



Machine Learning Intro

- 1. 머신러닝 개요
- 2. AI/머신러닝/딥러닝
- 3. 머신러닝 쓰이는 분야
- 4. 학습분류
- 5. 학습 데이터셋



머신러닝이란



Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

- Arthur Samuel (1959)

명시적으로 프로그램을 작성하지 않고 컴퓨터에 학습할 수 있는 능력을 부여하기 위한 연구 분야



머신러닝이란

Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure by P , improves with experience E

- Tom Mitchell (1998)

어떤 프로그램이 T (Task)라고 하는 작업을 수행하고 P (Performance Measure)라고 하는 성능 측정 결과를 통해 E (Training Experience)라고 하는 경험을 축적, 성능 개선을 한다면 프로그램은 E 라는 경험에서 학습한다고 할 수 있다.

머신러닝이란

"A computer program is said to **learn** from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**" – *T. Michell (1997)*



Example: A program for soccer tactics

T : Win the game

P : Goals

E : (x) Players' movements

(y) Evaluation

1. 머신러닝 개요

- ▶ 머신러닝 = 기계학습
- ▶ 기존의 컴퓨터 공학에서는 프로그래밍 위주
- ▶ 4차 산업혁명 시대에서는 AI 및 머신러닝 위주
- ▶ 머신러닝은 기본적으로 데이터를 이용하는 것
- ▶ Training 데이터로 모델을 학습시키고, 새로운 데이터로 모델의 예측값을 뽑아내는 것
- ▶ 사람의 개입 없이 모델이 스스로 학습한다



그림. 머신러닝이란

1. 머신러닝 개요

▶ 프로그래밍

- ▶ 주어진 코드나 알고리즘 대로 똑같이 수행된다
- ▶ Input, output 보다는 내부적 알고리즘이 중요



▶ 머신러닝

- ▶ 같은 구조의 모델이어도 데이터셋이 다르면 학습된 모델이 달라짐
- ▶ 데이터가 중요. 데이터 중심 & 모델 중심



2. AI/머신러닝/딥러닝

▶ AI > 머신러닝 > 딥러닝

▶ AI란

- ▶ 인공지능은 인간의 지능, 사고, 감각을 재현
- ▶ 인간과 비슷하거나 더 뛰어난 인지능력 및 지능을 가지고 있으며, 특정한 task가 아닌 일반적이고 새로운 task를 알아서 수행함

