Automated Machine Learning (AutoML)

- Introduction

Dr. Quanming Yao

Assistant professor, EE Tsinghua

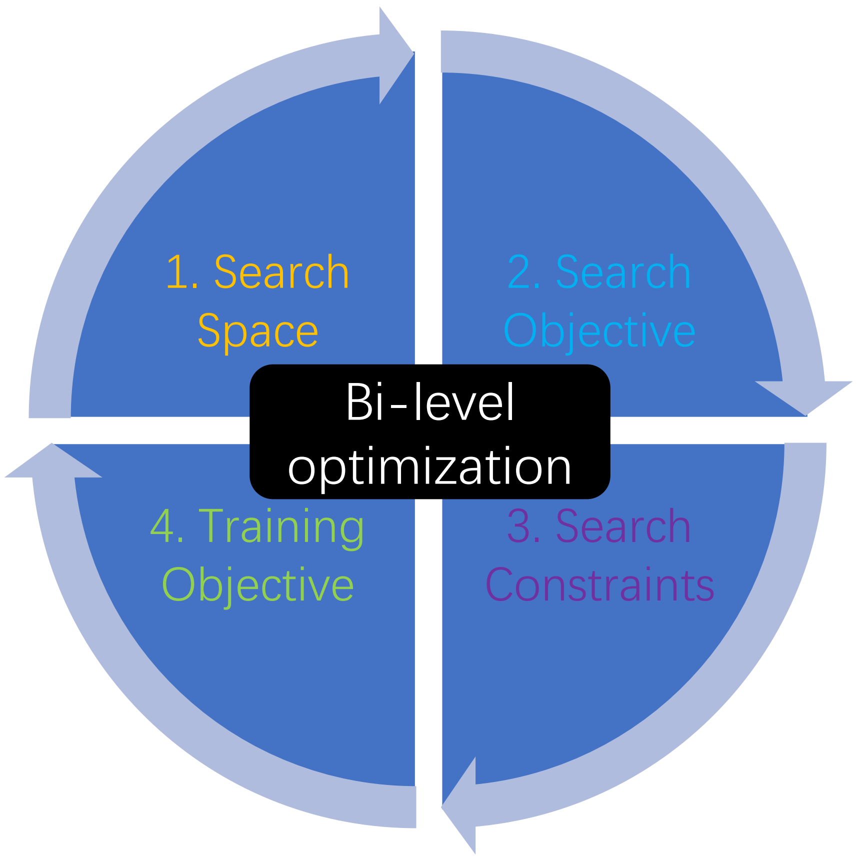
qyaoaa@tsinghua.edu.cn

https://lars-group.github.io/index.html

Little About Me

- Join EE, Tsinghua University at 2021.6
- Focus on making machine learning easier, faster and more robust
  - Work on automated machine learning, neural architecture search, graph neural networks, knowledge graph learning
- Published 60+ papers
  - include Cell Patterns, JMLR, TPAMI, ICML, NeurIPS and ICLR etc
  - ~4000 citations by Google SC
  - Top 10 cited paper in NeurIPS 2018
  - Wining solution on three tasks of OGB leaderboard
- Area chair of ICML, NeurIPS and ICLR.

All about AutoML



1. Define an AutoML problem

- Derive a search space from insights in specific domains
- Search objective is usually validation performance
- Search constraint is usually resource budgets
- Training objective usually comes from classical learning models

Search Space

$$\min_{\boxed{\lambda\in\mathcal{S}}} \boxed{M(F(w^*;\lambda), D_{\text{val}})}$$

Search Objective

$$\text{s.t.} \begin{cases} \boxed{\min\limits_{w} L(F(w;\lambda), D_{\text{tra}})} \\ \boxed{G(\lambda) \leq C} \end{cases}$$

Training Objective

Search Constraints

2. Design or select proper search algorithm

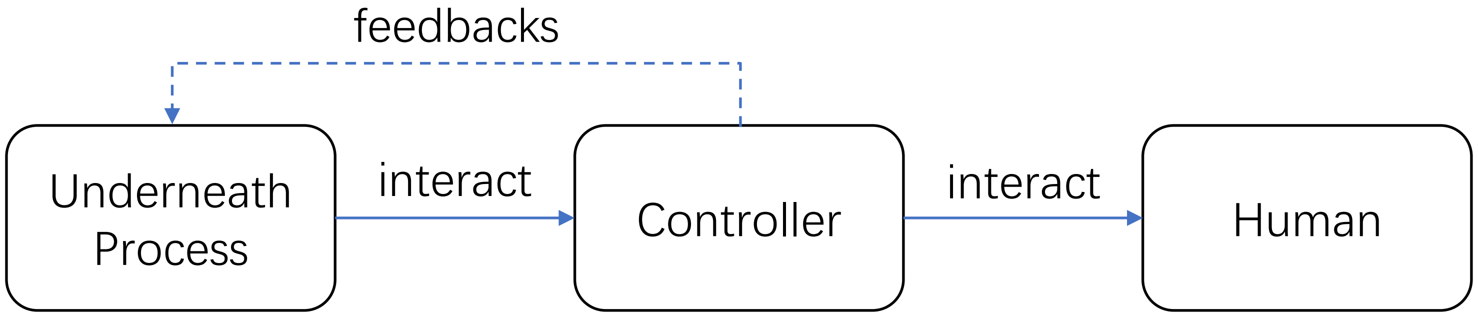
- Reduce model training cost (time to get w^*)

Outline

- What's Automated Machine Learning (AutoML)
- Examples of AutoML
- Trends in Machine Learning
- Future Works and Summary

What is Automation?

Automation is the technology by which a process or procedure is performed with minimal human assistance



Automation:

- Fewer and more understandable interface exposed to **human**
- The **controller** interacts with underneath process in a more robust and stable way

F. Golnaraghi and B. Kuo. Automatic control systems. (tenth edition) 2018 (Chapter 1)

What is Machine Learning (ML)?

Applications

Search Engine
Recommender
Systems
Loss Assessment

Image Classification

Predict the class of the object

Security Monitoring
Bio-payment
Flow Statistics

Face Recognition

Who is the person

COVID Simulation
Petroleum Exploration
Drug Discovery

Drug Design

Learn to make decisions

Better Performance

Higher Efficiency

Definition

Structure Samples

$$\min_x F(\, x \,; D \,)$$

Parameters

(iterative)
optimization

Prediction
Accuracy

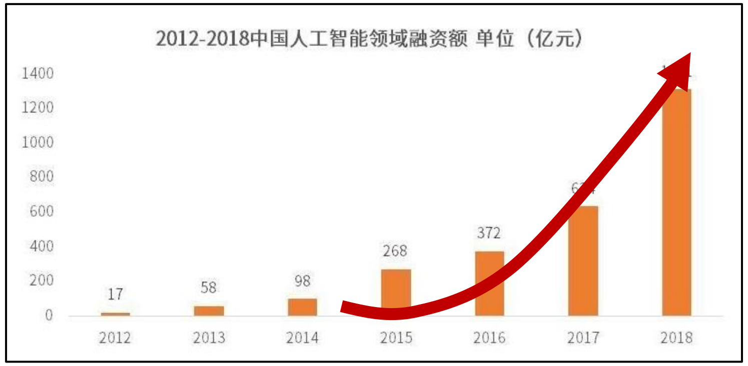
[1]. Machine Learning, Tom Mitchell, McGraw Hill, 1997.
[2]. 周志华 著. 机器学习, 北京: 清华大学出版社, 2016年

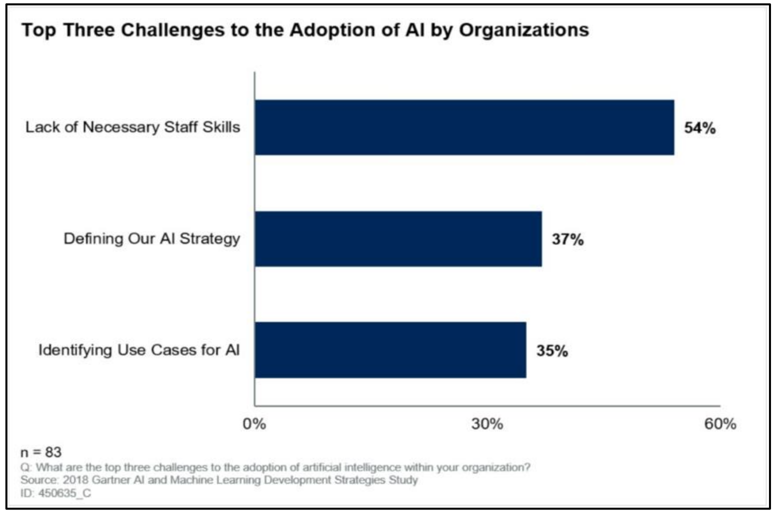
What is AutoML?

Automated machine learning (AutoML)

- Atomize a learning process into different blocks, and try to re-combine these blocks with optimization approaches.
- Why we need it?
  - Human participations can be naturally replaced by computation power
  - Understand the design of learning methods on a system-level

Why need AutoML?



