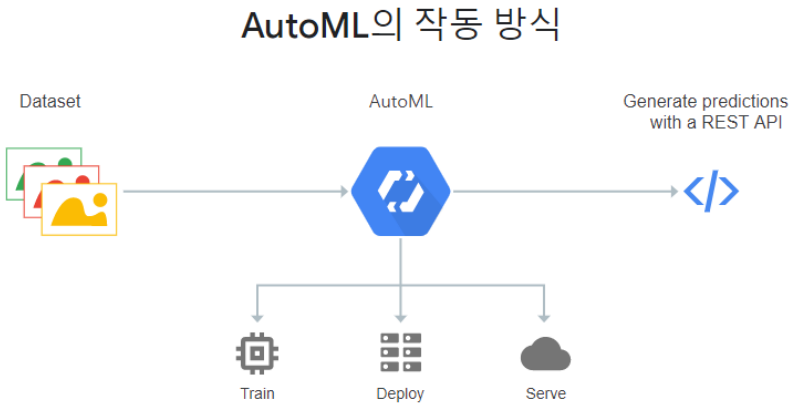
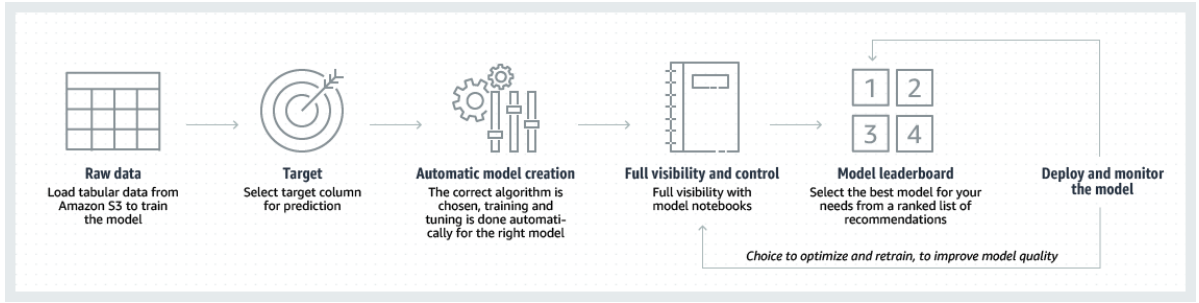
Chapter 08. 좋은 딥러닝 구조를 찾아내는 딥러닝 (Neural Architecture Search)

NAS란?

AutoML Automated Machine Learning





Google Cloud AutoML

Amazon SageMaker AutoPilot

컴퓨터가 자동으로 머신 러닝을 해 준다면? 인공지능 개발자에게는 천국일까 악몽일까?

What's in AutoML?

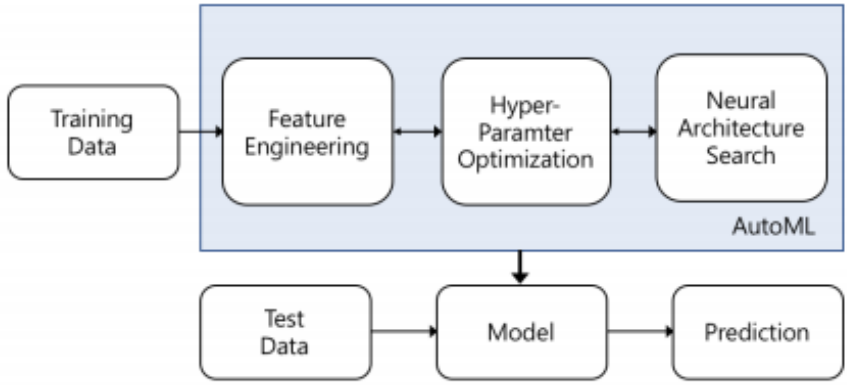


그림 1 자동 기계학습 시스템

NAS를 비롯하여 여러가지 기술이 포함되어 있다. NAS가 AutoML 분야에서 가장 주목받으면서, 최근에는 NAS라고 하면 AutoML 전반을 모두 포함하는 경우가 많다.

