choose for a particular dataset

Algorithms

“algorithm” used in a very loose sense

- different learners
- different parameterizations of the same learner
- different ensembles, boosted learners
- different machine learning workflows/pipelines
- ...

Evaluation of Portfolios

- single best algorithm
  - algorithm with the best performance across all datasets
  - lower bound for performance of portfolio – hopefully we are better!
- virtual best algorithm
  - choose the best algorithm for each dataset
  - corresponds to oracle predictor or overhead-free parallel portfolio
  - upper bound on portfolio performance

Parallel Portfolios

Why not simply run all algorithms in parallel?

- not enough resources may be available/waste of resources
- algorithms may be parallelized themselves
- memory contention
- . . .

Building an Algorithm Selection System

- most approaches rely on (meta-)machine learning
- train with representative data, i.e. performance of all algorithms in portfolio on representative datasets
- evaluate performance on separate set of datasets
- potentially large amount of prep work
- existing repositories of machine learning performances (e.g. OpenML) can help

Choosing Datasets

- we want selectors that generalize, i.e. good for more than one dataset
- split datasets into training set (which we learn a selector on) and test set (which we only evaluate performance on)
- need to balance easy/hard datasets in both sets
- may need a lot of data

Key Components of an Algorithm Selection System

- feature extraction
- performance model
- prediction-based selector

optional:

- presolver
- secondary/hierarchical models and predictors (e.g. for feature extraction time to avoid spending a long time for small performance gains)