Towards Assessing the Impact of Bayesian Optimization's own Hyperparameters

Marius Lindauer, Matthias Feurer,
Katharina Eggensperger,
André Biedenkapp & Frank Hutter

Motivation

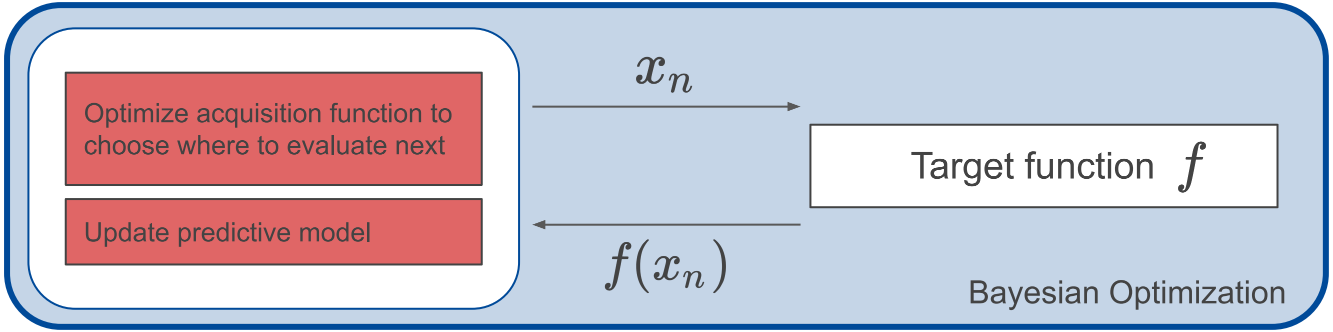
1 Hyperparameter optimization is crucial to achieve peak performance!

Motivation

1 Hyperparameter optimization is crucial to achieve peak performance!

2 Bayesian optimization is a successful approach for that!

Quick Recap on Bayesian Optimization



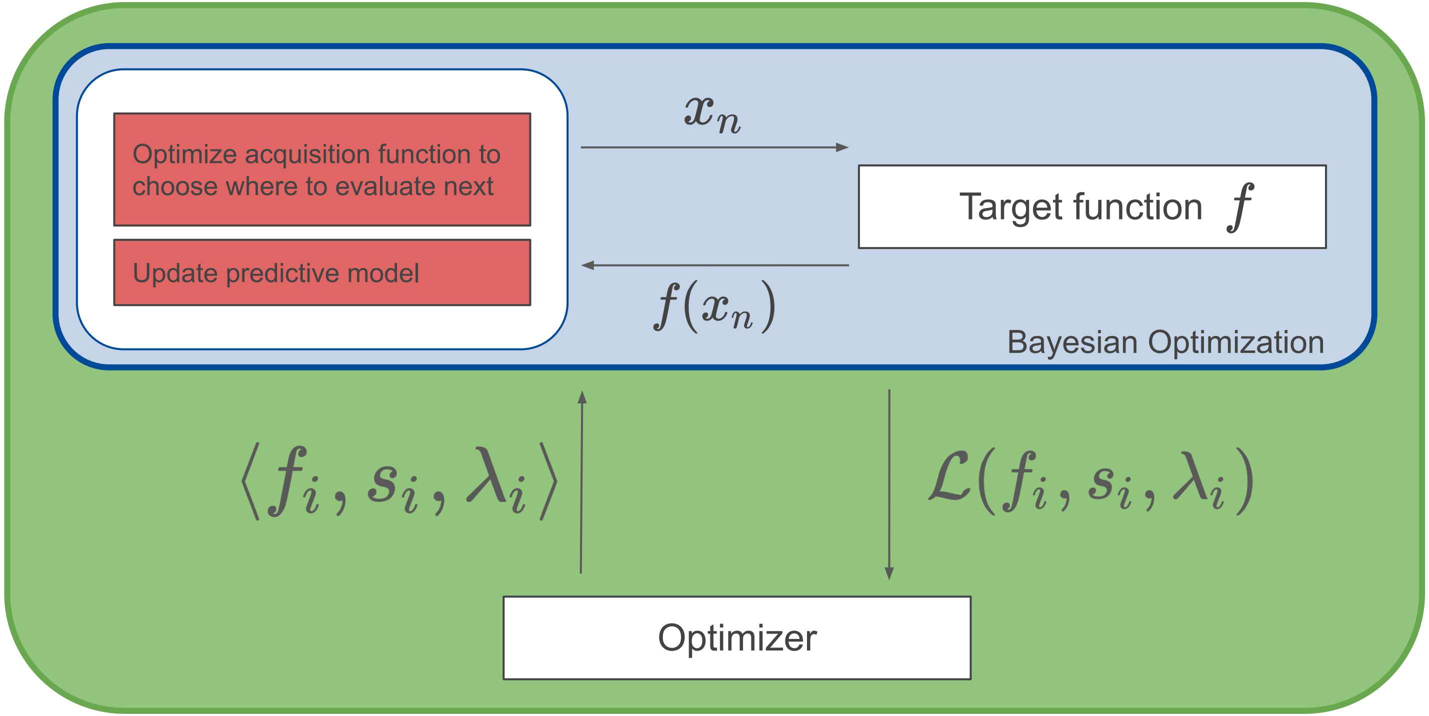
Related Work

Bayesian optimization can be improved with:

