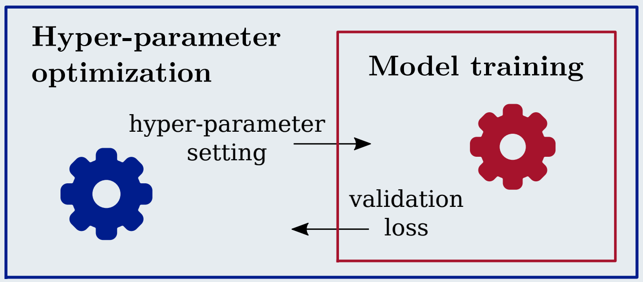
Model Performance Approximation in Hyper-parameter Optimization

Author: Markéta Jůzlová Supervisor: Ing. Tomáš Borovička

FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Motivation

Hyper-parameters of a machine learning algorithm are free parameters that need to be set before training the model (e.g. learning rate for artificial neural networks). These parameters have **significant impact** on algorithm behavior and in consequence on the **performance** of the trained model. However, finding an optimal hyper-parameter setting is **non-intuitive** and **problem-dependent**. Moreover, in order to evaluate a given setting, one needs to actually train the model. Training a model is **expensive** operation which may require a lot of computational resources. **The aim of automatic hyper-parameter optimization is to efficiently select a good setting without human assistance**.



Problem formulation

We formulate hyper-parameter optimization (HPO) as an optimization problem where we seek a hyper-parameter setting that minimizes expected loss (e.g. classification error) of the trained model. Our work reviews several state-of-the-art hyper-parameter optimization methods with a focus on methods that use a so-called surrogate model to speed up the optimization.

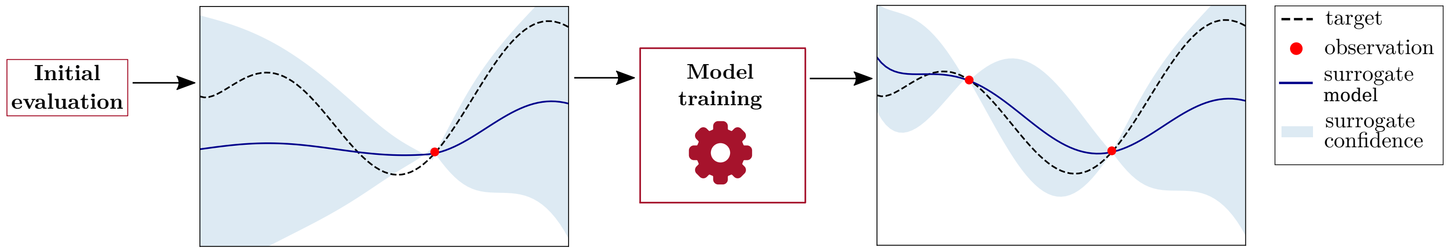


Figure 1: Hyper-parameter optimization with a surrogate model. The surrogate is used to select next point to evaluate. After evaluation, the surrogate is updated using the observed target function value.

Hyper-parameter optimization methods

Based on a literature review, we identified and described two traditional HPO methods – grid search and random search and two state-of-the-art methods – sequential model-based optimization and the Hyperband algorithm.

Sequential model-based optimization, also known as Bayesian optimization, uses surrogate model which approximates optimized function. The surrogate model is built using observations from previous evaluations of the target function. The surrogate is then used to select the next point to evaluate (see Figure 1). We identified four surrogate models as the most promising for HPO:

- **Gaussian processes**,
- **random forests**,
- **tree-structured parzen estimator**,
- **artificial neural networks**.

Hyperband evaluates hyper-parameter settings using only a part of the resources (e.g. a fraction of the training time). The settings then compete among themselves. The settings that performed well get allocated more resources and they are evaluated again with these resources. Other settings are discarded.

Experiments

The goal of the experiments was to compare selected HPO methods and surrogates with focus on deep understanding of methods' behavior. We optimized various hyper-parameters of neural network on two datasets. Highlights of the results of the experiments:

- Hyperband outperforms other methods owing its success to a large number of generated settings and discarding the poorly performing ones.
- Model-based optimization outperforms random search (which we use as a baseline). Surrogate models positively bias the search to promising regions.

Conclusion

Our experiments demonstrated that automatic HPO methods are efficient in finding a well-performing hyper-parameter setting. We proved that those methods significantly simplify the application of machine learning making it independent on a manual effort of a human expert. The experiments confirmed two promising techniques – partial evaluation and surrogate modeling. Future work will focus on a combination of these techniques.