Speedup Techniques for Hyperparameter Optimization

Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

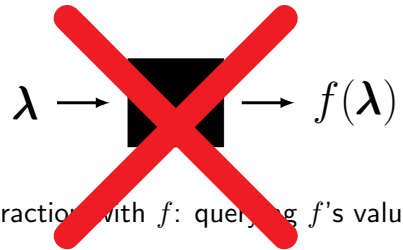
Beyond Black-box Optimization

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:



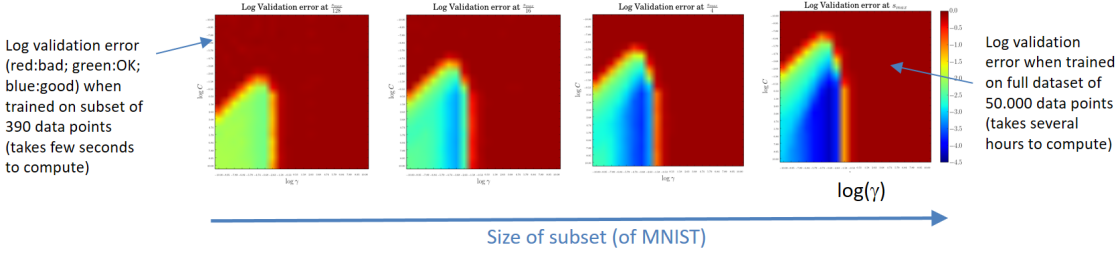
Only mode of interaction with f: querying f's value at a given $\boldsymbol{\lambda}$

Too slow for tuning expensive models

Methods for Going Beyond Blackbox Bayesian Optimization

- One possible cheap approximation of an expensive function: use a data subset
  - Many cheap evaluations on small subsets
  - Few expensive evaluations on the full data

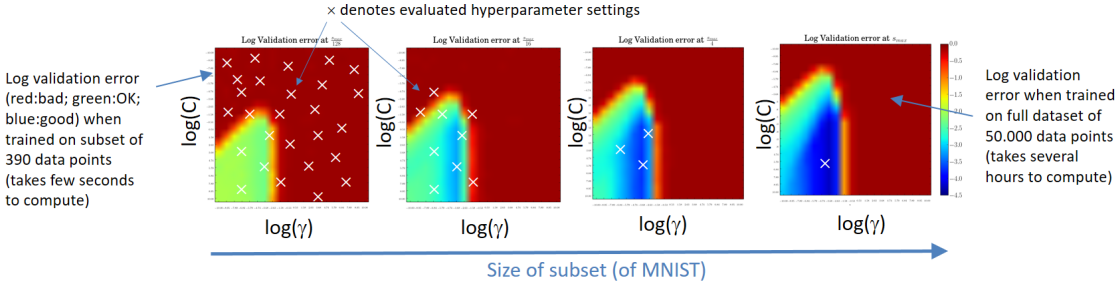
- E.g.: Support Vector Machines (SVM) on MNIST dataset (hyperparameters: C, γ)



Methods for Going Beyond Blackbox Bayesian Optimization

- One possible cheap approximation of an expensive function: use a data subset
  - Many cheap evaluations on small subsets
  - Few expensive evaluations on the full data

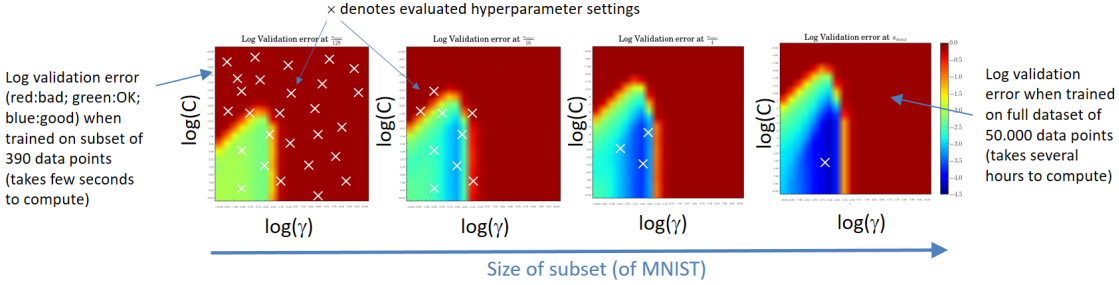
- E.g.: Support Vector Machines (SVM) on MNIST dataset (hyperparameters: C, γ)



Methods for Going Beyond Blackbox Bayesian Optimization

- One possible cheap approximation of an expensive function: use a data subset
  - Many cheap evaluations on small subsets
  - Few expensive evaluations on the full data

- E.g.: Support Vector Machines (SVM) on MNIST dataset (hyperparameters: C, γ)



→ up to 1000x speedups over blackbox optimization on full data [Klein et al, AISTATS 2017]

Learning Goals of this Lecture

After this lecture, students can ...

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