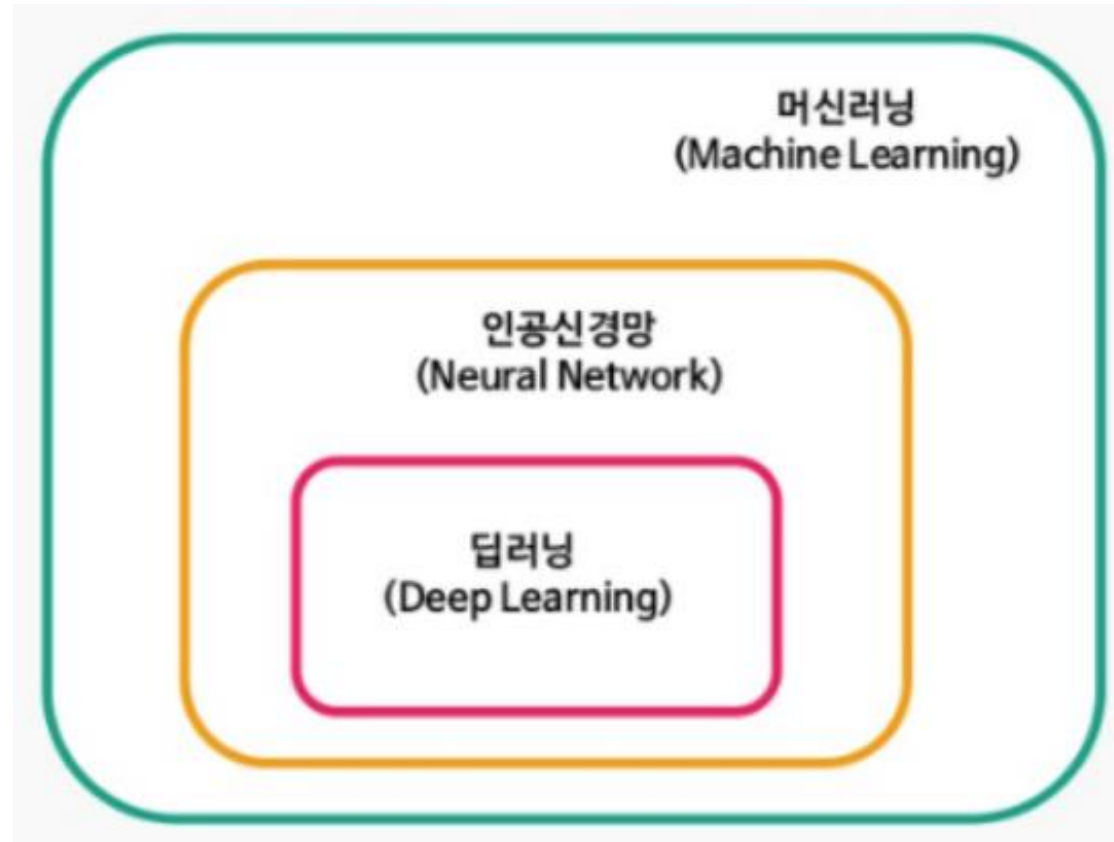
▶ 머신러닝이란

- ▶ 데이터를 이용해 인간 개입 없이 모델을 학습시키고, 새로운 데이터의 결과를 예측함
- ▶ 데이터를 이용하여 데이터의 특성과 패턴을 학습하여 그 결과를 바탕으로 새로운 어떤 데이터가 들어왔을때 그것의 결과값을 예측하는 것.

▶ 딥러닝이란

- ▶ 머신러닝 중 인공 신경망의 층이 아주 깊어지면 딥러닝이라고 함
- ▶ 인공 신경망은 인간 뇌의 뉴런(Neuron)을 흉내낸 것임

2. AI/머신러닝/딥러닝



3. 머신러닝 쓰이는 분야



인공지능 스피커



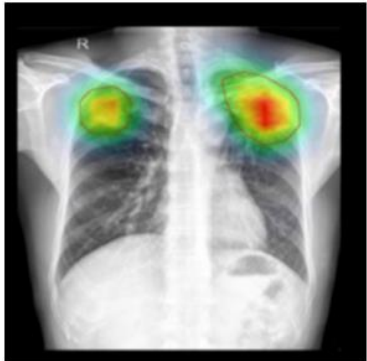
검색 엔진



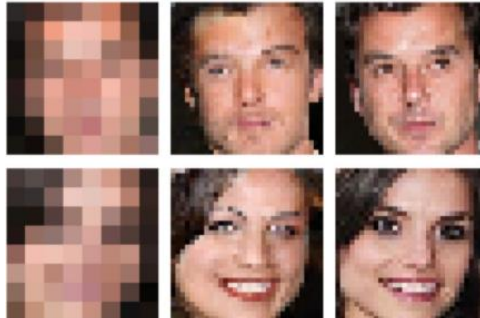
기계 번역



자율 주행



의료



이미지 복원



추천 엔진



로보틱스

4. 머신러닝 학습분류

- ▶ 지도학습 (Supervised Learning)

