

Artificial Intelligence

AI, ML, DNN and XAI

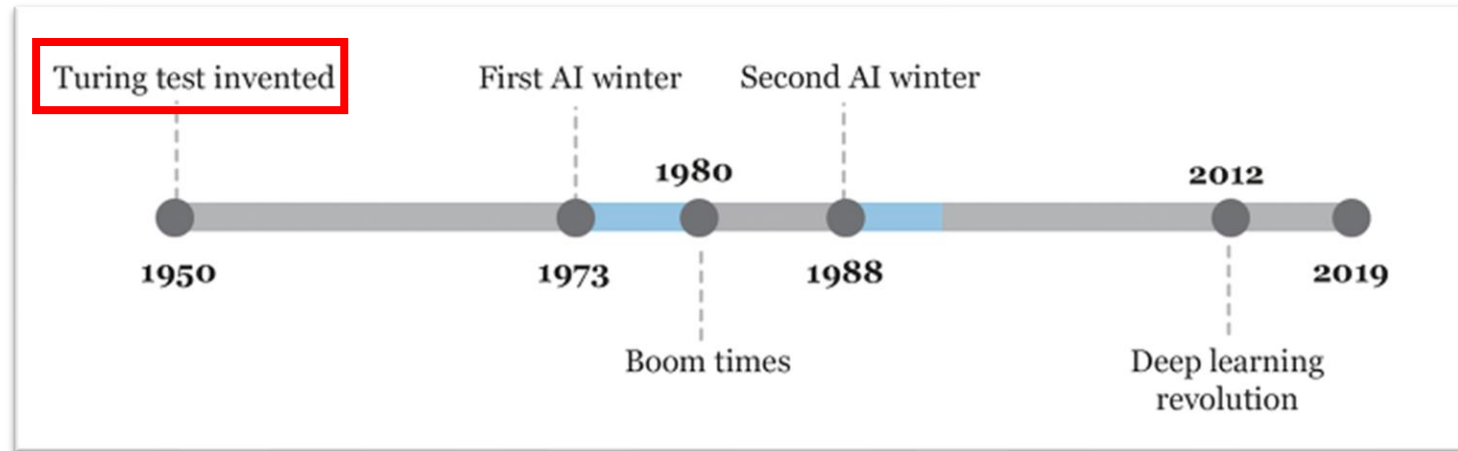
숭실대학교 컴퓨터학부

20182613 박상준

Contents

- History of AI
- AI, ML, DL
- Machine Learning
- Deep Neural Network
- XAI, black-box

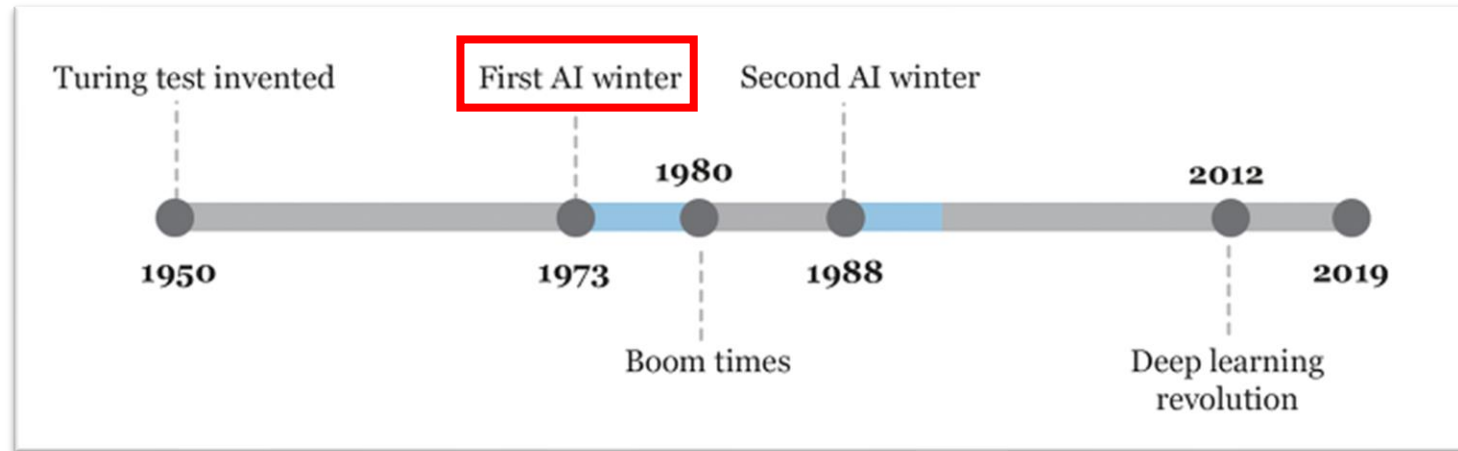
History of AI



Turing test:

기계가 인간과 얼마나 비슷하게 대화할 수 있는지를
기준으로 기계의 지능을 판별하는 test 기법.

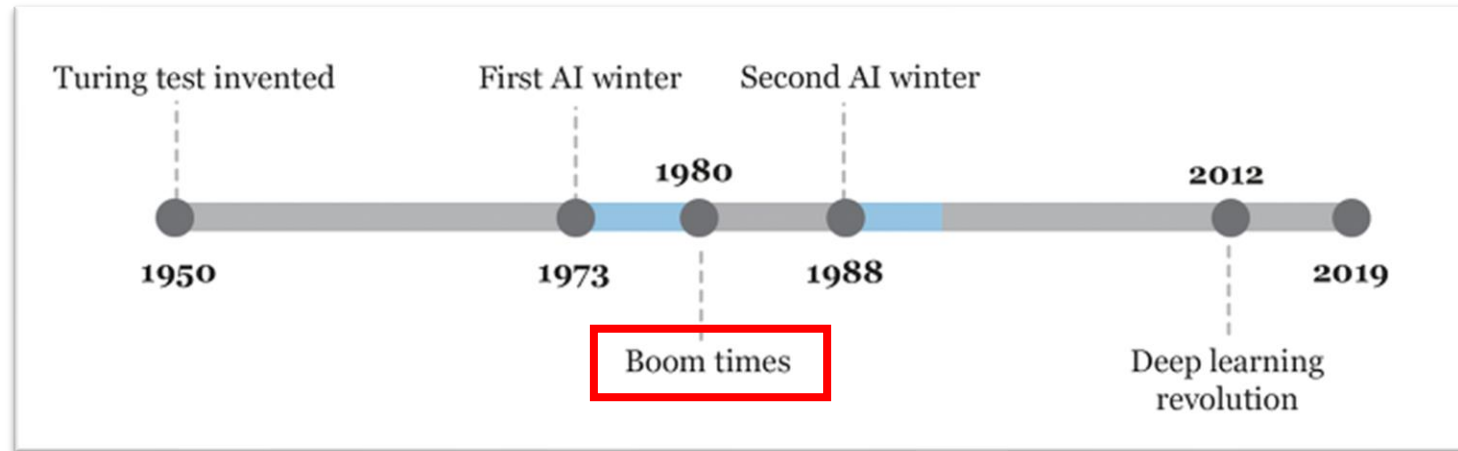
History of AI



First AI winter [첫 번째 암흑기]:

큰 기대를 가져왔지만, 기대만큼의 성과가 없어 투자가 중단됨.

