ORIE 4741: Introduction to AutoML

Chengrun Yang

October 28, 2019

Outline

Motivation

Some AutoML systems

Demo!

Outline

Motivation

Some AutoML systems

Demo!

What would we do to select models?

Model: algorithm + hyperparameter settings

e.g. ridge regression with $\lambda = 1$

In supervised learning, given a training set $\{(x_i, y_i)\}$ and test points $\{x_j\}$, how would we get $\{y_j\}$?

What would we do to select models?

Model: algorithm + hyperparameter settings

e.g. ridge regression with $\lambda = 1$

In supervised learning, given a training set $\{(x_i, y_i)\}$ and test points $\{x_j\}$, how would we get $\{y_j\}$?

- linear regression?
- random forest?
- gradient boosting?
- ...
- try all the models in scikit-learn [PVG+11], or all the available neural network architectures?

Model selection is important

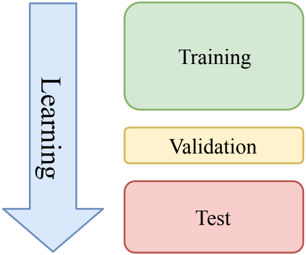
A naive exhaustive search that runs all models wastes

- programmer time
- computational time
    - takes long on small datasets
    - is impossible on large datasets

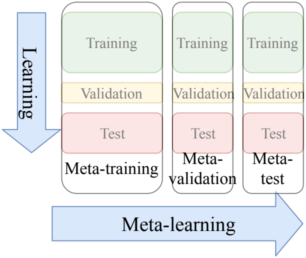
Approaches to avoid exhaustive search:

- single dataset: surrogate models to predict performance
    - Gaussian processes
    - genetic programming
- learn across datasets: meta-learning

Learning vs meta-learning



(a) Learning



(b) Meta-learning

Outline

Motivation

Some AutoML systems

Demo!

AutoML frameworks for general models

- Bayesian optimization frameworks [SLA12, KFB$^+$17, ZBSS16, THHLB13] ...
- auto-sklearn [FKE$^+$15]: meta-learning + Bayesian optimization
- TPOT [OUA$^+$16]: genetic programming
- H2O AutoML
- Hyperband [LJD$^+$18]
- Probabilistic matrix factorization (PMF) [FSE18]
- Oboe: matrix factorization + experiment design [YAKU19]
- ...

AutoML for neural architecture search

Techniques in use:

- reinforcement learning [ZL17] . . .
- genetic programming [SSN17] . . .
- Bayesian optimization [JSH19] . . .
- . . .

Outline

Motivation

Some AutoML systems

Demo!

Question: When does AutoML overfit?

Recall overfitting: low training error and high test error

References I

Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter.
Efficient and robust automated machine learning.
In *Advances in Neural Information Processing Systems*, pages 2962–2970, 2015.

Nicolo Fusi, Rishit Sheth, and Melih Elibol.
Probabilistic matrix factorization for automated machine learning.
In *Advances in Neural Information Processing Systems*, pages 3352–3361, 2018.

References II

- Haifeng Jin, Qingquan Song, and Xia Hu.
  Auto-keras: an efficient neural architecture search system.
  In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pages 1946–1956, New York, NY, USA, 2019. ACM.

- Aaron Klein, Stefan Falkner, Simon Bartels, Philipp Hennig, and Frank Hutter.
  Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets.
  In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 528–536, Fort Lauderdale, FL, USA, 20–22 Apr 2017. PMLR.

References III

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar.
Hyperband: A novel bandit-based approach to hyperparameter optimization.
Journal of Machine Learning Research, 18(185):1–52, 2018.

Randal S. Olson, Ryan J. Urbanowicz, Peter C. Andrews, Nicole A. Lavender, La Creis Kidd, and Jason H. Moore.
Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30 – April 1, 2016, Proceedings, Part I, chapter Automating Biomedical Data Science Through Tree-Based Pipeline Optimization, pages 123–137.
Springer International Publishing, 2016.

References IV

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.
Scikit-learn: Machine Learning in Python.
Journal of Machine Learning Research, 12:2825–2830, 2011.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams.
Practical Bayesian optimization of machine learning algorithms.
In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

References V

Masanori Suganuma, Shinichi Shirakawa, and Tomoharu Nagao.
A genetic programming approach to designing convolutional neural network architectures.
In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 497–504. ACM, 2017.

Chris Thornton, Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown.
Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms.
In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 847–855. ACM, 2013.

References VI

Chengrun Yang, Yuji Akimoto, Dae Won Kim, and Madeleine Udell.
Oboe: Collaborative filtering for automl model selection.
In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1173–1183. ACM, 2019.

Yuyu Zhang, Mohammad Taha Bahadori, Hang Su, and Jimeng Sun.
FLASH: fast Bayesian optimization for data analytic pipelines.
In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 2065–2074. ACM, 2016.

References VII

Barret Zoph and Quoc V. Le.
Neural architecture search with reinforcement learning.
2017.