

Hyper-parameter search methods for better model recommendation

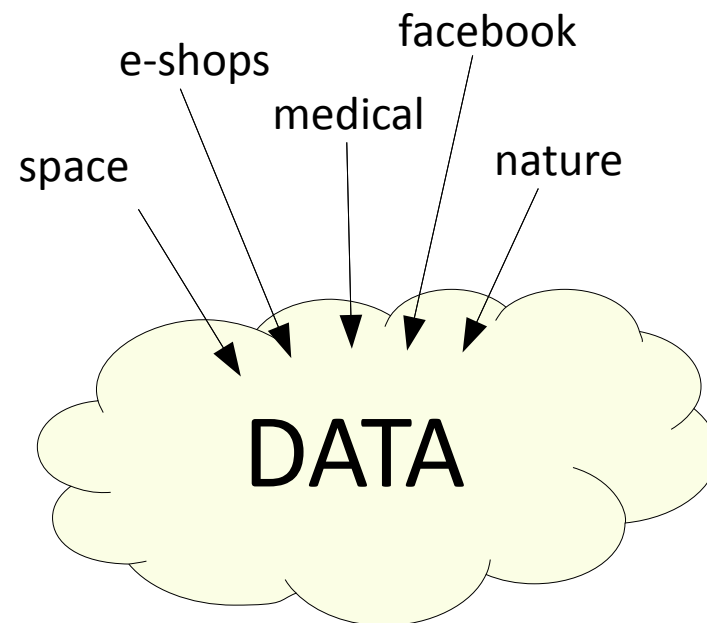
Klára Pešková¹, Roman Neruda²

¹Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, klara@pisecko.cz

²Institute of Computer Science, Czech Academy of Sciences, Prague, Czech Republic

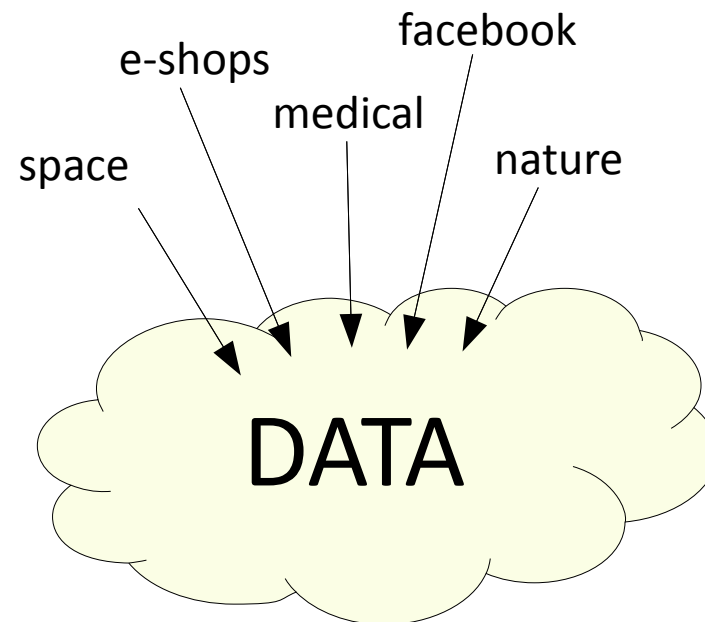
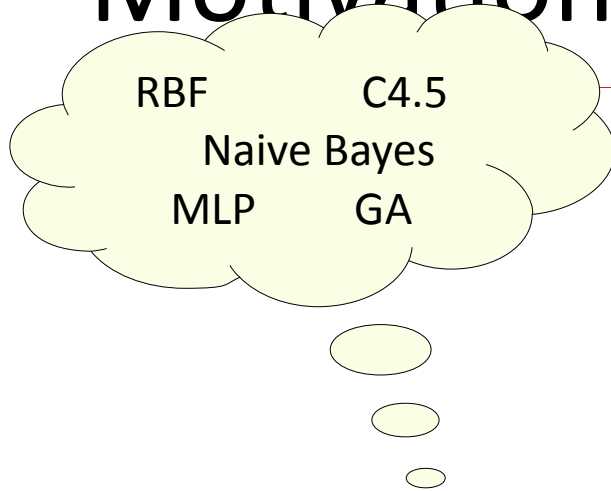
Motivation

[WHY]