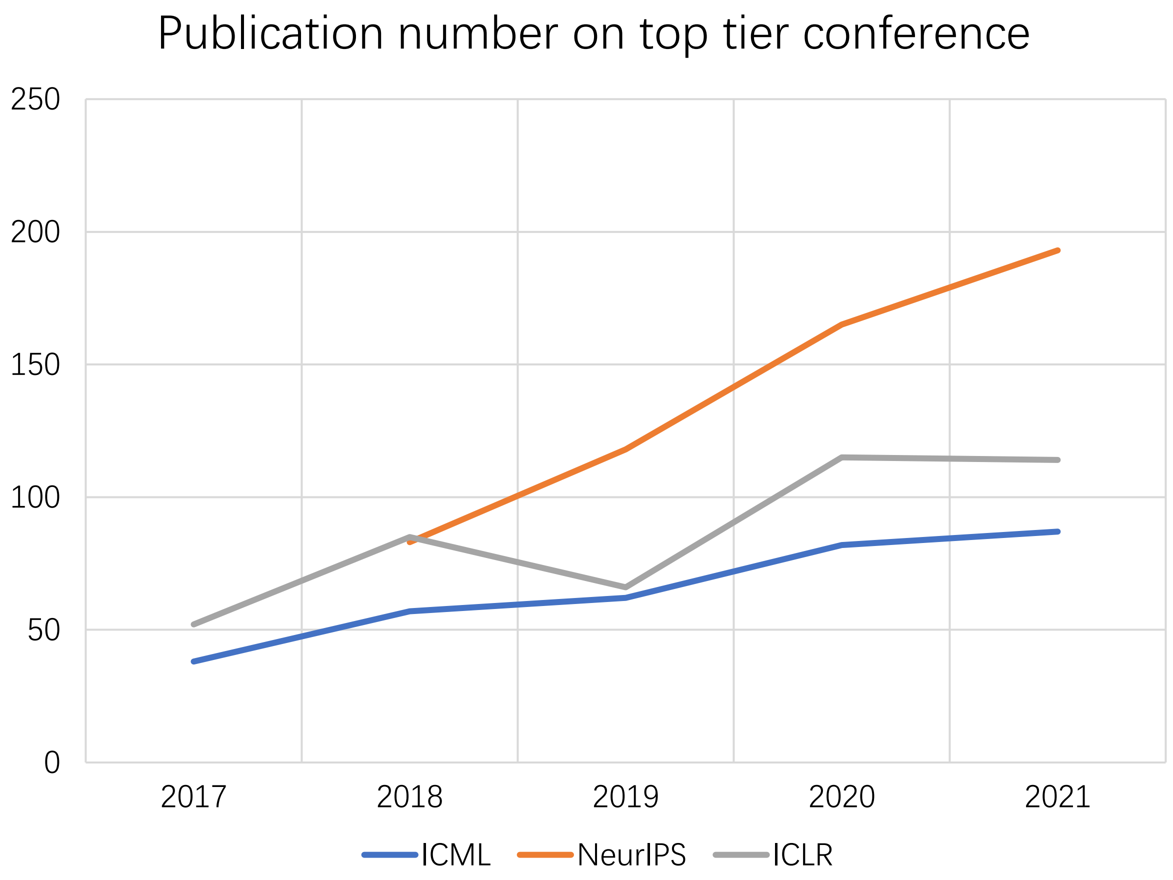


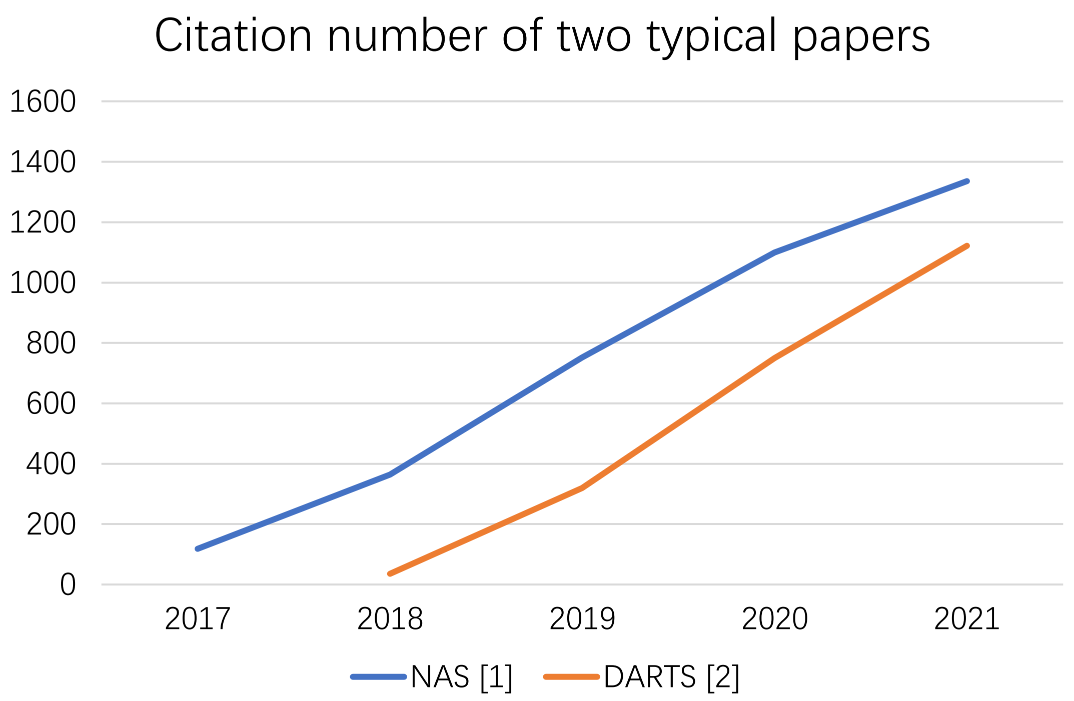
- **Industry** – reduce the expense, increase usage coverage – huge market value [1]
- **Academy** – understanding data science on a higher level – great intelligence value [2]

[1]. Gartner: https://www.forbes.com/sites/janakirammsv/2020/03/02/key-takeaways-from-the-gartner-magic-quadrant-for-ai-developer-services/#a95b99ee3e5e

[2]. Yoshua Bengio: From System 1 Deep Learning to System 2 Deep Learning | NeurIPS 2019

Recent Publications (NAS)



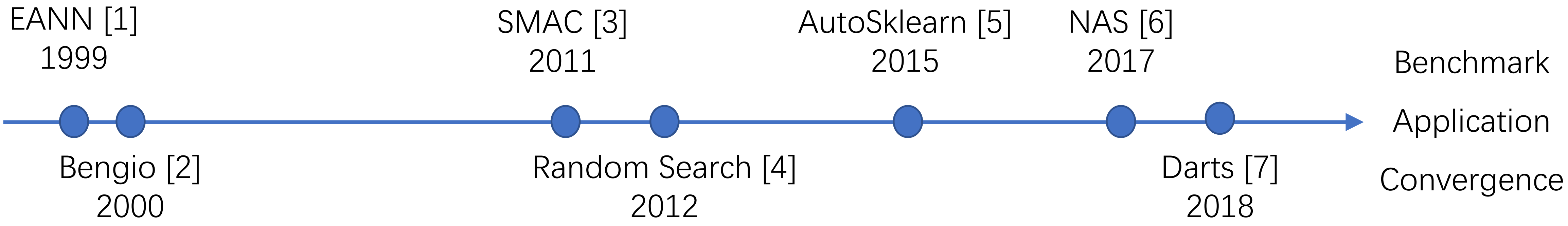


- [1] is the pioneering work of NAS
- [2] makes NAS fast

[1]. Zoph, Barret, and Quoc V. Le. Neural architecture search with reinforcement learning. ICLR 2017

[2]. Hanxiao Liu, Karen Simonyan, Yiming Yang. DARTS: Differentiable architecture search. ICLR 2019

Timeline



[1]. Yao, Xin. Evolving artificial neural networks. Proceedings of the IEEE 87.9 (1999)

[2] Bengio, Yoshua. Gradient-based optimization of hyperparameters. Neural computation 12.8 (2000)

[3]. Hutter, Frank, Holger H. Hoos, and Kevin Leyton-Brown. Sequential model-based optimization for general algorithm configuration. LION 2011

[4] Bergstra, James, and Yoshua Bengio. Random search for hyper-parameter optimization. Journal of machine learning research 13.2 (2012)

[5] Feurer, Matthias, et al. Efficient and robust automated machine learning. NIPS 2015

[6] Zoph, Barret, and Quoc V. Le. Neural architecture search with reinforcement learning. ICLR 2017

[7]. Hanxiao Liu, Karen Simonyan, Yiming Yang. DARTS: Differentiable architecture search. ICLR 2019 (appear in 2018)

