

Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms

Chris Thornton

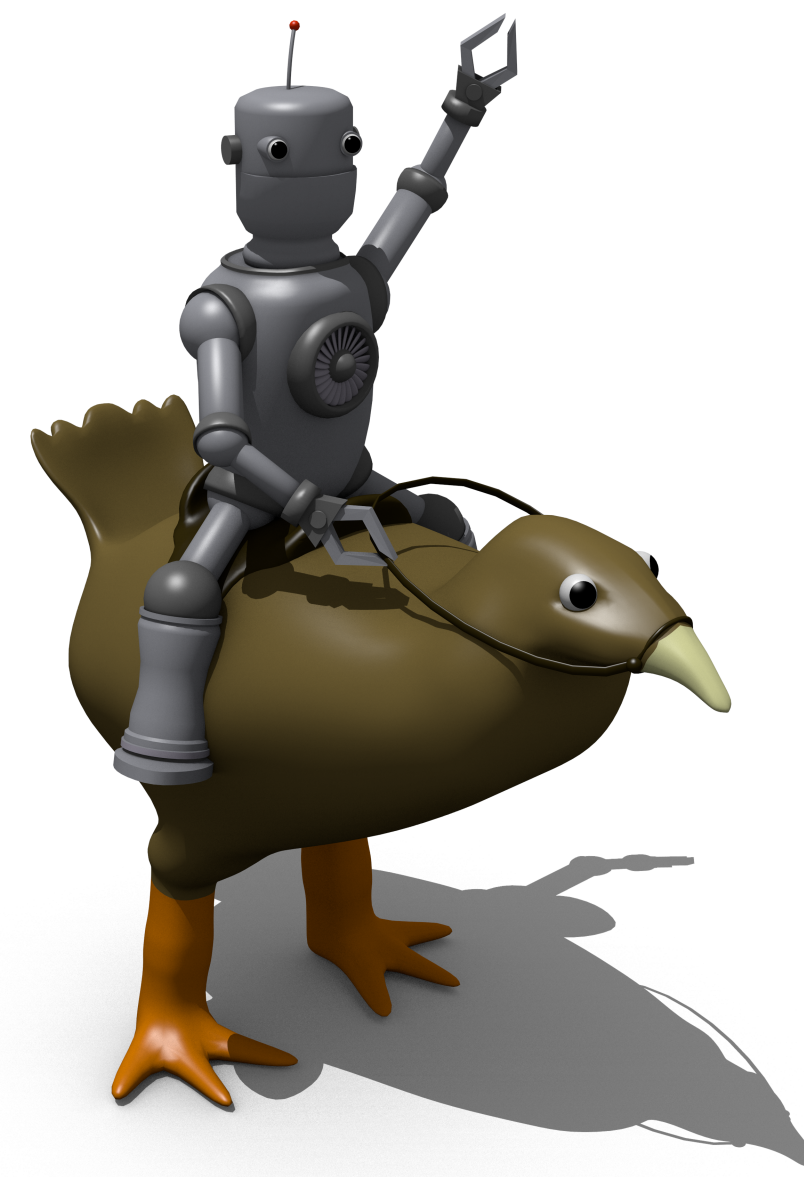
Frank Hutter

Holger H. Hoos

Kevin Leyton-Brown

University of British Columbia, Vancouver, Canada

The 15 Second Version



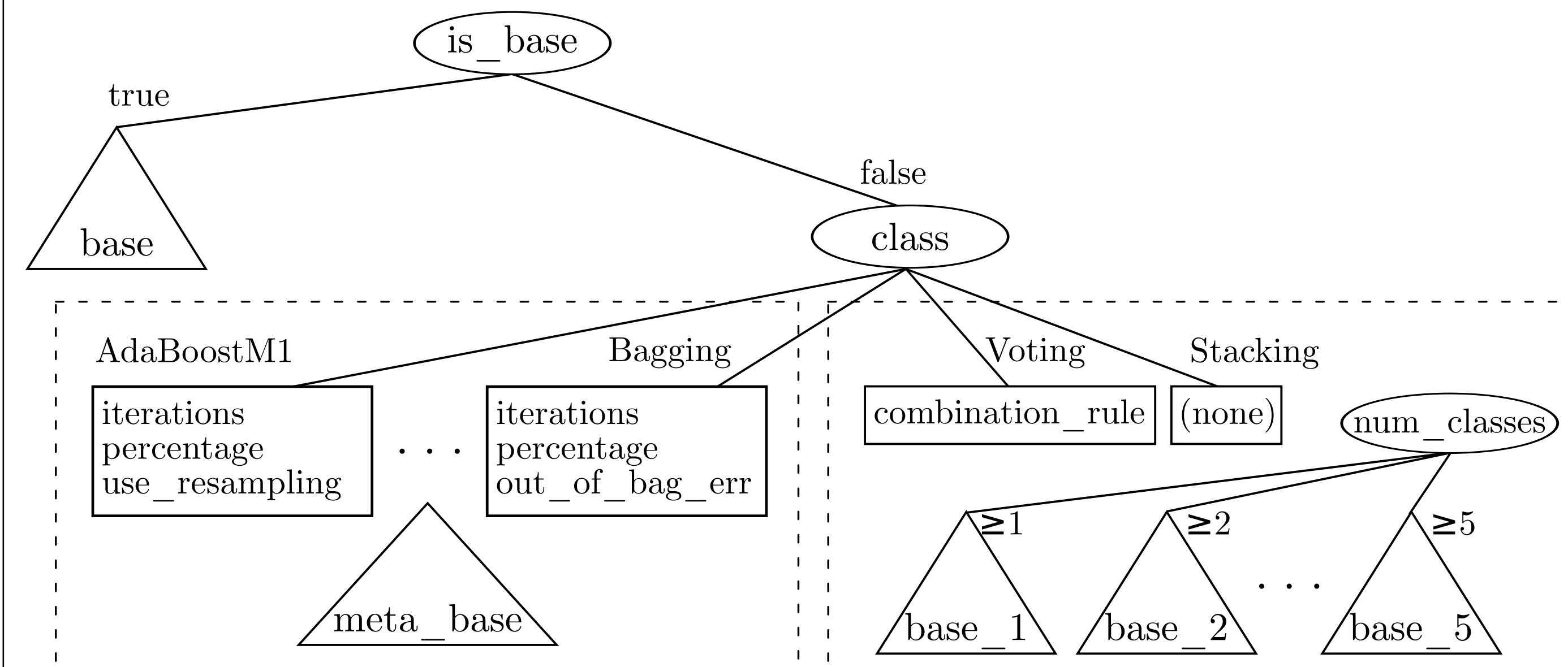
- Auto-WEKA can automatically solve your problem of selecting a classification algorithm and hyperparameters that achieve good performance on your dataset.
- You only have to set the time budget/memory limits, and let Auto-WEKA do the rest.
- In experiments on 21 benchmark datasets, Auto-WEKA showed improvements over baselines in 15 cases, reducing error by up to 5.5%.