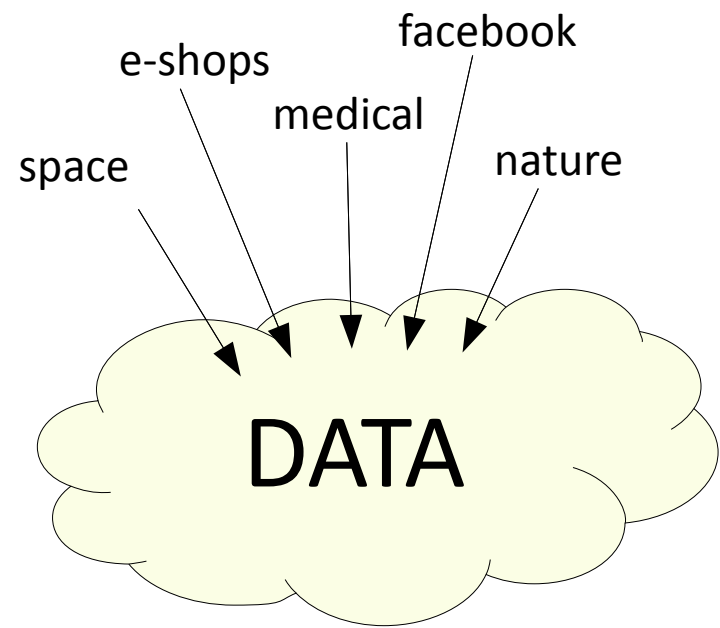
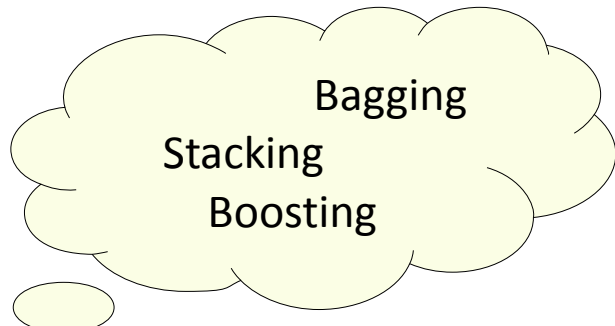
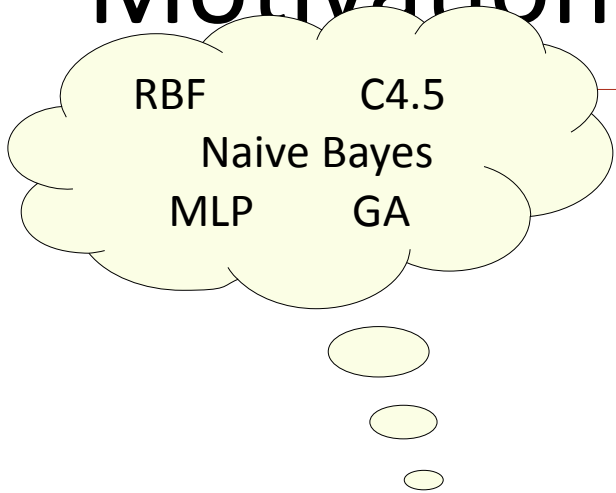
Motivation

[WHY]



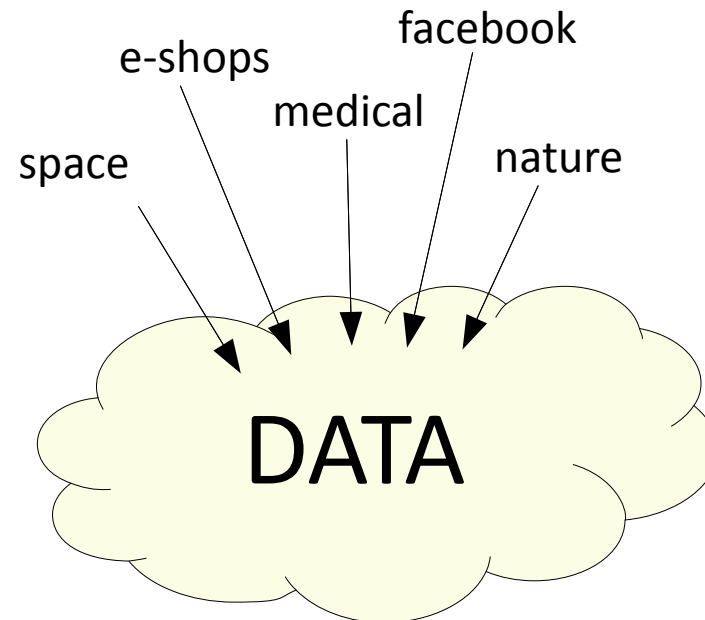
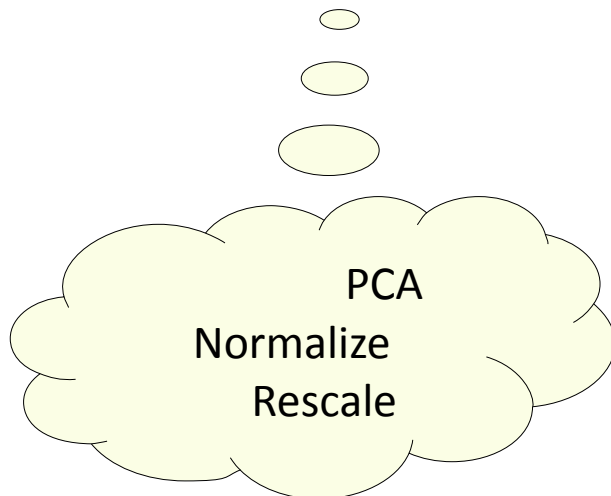
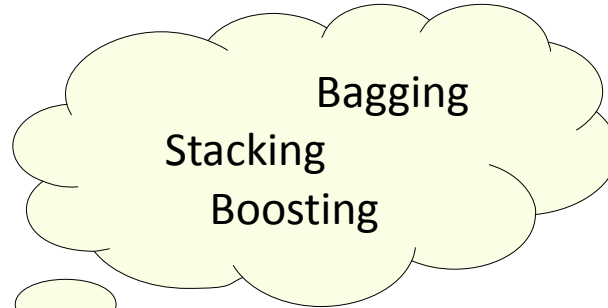
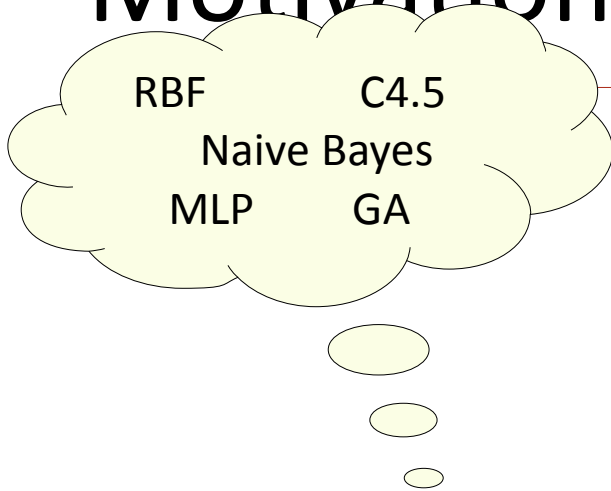
Motivation