- ▶ 데이터를 학습시킬 때 답을 알려주는 것

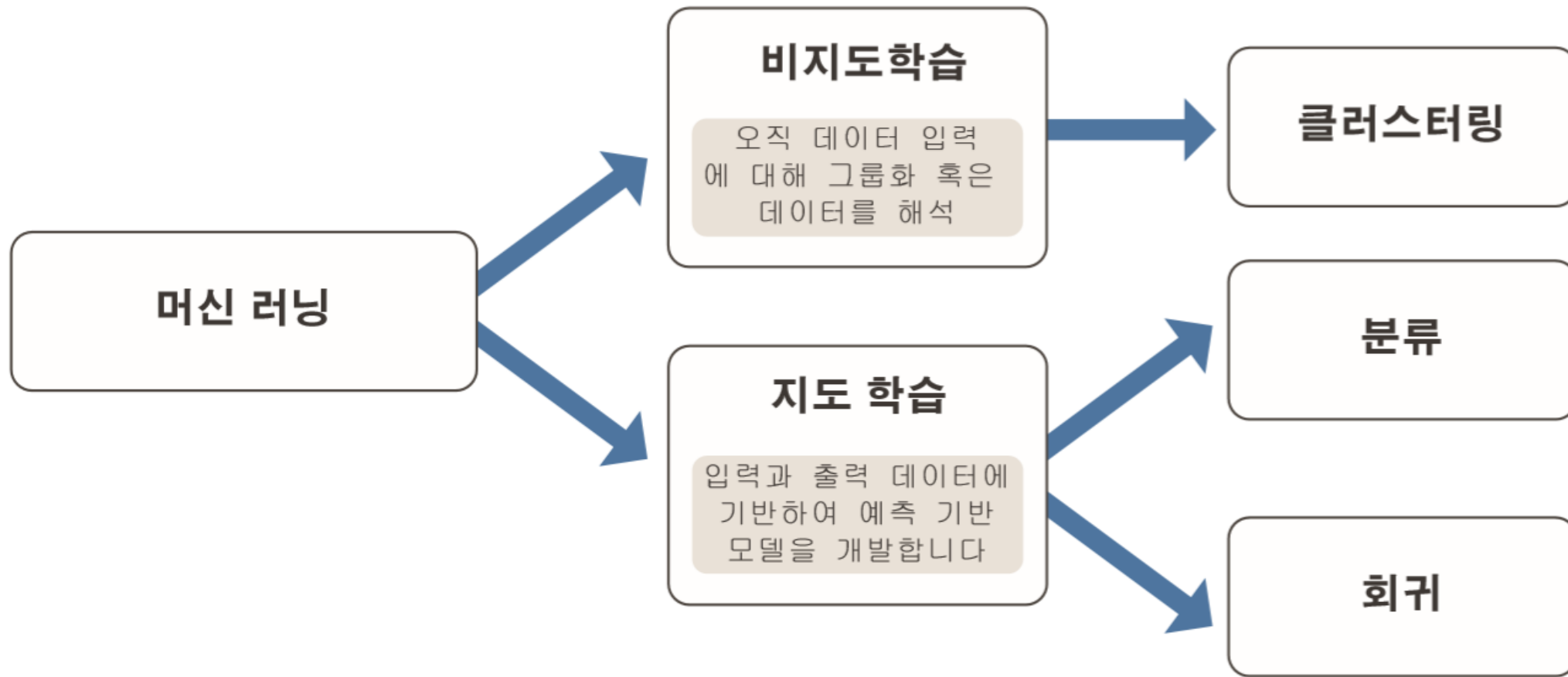
- ▶ 비지도학습 (Unsupervised Learning)

- ▶ 데이터를 학습시킬 때 답을 알려주지 않고, 모델이 스스로 데이터의 특성에 따라 유의미한 결과를 뽑아내도록 하는 것

- ▶ 강화학습 (Reinforcement Learning)

- ▶ Agent가 현재 상태에서 선택 가능한 행동들 중 보상을 최대화하는 방식으로 학습
- ▶ 모델이 여러 번 시행착오를 겪으며 보상이 최대화되는 행동을 하도록 학습함
- ▶ 인간이 어렸을 때 가정교육을 받는 방식과 비슷함

4. 머신러닝 학습분류



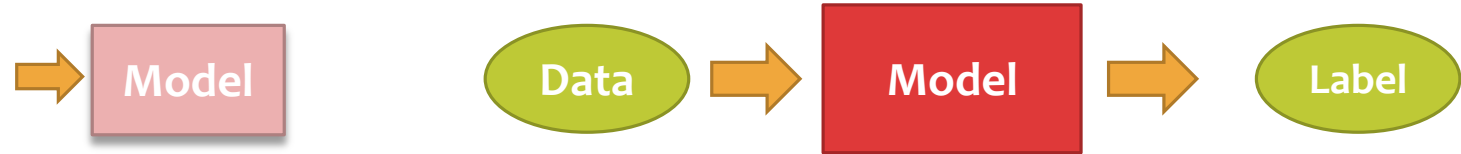
4. 머신러닝 학습분류 - 지도학습



4. 머신러닝 학습분류 - 비지도학습

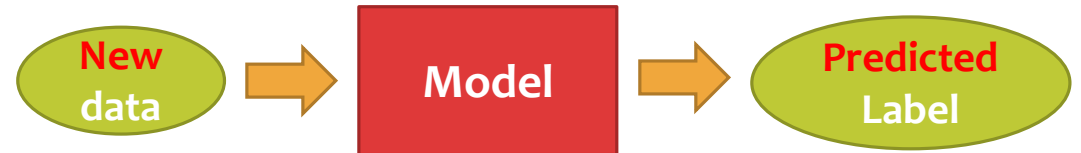


▶ Training(학습)



- ▶ 라벨이 붙어있는 데이터를 이용하여 모델을 학습시킴
- ▶ 이때 모델은 N개의 클래스로 분류하는 모델 (ex: 고양이, 자동차, 트럭, 비행기 등)