• Get it at <http://www.cs.ubc.ca/labs/beta/Projects/autoweika/>

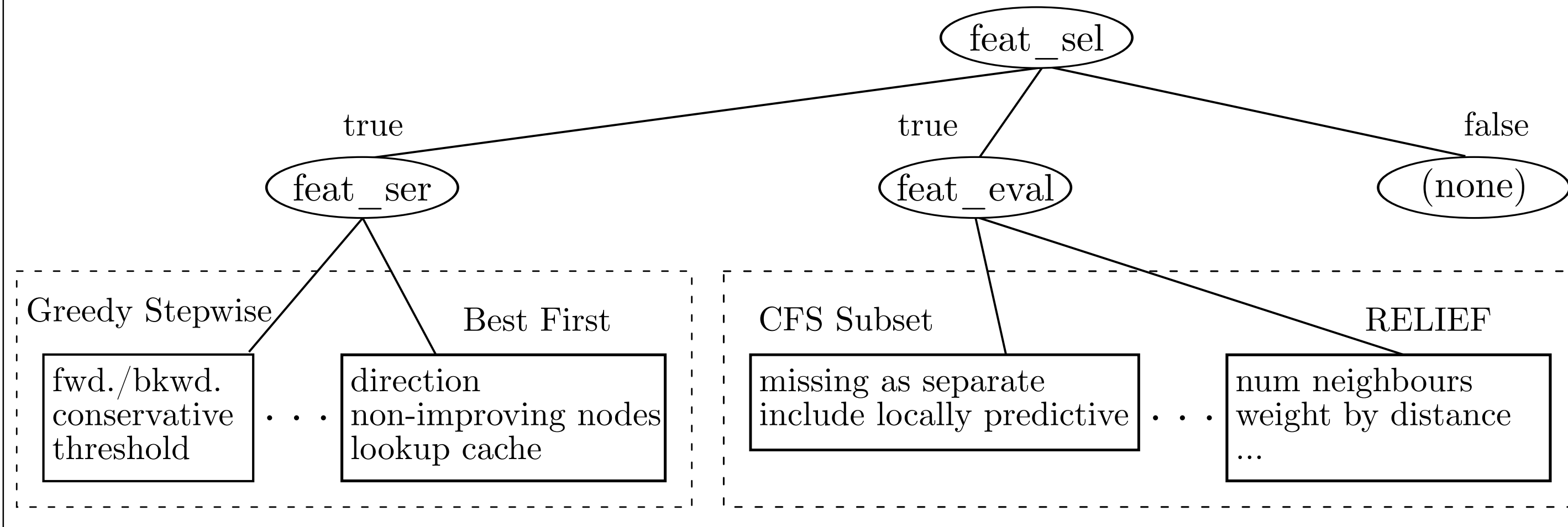
Auto-WEKA

- Optimizes over the 39 classifiers in WEKA (27 base, 10 meta and 2 ensemble methods), in addition to feature/attribute selection.
 - Search space has 786 parameters.
- Optimizes 10-fold CV error on training data using any general purpose algorithm configuration method.
 - We use two existing Sequential Model Based Optimization (SMBO) methods.

