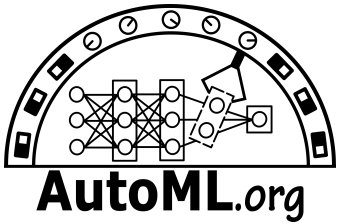
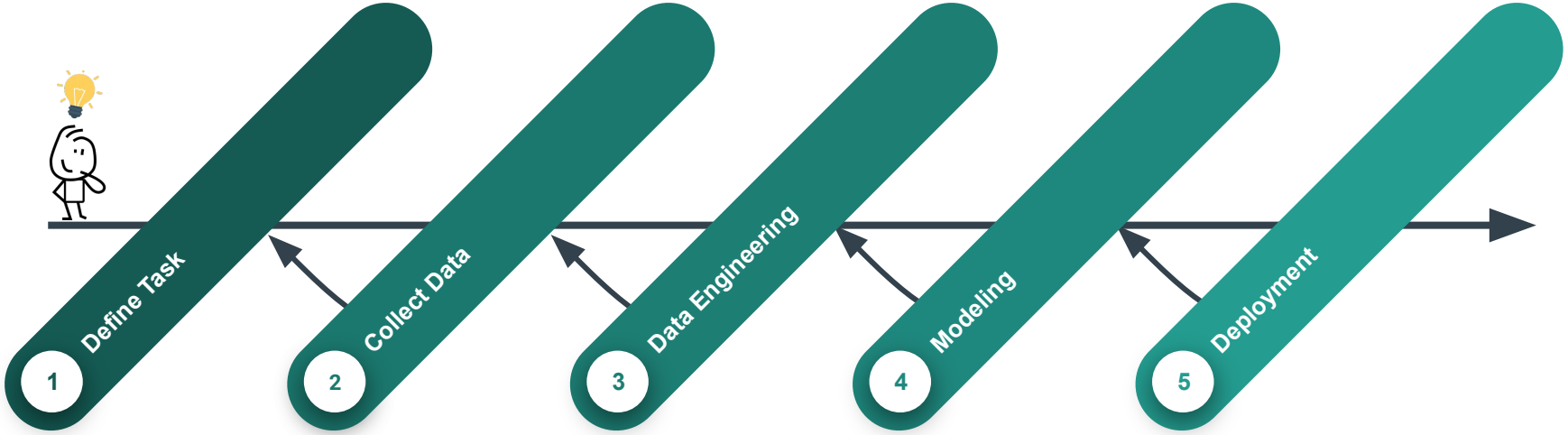
AutoML with Auto-Sklearn

Matthias Feurer /__mfeurer__
Katharina Eggensperger **/KEggensperger**
Eddie Bergman **/Edberg_Wardman**
special guest today: **Aron Bahram (Hiwi working on the code)**

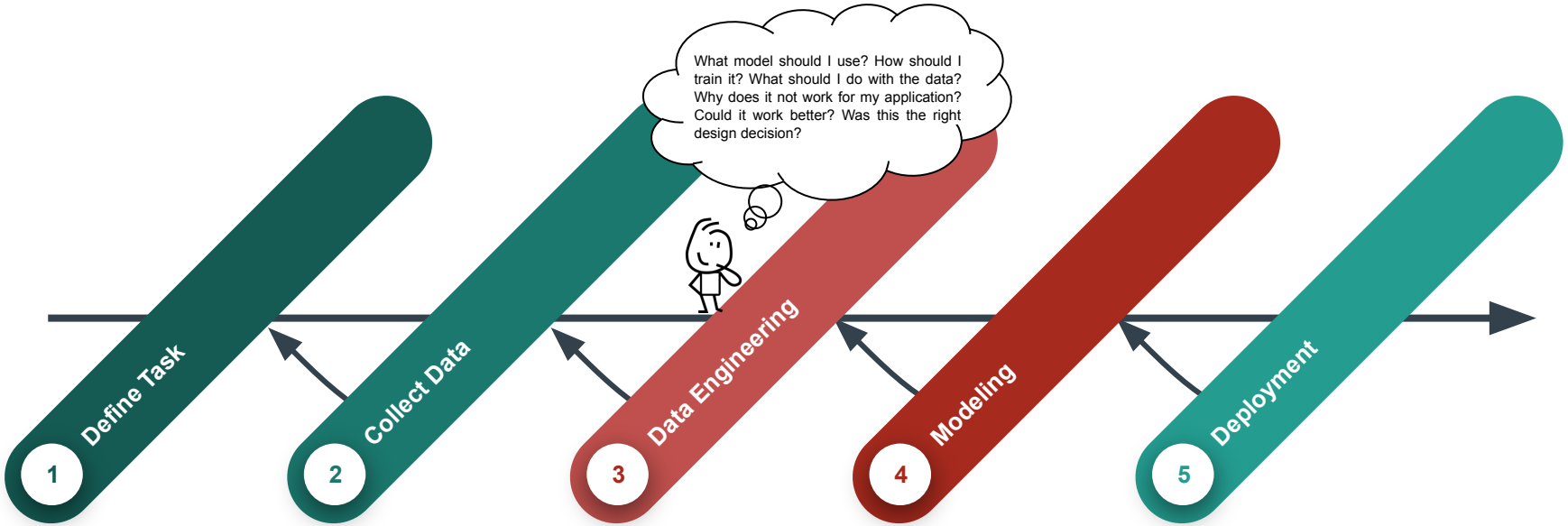


Find slides, notebooks and more here:
https://github.com/automl/auto-sklearn-talks

What is AutoML (and how can it help you)?



What is AutoML (and how can it help you)?



→ automates parts of the ML workflow
→ makes ML easy to use

"Machine Learning for everyone"

What is AutoML (and how can it help you)?

Popular Approaches?

- Hyperparameter Tuning
- Neural Architecture Search
- **AutoML Systems**

Focus of today

Target User?

- non ML experts
- ML researchers
- ML practitioners

Use cases?

- fast prototyping
- strong baselines & fair comparisons
- avoid tedious manual tuning
- get the best out of a model
- study ML algorithms for new problems

“Machine Learning for everyone”

Our Vision

Machine Learning for everyone

in 4 lines of code

```
import autosklearn.classification
>>> cls = autosklearn.classification.AutoSklearnClassifier()
>>> cls.fit(X_train, y_train)
>>> predictions = cls.predict(X_test)
```

How it started?

ChaLearn Automatic Machine Learning Challenge (AutoML)

Organized by automl.chalearn - Current server time: June 5, 2023, 7:14 a.m. UTC
Reward $30,000

First phase	End
Tweakathon0	Competition Ends
Dec. 8, 2014, midnight UTC	June 25, 2016, midnight UTC

Joaquin Vanschoren @joavanschoren · Jun 25, 2016
Congratulations to the **AutoML Challenge** winners!@__mfeurer__ @abhi1thakur @narnars0 and Lisheng Sun

5 8

PAKDD 2018 AUTOML Challenge 2018

AutoML 2018 challenge :: PAKDD2018

Organized by hugo.jair - Current server time: June 5, 2023, 7:21 a.m. UTC
Reward $3,000

First phase	End
Feedback	Competition Ends
Nov. 30, 2017, midnight UTC	March 31, 2020, midnight UTC

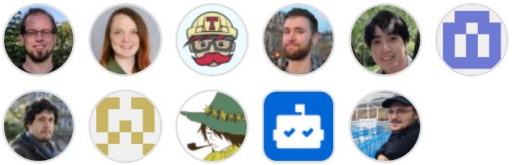
Where we are now

automl / **auto-sklearn** Public | Edit Pins | Unwatch 213 | Fork 1.2k | Starred 7k

Code | Issues 150 | Pull requests 5 | Discussions | Actions | Projects 1 | Wiki | ...