- Changing **transformations of the target function**[2]
- Changing its **initial design**[2,4]
- **Tuning the model** on- and offline[1,3]
- Changing the **acquisition function**[4,5]

[1] G. Malkomes and R. Garnett. *Automating Bayesian optimization with Bayesian optimization.* NeurIPS 2018
[2] D. Jones et al. *Efficient global optimization of expensive black box functions*. JGO 1998
[3] J. Snoek et al. *Scalable Bayesian optimization using deep neural networks.* ICML 2015
[4] D. Brockhoff et al. *The impact of initial designs on the performance of matsumoto on the noiseless BBOB-2015 testbed: A preliminary study.* GECCO 2015
[4] V. Picheny et al. *A benchmark of kriging-based infill criteria for noisy optimization.* Structural and Multidisciplinary Optimization 2013
[5] M. Hoffman et al. *Portfolio allocation for Bayesian optimization.* UAI'11

Goal: Meta-Optimization



Similar to N. Dang, L. Pérez Cáceres, P. De Causmaecker, and T. Stützle. *Configuring irace using surrogate configuration benchmarks.* GECCO'17

Research Questions

1 How large is the impact of tuning Bayesian optimization's own hyperparameters?

Research Questions

1 How large is the impact of tuning Bayesian optimization's own hyperparameters?

2 How well does this transfer to similar target functions?

3 How well does this transfer to different target functions?

Research Questions

1 How large is the impact of tuning Bayesian optimization's own hyperparameters?

2 How well does this transfer to similar target functions?

3 How well does this transfer to different target functions?

4 Which hyperparameters are actually important?

Research Questions

1 **How large is the impact of tuning Bayesian optimization's own hyperparameters?**

2 **How well does this transfer to similar target functions?**

3 How well does this transfer to different target functions?

4 **Which hyperparameters are actually important?**

What do we need to tune BO's hyperparameters?

1 Search Space

2 Target functions

3 Meta-loss function to be optimized

4 Optimizer

Ingredients

1 Search Space

RF
+model hyperparameter
+initial design
+acquisition function
+transformation

GP-ML
+model hyperparameter
+initial design
+acquisition function
+transformation

GP-MAP
+model hyperparameter
+initial design
+acquisition function
+transformation

2 Target functions

3 Meta-loss function to be optimized

4 Optimizer

Ingredients

1 Search Space

2 Target functions

→ Meta-optimization is quite expensive
→ Use artificial functions
→ Surrogate benchmark problems

3 Meta-loss function to be optimized

4 Optimizer

SVMs
- 10 datasets
- 3 continuous hyperparameters
- 1 categorical hyperparameter

NNs
- 6 datasets
- 6 continuous hyperparameters

Artificial functions
- 10 functions
- 2-6 continuous hyperparameter

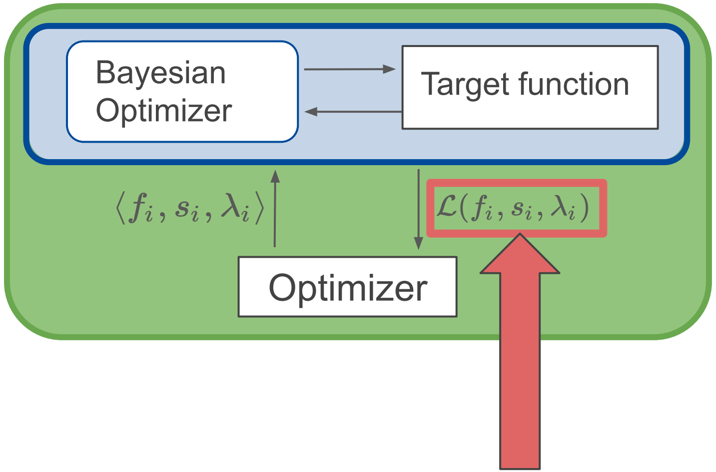
Ingredients

1 Search Space

2 Target functions

3 Meta-loss function to be optimized

- Measure good anytime performance
- Compare across multiple functions
- Hit optimum accurately



