INFORMATIONSFABRIK

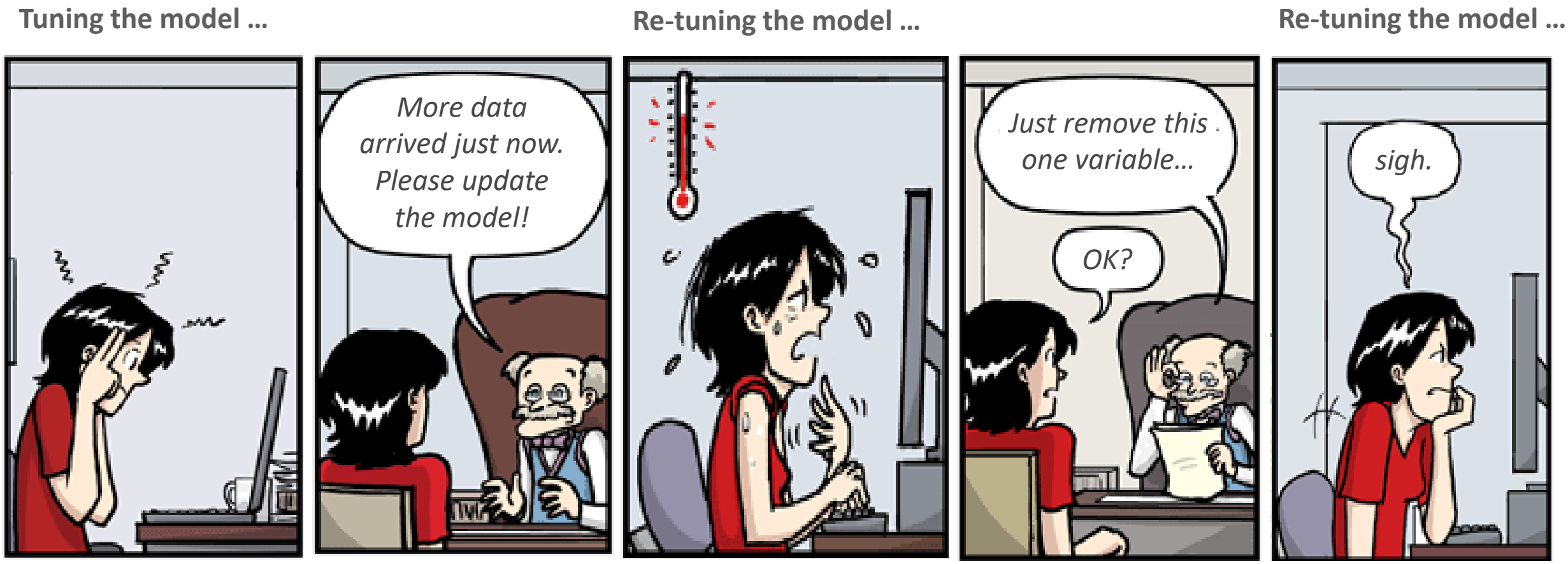
DATEN VERSTEHEN, ENTSCHEIDUNGEN TREFFEN

Hyperparameter Tuning
- brute force or smart algorithms?

Tobias Rippel

Data Science Meetup, 07.06.2018

A *week* in the life of a data scientist



Manually tuning Machine Learning models is tedious

Manually truning

- Time consuming, even boring
- Got more data? - *Good for you, now start over!*
- *Fear of missing out*

Automated tuning!