History of AI

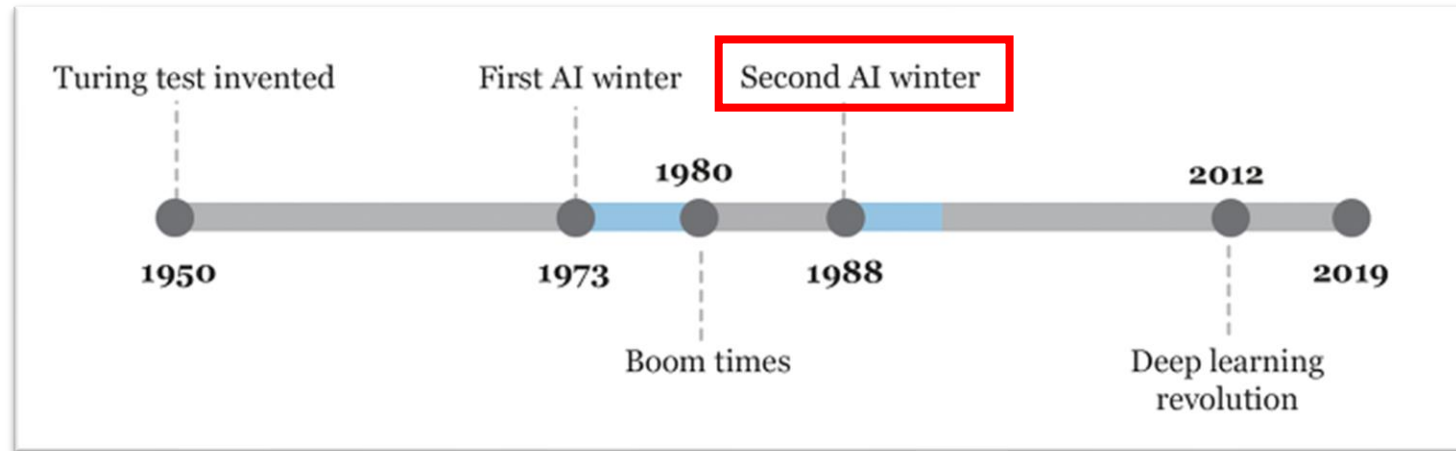


Boom times [발전기]:

Single layer perceptron → Multi layer perceptron

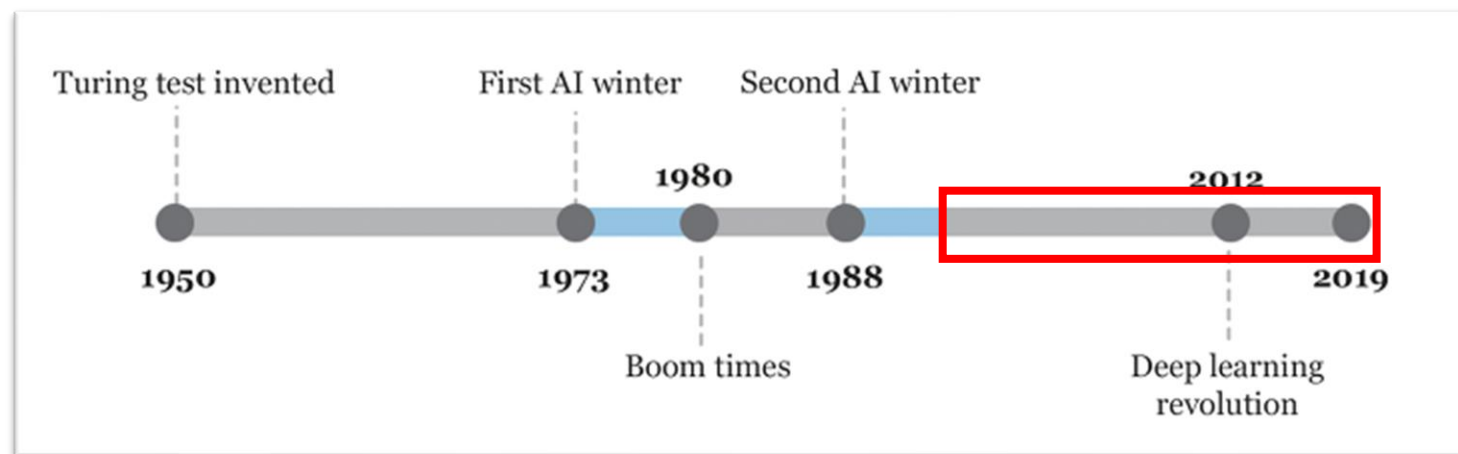
Back propagation 제안.

History of AI



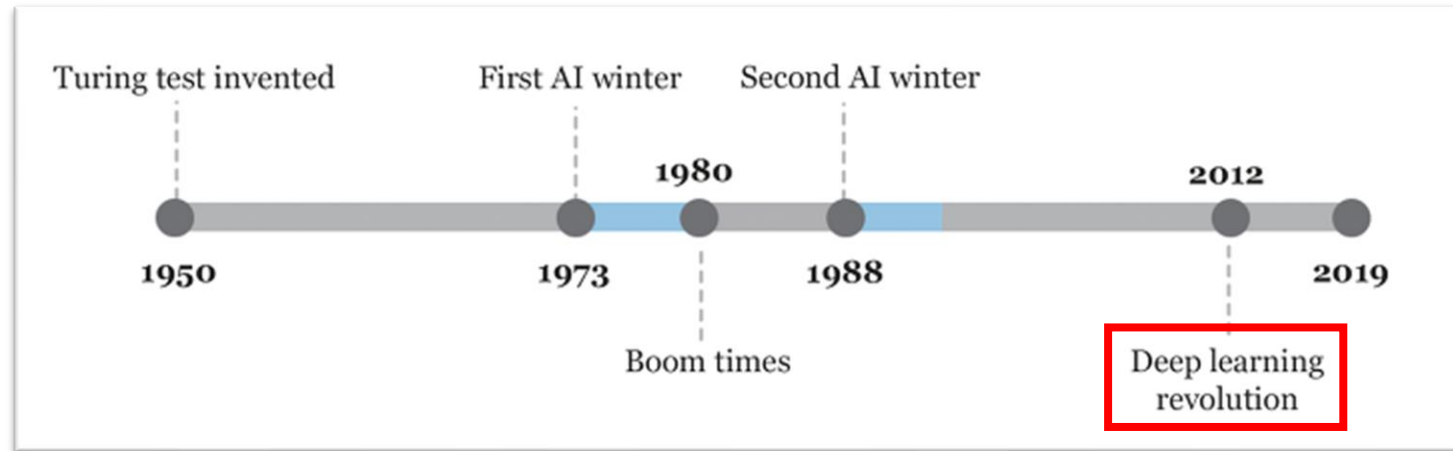
Second AI winter (두 번째 암흑기):
Multi layer perceptron - Too slow (resource)