[WHY]



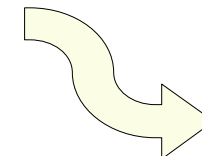
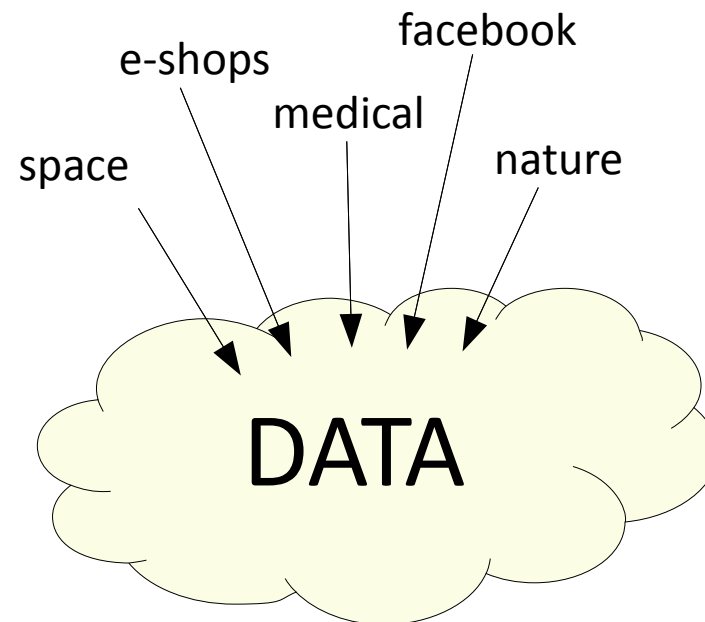
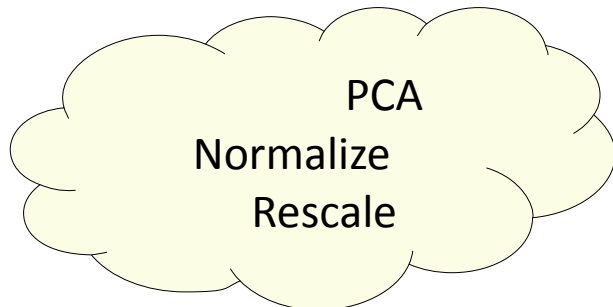
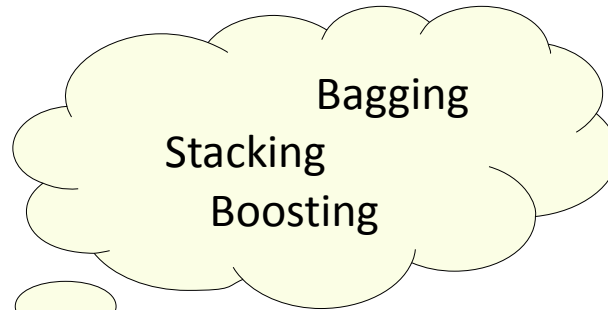
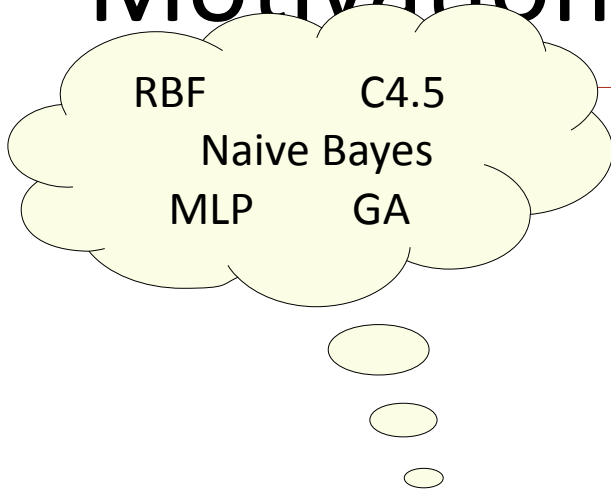
Motivation

[WHY]