▶ Inference(추론, 예측)



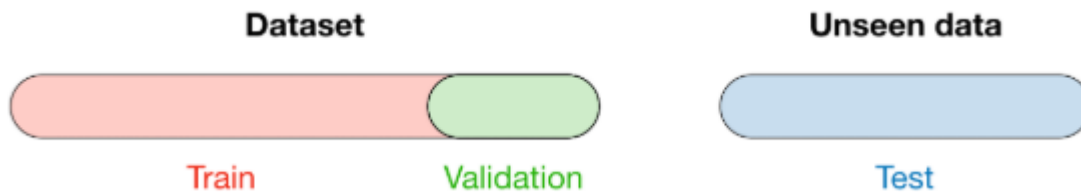
- ▶ 학습 때 사용하지 않은 새로운 데이터를 모델에 넣고, 결과값을 예측하는 과정
- ▶ 예측한 것 중 얼마나 맞췄는지에 따라 그 모델의 성능을 평가함

▶ Training error (학습오차)

- ▶ 라벨이 붙어있지만 모든 데이터를 다 perfect하게 맞추는 모델은 아님

▶ Test error (예측오차)

- ▶ 새로운 데이터를 넣어서 결과를 예측했을 때의 오차
- ▶ 예측오차가 작으면, 일반화가 잘 되는 좋은 모델임



학습데이터에서 측정한 error를 우리는 training error라고 합니다. 관측하지 못한 새로운 input에 대한 error의 기댓값(expectation)을 우리는 generalization error=test error라고 부릅니다. generalization error를 최소화하는 것이 머신러닝의 궁극적인 목적이라고 할 수 있습니다.