/automl/auto-sklearn

Contributors 80



Used by 425

Goals & Outline

Goals

Understand how Auto-sklearn works & Apply Auto-Sklearn

Outline

1. Interactive Discussion (15 mins)
2. Introduction to Auto-sklearn (45 mins)
3. Hands-On (1.5h)
   - **Task 1**: BYOP
   - **Task 2**: ASKL
   - **Task 3**: EXPLAIN

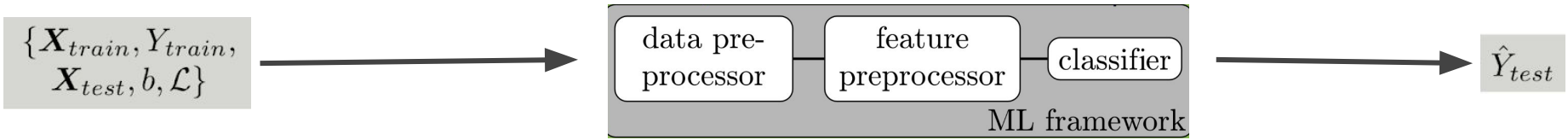
   Bonus Tasks - CUSTOM METRICS, MOO, ASKL 2.0, CUSTOM MODEL
4. Outroduction (20 mins)

Now: Your input

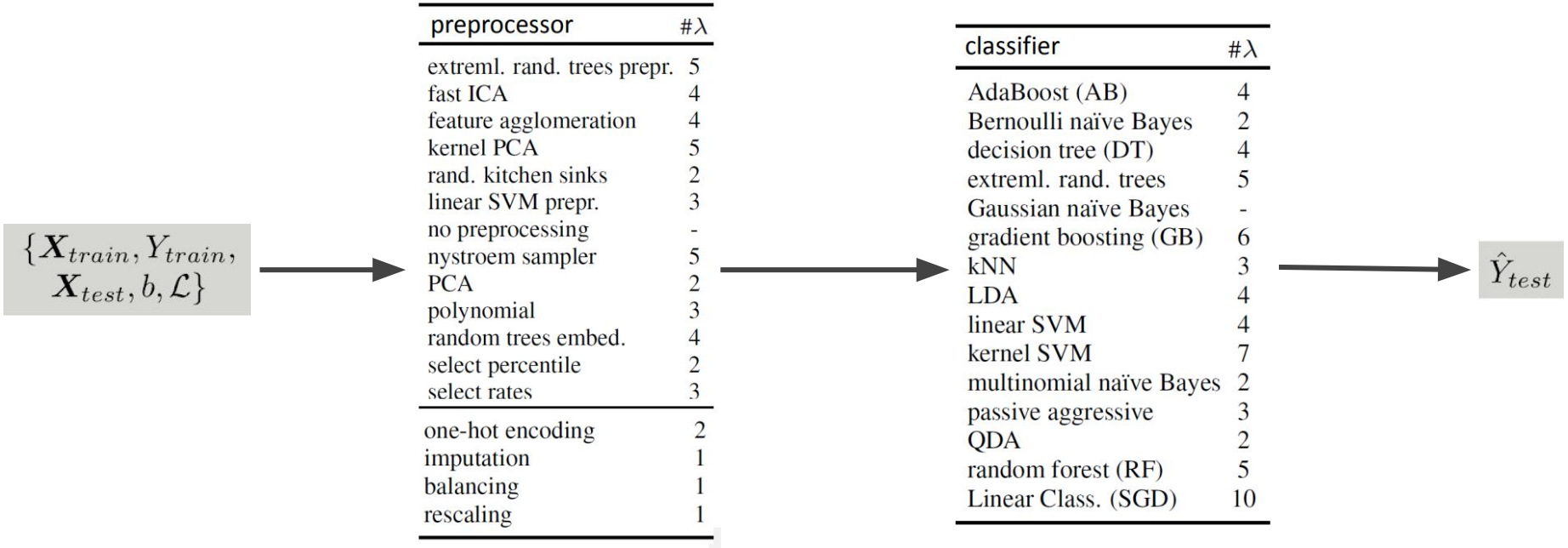
Back to

What is Auto-Sklearn?

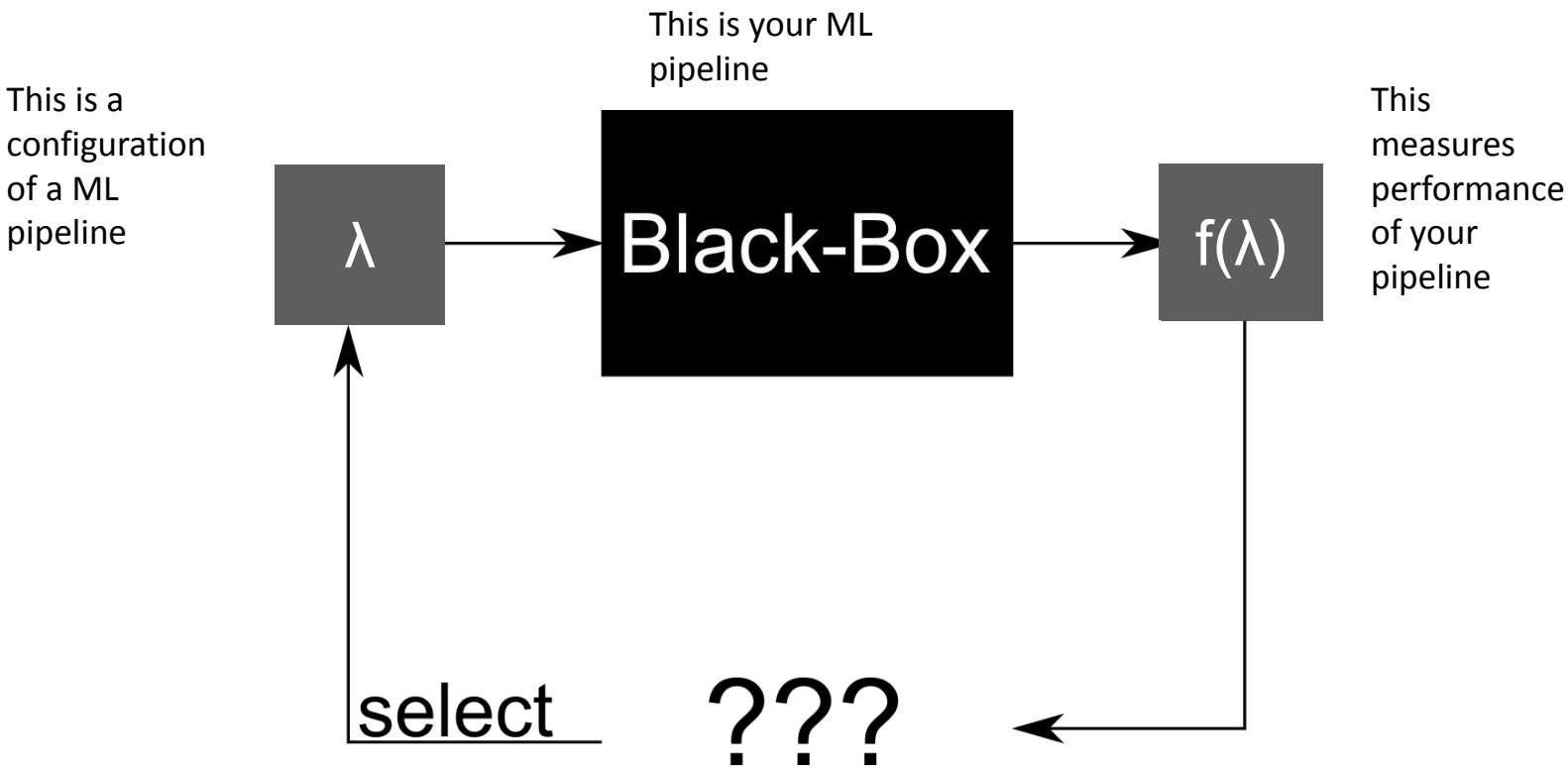
Design Space: Traditional ML with scikit-learn