https://ettrends.etri.re.kr/ettrends/178/0905178004/34-4_32-42.pdf

NAS Neural Architecture Search

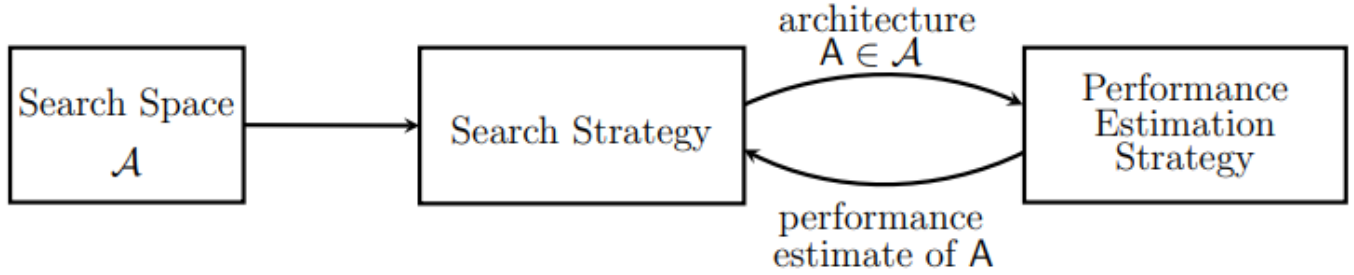


Figure 1: Abstract illustration of Neural Architecture Search methods. A search strategy selects an architecture A from a predefined search space $\mathcal{A}$. The architecture is passed to a performance estimation strategy, which returns the estimated performance of A to the search strategy.

NAS를 수행하는 과정은 Straight Forward하다. 각 과정의 연산 시간을 최소화 하는 것이 주된 목표이다.

http://www.jmlr.org/papers/volume20/18-598/18-598.pdf

Search Space

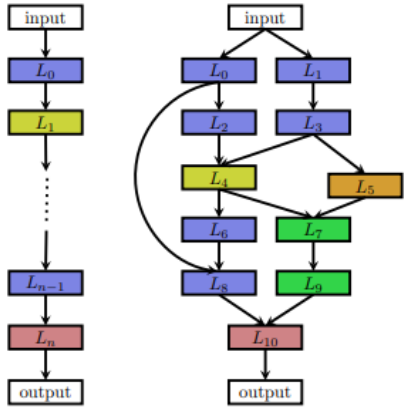


Figure 2: An illustration of different architecture spaces. Each node in the graphs corresponds to a layer in a neural network, e.g., a convolutional or pooling layer. Different layer types are visualized by different colors. An edge from layer L_i to layer L_j denotes that L_j receives the output of L_i as input. Left: an element of a chain-structured space. Right: an element of a more complex search space with additional layer types and multiple branches and skip connections.