Motivation

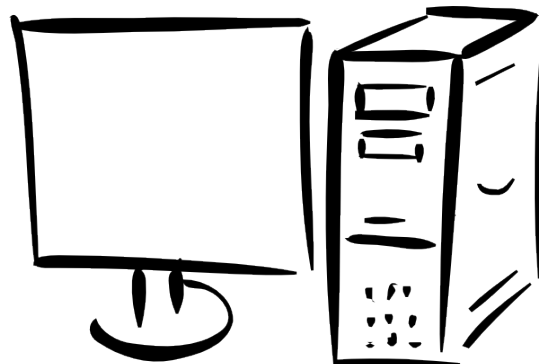
[WHY]



knowledge

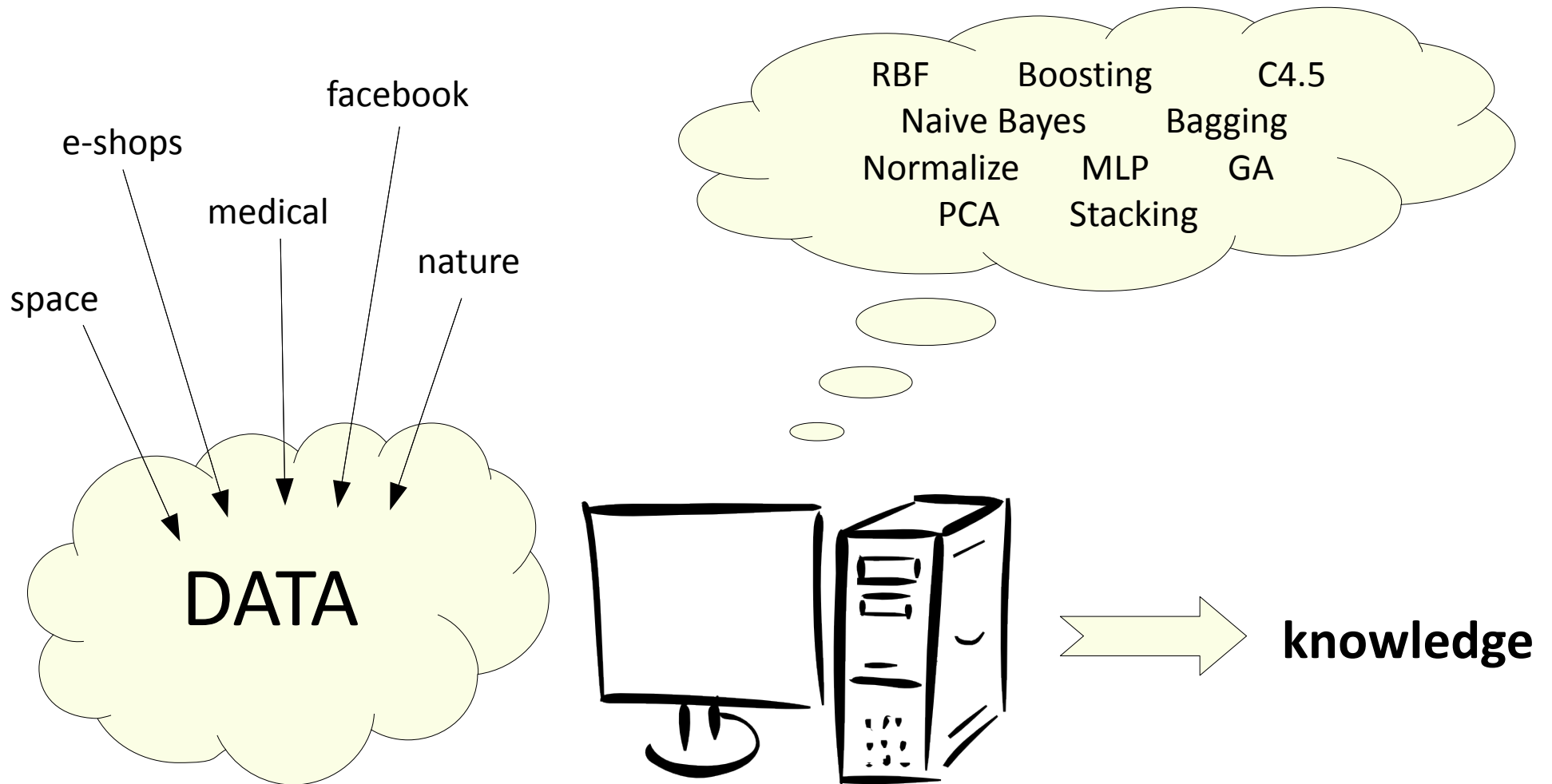
Solution

[WHAT]



Solution

[WHAT]



Metalearning challenges

[WHAT]

- **Two main tasks of automated data mining:**
 - **data dependent model recommendation (for previously unseen dataset)**
 - **model's hyper-parameter recommendation.**

Metalearning challenges

[WHAT]

- **Two main tasks of automated data mining:**
 - **data dependent model recommendation (for previously unseen dataset)**
 - **model's hyper-parameter recommendation.**
- **Combination of these two approaches**

Metalearning challenges

[WHAT]

- **Two main tasks of automated data mining:**
 - **data dependent model recommendation (for previously unseen dataset)**
 - Metadata with suitable metrics
 - History of experiments (method, dataset, results)
 - Computational intelligence methods on metadata
 - **model's hyper-parameter recommendation.**