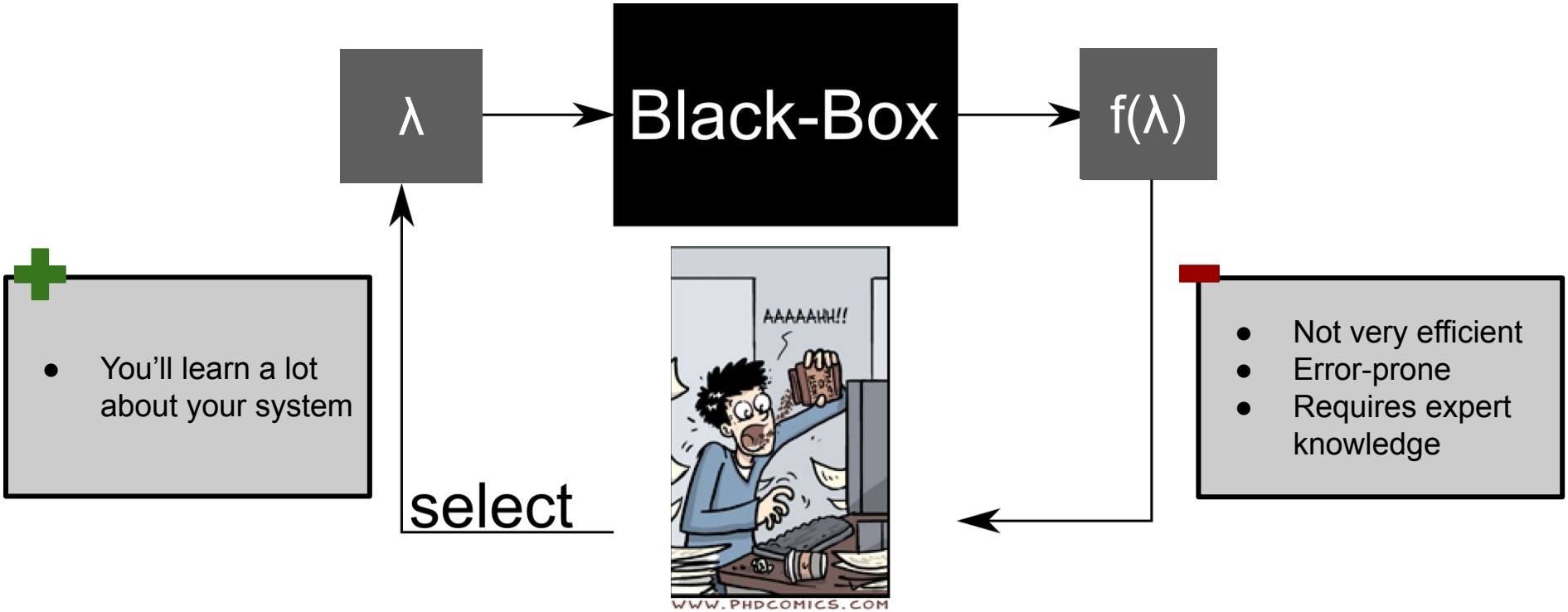
Design Space: Traditional ML with scikit-learn



Black Box Optimization



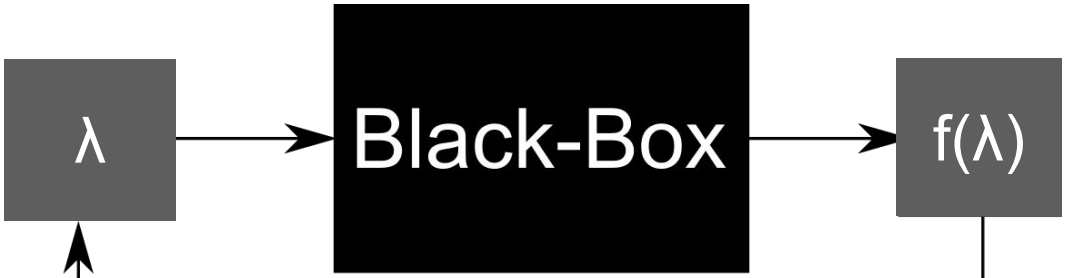
Black Box Optimization: The Human Optimizer



- You'll learn a lot about your system

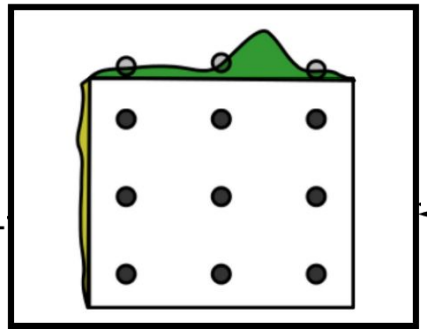
- Not very efficient
- Error-prone
- Requires expert knowledge

