Speedup Techniques for Hyperparameter Optimization

Overview

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

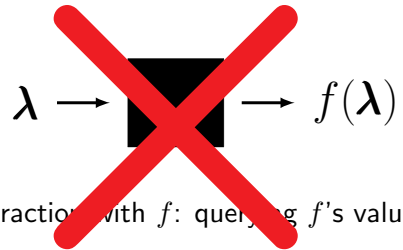
Beyond Black-box Optimization

Recall general blackbox optimization:

Only mode of interaction with f: querying f's value at a given $\boldsymbol{\lambda}$

Beyond Black-box Optimization

Recall general blackbox optimization:

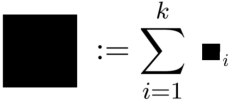


Only mode of interaction with f: querying f's value at a given $\boldsymbol{\lambda}$

Too slow for tuning expensive models

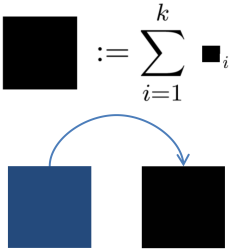
Methods for Going Beyond Blackbox Bayesian Optimization

- Sum of little black boxes
  - Each little black box is fast but only yields a noisy estimate
  - SMAC [Hutter et al. 2011] directly solves $\arg\min_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \sum_{i=1}^{N} c(\boldsymbol{\lambda}, i)$
  - Auto-WEKA [Thornton et al, 2013] used this to optimize 10-fold cross-validation performance



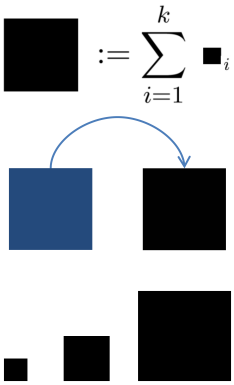
Methods for Going Beyond Blackbox Bayesian Optimization

- Sum of little black boxes
  - Each little black box is fast but only yields a noisy estimate
  - SMAC [Hutter et al. 2011] directly solves $\arg\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \sum_{i=1}^{N} c(\boldsymbol{\lambda}, i)$
  - Auto-WEKA [Thornton et al, 2013] used this to optimize 10-fold cross-validation performance
- Meta-learning across problems / datasets



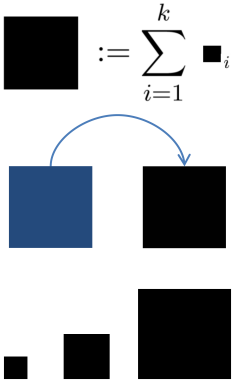
Methods for Going Beyond Blackbox Bayesian Optimization

- Sum of little black boxes
  - Each little black box is fast but only yields a noisy estimate
  - SMAC [Hutter et al. 2011] directly solves $\arg\min_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \sum_{i=1}^{N} c(\boldsymbol{\lambda}, i)$
  - Auto-WEKA [Thornton et al, 2013] used this to optimize 10-fold cross-validation performance
- Meta-learning across problems / datasets
- Multi-fidelity optimization



Methods for Going Beyond Blackbox Bayesian Optimization

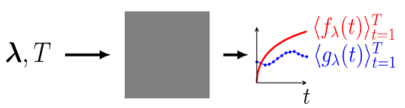
- Sum of little black boxes
  - Each little black box is fast but only yields a noisy estimate
  - SMAC [Hutter et al. 2011] directly solves $\arg\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \sum_{i=1}^{N} c(\boldsymbol{\lambda}, i)$
  - Auto-WEKA [Thornton et al, 2013] used this to optimize 10-fold cross-validation performance



- Meta-learning across problems / datasets

- Multi-fidelity optimization

- Graybox optimization / learning curve prediction



Learning Goals of this Lecture

After this lecture, students can ...

- Describe many different ways of using meta-learning to speed up HPO
- Explain the concept of multi-fidelity optimization to speed up HPO
- Explain the Successive Halving and Hyperband algorithms
- Explain how to combine Bayesian optimization and Hyperband in BOHB
- Describe how to exploit multiple fidelities in Bayesian optimization
- Discuss several ways of predicting learning curves
- Discuss success stories of speeding up Bayesian optimization