Towards an Empirical Foundation for Assessing Bayesian Optimization of Hyperparameters

Katharina Eggensperger, Matthias Feurer, Frank Hutter
Freiburg University
{eggenspk,feurerm,fh}@informatik.uni-freiburg.de

James Bergstra
University of Waterloo
james.bergstra@uwaterloo.ca

Jasper Snoek
Harvard University
jsnoek@seas.harvard.edu

Holger H. Hoos and Kevin Leyton-Brown
University of British Columbia
{hoos,kevinlb}@cs.ubc.ca

… in 15 sec

Bayesian Optimization techniques are becoming readily applicable for hyperparameter tuning.

Many approaches exist, but it remains unclear which one performs best in which case.

For further progress we need an easy-to-use **benchmark library** and a **baseline comparison**. We provide both.

Contribute

Our **publicly available** software wraps various benchmarks and the Bayesian optimizers SMAC, Spearmint and TPE. This is just a first step towards a canonical benchmark library. You can help us:

- **Contribute your algorithm**
  to add another benchmark for future research
- **Include your optimizer**
  to find out its strengths and weaknesses

Ready To Use

SMAC (Sequential Model-based algorithm configuration) is based on random forests and can handle continuous, discrete and conditional hyperparameters.

[Hutter, Hoos, and Leyton-Brown, 2011]

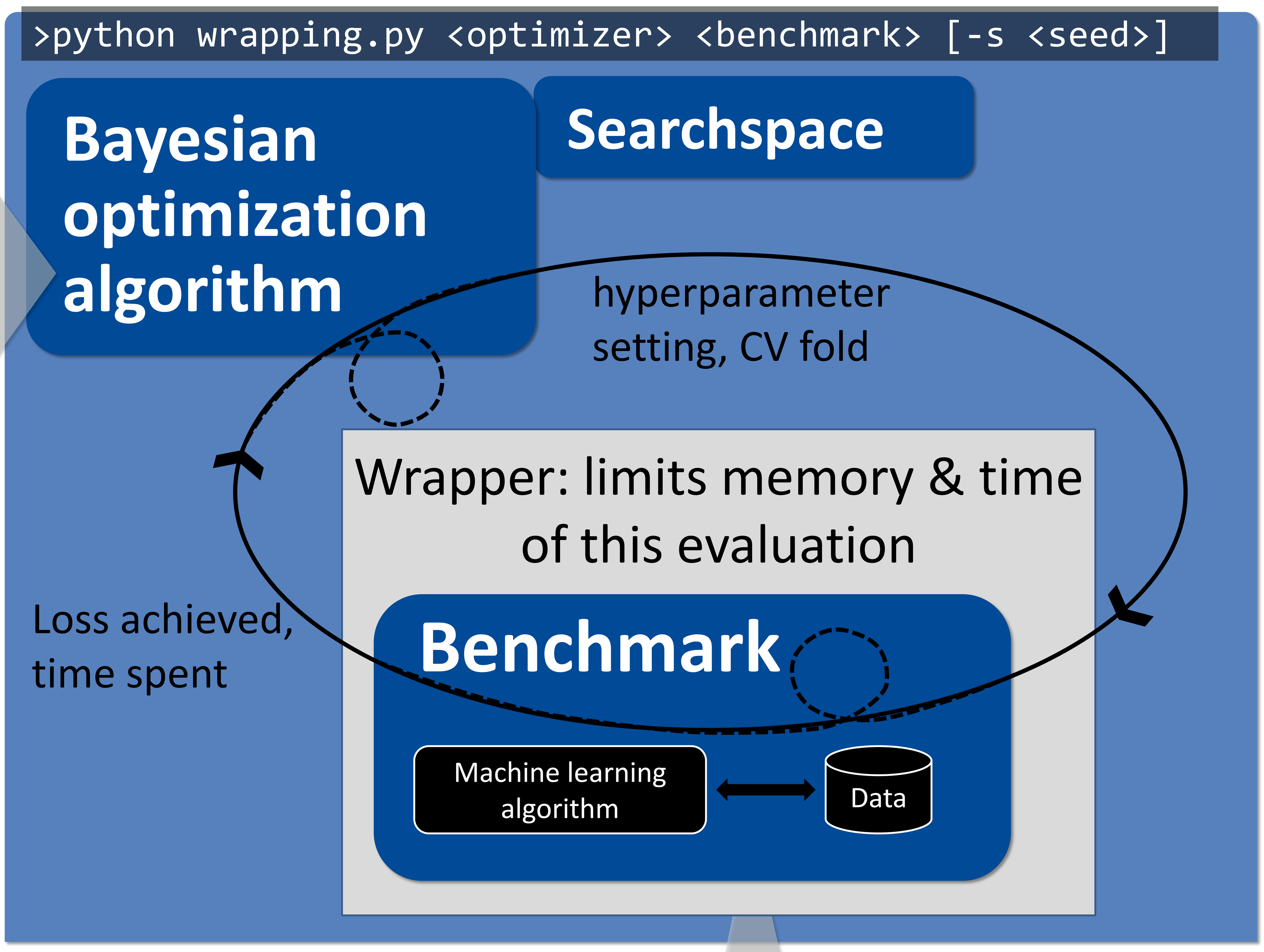
Spearmint uses Gaussian Process (GP) models and performs slice sampling over the GP's hyperparameters. Can handle continuous and discrete hyperparameters.