History of AI



Computing Performance (resource) 의 엄청난 향상

History of AI



Deep learning revolution :

Google - DNN을 구현한 고양이 영상인식(CNN) : 2012

Facebook - DeepFace 얼굴인식 알고리즘: 2014

Google Alpha Go - 이세돌 바둑 대결 : 2016

AI, ML, DL

AI :

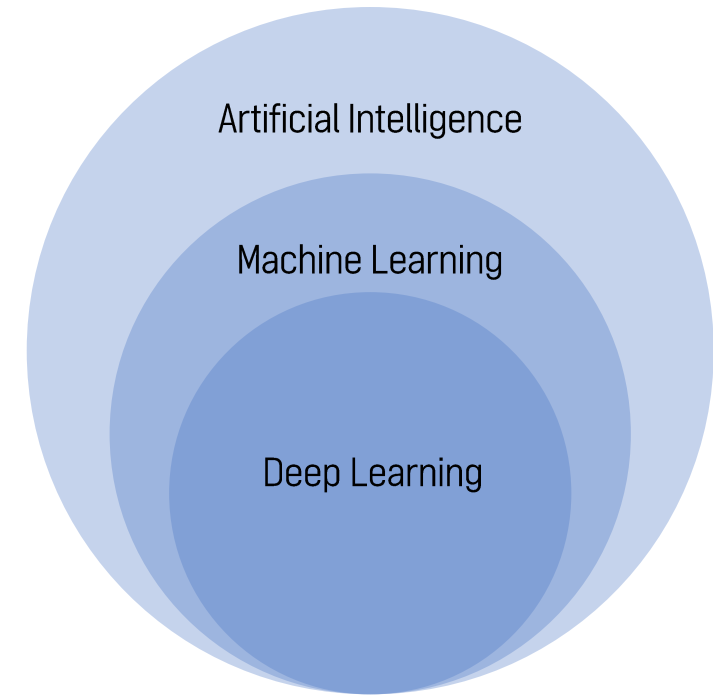
인간의 학습/추론/지각 능력 등 인간이 가진 지적 능력을 구현한 컴퓨터 프로그램 또는 컴퓨터 시스템

ML :

경험을 통해 자동으로 개선하는 컴퓨터 알고리즘의 연구. 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야

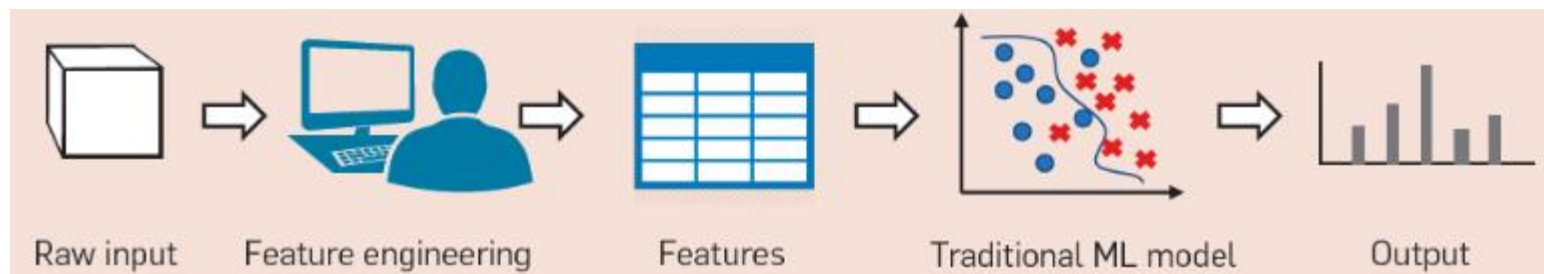
DL :

여러 비선형 변환기법의 조합을 통해, 높은 수준의 추상화를 시도하는 기계학습 알고리즘의 집합.