Outline

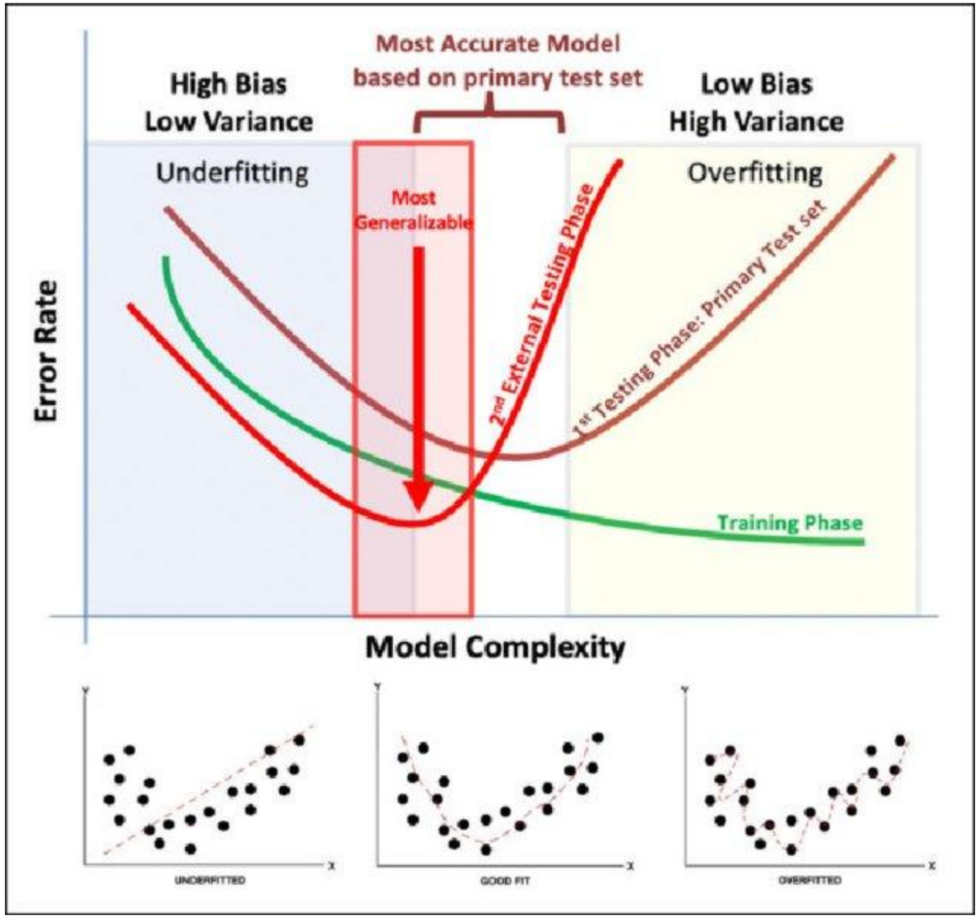
- What's Automated Machine Learning (AutoML)
- Examples of AutoML
  - Practical view
  - Theoretical view
- Trends in Machine Learning
- Future Works and Summary

Simple Example – Tune hyper-parameter

Search λ

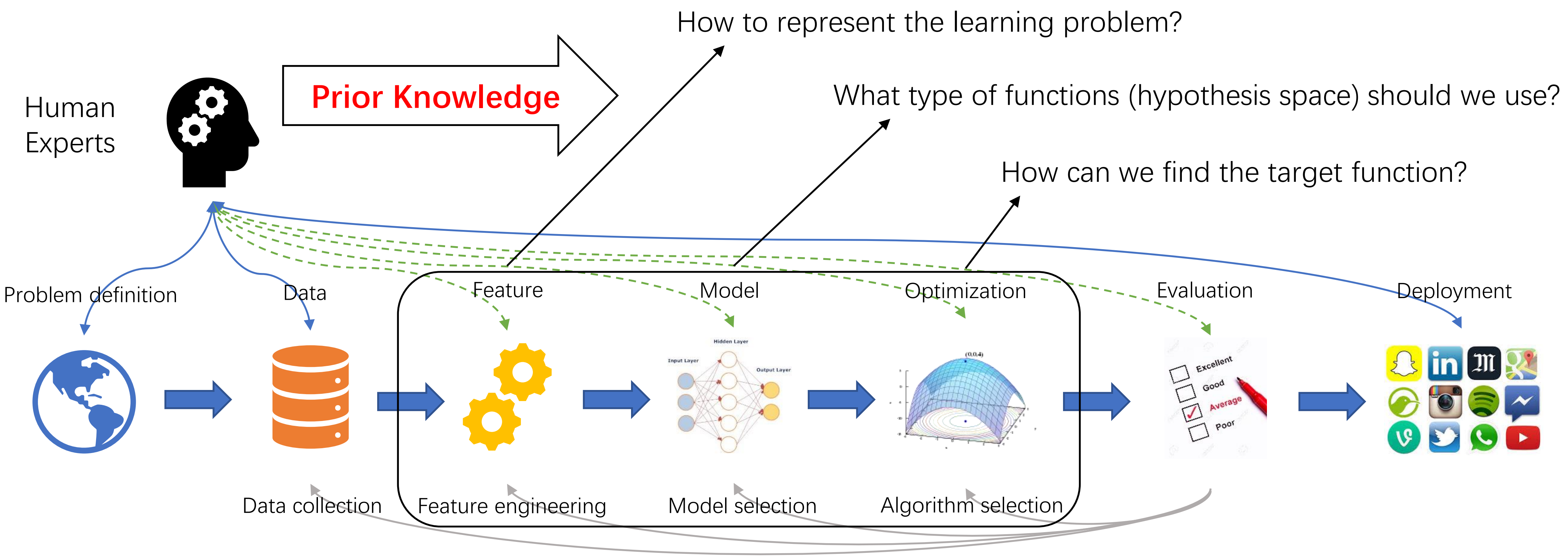
$$\underbrace{\max_{\lambda} \sum_j h(x_j; w^*)}_{\text{Validation Performance}} \quad \text{s.t.} \quad \underbrace{w^* = \min_{w} \sum_i f(x_i; w) + \overset{\text{Hyper-parameter}}{\lambda} \|w\|_1}_{\text{Training objective}}$$

- Seek proper λ to maximize performance
- Grid search: enumerating $\lambda \in \{1,2,4,8, \dots\}$



[1]. Image source: Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods.

How to use ML well?

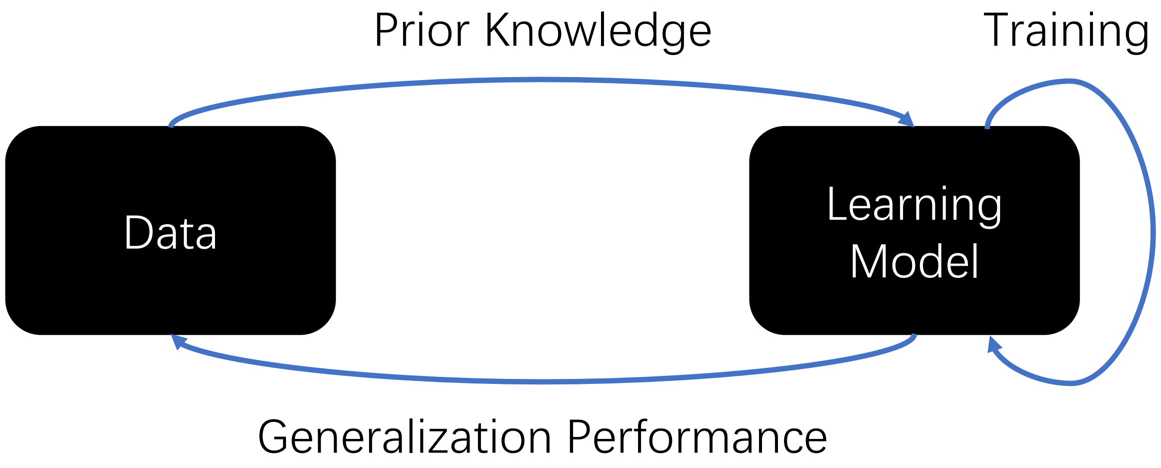


A trial-and-error process

Also see. T. Mitchell. Machine learning. 1997 (Chapter 1.3.1)

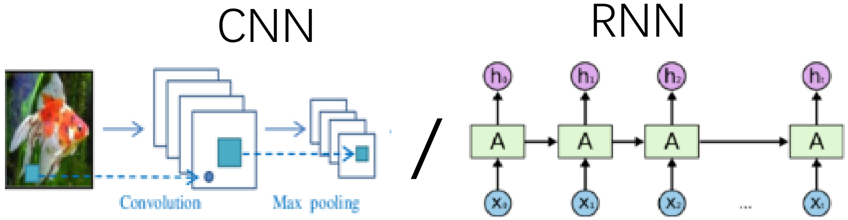
Figure is from Q. Yao et.al. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. arXiv 2018

How to use ML well?