Black Box Optimization: Grid Search



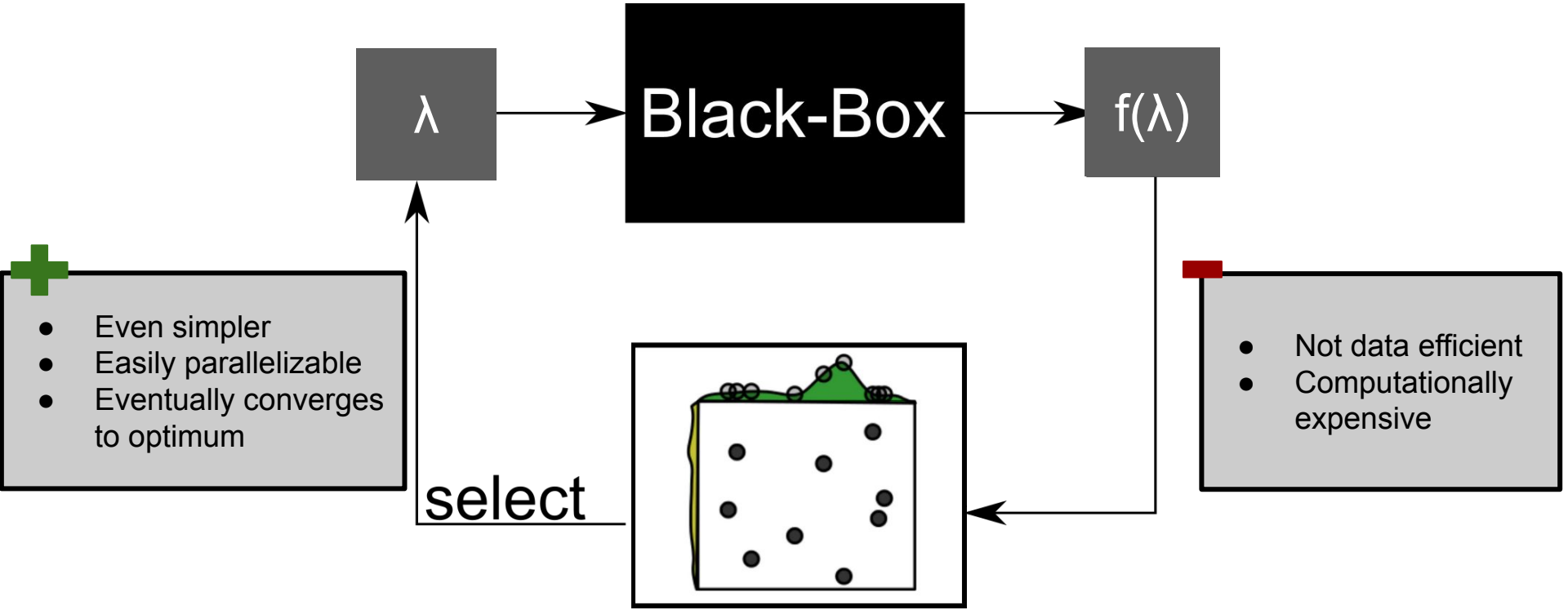
- Very simple approach
- Easily parallelizable
- Can be used to study the problem

- Does not scale to high dimensions
- Grid needs to be defined

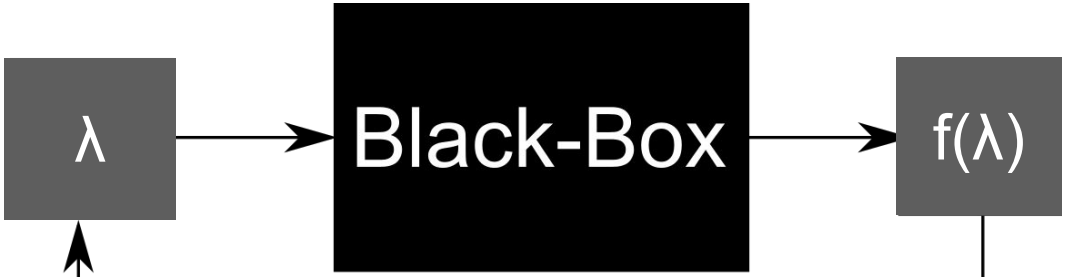


select

Black Box Optimization: Random Search



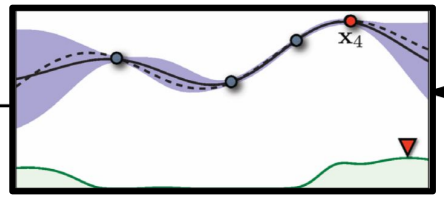
Black Box Optimization: Bayesian Optimization



- Trade-off Exploration & Exploitation
- Data efficient
- State-of-the-art

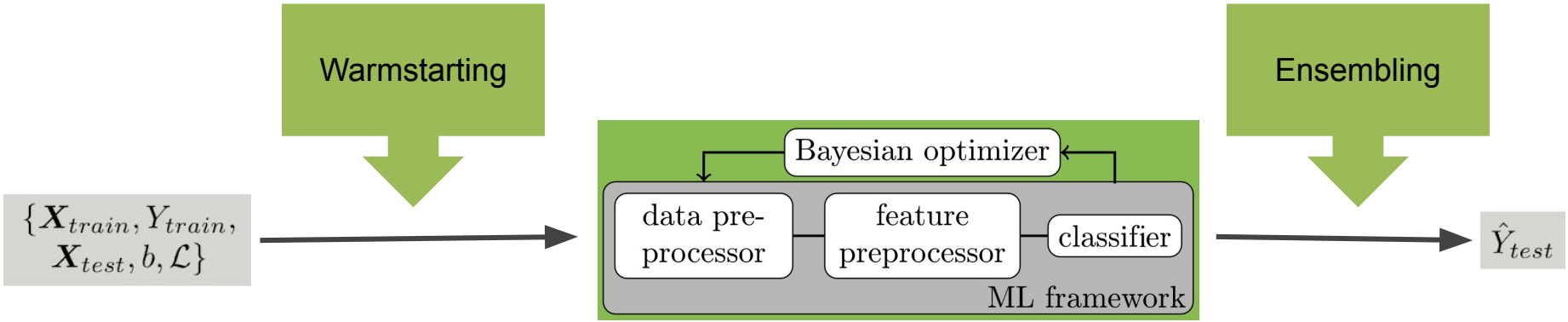
- (Does not scale well with parallel resources)

select



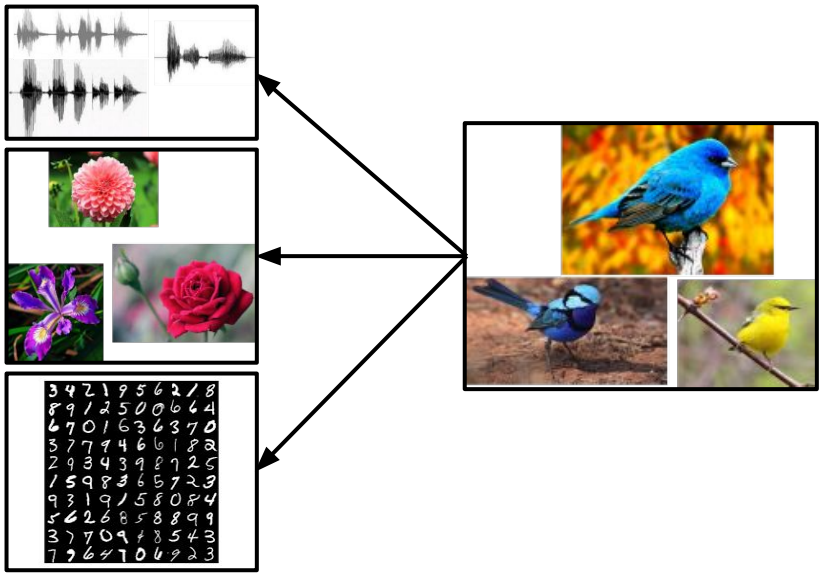
A tiny intro to Bayesian Optimization:
https://distill.pub/2020/bayesian-optimization/

Design Space: Traditional ML with scikit-learn



How to reuse previous experience?