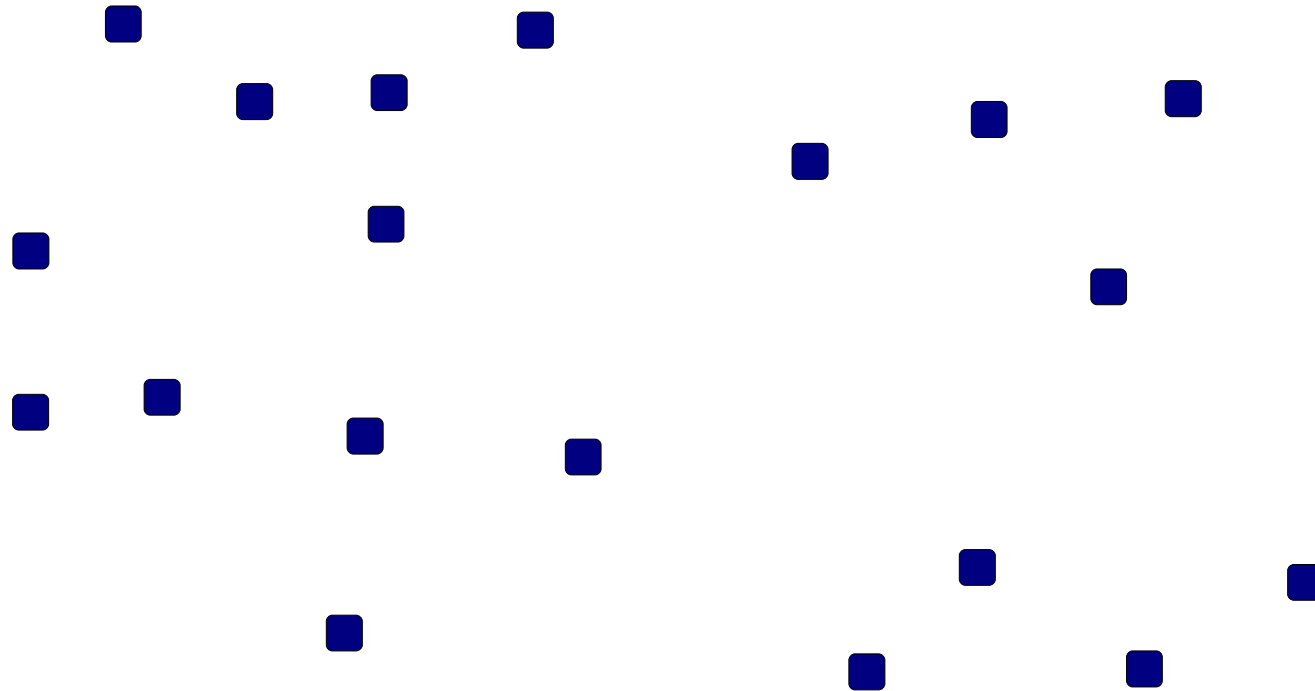
- **Combination of these two approaches**

Metalearning challenges

[WHAT]

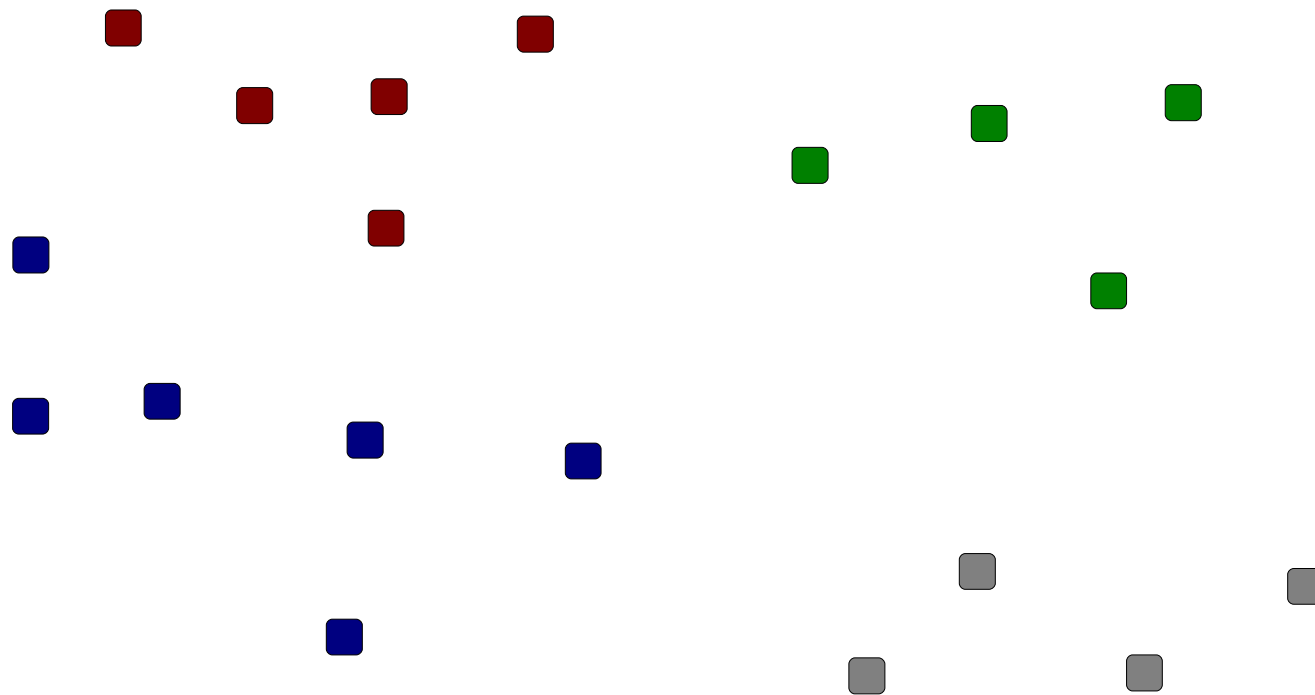
- **Two main tasks of automated data mining:**
 - **data dependent model recommendation (for previously unseen dataset)**
 - Metadata with suitable metrics
 - History of experiments (method, dataset, results)
 - Computational intelligence methods on metadata
 - **model's hyper-parameter recommendation.**
 - History of experiments
 - Search parameter spaces
- **Combination of these two approaches**