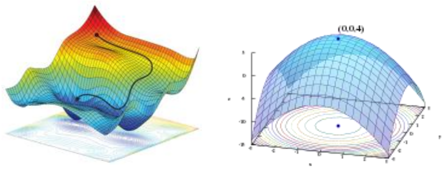


The Advancement of Learning

- An iteration between theory and practice
- A feedback loop



Generalization: What kind of f should we use?



SGD v.s. Adagrad[1]

Optimization: How can we find such f?

Prior knowledge

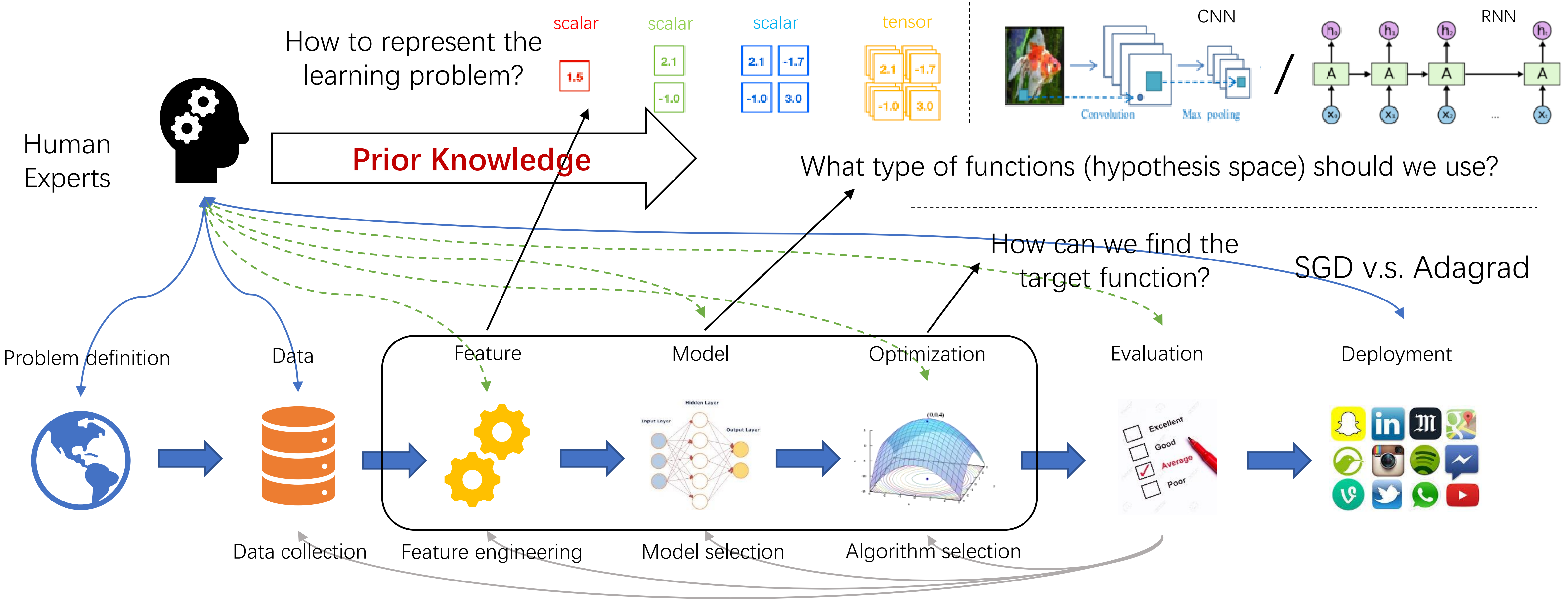
"All models are wrong, but some are useful"[2]

Better understanding of prior knowledge → Better hypothesis → Better generalization performance

[1]. Image Source: A. Amini et al. "Spatial Uncertainty Sampling for End-to-End Control". NeurIPS Bayesian Deep Learning 2018
[2] G. Box, Science and statistics, JASA 1976

Figure is from Q. Yao et.al. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. arXiv 2018

What is AutoML? – User perspective



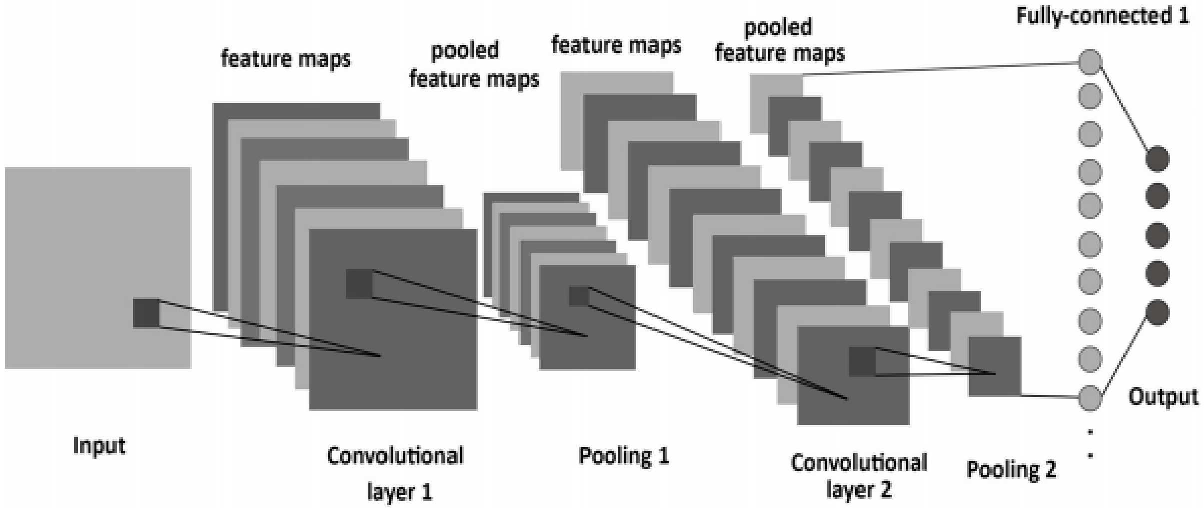
Re-combine learning blocks with optimization approaches

As a consequence

- Human participations can be naturally replaced by computation power
- Design of learning methods can be understand on a system-level

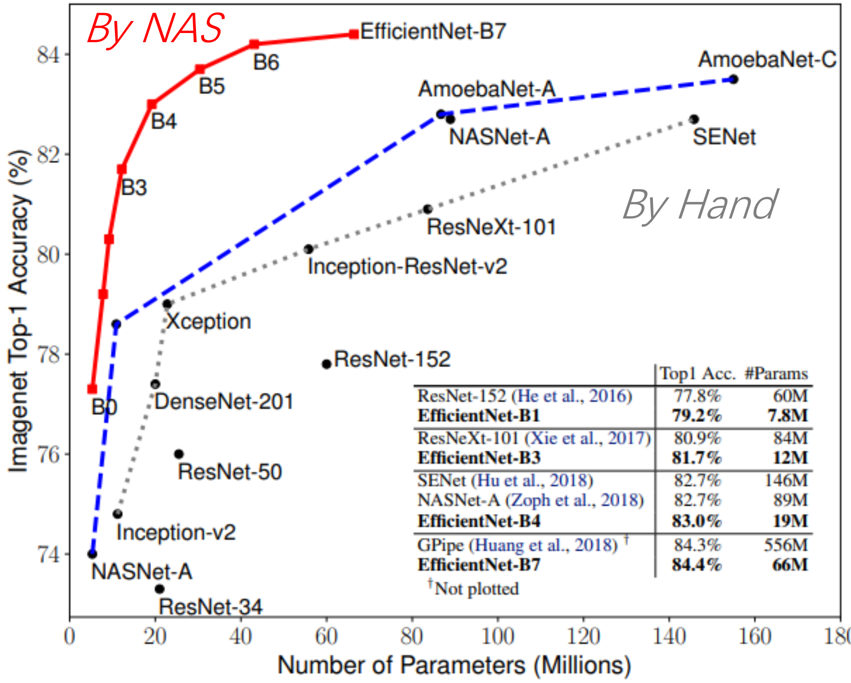
AutoML – Research example

Architecture of networks are critical to deep learning's performance but hard to fine-tune



Design choice in each layer

- number of filters
- filter height
- filter width
- stride height
- stride width
- skip connections



Much better than hand-designed ones

Neural Architecture Search (**NAS**) tries to directly optimize network architecture using validation data sets

Right figure: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML 2019
Left figure: Gradient-based learning applied to document recognition. 1998