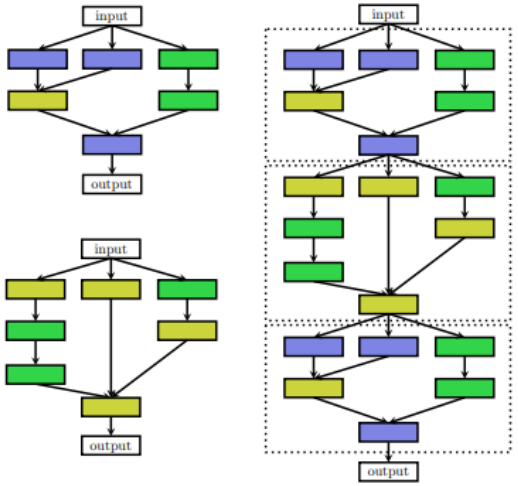
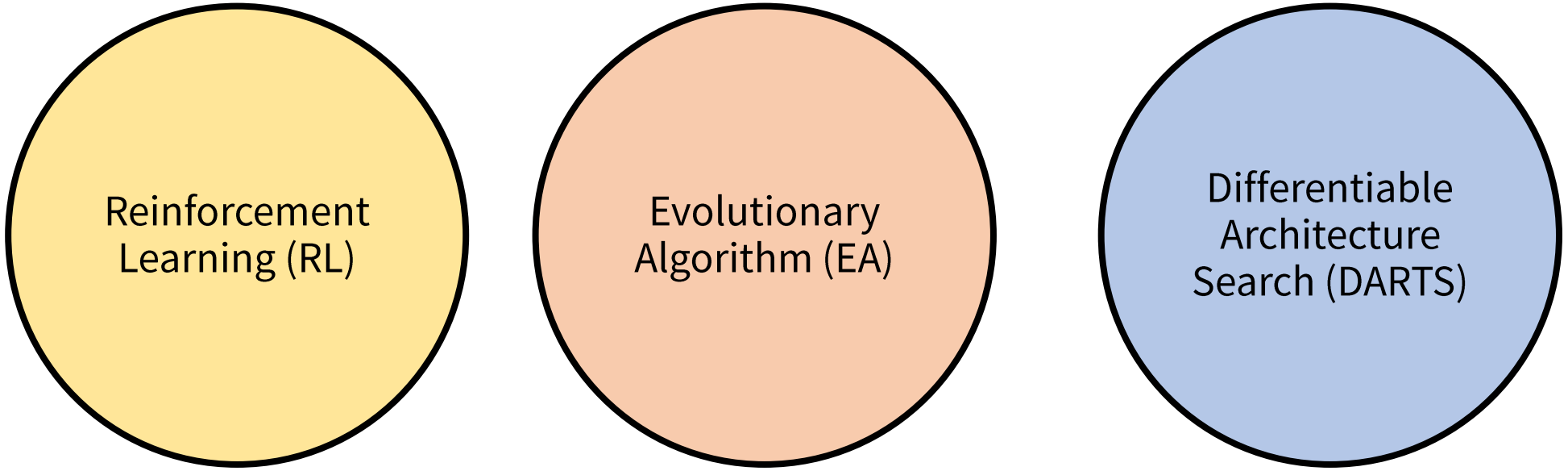


Figure 3: Illustration of the cell search space. Left: Two different cells, e.g., a normal cell (top) and a reduction cell (bottom) (Zoph et al., 2018). Right: an architecture built by stacking the cells sequentially. Note that cells can also be combined in a more complex manner, such as in multi-branch spaces, by simply replacing layers with cells.

서로 다른 Search Space를 가지는 두 Neural Network 구조와 Cell Search Space.

http://www.jmlr.org/papers/volume20/18-598/18-598.pdf

Search Strategy



다양한 Search Strategy가 나와있다. 전통적인 EA, 최근에 효시 역할을 한 RL, 현 SOTA인 DARTS까지.

http://www.jmlr.org/papers/volume20/18-598/18-598.pdf

Performance Estimation

Speed-up method	How are speed-ups achieved?	References
Lower fidelity estimates	Training time reduced by training for fewer epochs, on subset of data, downscaled models, downscaled data, ...	Li et al. (2017), Zoph et al. (2018), Zela et al. (2018), Falkner et al. (2018), Real et al. (2019), Runge et al. (2019)
Learning Curve Extrapolation	Training time reduced as performance can be extrapolated after just a few epochs of training.	Swersky et al. (2014), Domhan et al. (2015), Klein et al. (2017a), Baker et al. (2017b)
Weight Inheritance/ Network Morphisms	Instead of training models from scratch, they are warm-started by inheriting weights of, e.g., a parent model.	Real et al. (2017), Elsken et al. (2017), Elsken et al. (2019), Cai et al. (2018a,b)
One-Shot Models/ Weight Sharing	Only the one-shot model needs to be trained; its weights are then shared across different architectures that are just subgraphs of the one-shot model.	Saxena and Verbeek (2016), Pham et al. (2018), Bender et al. (2018), Liu et al. (2019b), Cai et al. (2019), Xie et al. (2019)

Table 1: Overview of different methods for speeding up performance estimation in NAS.

다양한 방법을 이용해 선택된 구조의 성능을 예측한다.

http://www.jmlr.org/papers/volume20/18-598/18-598.pdf