→ Warmstart Bayesian Optimization



Offline / Before:

1) Collect >200 datasets
2) Find the best pipeline on each dataset

Online / For a new dataset:

1) Compute 38 meta-features, select 25 most similar previous datasets
2) Initialize optimization with best pipelines on those datasets

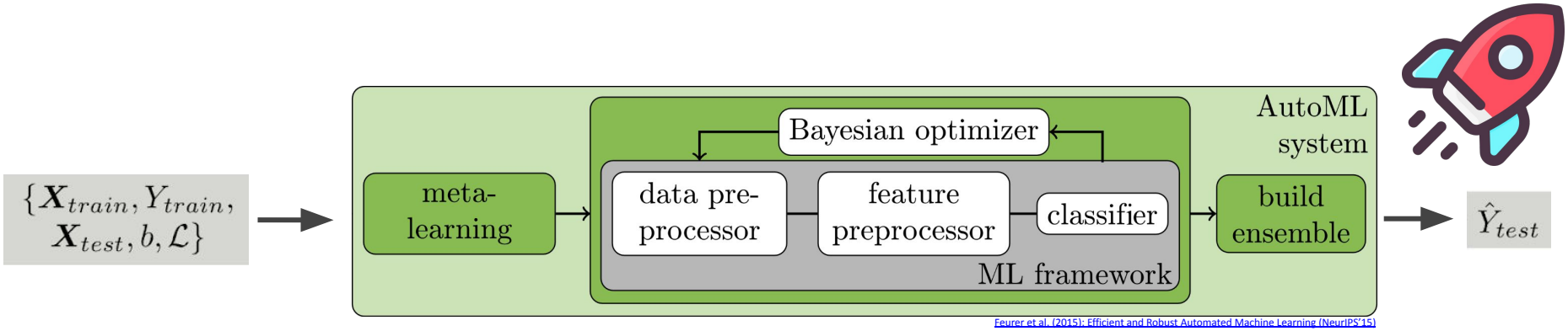
More II: Ensembling

How to get the best out of all evaluated models?

→ **Build an ensemble**

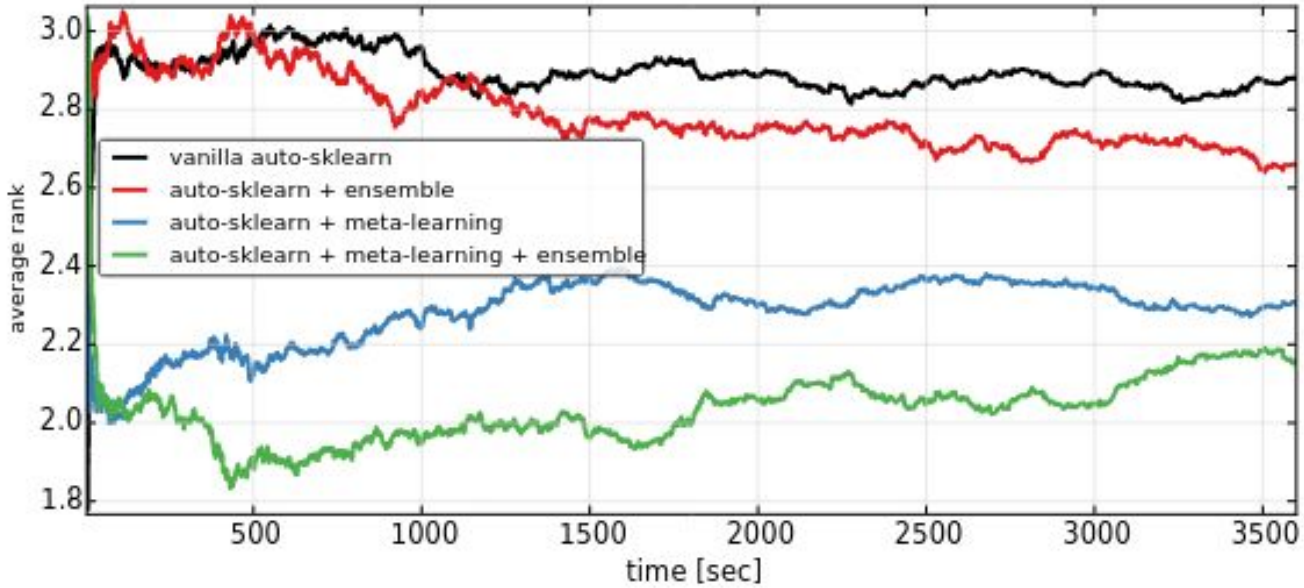
Image credit: Photo by Denisse Leon

Auto-Sklearn 1.0

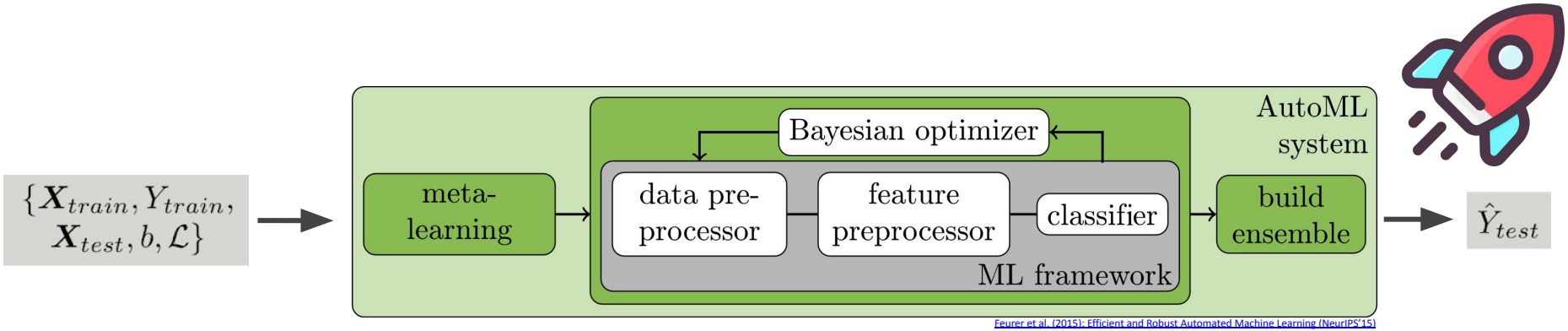


Feurer et al. (2015): Efficient and Robust Automated Machine Learning (NeurIPS'15)