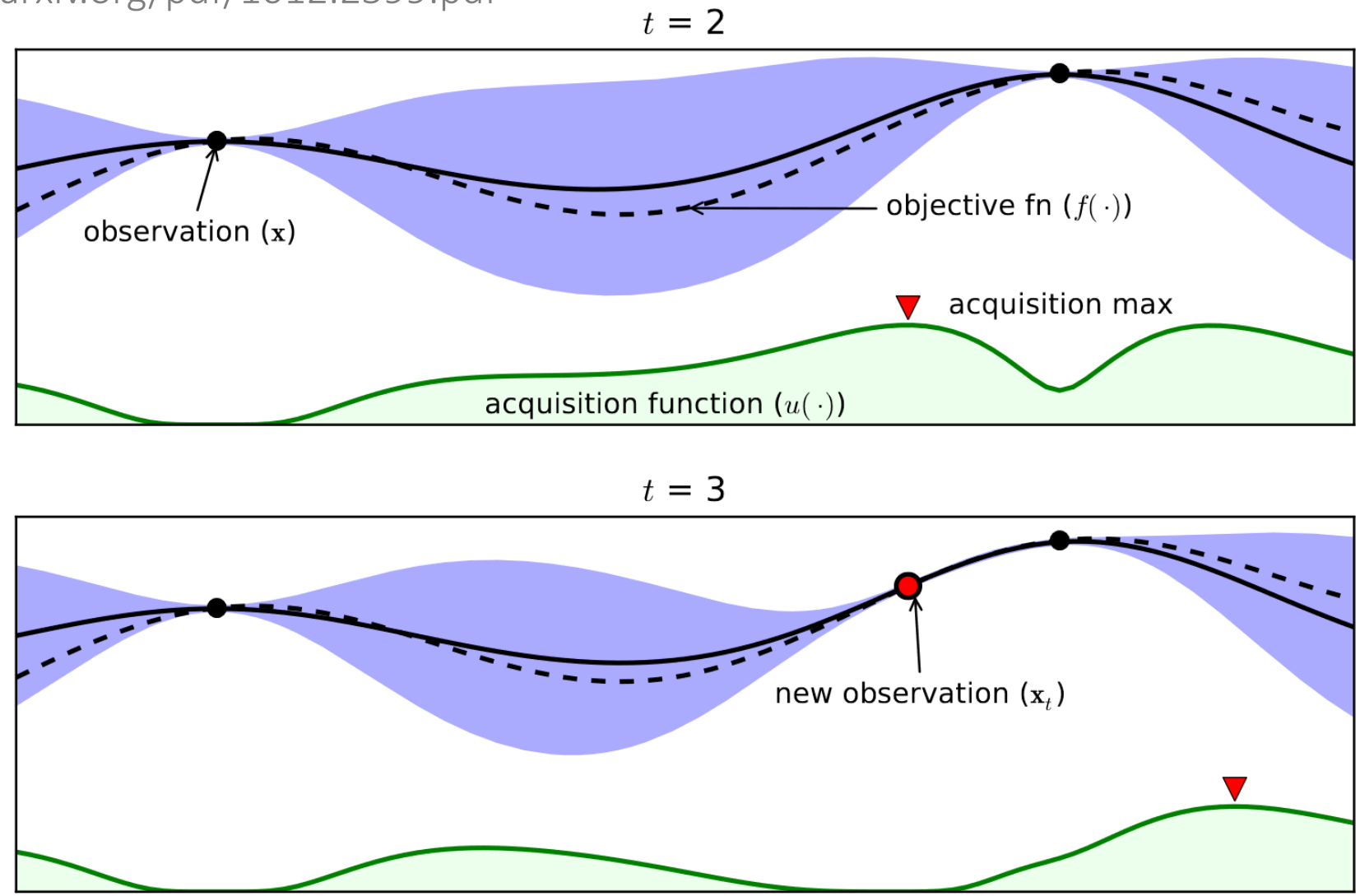
- Test all parameter settings? → . . .
- Grid Search → Takes too long
- Random Search → Better, but is that all?
- Smart Algorithms → Sound‘s interesting!