Algorithm/Hyperparameter Search Space

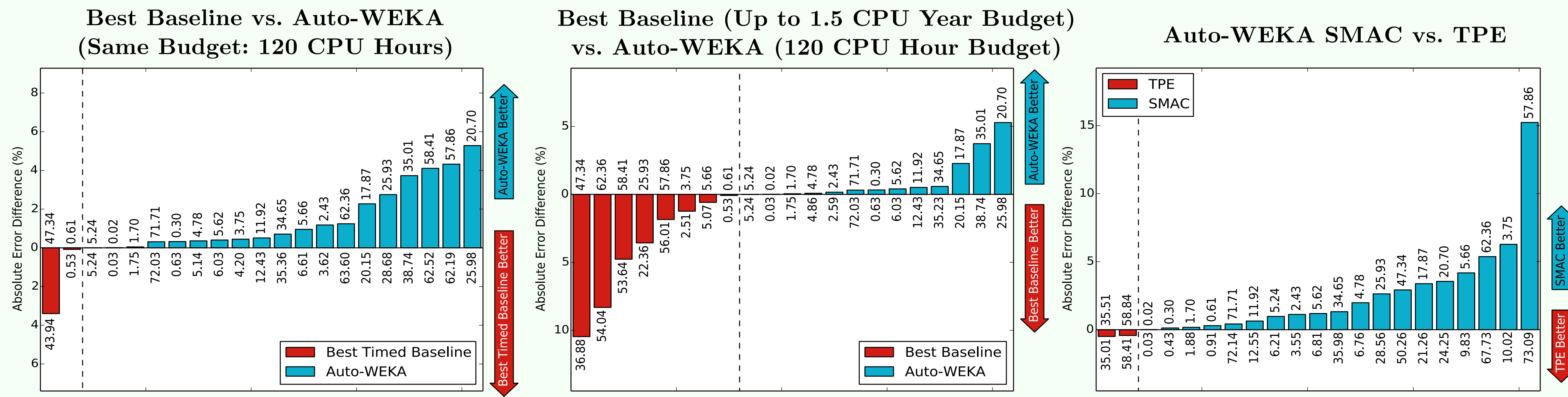


Feature/Attribute Selection Search Space

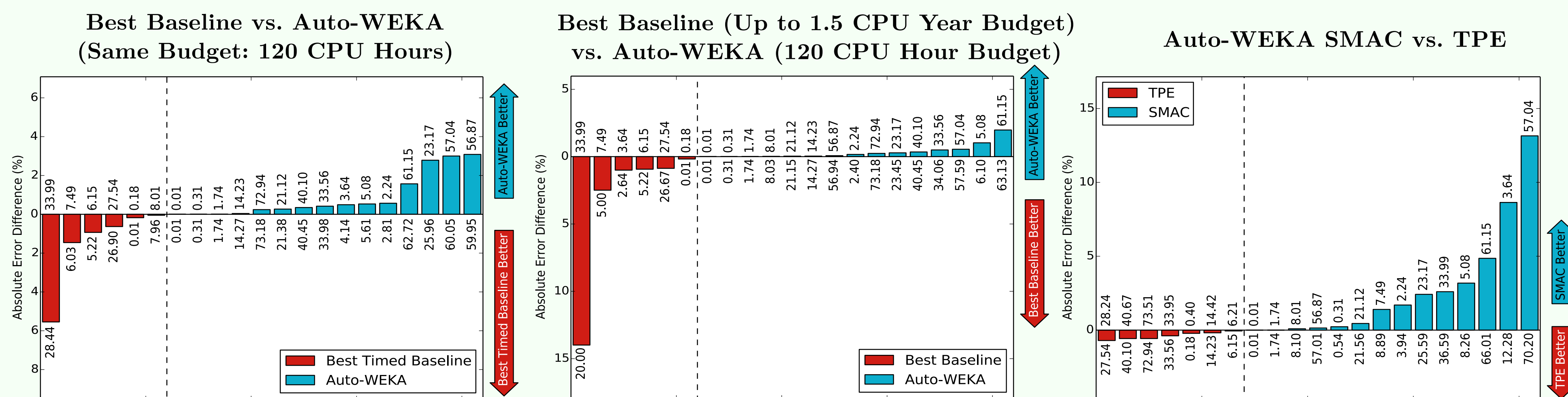


Results

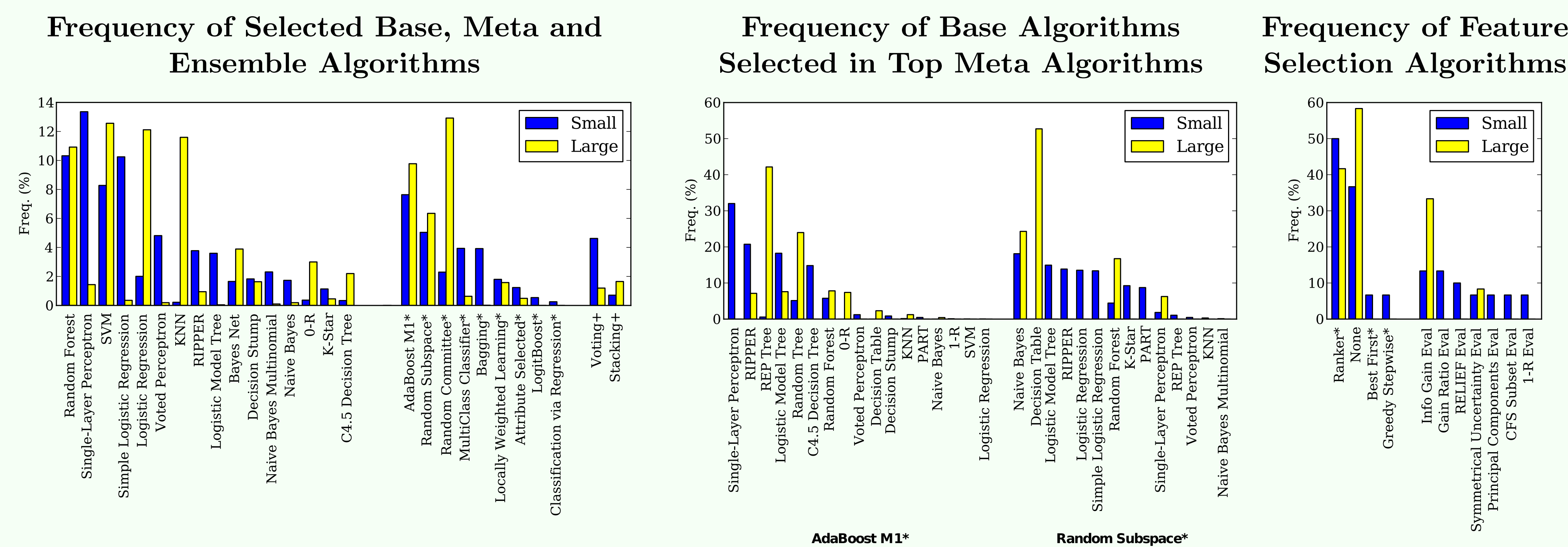
10-Fold CV (Training) Performance



Generalization (Testing) Performance



Selected Algorithms



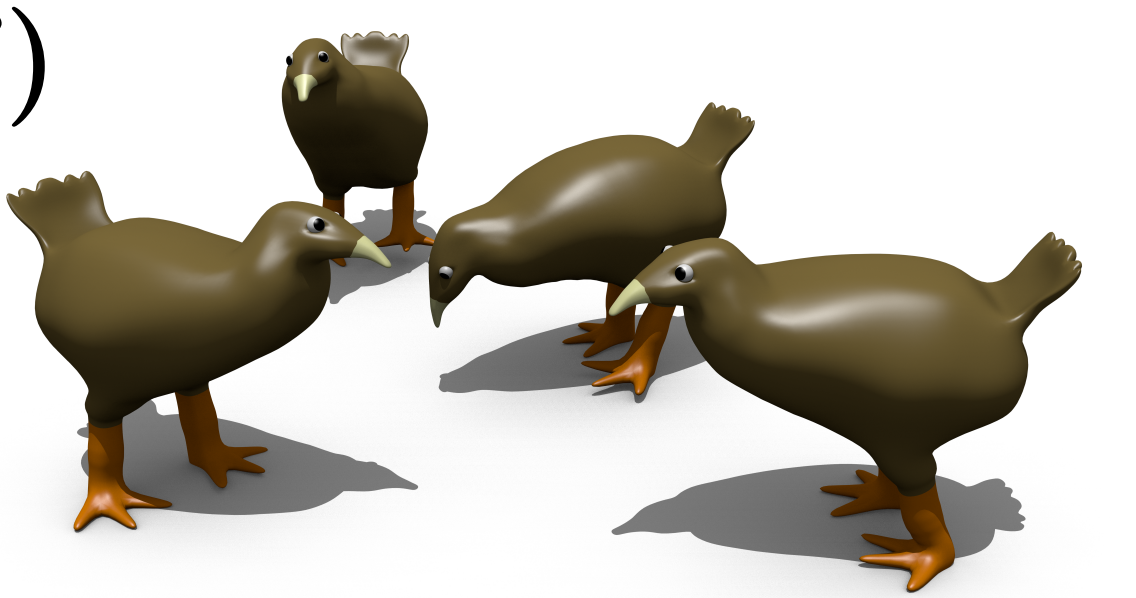
Detailed Results on Largest Datasets

Dataset	Oracle Perf. (%)				10-Fold C.V. Performance (%)				Test Performance (%)				