Impact of Meta-Learning and Ensembling



Auto-Sklearn 1.0



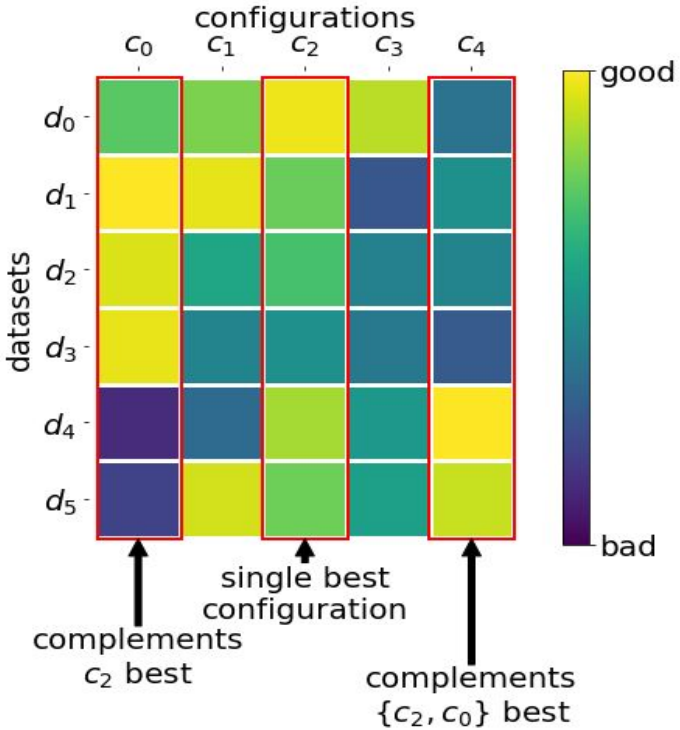
However, some things to be improved

- meta-features can be expensive to compute
- large datasets can be an issue

Even More I: Portfolios

Goal: Meta-Learning without meta-features

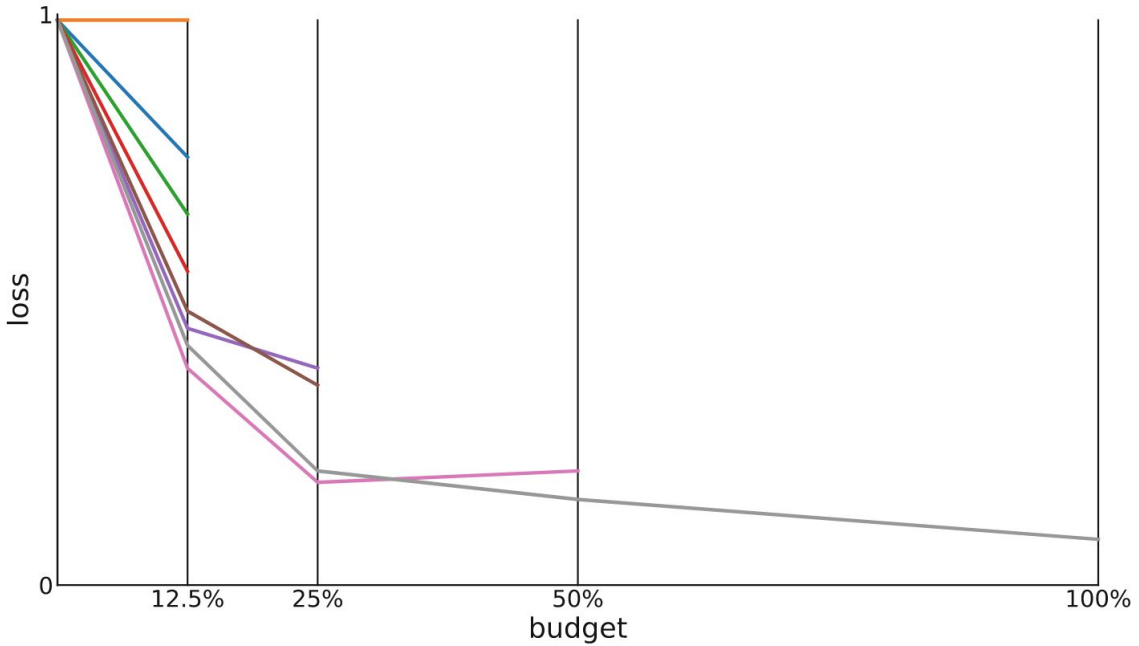
Idea: Construct a Portfolio (a list of diverse pipelines)



Even More II: Successive Halving

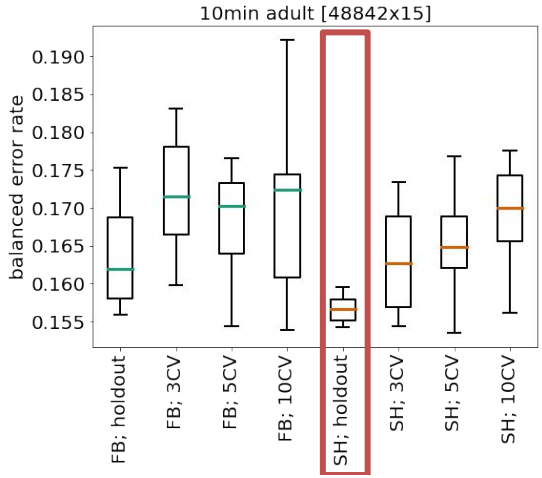
Goal: Scale to large datasets.

Idea: Allocate more resources to promising pipelines



But what about small datasets?

Impact of the Optimization Strategy



But wait … did we make it worse?

Can we automatically select an optimization policy?