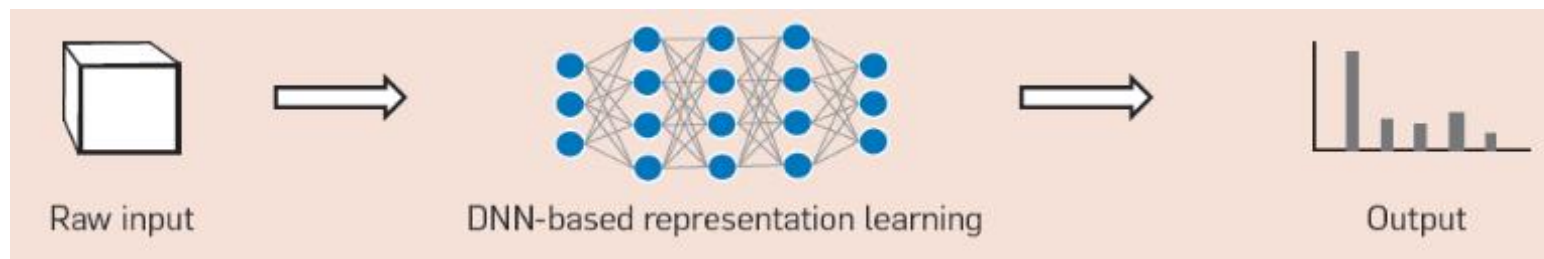


AI, ML, DL

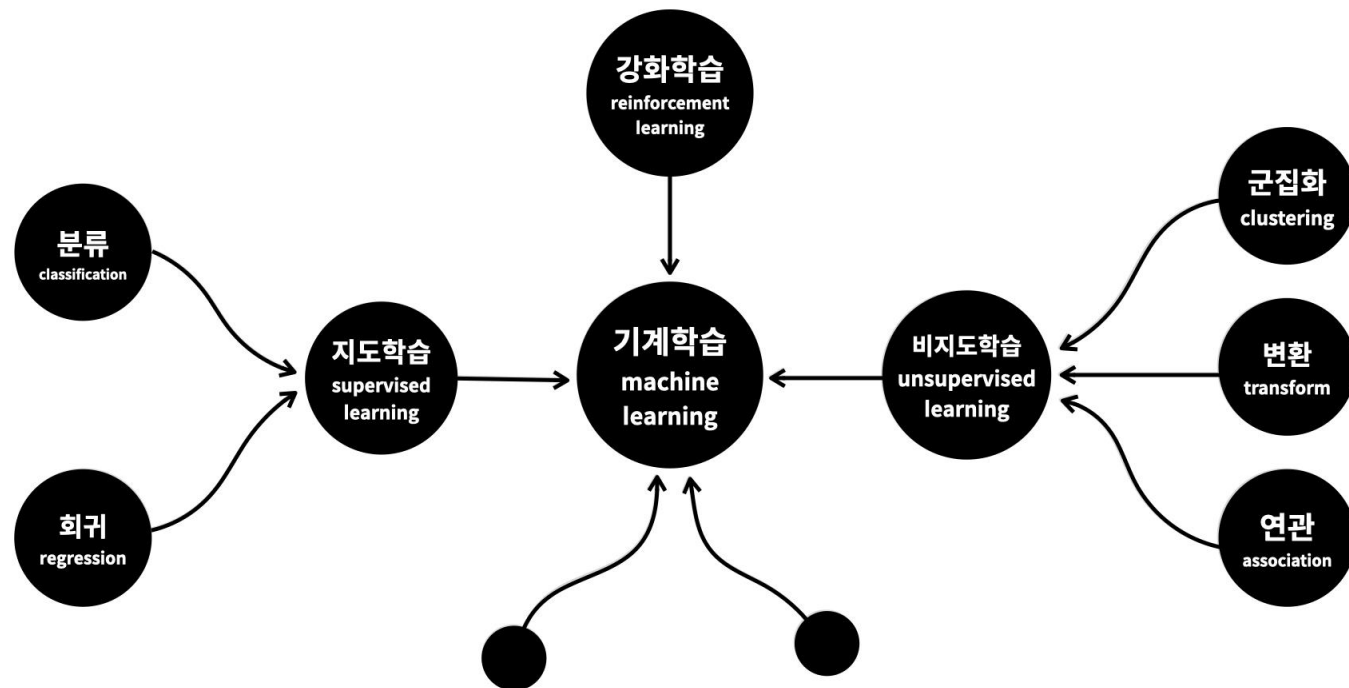
Machine Learning



Deep Learning



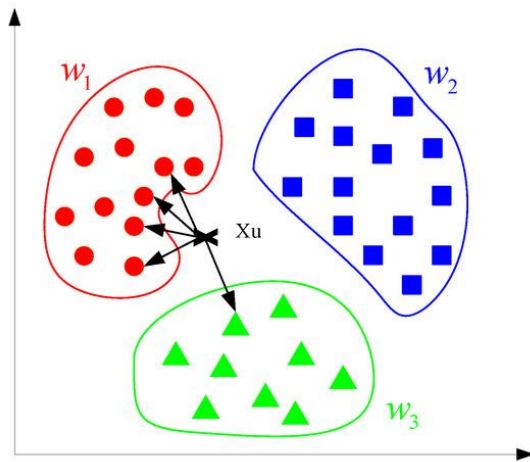
Machine Learning



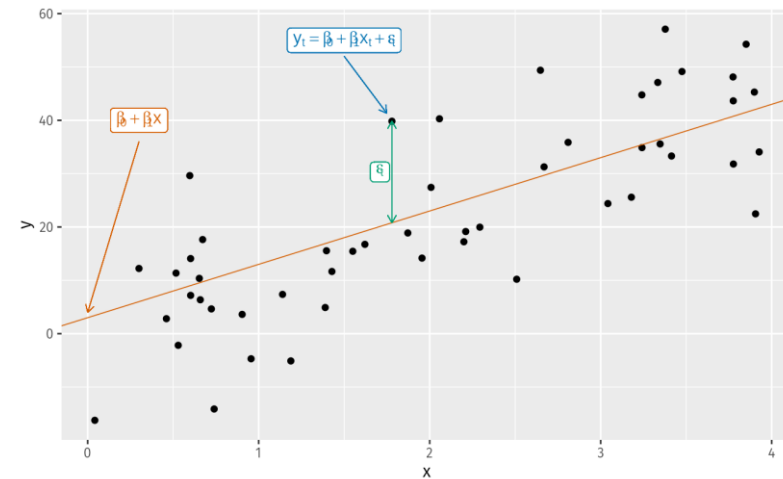
Machine Learning

>> Supervised

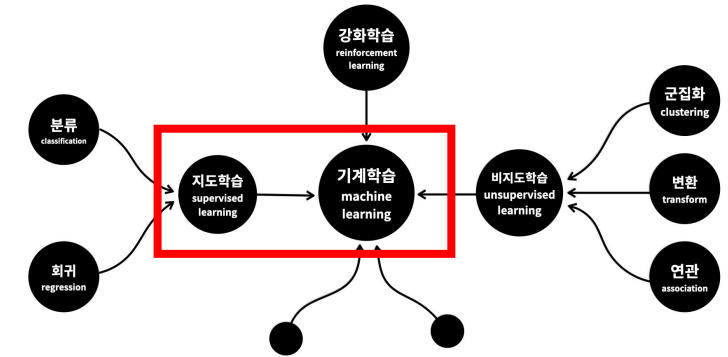
- 정답이 있는 데이터를 통한 학습
Label
- 분류 [Classification] 와 회귀 [Regression]
Discrete Value Continuous Value



KNN (K-Nearest Neighbor)



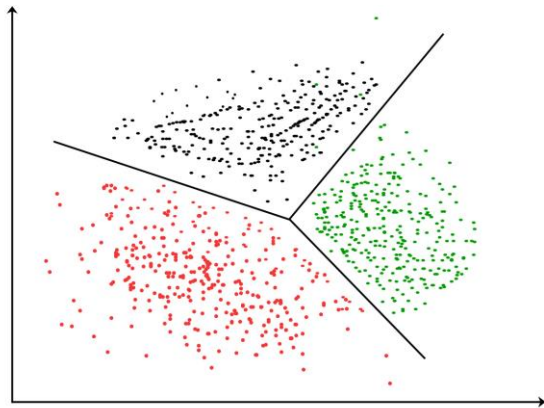
Linear Regression



Machine Learning

>> Unsupervised

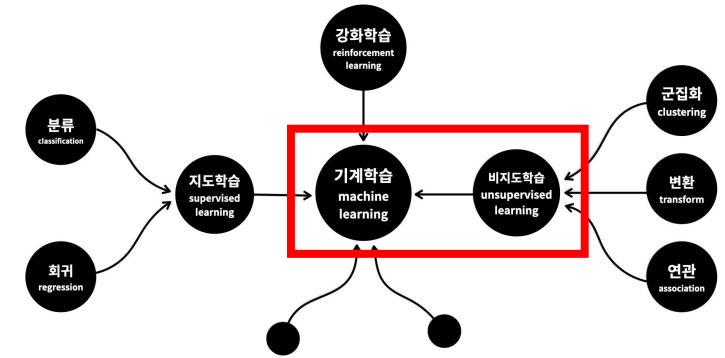
- 정답이 없는 데이터를 통한 학습
- 군집화 (Clustering), 연관 (Association)