	Ex-Def		Grid Search		Ex-Def	Grid Search	Rand. Search	Auto-WEKA	Ex-Def	Grid Search	Rand. Search	Auto-WEKA	
	Best	Worst	Best	Worst	Ex-Def	Grid Search	TPE	SMAC	Ex-Def	Grid Search	TPE	SMAC	
CIFAR-10-Small	65.91	90.00	52.16	90.36	66.59	53.64	67.33	58.41	58.84	65.91	56.94	66.12	57.01
MNIST Basic	5.19	88.75	2.58	88.75	5.12	2.51	5.05	10.02	3.75	5.19	2.64	5.05	12.28
Rot. MNIST + BI	63.14	88.88	55.34	93.01	66.15	56.01	68.62	73.09	57.86	63.14	57.59	66.40	70.20
Shuttle	0.0138	20.8414	0.0069	89.8207	0.0328	0.0361	0.0345	0.0251	0.0224	0.0138	0.0414	0.0157	0.0130
KDD09-Apptency	1.7400	6.9733	1.6332	54.2400	1.8776	1.8735	1.7510	1.8776	1.7038	1.7405	1.7400	1.7400	1.7381
CIFAR-10	64.27	90.00	55.27	90.00	65.54	54.04	69.46	67.73	62.36	64.27	63.13	69.72	66.01

Baselines and SMBO Methods

Exhaustive Default (Ex-Def)

- Use default hyperparameters on all WEKA classifiers, pick the best.

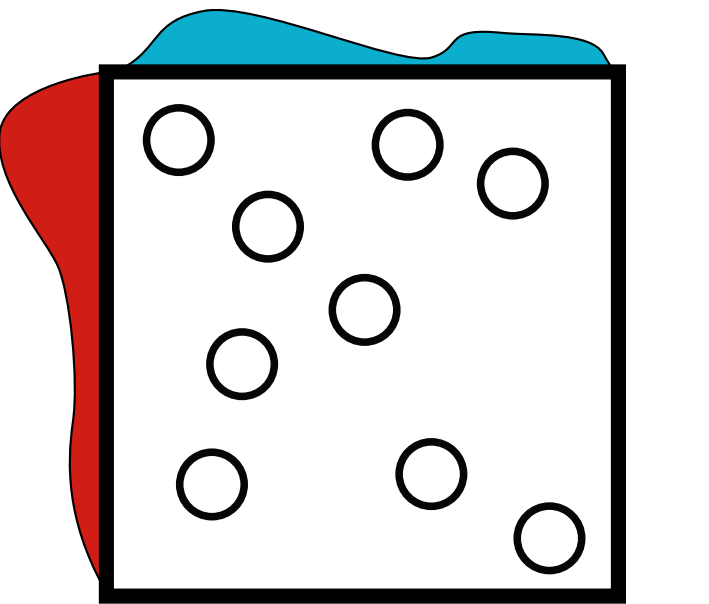


Grid Search

- Discretize hyperparameters of base classifiers.
- Use the 10-fold CV performance to select the classifier and hyperparameters.
- High computational cost - an average of 6 CPU months.