Smart Algorithm: Basic ideas

- Gaussian Process
- Tree-structured Parzen Estimator (TPE)
- …

Gaussian Process

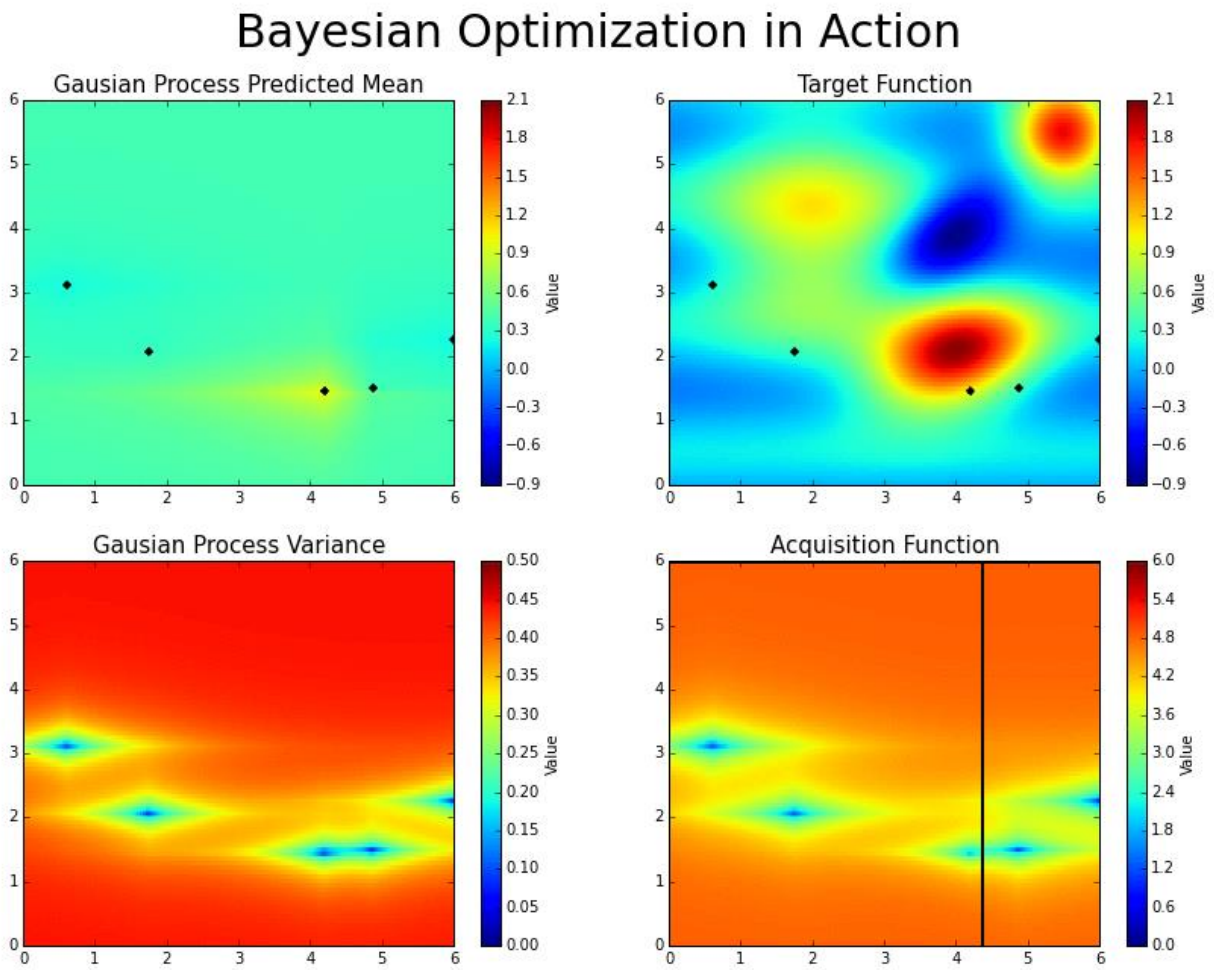
Source: https://arxiv.org/pdf/1012.2599.pdf