4 Optimizer

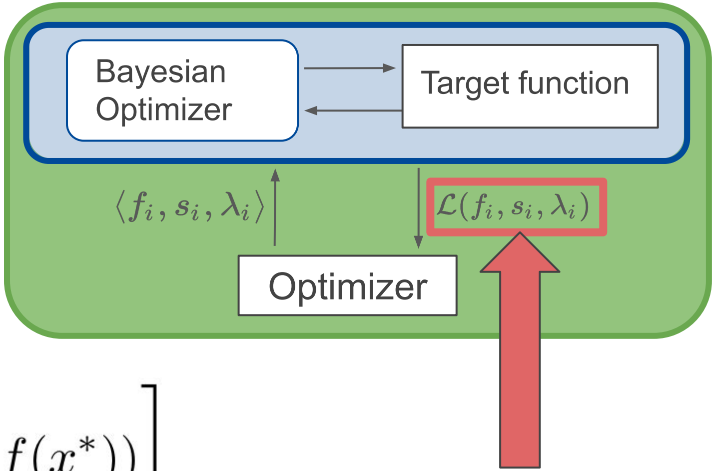
Ingredients

1 Search Space

2 Target functions

3 Meta-loss function to be optimized

- Measure good anytime performance
- Compare across multiple functions
- Hit optimum accurately



$$\mathcal{L}(\lambda) = \mathbb{E}_{f\sim\mathcal{F}}\left[\frac{1}{T}\sum_{t=1}^{T}\min_{\hat{x}\in\mathbf{x}(\lambda)_{1:t}}\log\left(f(\hat{x}) - f(x^*)\right)\right]$$

4 Optimizer

Ingredients

1 Search Space

2 Target functions

3 Meta-loss function to be optimized

4 Optimizer

→ Algorithm configuration

$$\lambda^* \in \arg\min_{\lambda \in \Lambda} \mathbb{E}_{\pi \sim \Pi} \left[c(\lambda, \pi) \right]$$

How Large is the Impact of Tuning

	DEF		DEF		DEF	
		RF		GP-ML		GP-MAP
artificial	−0.19		−2.35		−2.41	
SVM	−2.65		−2.90		−2.87	
ParamNet	−2.12		−2.15		−2.25	

Average log-regret (lower is better).

How Large is the Impact of Tuning

	DEF	LOFO	DEF	LOFO	DEF	LOFO
	RF		GP-ML		GP-MAP	
artificial	−0.19	**−0.95**	−2.35	−2.50	−2.41	−2.43
SVM	−2.65	−2.73	−2.90	**−3.11**	−2.87	−2.87
ParamNet	−2.12	**−2.37**	−2.15	**−2.36**	−2.25	**−2.32**

Average log-regret (lower is better).

LOFO: Running the Meta-optimizer on all but one function from a family, rerun the best found configuration on the left out function

Important Hyperparameters

Ablation[1] showed:

→ Only a **small set** of hyperparameters is important

→ Which hyperparameters **depend on the model**

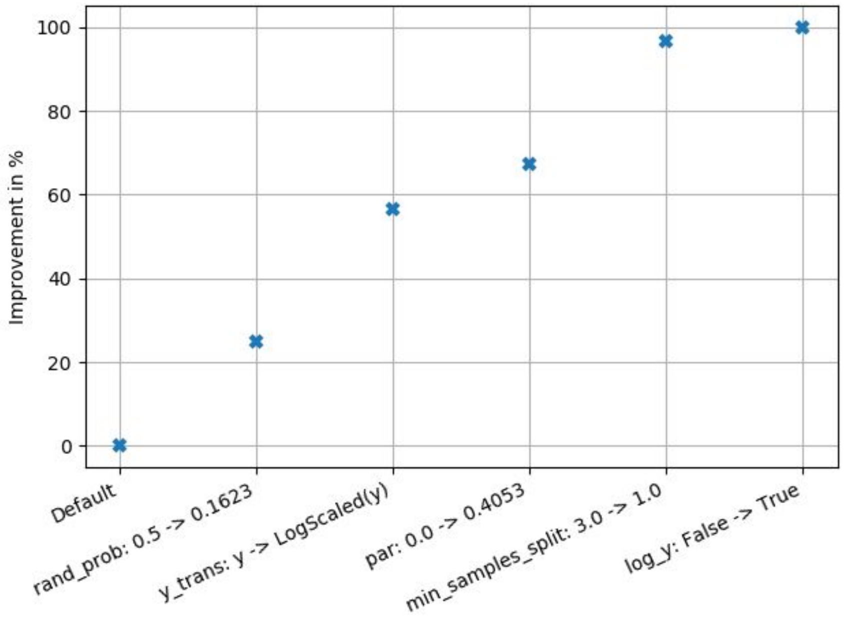


Figure: Most important hyperparameters according to ablation for **Bayesian optimization with Random Forests** on the artificial function family.

[1] C. Fawcett, H. H. Hoos. *Analysing differences between algorithm configurations through ablation*. J. Heuristics 2016

Wrap-Up

→ Hyperparameter optimization for Bayesian optimization is important

Open questions and future work:

- How to handle this in practice?
- Measure similarity of target functions