AutoML – Commercialized examples

AutoSklearn

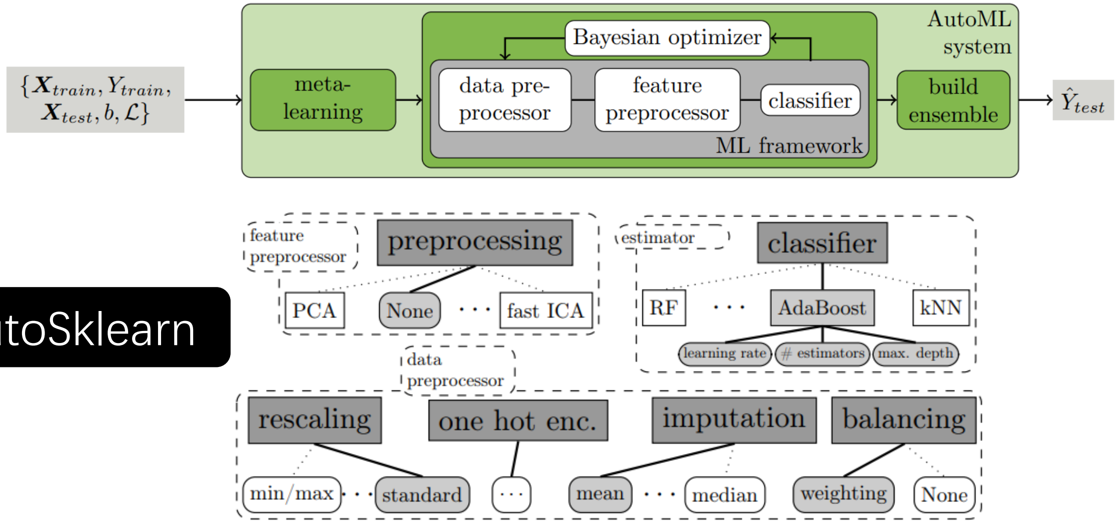
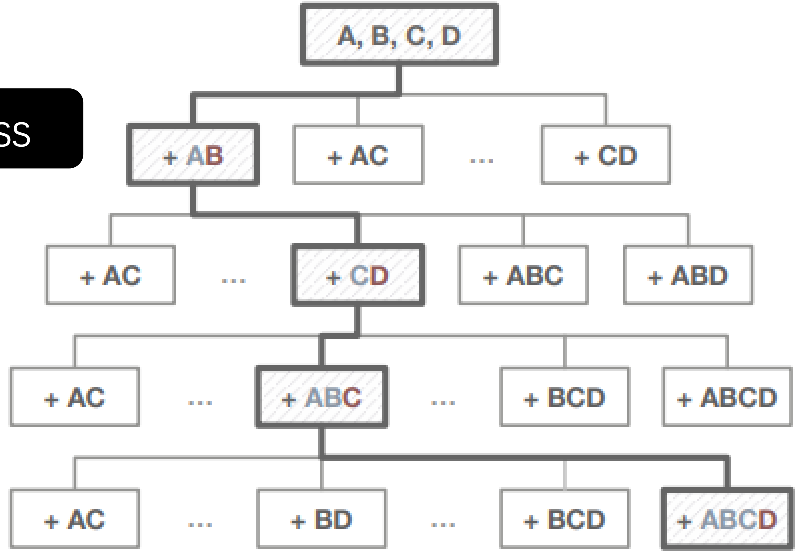


Figure 2: Structured configuration space. Squared boxes denote parent hyperparameters whereas boxes with rounded edges are leaf hyperparameters. Grey colored boxes mark active hyperparameters which form an example configuration and machine learning pipeline. Each pipeline comprises one *feature preprocessor*, *classifier* and up to three *data preprocessor* methods plus respective hyperparameters.

Tuning few hyper-parameters [1]

AutoCross

$$\mathbf{c}_{i,j,\cdots,k} = \mathrm{vec}\left(\mathbf{f}_i \otimes \mathbf{f}_j \otimes \cdots \otimes \mathbf{f}_k\right),$$



Cross-product feature generation [2]

[1]. F. Matthias et.al. Efficient and Robust Automated Machine Learning. NIPS 2015
[2]. Y. Luo, et.al. AutoCross: Automatic Feature Crossing for Tabular Data in Real-World Applications. KDD 2019

AutoML – Commercialized examples (video)

Y. Luo, et.al. AutoCross: Automatic Feature Crossing for Tabular Data in Real-World Applications. KDD 2019

Is AutoML Expensive?

- YES: at early stage
- NO: With proper usage of prior knowledge
- 1GPU Card is enough for you to play with CIFAR

Results on ImageNet

Method	Data Augmentation	#Params	Err	GPU days
NASNet-A (Zoph et al., 2018)	cutout	3.3M	2.65	2000
AmoebaNet-B-small (Real et al., 2019)	cutout	2.8M	2.50±0.05	3150
AmoebaNet-B-large (Real et al., 2019)	cutout	34.9M	2.13±0.04	3150
AlphaX (Wang et al., 2019b)	cutout	2.83M	2.54±0.06	1000
NAO (Luo et al., 2018)	cutout	3.2M	3.14±0.09	225
DARTS (Liu et al., 2019b)	cutout	3.3M	2.76±0.09	1
P-DARTS (Chen et al., 2019a)	cutout	3.4M	2.5	0.3
PC-DARTS (Xu et al., 2020)	cutout	3.6M	2.57±0.07	0.3
Fair-DARTS (Chu et al., 2019b)	cutout	3.32M	2.54±0.05	3
BayeNAS (Zhou et al., 2019)	cutout	3.4M	2.81±0.04	0.2
CNAS (Lim et al., 2020)	cutout	3.7M	2.60±0.06	0.3
MergeNAS (Wang et al., 2020)	cutout	2.9M	2.68±0.01	0.6
ASNG-NAS (Akimoto et al., 2019)	cutout	3.32M	2.54±0.05	0.11
XNAS (Nayman et al., 2019)	cutout + autoaug	3.7M	1.81	0.3
one-shot REA	cutout + autoaug	3.5M	2.02±0.03	0.75
one-shot LaNas (Wang et al., 2019a)	cutout + autoaug	3.6M	1.68±0.06	3
few-shot DARTS-Small	cutout	3.8M	**2.31±0.08**	1.35
few-shot DARTS-Large	cutout	45.5M	**1.92±0.08**	1.35
few-shot DARTS-Small	cutout + autoaug	3.8M	**1.70±0.08**	1.35
few-shot DARTS-Large	cutout + autoaug	45.5M	**1.28±0.08**	1.35
few-shot REA	cutout + autoaug	3.7M	**1.81±0.05**	0.87
few-shot LaNas	cutout + autoaug	3.2M	**1.58±0.04**	3.8

Outline

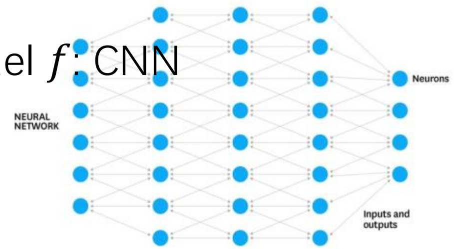
- What’s Automated Machine Learning (AutoML)
- Examples of AutoML
  - Practical view
  - Theoretical view
- Trends in Machine Learning
- Future Works and Summary

Once for All Solution?

Image Classification

Optimization ⇩

Model f: CNN



Generalization ⇩

Accuracy

Design a **model f** to perform the learning task

Target	Concept	Issue
Better Performance (more important)	Generalization	What kind of f should we use?
Higher Efficiency	Optimization	How can we find such f?

Not everything can be learnt

PAC-Learning (Definition 2.3 in [1]): What kind of problems can be solved in polynomial time

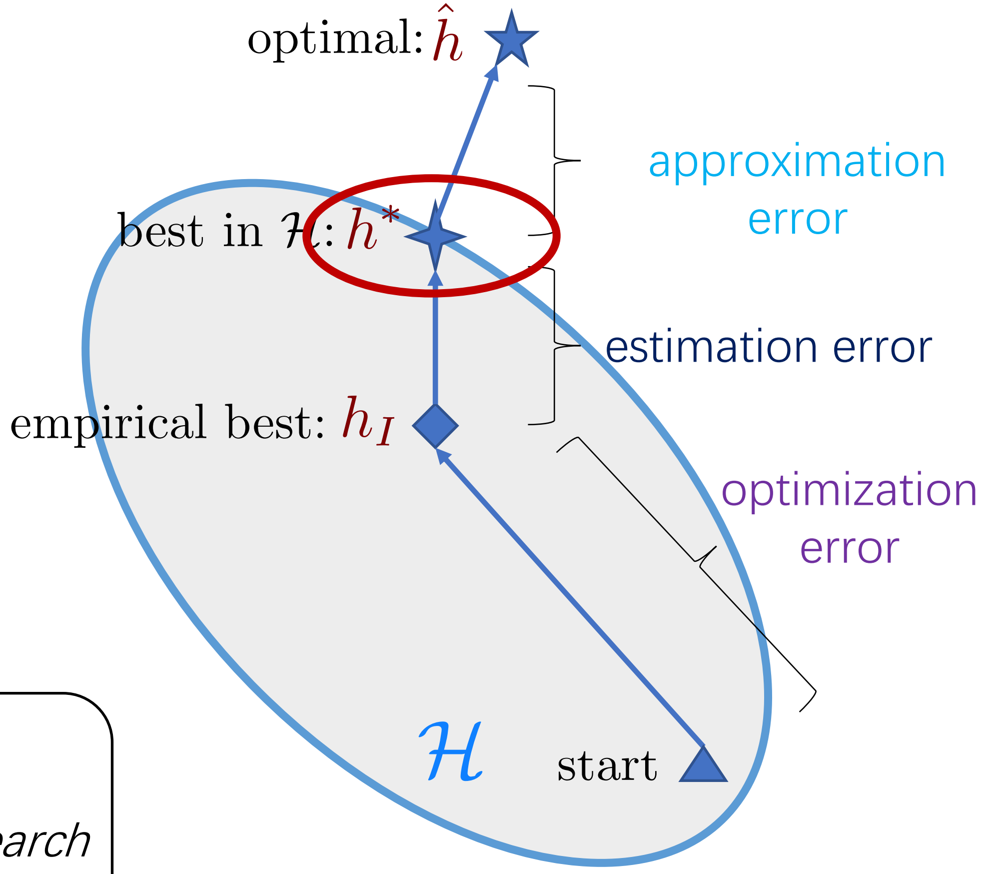
No Free Lunch Theorem (Appendix B [2]): No single algorithm can be good on all problems