Gaussian Process

Source: https://github.com/fmfn/BayesianOptimi

Bayes Estimation of Truth



Truth

Variance Estimation

Data Point Selector

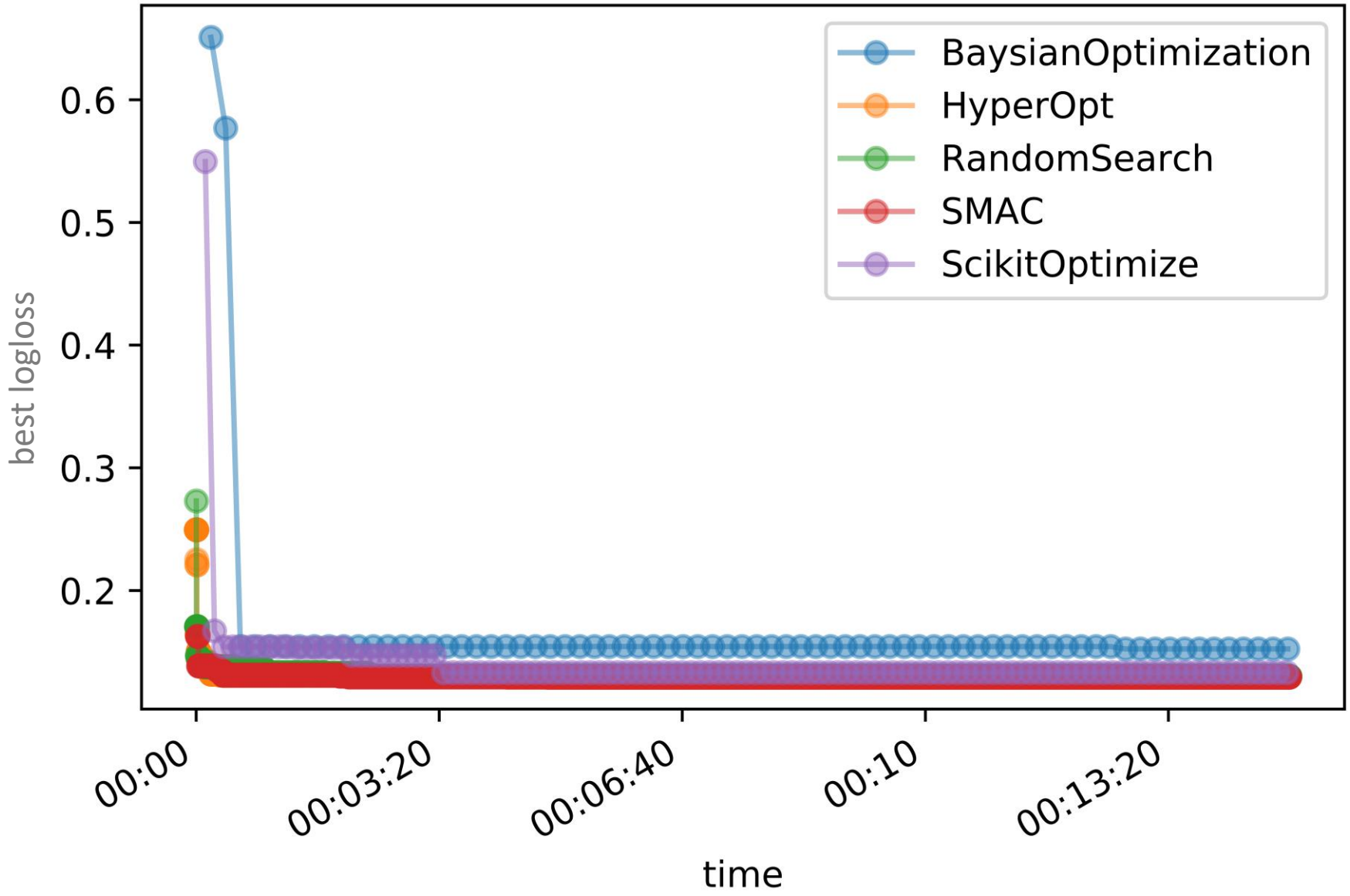
Benchmark

- Our benchmark approach:
  - Limited time (1h)
  - Same resources (32 core machine)
  - Same ml algorithm (xgboost) optimized on same loss function (log-loss)
  - Same datasets (Iris, Real Life Dataset (130k rows, 20 cols)) -> Classification

Category	Hyperparameter Tuning Algorithms
Benchmark	Random Search
Gaussian Process	BayesianOptimization (https://github.com/fmfn/BayesianOptimi)
	Scikit-Optimize (https://github.com/scikit-optimize/scikit-optimize)
Tree-structured Parzen Estimator (TPE)	Hyperopt (https://github.com/hyperopt/hyperopt)
	SMAC (https://github.com/automl/SMAC3)