Clustering



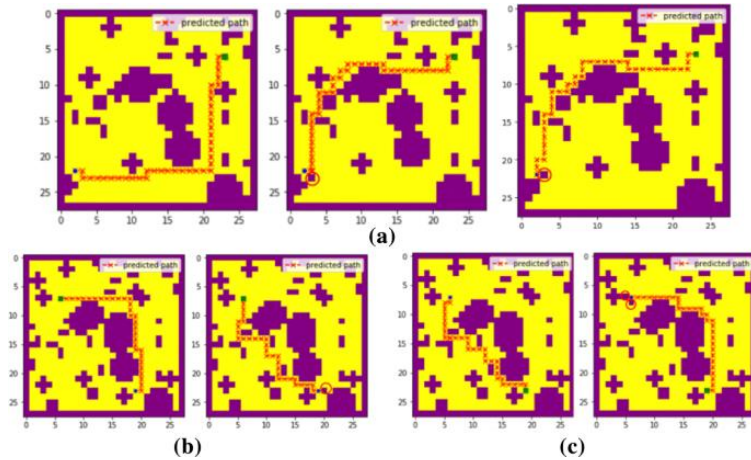
Association



Machine Learning

>> Reinforcement

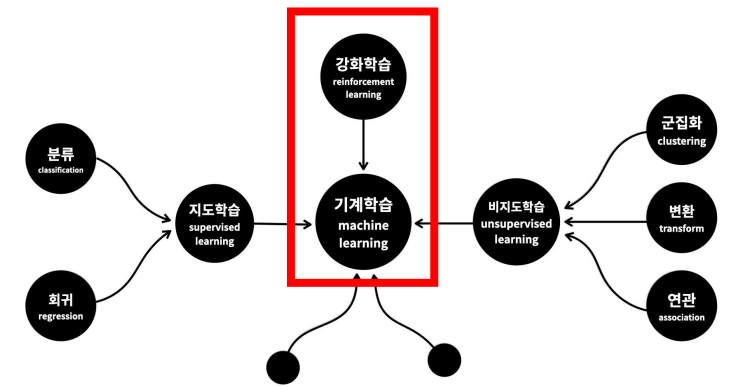
- Reward & Punishment 를 통한 학습
- Delayed Reward



Finding Path



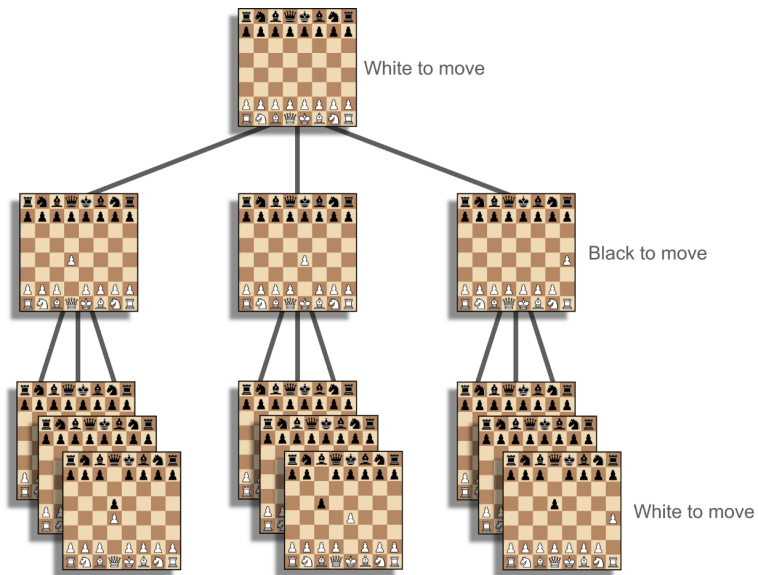
Chess game



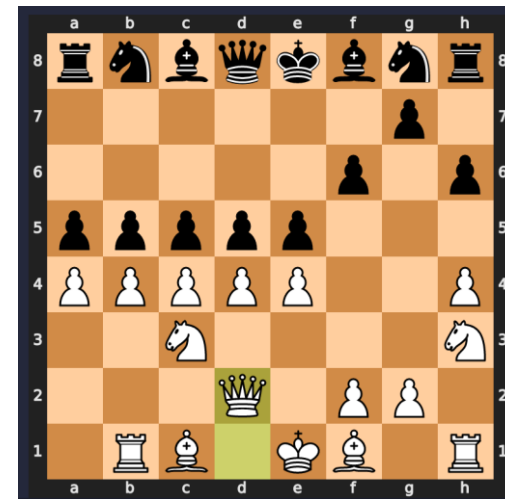
Machine Learning

>> Reinforcement

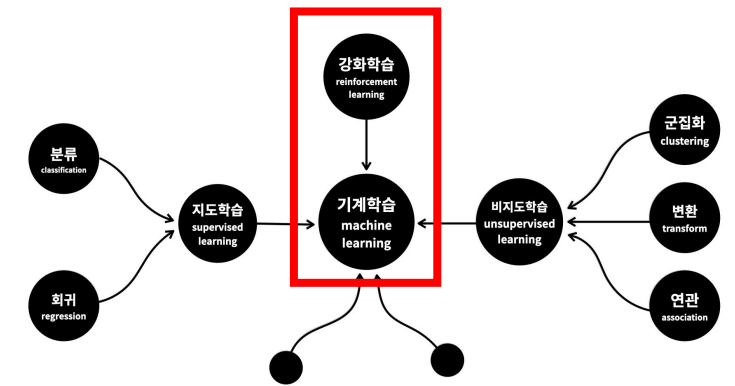
- Reward & Punishment 를 통한 학습
- Delayed Reward



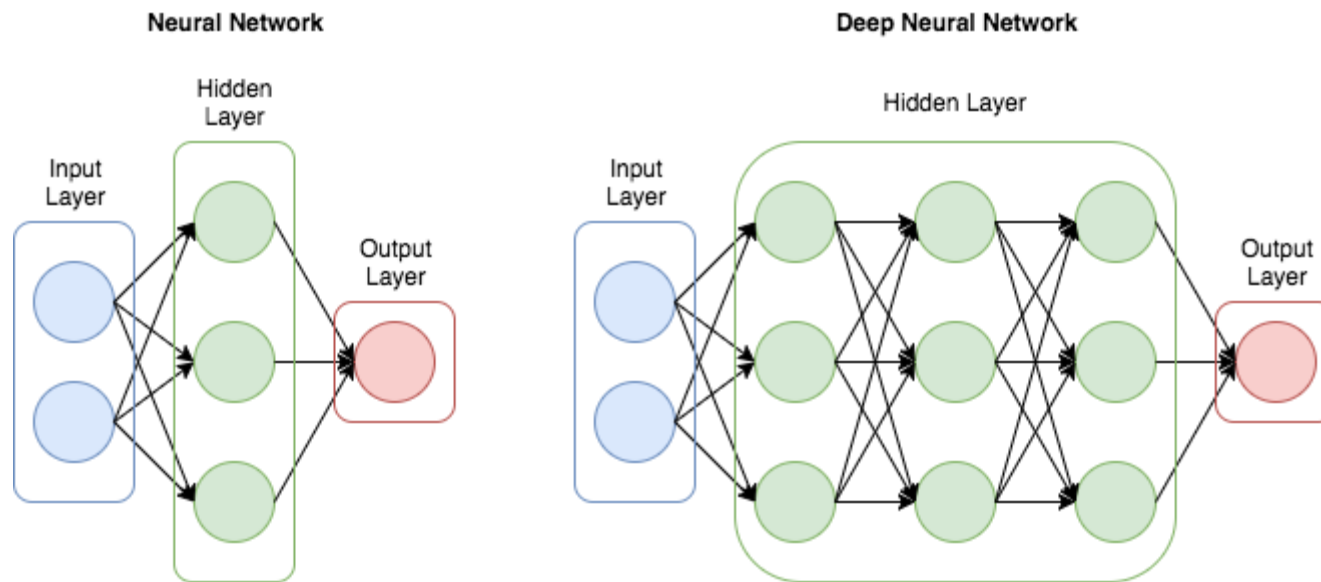
Real world: Optimal policy를 모름.
'Q Learning'