[1]. M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of machine learning. 2018
[2]. O. Bousquet, et.al. Introduction to Statistical Learning Theory. 2016

Look Inside Error Decomposition

Automatically find h^* by searching hyper-parameters

$$\underbrace{\max_{\lambda} \sum_j h(x_j; w^*)}_{\text{Validation Performance}} \quad \text{s.t.} \quad w^* = \underbrace{\min_{w} \sum_i f(x_i; w) + \lambda \|w\|_1}_{\text{Training objective}}$$



How to further improve the performance in an automatic manner (i.e., reduce the approximation error)?

- Feature can be weak → *Automatic feature engineering*
- Linear predictor can be too restrictive → *Neural architecture search*
- Grid search can be slow → *Search in a supernet*

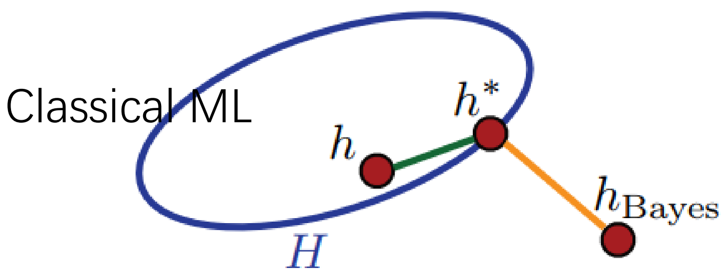
AutoML

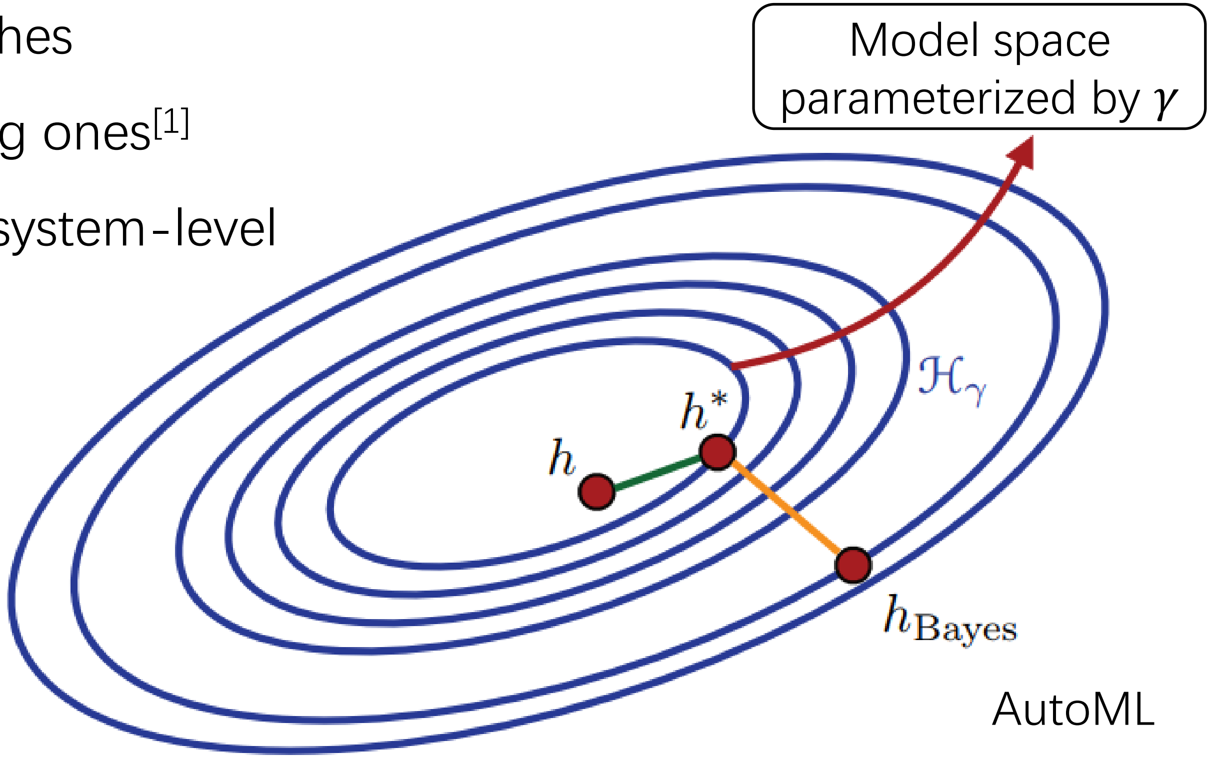
Figure 4.1 & 4.2. M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of machine learning. 2018
[1]. P. Battaglia. et.al. Relational inductive biases, deep learning, and graph networks. arXiv 2018.

What is AutoML – Producer perspective

Re-combine these blocks with optimization approaches

- combinatorial generalize new models from existing ones[1]
- Understand the design of learning methods on a system-level





Parameterize (low-level) prior knowledge in the usage and design of machine learning

Outline

- What's Automated Machine Learning (AutoML)
- Examples of AutoML
- Trends in Machine Learning
- Future Works and Summary