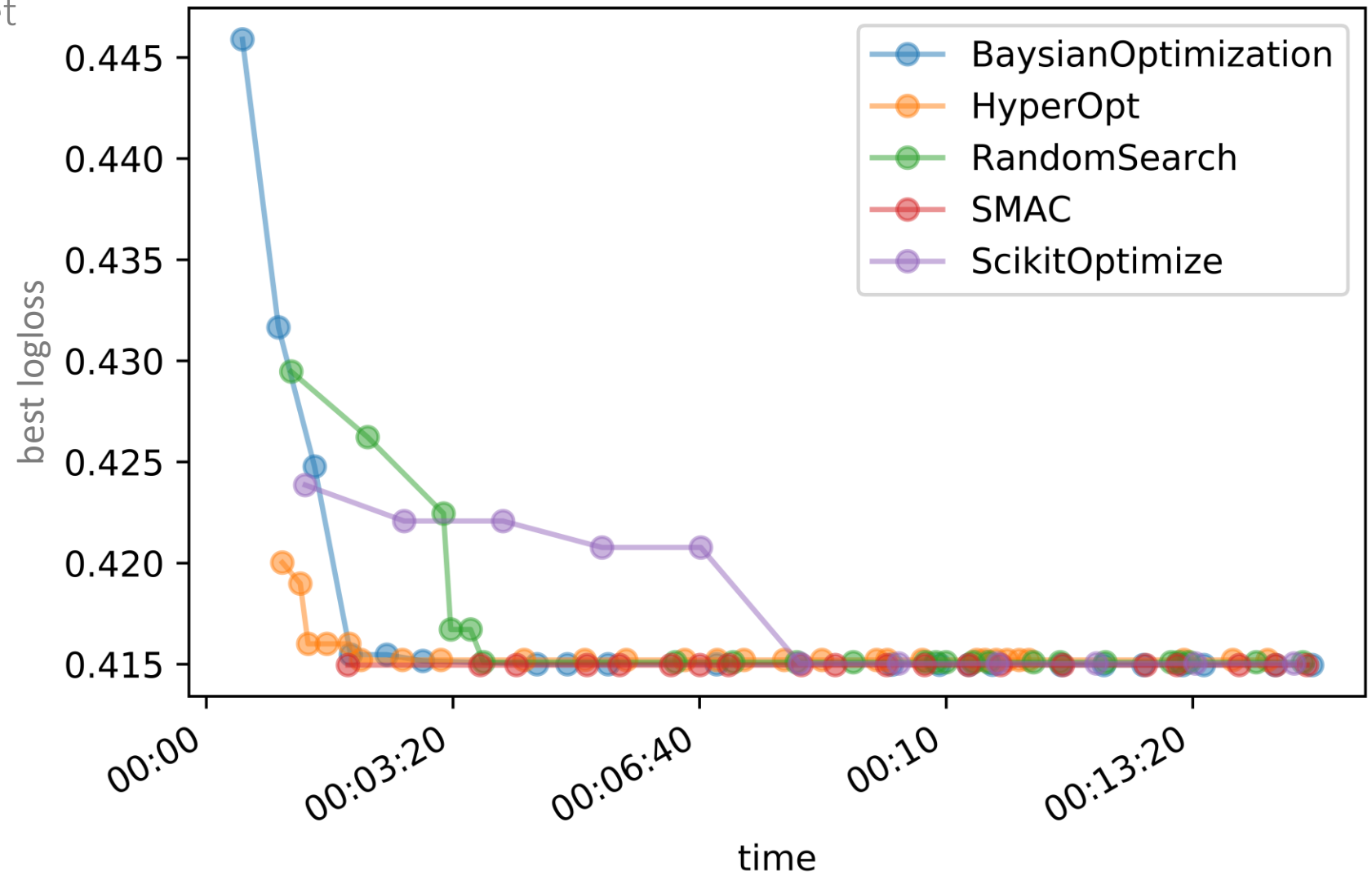
Results

Iris Dataset



Results

Real-Life Dataset



Summary + Take aways

Initial results

- Chosen Smart Algorithms have limited/no benefits over Random search (on our datasets)
- Smart Algorithms have to be parameterized themselves… (BOHHHHH!)

Further experiments

- Test more comlpex probelms (datasets)
- Compare a human expert to these results

Take aways

- Only use Hyperparameter Tuning after you have all the data!
- Spend your time wisely! -> Feature Engineering/ more data are oftentime more desirable