[Snoek, Larochelle, and Adams, 2012]

TPE (Tree Parzen Estimator) is based on Gaussian Mixture Models. Supports conditional, continuous and discrete parameters and also priors over them.

[Bergstra, Bardenet, Bengio, and Kégl, 2011]

Add your **optimizer**, executable from command line interface



Empirical Results

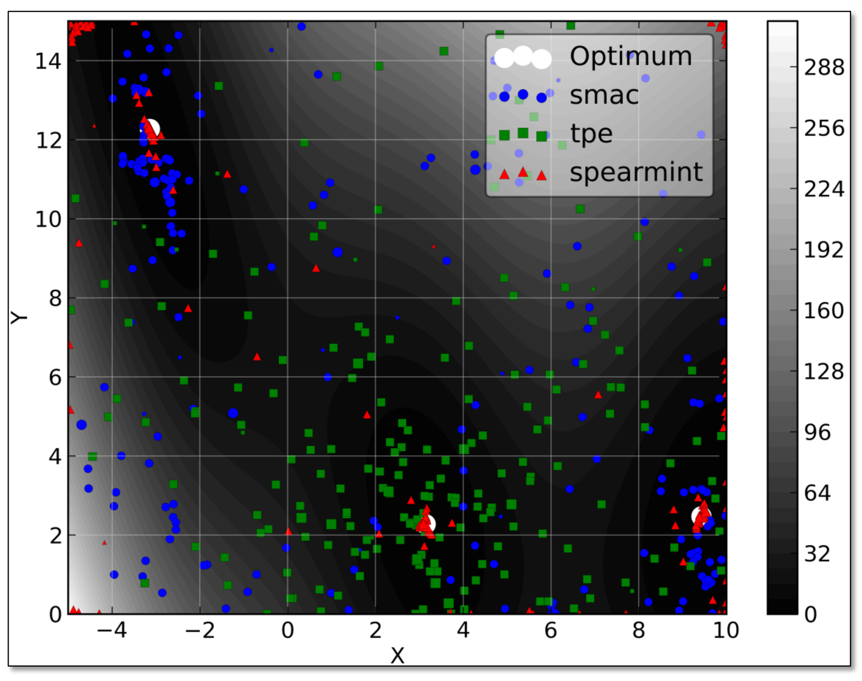
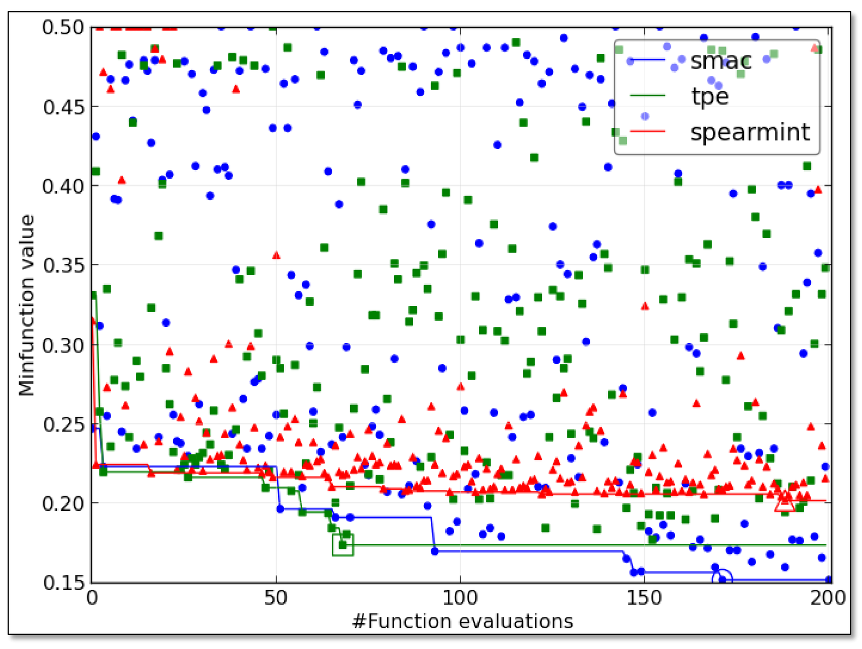