Chess game



Deep Neural Network



모든 Neural Network 기반의 알고리즘의 기본 형태

Deep Neural Network

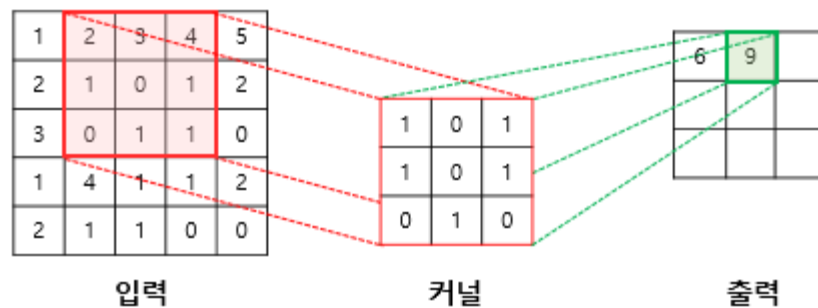
>> CNN

정의 [편집]

두 개의 함수 f 와 g 가 있을 때, 두 함수의 합성곱을 수학 기호로는 $f * g$ 와 같이 표시한다.

합성곱 연산은 두 함수 f, g 가운데 하나의 함수를 뒤집기(reverse), 전이(shift)시킨 다음, 다른 하나의 함수와 곱한 결과들을 적분하는 것을 의미한다. 이를 수학 기호로 나타내면 다음과 같다.

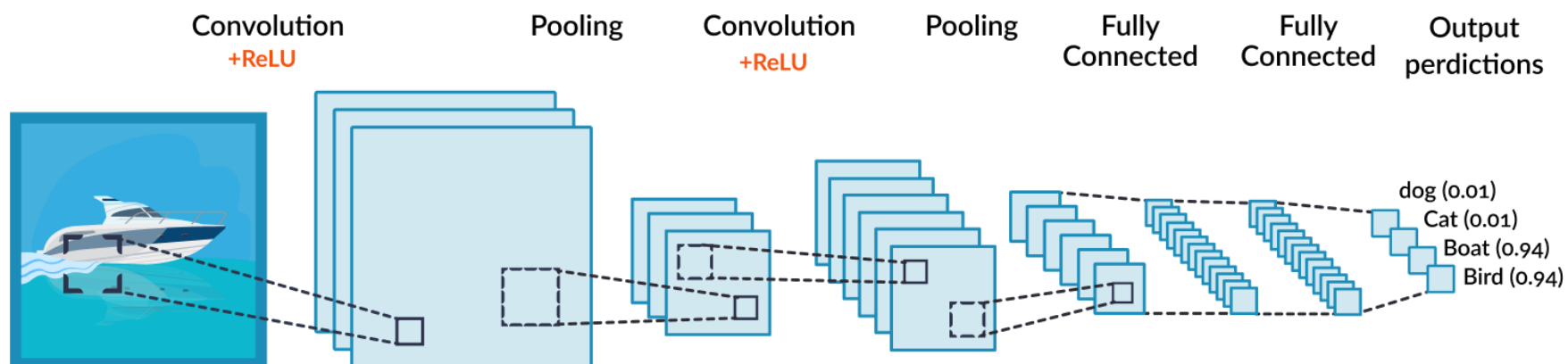
$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau) d\tau$$