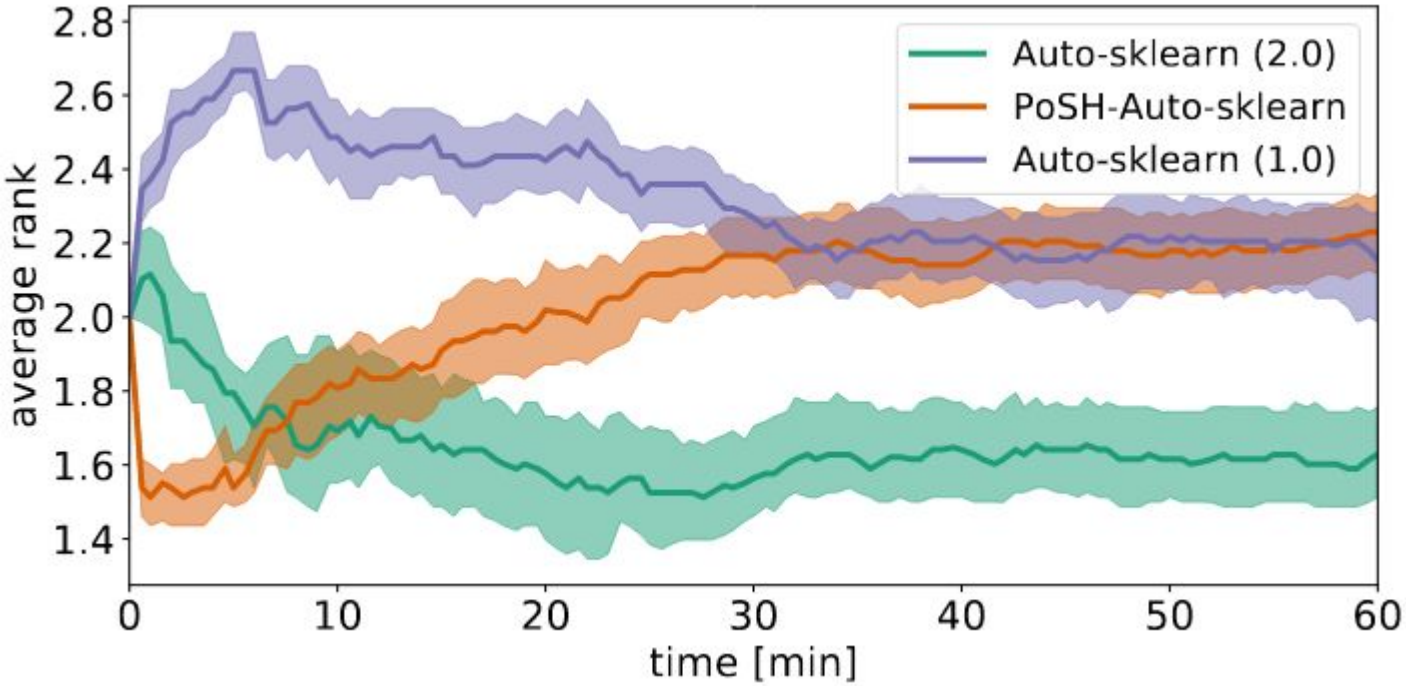
→ Auto-sklearn 2.0

Image Credit - CC BY-NC-ND 2.0; by Beagle Mama

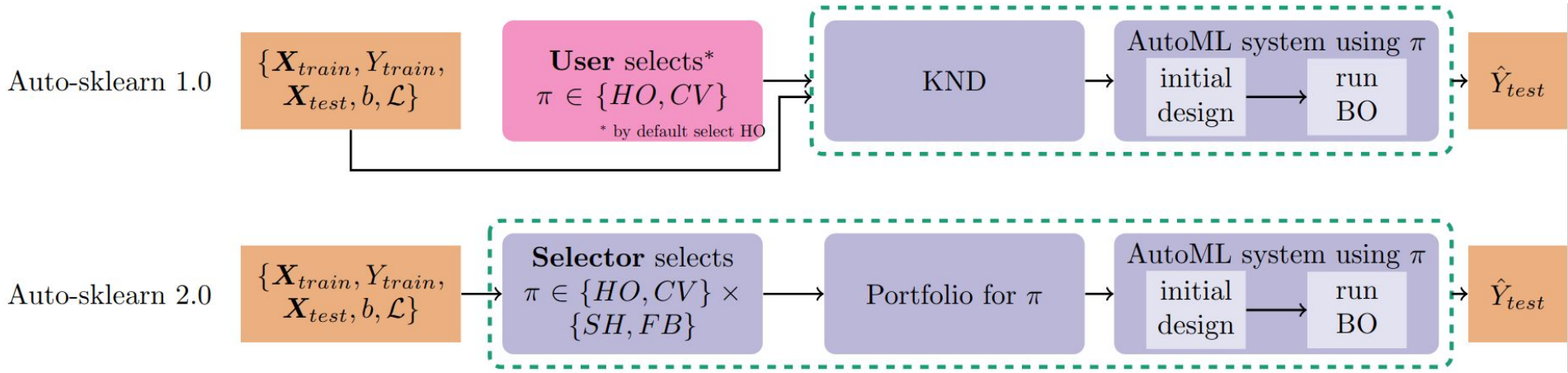
Yes, with a learned selector!

For more details see "Feurer et al. (2022): Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning"

Improvement over Auto-sklearn 1.0



Autosklearn 1.0 vs Auto-sklearn 2.0



Practical features


Simple → follows scikit-learn API


Extensible → to your own algorithms


Efficient → parallelization using Dask


Robust → strict resource limitations

… and many more:

- configurable
- compatible with Pandas/numpy arrays
- support for text features
- …

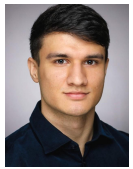
images: Flaticon.com

Summary

```
import autosklearn.classification
>>> cls = autosklearn.classification.AutoSklearnClassifier()
>>> cls.fit(X_train, y_train)
>>> predictions = cls.predict(X_test)
```

- based on **scikit-learn; simple & familiar API**
- integrates **latest research** (>**1K citations**)
- **>20K** downloads per month

- **BSD-3-Clause** License
- works best under **Linux**
- requires **Python>=3.7**



Aron Bahram
Hiwi @ ML Lab Freiburg

/automl/auto-sklearn

Matthias Feurer
Thomas Bayes Fellow and Substitute Professor @ MCML / LMU Munich

Katharina Eggensperger
Group Leader @ Uni Tübingen

Edward Bergman
Research Engineer @ ML Lab Freiburg

Prof. Dr. Marius Lindauer
Head of the ML Lab Hannover

Prof. Dr. Frank Hutter
Head of the ML Lab Freiburg

Now: **Hands-On Session**

- Form groups of 3-5 people
- Copy and open the Colab notebook (see github.com/automl/auto-sklearn-talks)
- Work through tasks

Feedback Time!

What else?

Hyperparameter Tuning

Goal: Find the best performing configuration:

$$\lambda^* \in \arg\min_{\lambda \in \Lambda} f\left(\mathcal{A}_\lambda\right)$$

```python
from ConfigSpace import ConfigurationSpace

cs = ConfigurationSpace(
    space={
        "lr": (0.00001, 1.),        # UniformFloat
        "num_neurons": (16, 512),   # UniformInt
        "opt": ["Adam", "SGD"],     # Categorical
    },
)
```

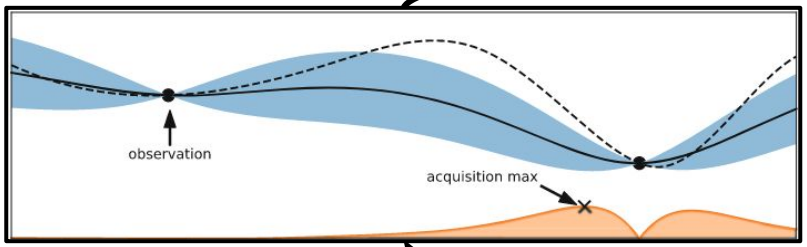
Hyperparameter Tuning

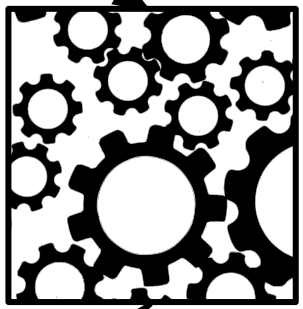
Goal: Find the best performing configuration:

$$\lambda^* \in \arg\min_{\lambda \in \Lambda} f\left(\mathcal{A}_\lambda\right)$$

λ_n

Bayesian optimization
(and its advanced variants)





target algorithm A

$f(A_{\lambda n})$

/automl/SMAC3

Neural Architecture Search (NAS)