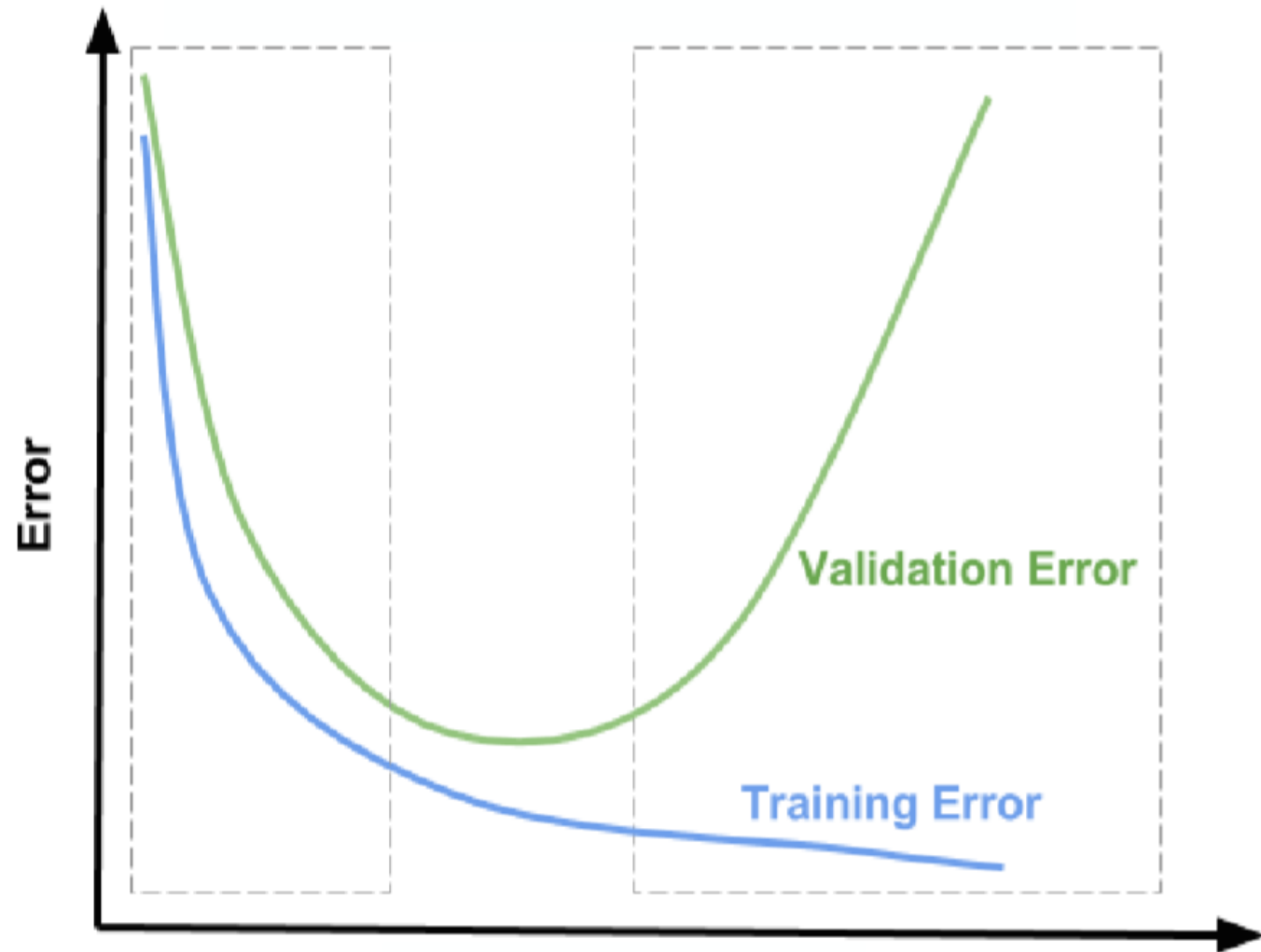


train vs test error

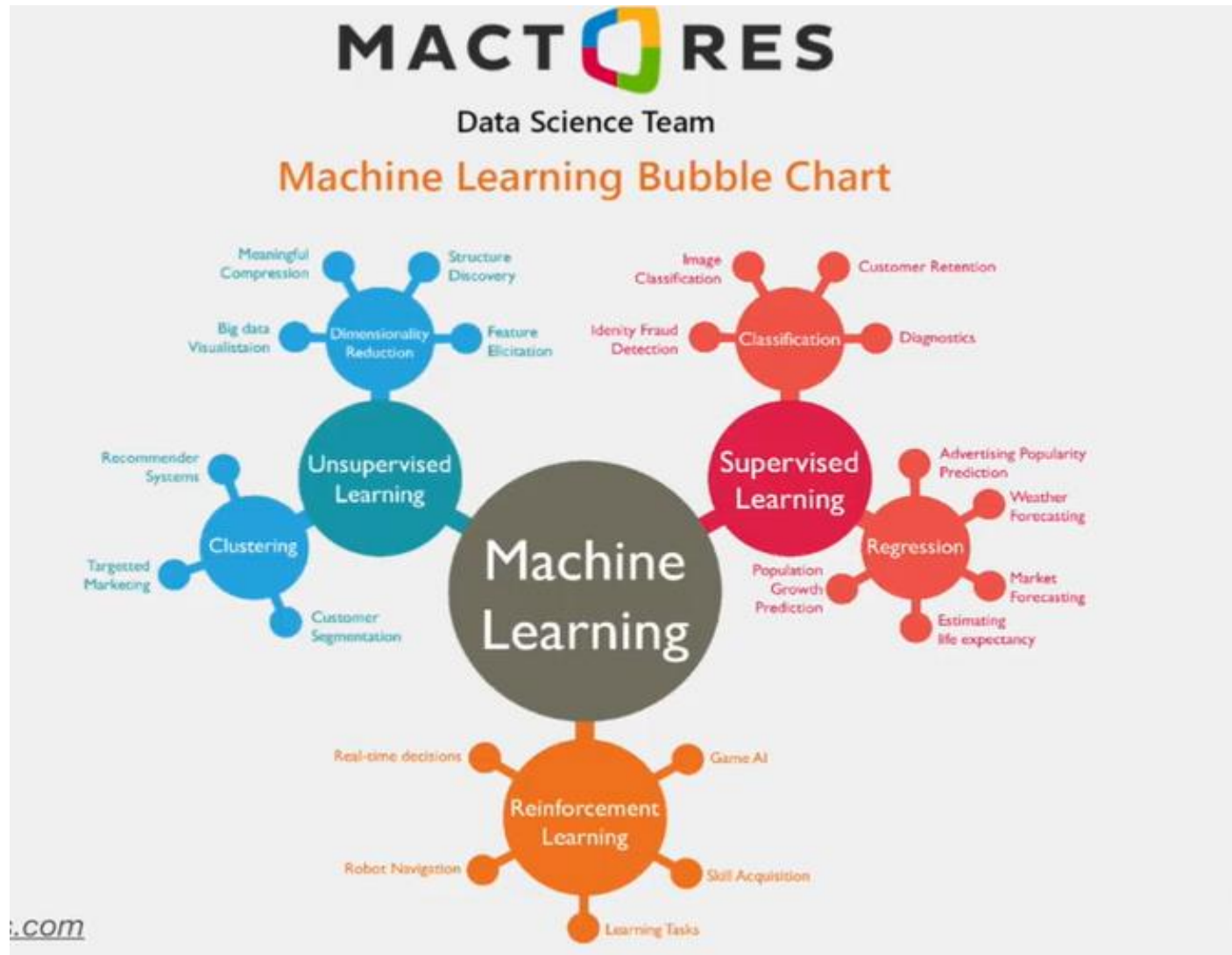


Underfitting
=(high train/test error)