Convolution: 합성곱

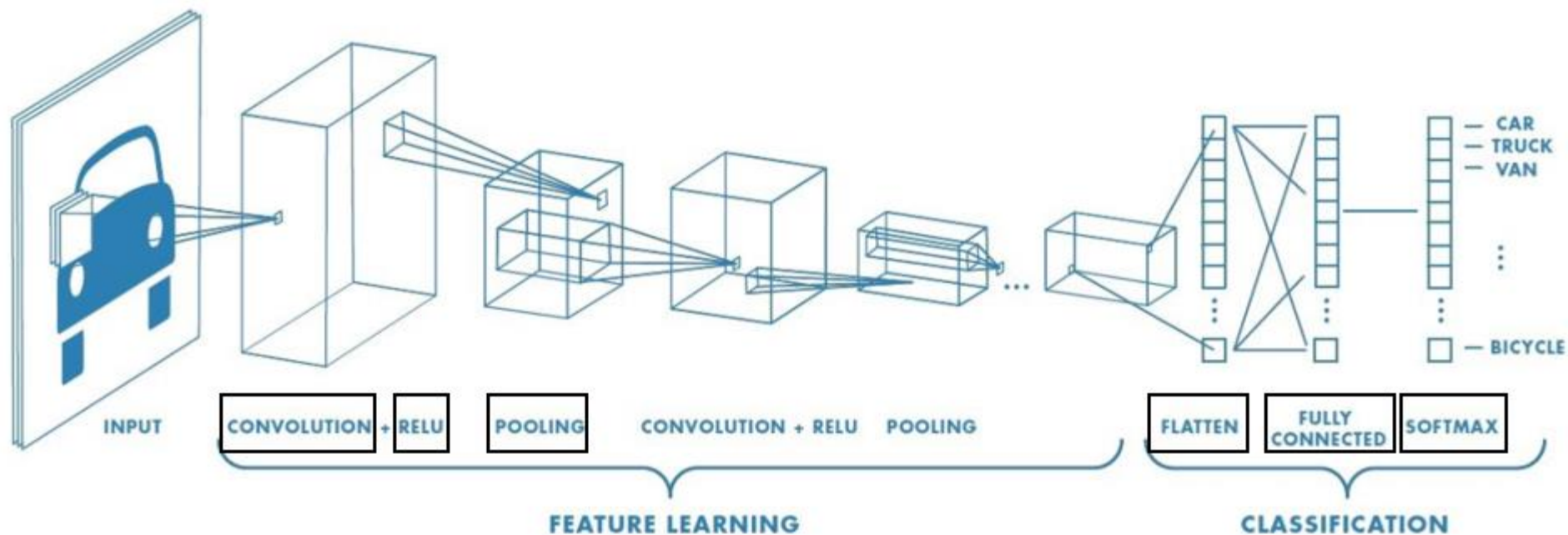
Deep Neural Network

>> CNN



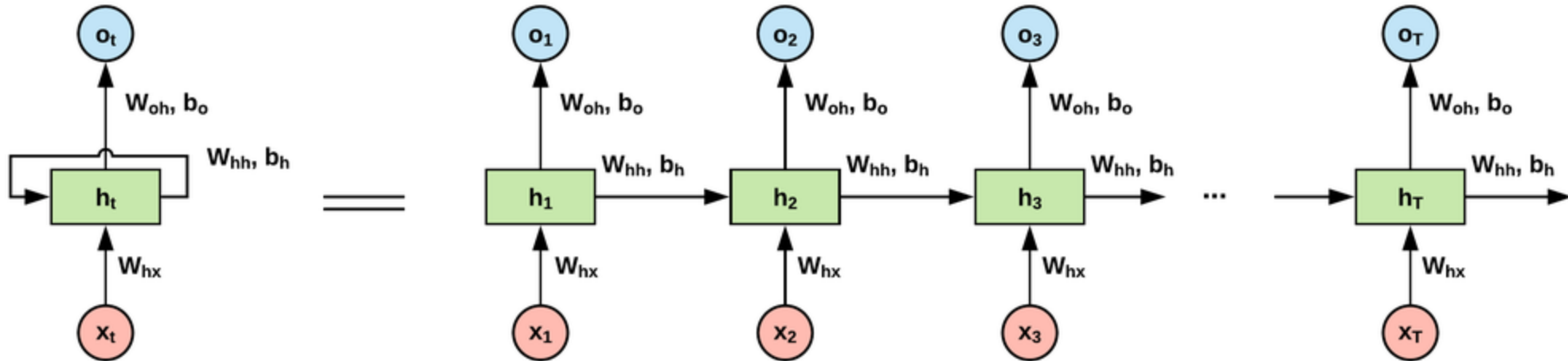
Deep Neural Network

>> CNN



Deep Neural Network

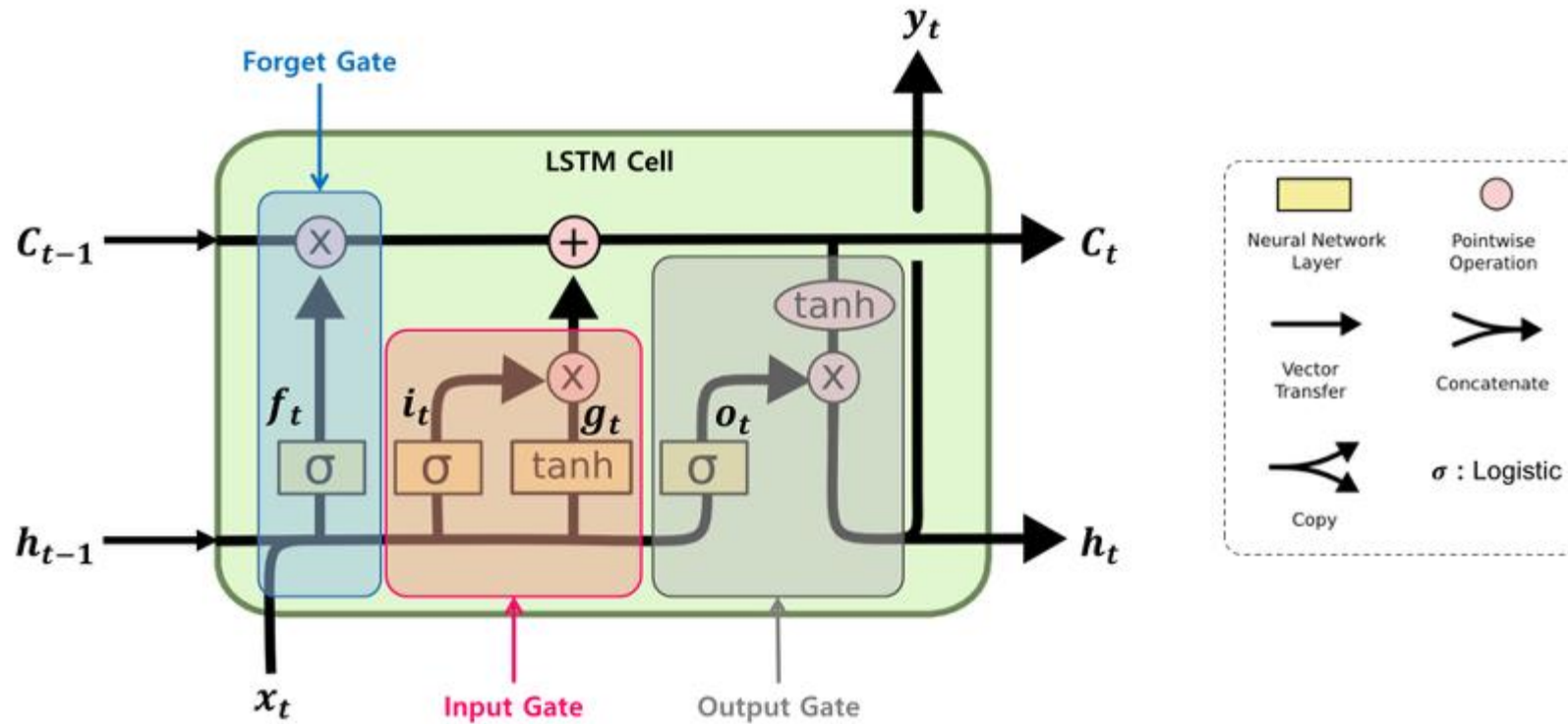
>> RNN



- 순서가 있는 (순서가 중요한) 데이터에 대한 학습 및 예측
- 이전의 값들을 대표하는 값을 현재 상태에서 사용
- Ex) NLP, 주가 예측, etc.

Deep Neural Network

>> LSTM, GRU



- RNN의 문제

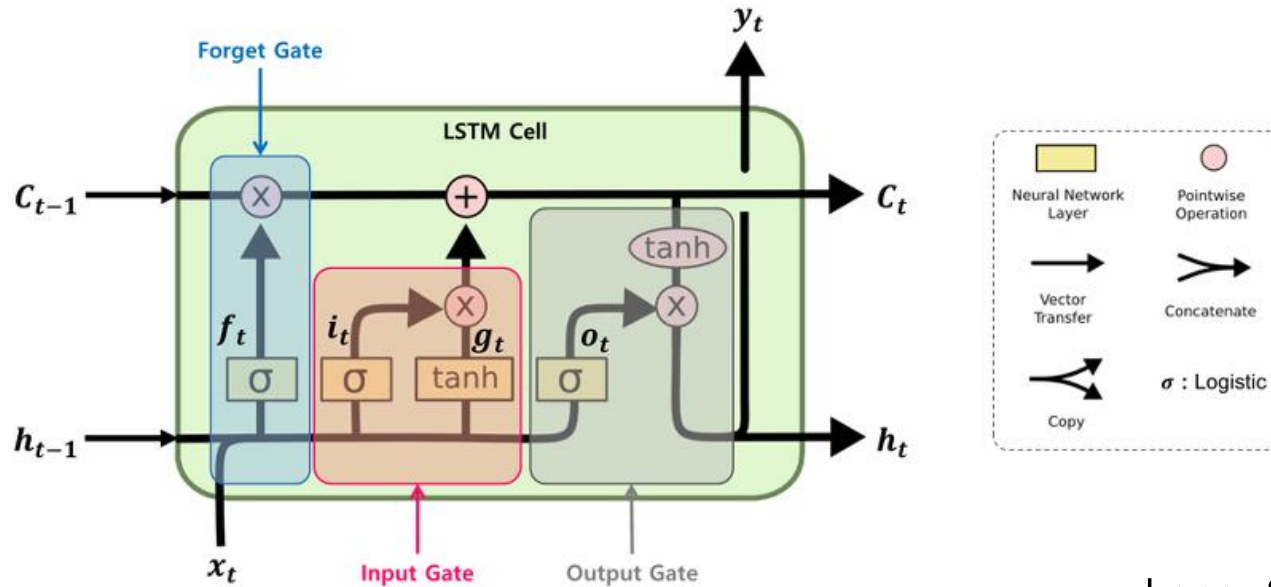
- $Wxx > 1 \rightarrow \text{INF}$: exploding gradient
- $Wxx < 0 \rightarrow 0$: vanishing gradient