AutoML – Successor of ML's trend

Rule-based
Association rules mining
1970s

Statistics-based
Support vector machine
1990s

Deep Learning-based
Convolutional neural networks
2010s

AutoML-based
Neural architecture search
2017

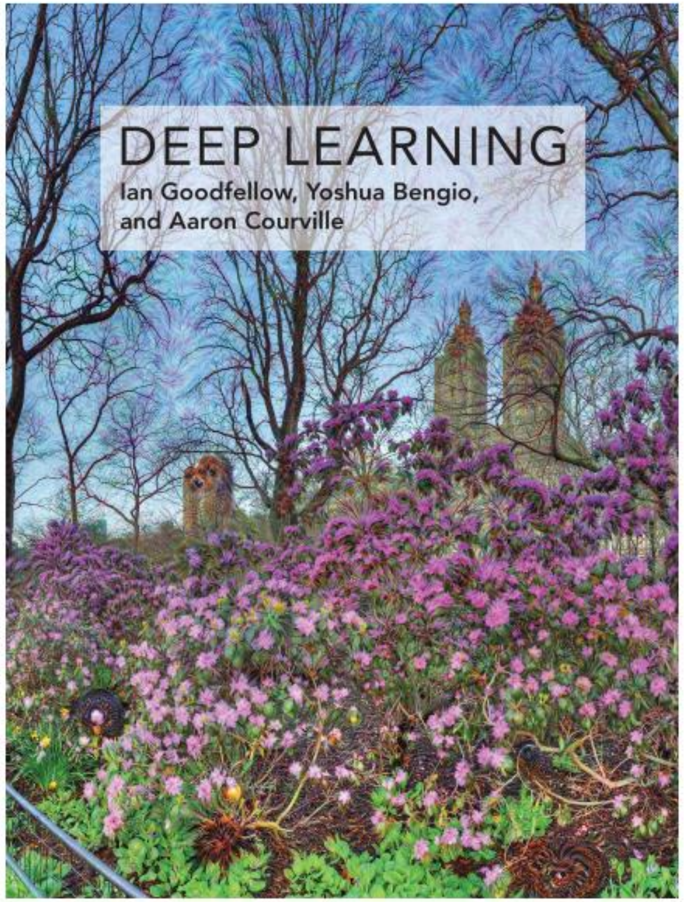
Easier to get better performance

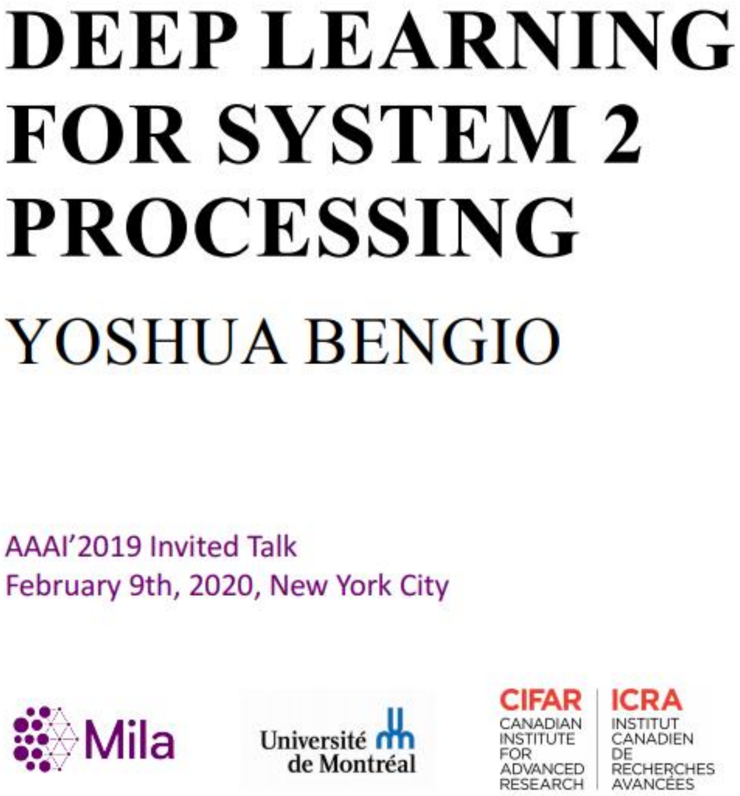
Continue the trends

- Larger hypothesis (more complex models) are being used.
- The prior knowledge is parameterized on a more abstract leve.

- Human participations can be naturally replaced by computation power
- Design of learning methods can be understand on a system-level

AutoML – Successor of ML's trend





Parameterized
prior knowledge
on a higher level

Related Areas

Sub-areas

- Neural architecture search
- Hyper-parameter search
- Automated feature engineering
- Algorithms selection
- Model selection

Related areas

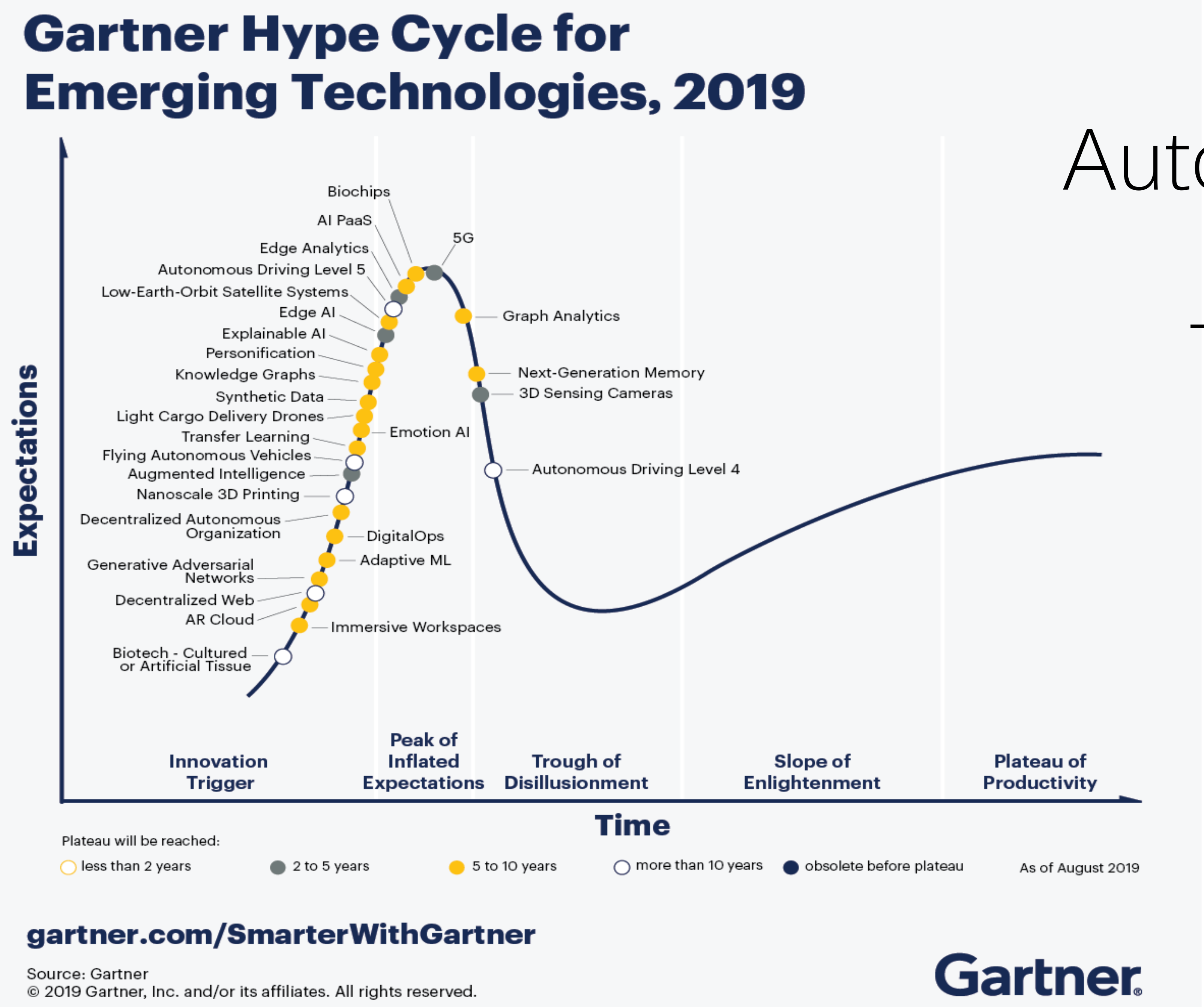
- Bi-level / Derivative-free optimization
  - Focus more on algorithm design
  - AutoML objective is one kind of objective where these algorithms can be applied
- Meta-learning
  - Focus on parameterize task distributions
  - Another kind of bi-level objective
  - Do not use validation set to update hyper-parameters

AutoML is Diverging – Disappointed

Feel painful with tuning meta-hyperparameters / Get enough with instability of results?

Industry	Academy
Automation	Machine Learning
Replace low-level prior knowledge with computation	Achieve better performance by parameterized low-level prior knowledge

V.S.



AutoML is Diverging

— Where is solution?

The world is big

Be more professional in your own area

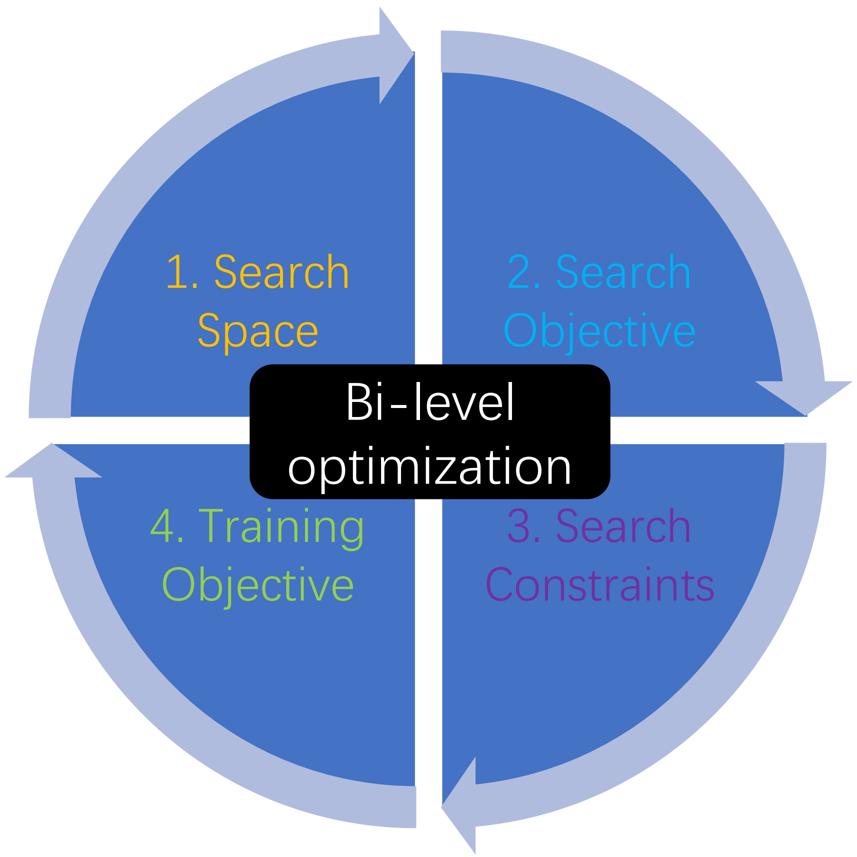
But keep your eyes to others

C. Christensen. The innovator's dilemma: when new technologies cause great firms to fail. Book 1997.

Outline

- What's Automated Machine Learning (AutoML)
- Examples of AutoML
- Trends in Machine Learning
- Future Works and Summary