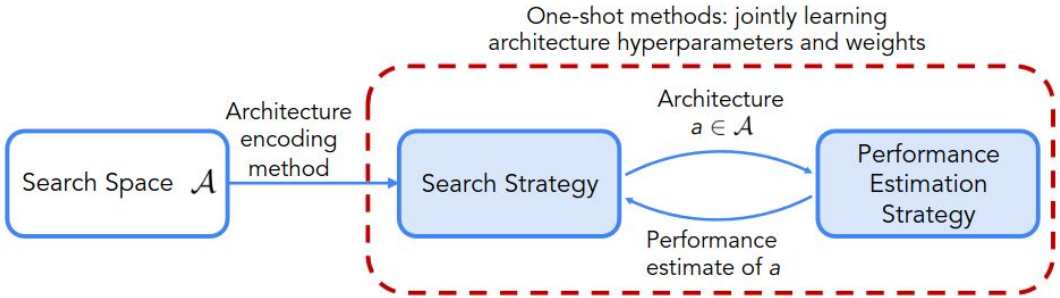


Image from “Colin White et al.: Neural Architecture Search: Insights from 1000 Papers. arXiv:2301.08727v2 [cs.LG]”

Want to build your own NAS optimizer?

/automl/NASlib

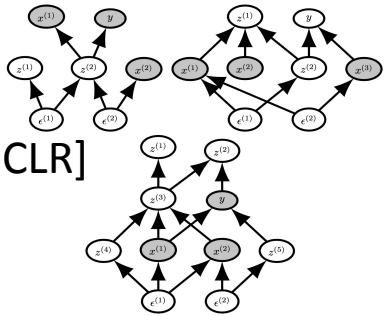
Want a network?

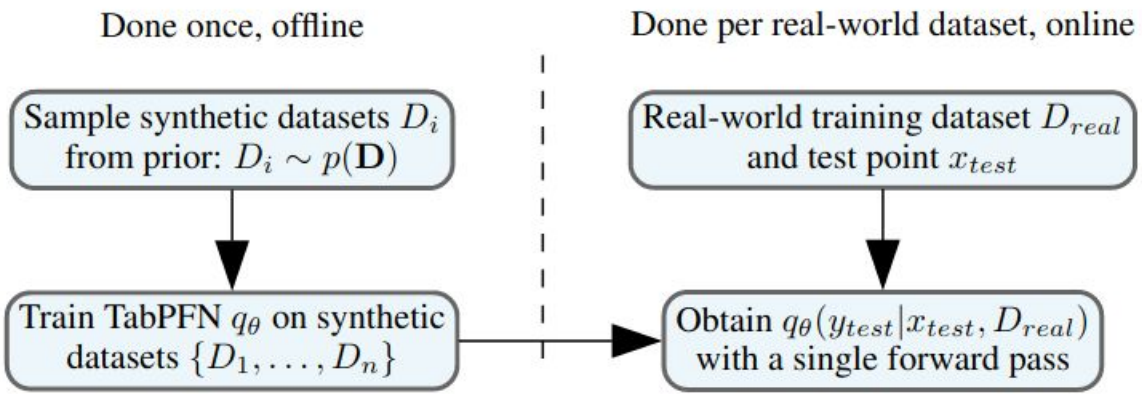
/automl/AutoPytorch

TabPFNs for small data

PFN = Prior-data Fitted Networks [Müller et al. 2022; ICLR]

TabPFN = PFNs trained with a prior for **tabular** data [Hollman et al. 2023; ICLR]





Limitation: 1K samples; 100 features; 10 classes

/automl/tabpfn

Other Tools?

	search	searchspace
AutoGluon	None	None
AutoKeras	Hill-climbing	Keras
Auto-PyTorch	BO (SMAC)	PyTorch
Auto-sklearn	BO (SMAC)	scikit-learn
AutoWEKA	BO (SMAC)	WEKA
FLAML	Custom	scikit-learn, XGBoost, CatBoost, LightGBM
H2O.ai	RS	H2O + XGBoost
HPO tools	BO, Bandit, DE, etc…	everything :)

AutoML DIY - A modular AutoML Toolkit

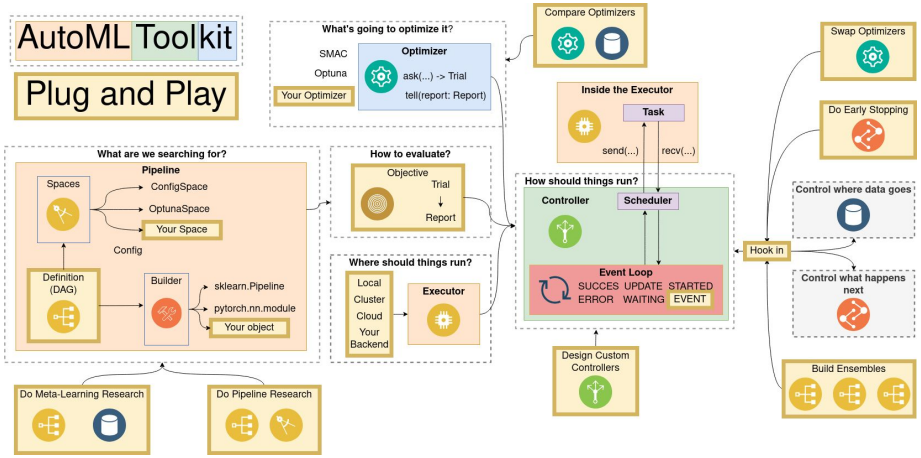
→ **Many AutoML systems exist. They're all awesome!**

But: Research on AutoML systems is hard.

- No common code framework
- Software-engineering heavy

What we want? A Modular AutoML Framework.
(or "Build-your-own-system-in-a-few-lines-of-code")

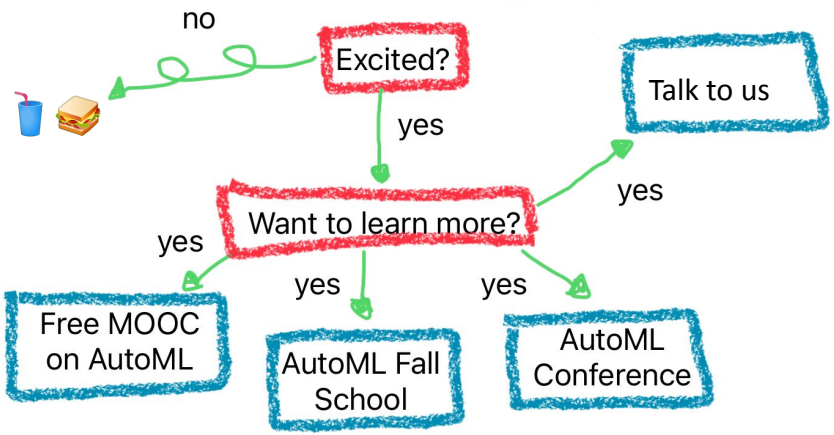
Who is this for? AutoML researcher & ML practitioners



Interested? Talk to us.

/automl/<talktous>

Thank you!



see: AutoML.org