Overfitting
=(low train/ high test error)



4. 머신러닝 학습분류



4. 머신러닝 학습분류

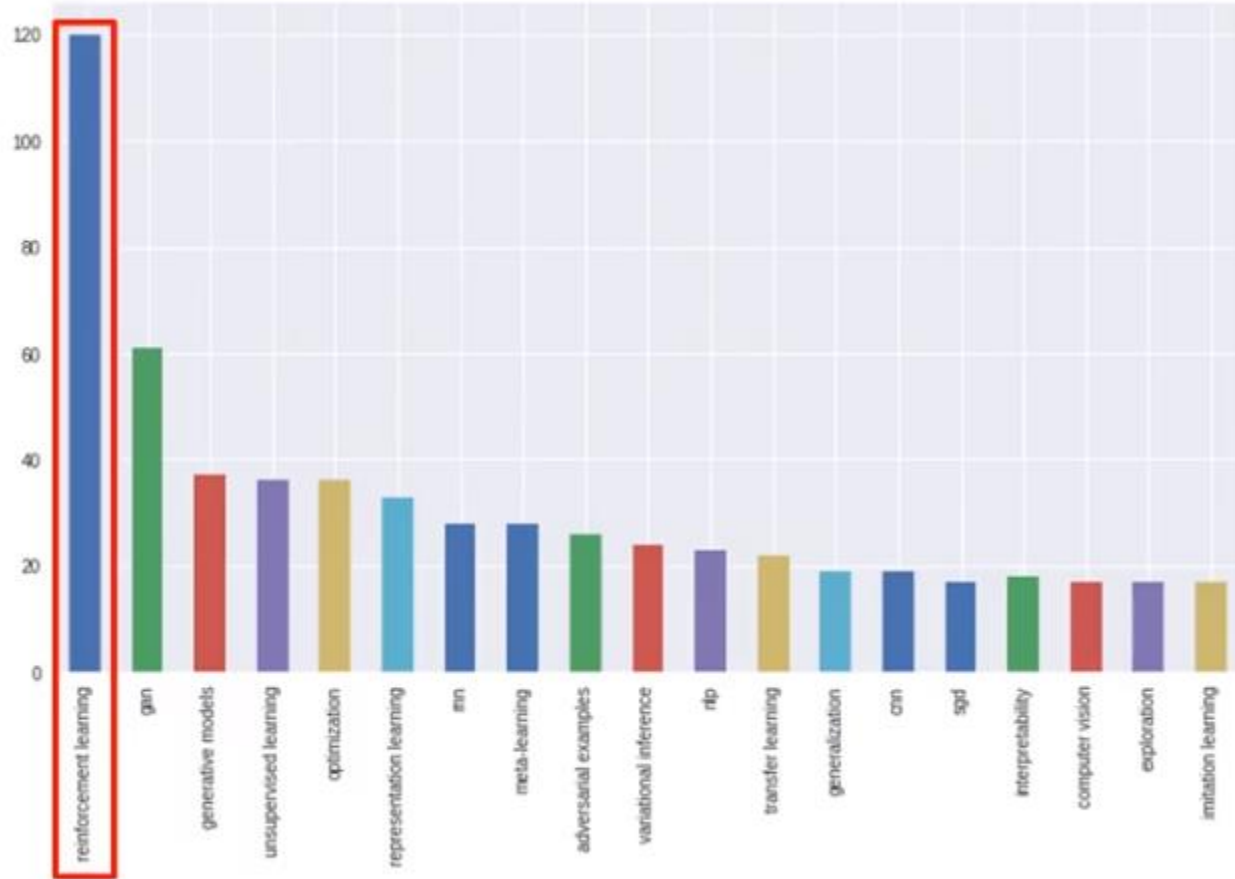
Unsupervised

- Clustering & Dimensionality Reduction
 - SVD
 - PCA
 - K-means
- Association Analysis
 - Apriori
 - FP-Growth
- Hidden Markov Model