How to use AutoML



1. Define an AutoML problem

- Derive a search space from insights in specific domains
- Search objective is usually validation performance
- Search constraint is usually resource budgets
- Training objective usually comes from classical learning models

Search Space

Search Objective

$$\min_{\lambda \in \mathcal{S}} \boxed{M(F(w^*;\lambda), D_{\text{val}})}$$

$$\text{s.t.} \begin{cases} \boxed{\min_{w} L(F(w;\lambda), D_{\text{tra}})} \\ \boxed{G(\lambda) \leq C} \end{cases}$$

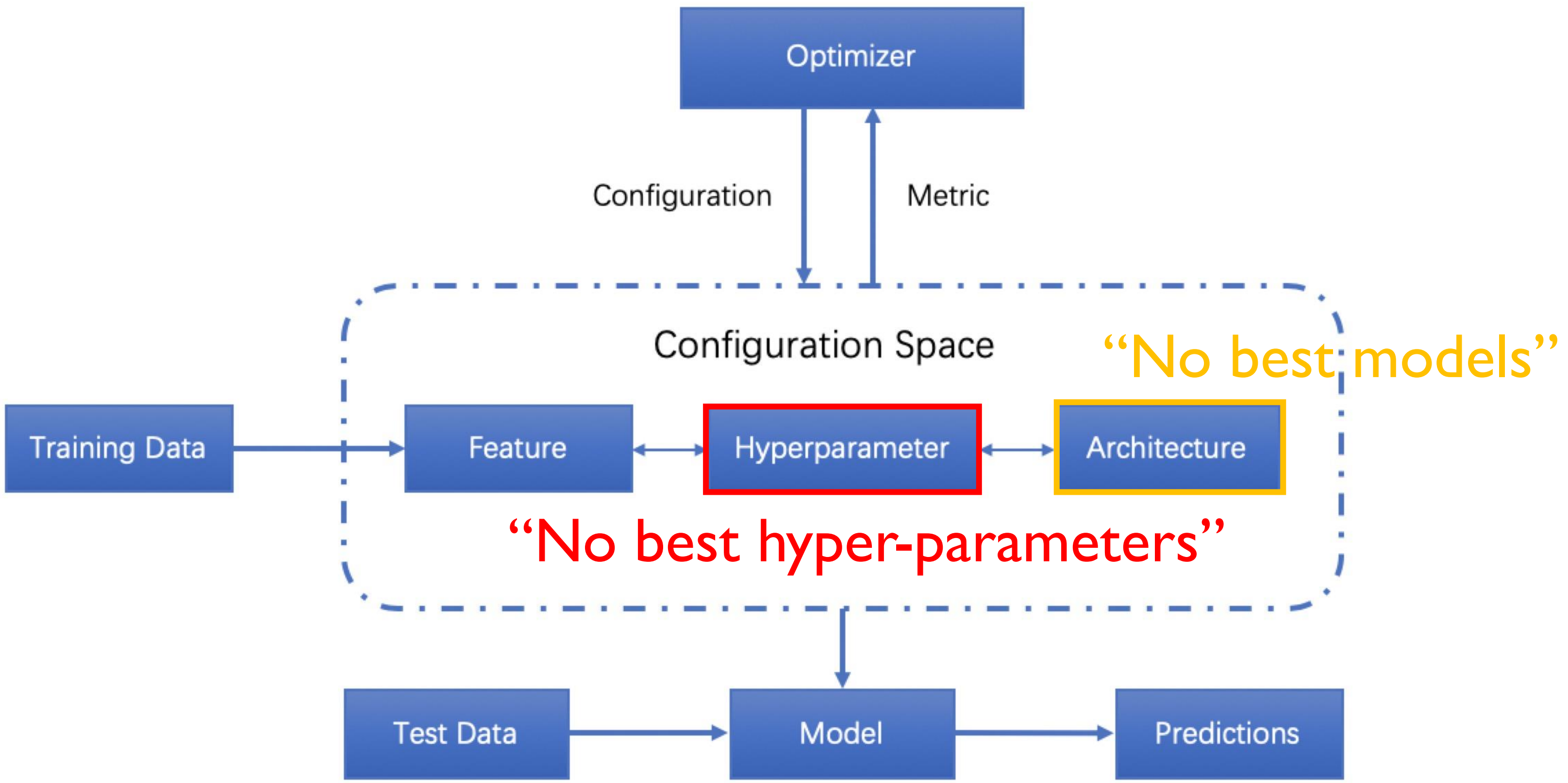
Training Objective

Search Constraints

2. Design or select proper search algorithm

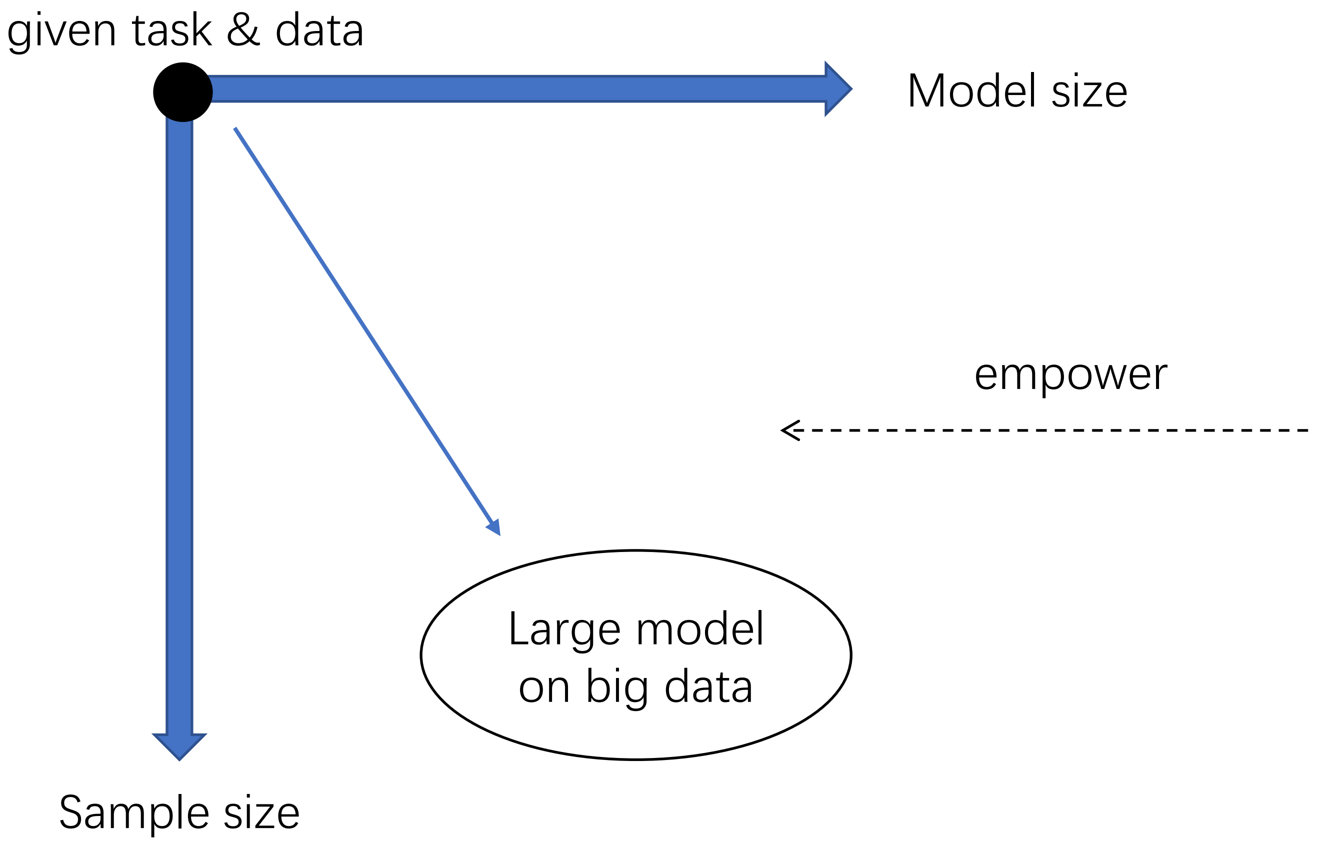
- Reduce model training cost (time to get w^*)

Current Scope of AutoML



https://awesomeopensource.com/project/hibayesian/awesome-automl-papers

AutoML – Research landscape



AutoML – Research focus

Space
- More application
- Space transformation

Objective
- New task / applications
- Function properties

Evaluation
- Understand parameter-sharing
- Evaluate without training

Constraint
- Hard-ware aware
- Architecture prior

Bi-level optimization

Convergence guarantee | **Generalization guarantee**

Summary

- Exploring prior knowledge is important in machine learning
  - Cost time and critical to generalization performance
- AutoML attempts to parameterize low-level prior knowledge
  - Human participations can be naturally replaced by computation power
  - total error can be reduced (generalization can be improved)
- To use well AutoML techniques
  - Exploring high-level domain knowledge when defining the AutoML problem
  - Reducing model training cost when design search algorithm