Method recommendation process



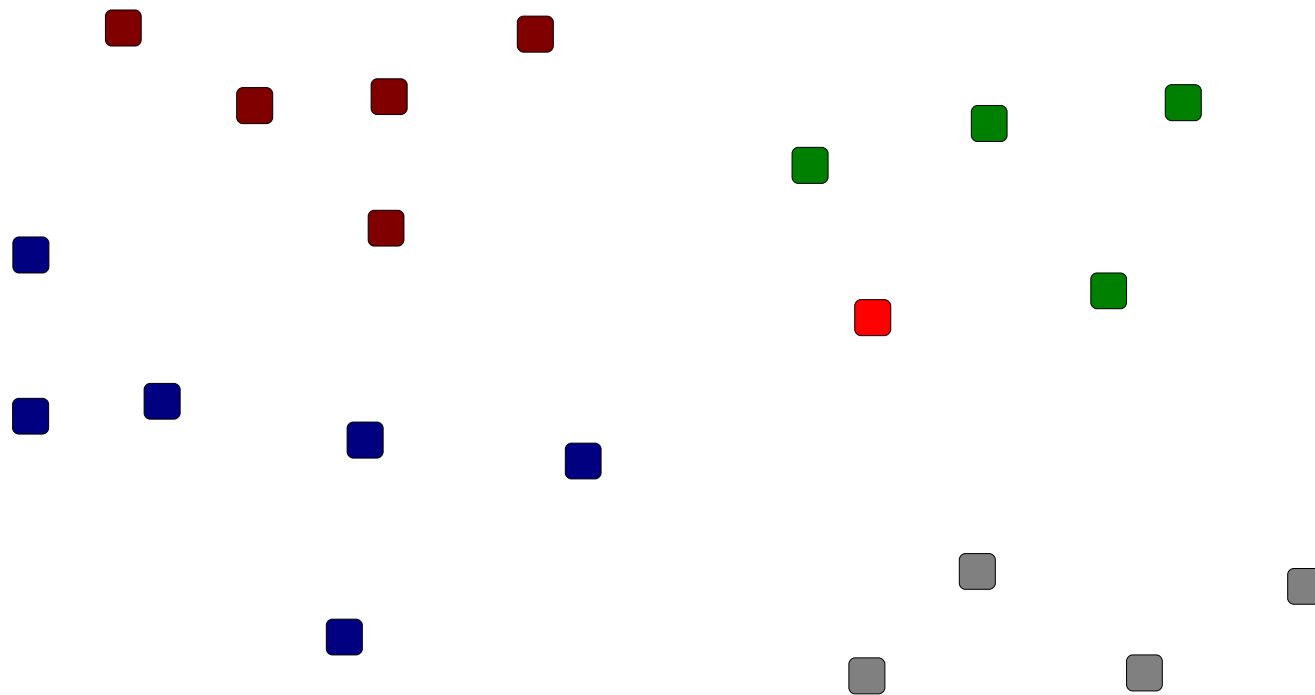
Method recommendation process

A. Training



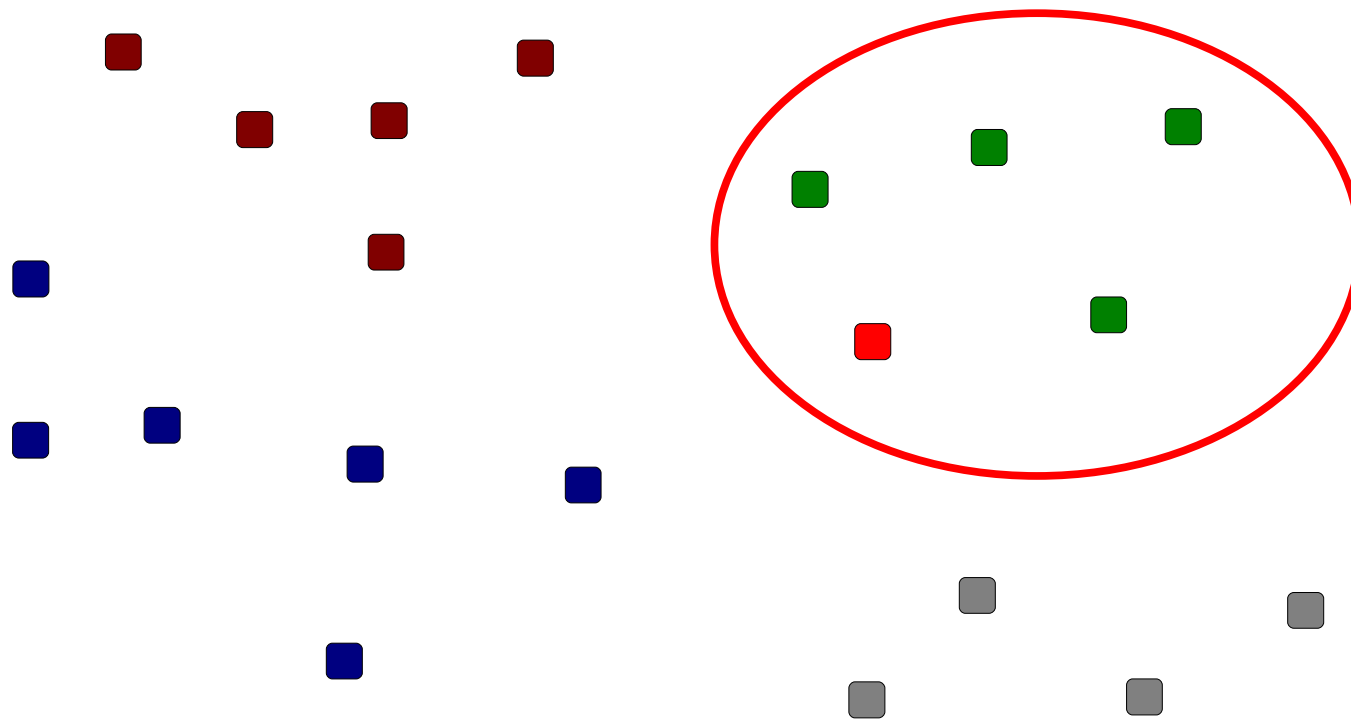
Method recommendation process

=> new dataset



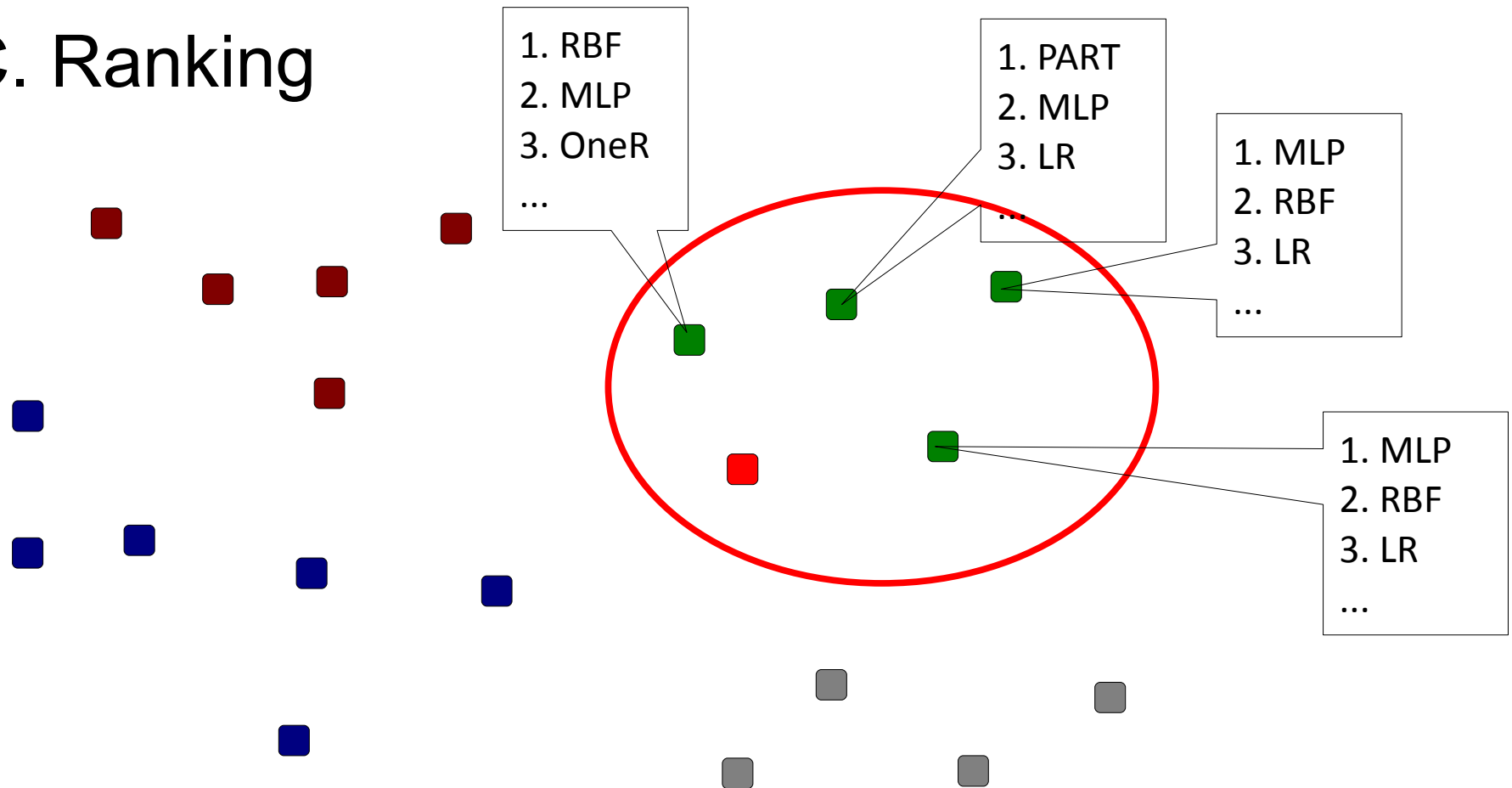
Method recommendation process

B. Zooming



Method recommendation process

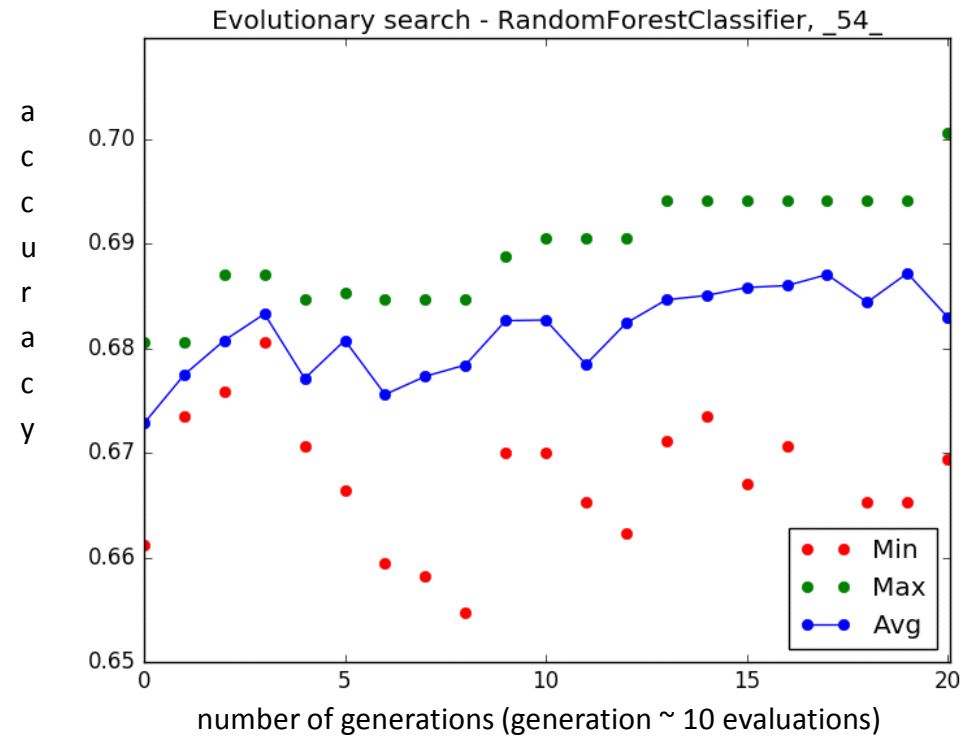
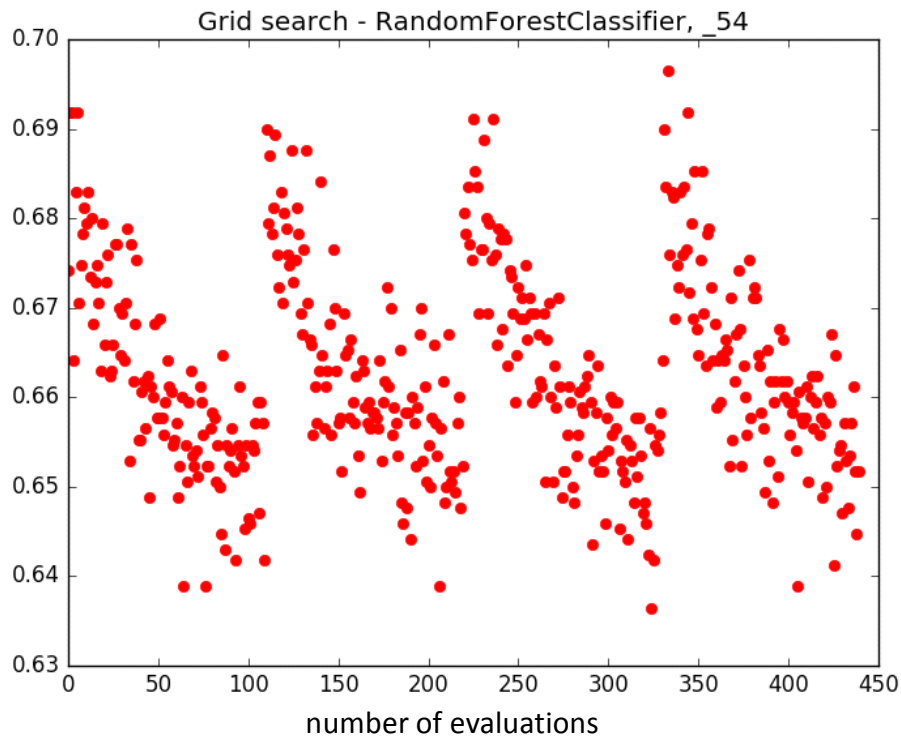
C. Ranking



Searching the parameter space [**HOW**]

- goal: to **optimize the parameters of the method**
- find a suitable **range** of the hyper-parameter values
- search procedures to **search** through the subset of hyper-parameter space

Examples of search algorithms



- **vehicle dataset** (classification task, 846 instances, 19 features, 4 classes)
- 4 parameters were optimized

Technical stuff

- OpenML - datasets, metadata
- Python implementation:
 - **scikit-learn** methods, visualization
 - **deap** evolutionary algorithms library



Thank you for your attention

Metadata, Metrics

Metadata

- categorical attributes:
 - number of attributes
 - number of instances
 - data type
 - missing values
- simple measures:
 - categorical / integer / real ratio
 - missing values ratio
- Information theoretic measures - data complexity

Metrics

- distance between datasets - weighted sum of distances between metadata attributes

$$d(m_1, m_2) = \sum_{i=1}^n w_i \cdot d_i(m_1[i], m_2[i])$$

- weights: default value (1.0), evolutionary optimization