

Autokit: automatic machine learning via representation and model search

Tadej Štajner¹ - tadej@tdj.si

¹Jožef Stefan International Postgraduate School

11.07.2015, AutoML Workshop at ICML 2015, Lille, France

Key ideas

Define a pipeline space for HYPEROPT that covers many orthogonal strategies.

- ▶ Non-linearities: kernel approximation
- ▶ Independence assumptions: Naive Bayes
- ▶ Feature interaction: Random Forests
- ▶ Preprocessing: normalization, scaling

Key ideas

Parameter and pipeline space as a function of the dataset:

- ▶ Number of latent components: prior is a log-uniform distribution up to $\min(30, \frac{\#features}{4})$
- ▶ Sparse data: skip decision trees, skip PCA

Implementation - AUTOKIT

- ▶ 3rd place (out of 68) on AutoML Round 1
- ▶ Code available at <https://github.com/tadejs/autokit>
- ▶ Based on HYPEROPT, HYPEROPT-SKLEARN