Fig. 1: Evaluated points for the 2-dimensional branin test function. The function's three global optima are visualized as white circles.

Fig. 2: Representative traces of loss function values over time when optimizing the 38 hyperparameters of the HP-NNET on the convex dataset.



Mean loss (typically: validation error rate) and standard deviation across 10 runs of each optimizer. Bold face indicates best value, underlined values are not statistically significant worse than the best value.

Ready To Use

Experiment	#evals	SMAC	Spearmint	TPE	# hyp.params (conditional)	continuous /discrete	Dataset	Citation
Branin (0.398)	200	0.655 ±0.27	**0.398**±0.0	0.526±0.12	2 (-)	2 / -	-	[Hedar]
Har6 (-3.322)	200	-2.98±0.11	**-3.13**±0.41	-2.82±0.18	6 (-)	6 / -	-	
Log. Regr.	100	8.6±0.9	**7.3**±0.2	8.2±0.6	4 (-)	4 / -	MNIST	[Snoek, Larochelle, and Adams, 2012] [LeCun, Bottou, Bengio, and Haffner, 1998] [Hoffman , Blei, and Bach, 2010] [Miller, Kumer, Packer, Goodman, and Koller, 2012]
LDA ongrid	50	**1270**±3	1273±10	1272±4	3 (-)	- / 3	Wikipedia Articles	
SVM ongrid	100	**24.2**±0.1	24.6±0.9	**24.2**±0.0	3 (-)	- / 3	UniProbe	
HP-NNET	100	**19.5**±1.5	20.6±0.3	**19.5**±1.6	14 (4)	7 / 7	convex	[Bergstra, Berdenet, Bengio, and Kégl, 2011] [Larochelle, Erhan, Courville, Bergstra, and Bengio, 2007]
HP-NNET	200	**18.3**±1.9	20.0±0.9	18.5±1.4				
HP-NNET	100	51.5±2.8	52.2±3.3	**50.0**±1.7	14 (4)	7/ 7	MRBI	
HP-NNET	200	**48.3**±1.8	51.4±3.2	48.9±1.4				
HP-DBNET	100	**16.4**±1.2	20.74±6.9	17.29±1.7	36 (27)	19 / 17	convex	
HP-DBNET	200	**15.4**±0.8	17.45±5.6	16.1±0.5				
Auto-WEKA	30h	**27.49**±4.9	40.64±7.2	35.33±2.93	786 (784)	296 / 490	convex	[Thornton, Hutter, Hoos, and Leyton-Brown, 2012] [Larochelle, Erhan, Courville, Bergstra, and Bengio, 2007] [Hall, Frank, Holmes, Pfahringer, Reutemann, and Witten, 2009]
Log. Regr. 5CV	500 folds	**8.1**±0.2	8.2±0.1	8.9±0.5	4 (-)	4 / -	MNIST	[Snoek, Larochelle, and Adams, 2012] [Bergstra, Berdenet, Bengio, and Kégl, 2011] [Larochelle, Erhan, Courville, Bergstra, and Bengio, 2007]
HP-NNET 5CV	500 folds	**18.2**±1.5	23.0±5.0	20.9±1.3	14 (4)	7 / 7	convex	
HP-NNET 5CV	500 folds	**47.9**±0.7	52.8±5.1	50.8±1.4	14 (4)	7 / 7	MRBI	

Add your **algorithm,** executable from command line interface

Implementation Publicly Available: www.automl.org/hpolib