Supervised

- Regression
 - Linear
 - Polynomial
- Decision Trees
- Random Forests
- Classification
 - KNN
 - Trees
 - Logistic Regression
 - Naive-Bayes
 - SVM

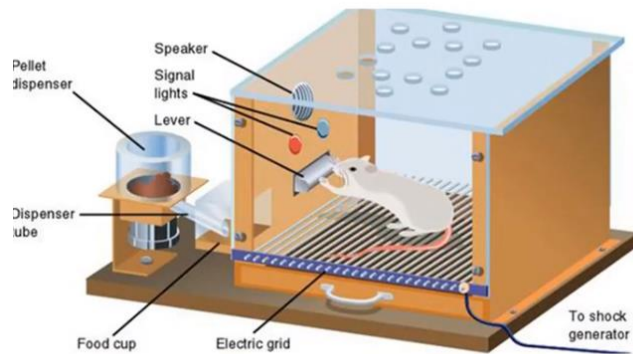
4. 머신러닝 학습분류 - 강화학습



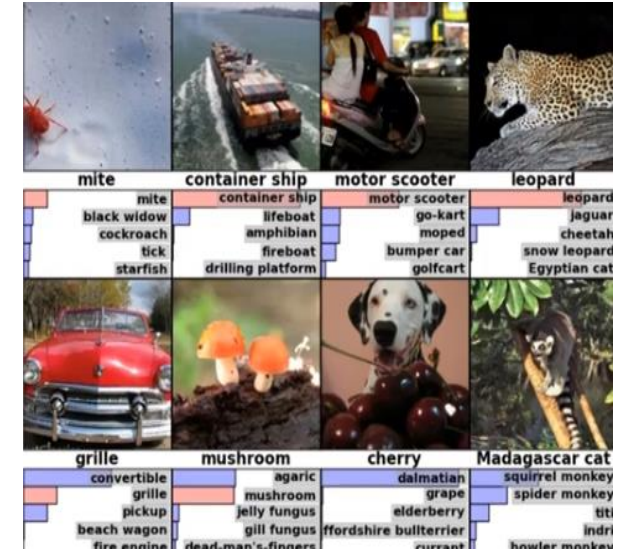
최근 딥러닝의 대표적인 학회는 ICLR에서 제출한 2019년 논문에 의하면 여러 학습분류들 중에서 강화학습이 다른 분야들을 제치고 압도적인 1위를 차지했습니다.

4. 머신러닝 학습분류 - 강화학습

머신러닝 분야에서 가장 대표적인 학습은 지도학습이지만 데이터 분류를 하기 위한 라벨을 만들어야 한다는 문제가 있습니다.
반면에 강화학습은 라벨이 필요 없습니다.
이런 발전 가능성 때문에 많은 관심을 받고 있습니다.