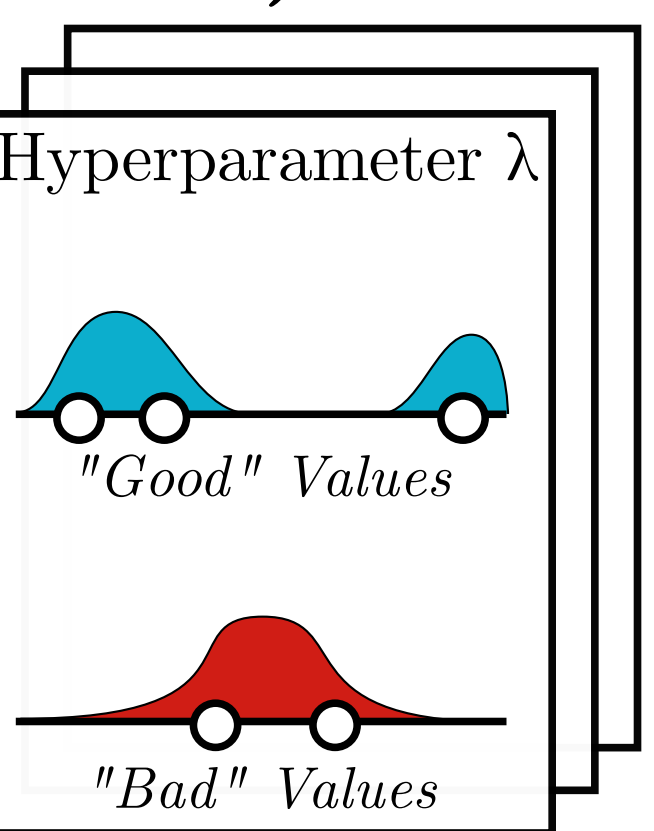
Random Search

- Select classifiers/hyperparameters at random.
- Shown to be competitive with grid search.



Tree-structured Parzen Estimator (TPE)

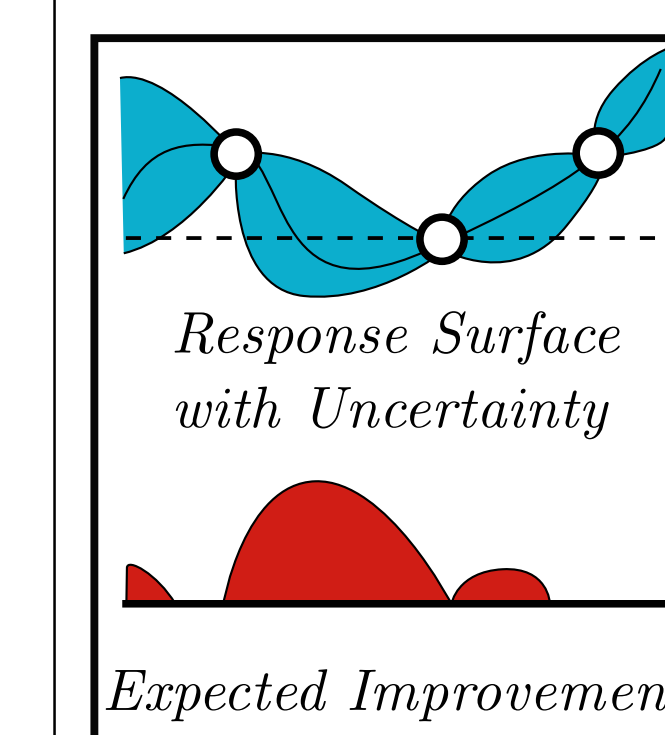
- 1-D Parzen estimator for each hyperparameter - models "good" vs "bad".
- Tree structure of estimators enforces hyperparameter conditionality.



[Bergstra, Barden, Bengio, Kégl, 2011]

Sequential Model-based Algorithm Configuration (SMAC)

- Builds a single regression forest model of performance.
- Alternates between random/model chosen hyperparameters.
- Only computes the performance of a single CV fold at a time - quickly rejects poor hyperparameters.



[Hutter, Hoos, Leyton-Brown, 2011]

Future Work

- More sophisticated use of the given dataset to exploit the performance gains on training data.
- Modify the SMBO methods to increase learned knowledge about base algorithm hyperparameters.