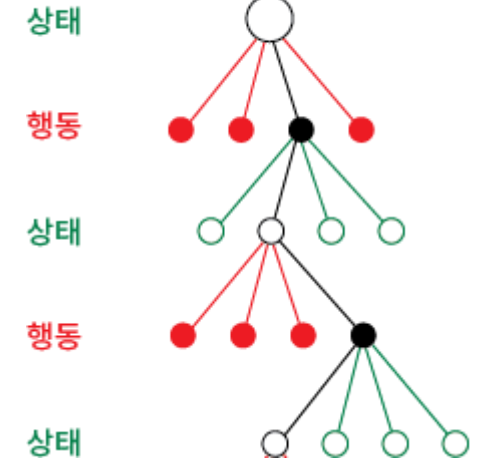
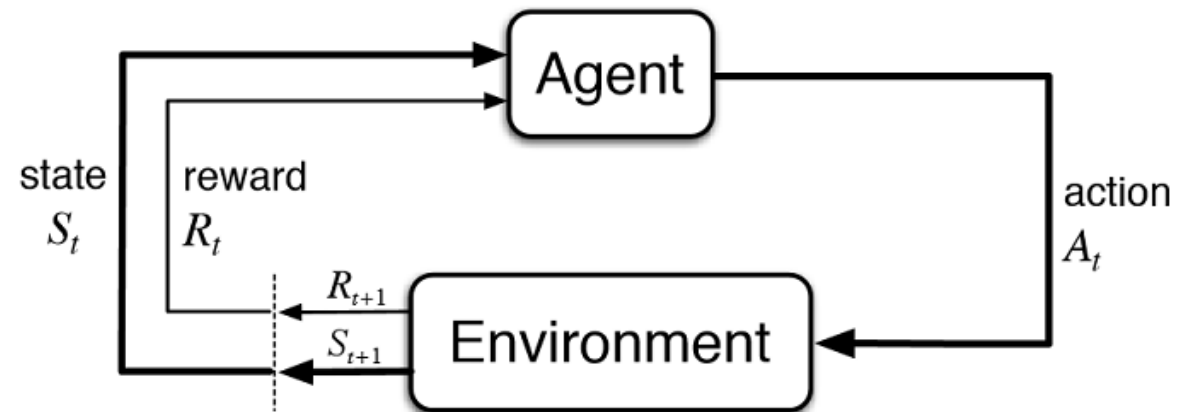
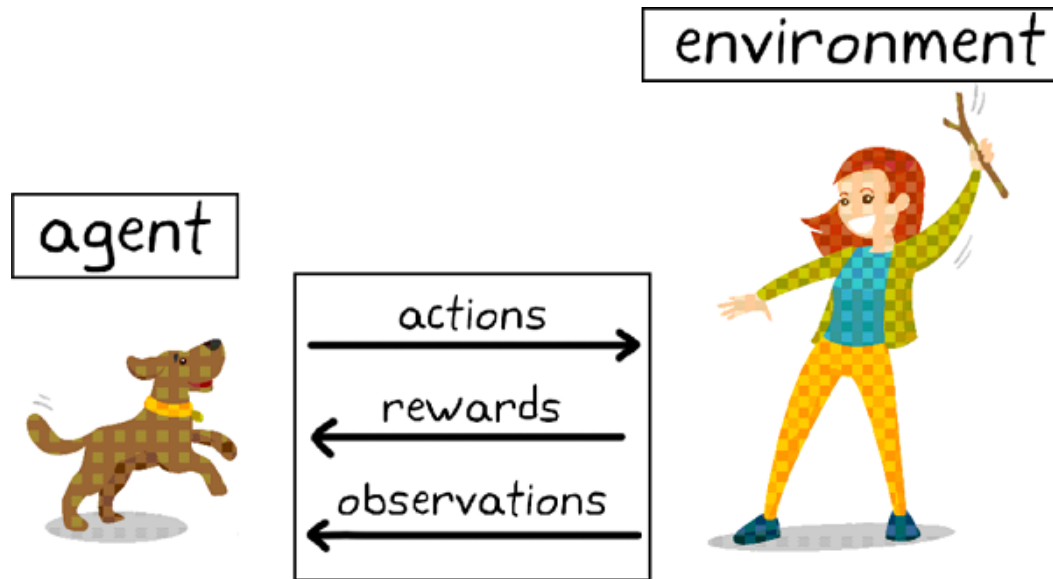
V=0.81	V=0.9	V=1	
V=0.73		V=0.9	
	V=0.73	V=0.81	V=0.73



4. 머신러닝 학습분류 - 강화학습

▶ 강화학습 (Reinforcement Learning)

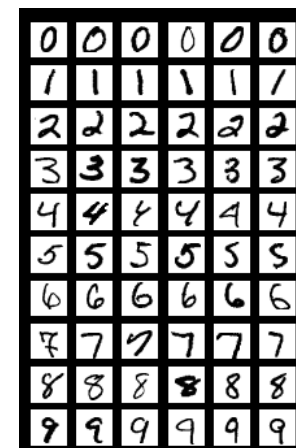
- ▶ Agent가 현재 상태에서 선택 가능한 행동들 중 보상을 최대화하는 방식으로 학습
- ▶ 모델이 여러 번 시행착오를 겪으며 보상이 최대화되는 행동을 하도록 학습함



5. 파이썬 기본모듈 – 유명한 데이터셋

▶ MNIST

- ▶ 0~9까지의 흑백 손글씨 이미지 데이터셋
- ▶ 각 데이터는 (28*28)의 사이즈의 이미지
- ▶ 6만 개의 training 데이터/ 1만 개의 test 데이터
- ▶ Fashion-MNIST: 10개 클래스의 흑백 패션이미지



▶ Cifar-10

- ▶ 머신러닝 및 컴퓨터 비전 쪽에서 널리 사용되는 데이터셋
- ▶ 10개의 서로 다른 클래스의 이미지들
- ▶ 총 6만 개의 (32*32) 사이즈의 다양한 이미지

비행기
자동차
새
고양이
사슴
개
개구리
말
배
트럭