- Long Short Term Memory

- Gradient flow을 제어하는 벨브 역할
- 4개의 Gate 추가

Deep Neural Network

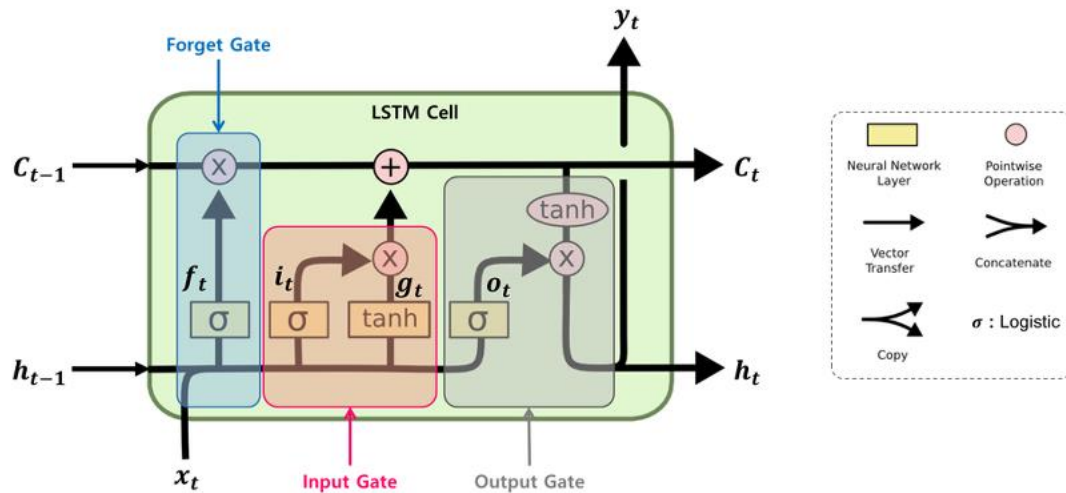
>> LSTM, GRU



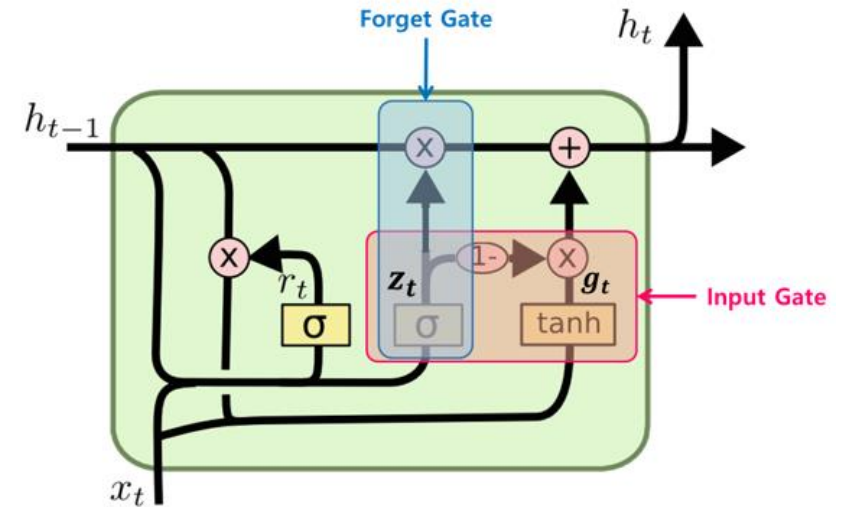
- Long-Short Term Memory
 - 1) 새로운 입력, 이전 상태를 얼마나 사용할 것인지 결정.
 - 2) 얼마나 잊어버릴 것인지 결정.
 - 3) 이 둘을 적절히 섞음.
 - 4) 섞은 정보와 각각의 정보를 통합하여 Output 결정.

Deep Neural Network

>> LSTM, GRU



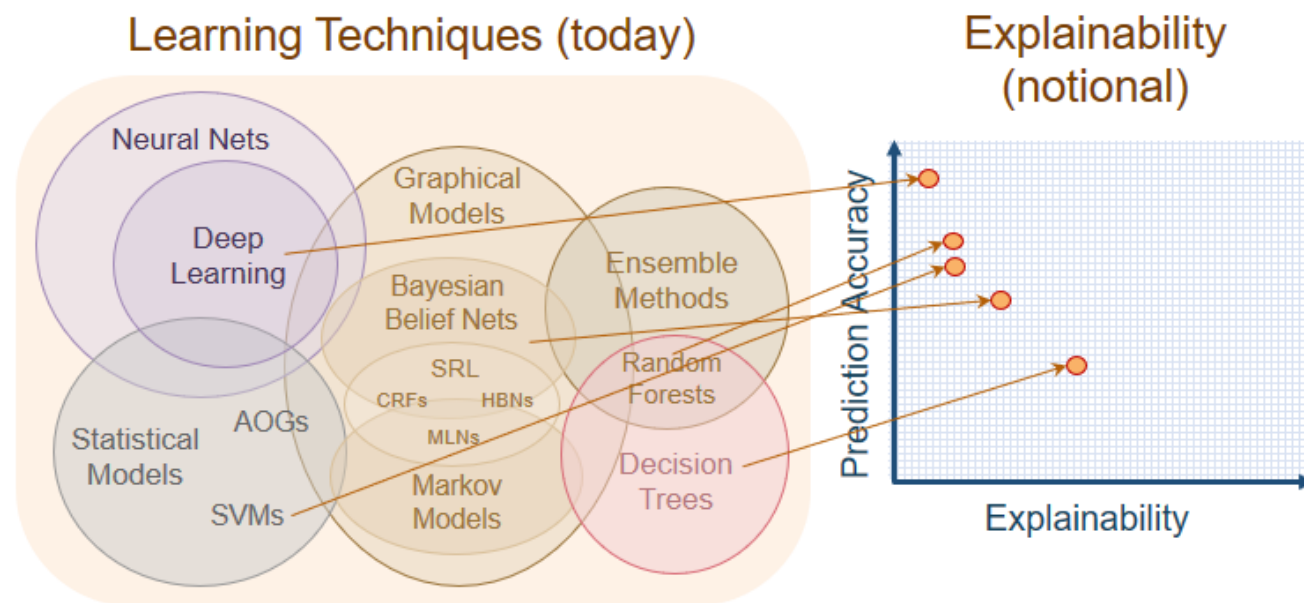
- LSTM: 4개의 gate로 구성



- GRU: Cell State가 없음. 나머지 LSTM과 동일

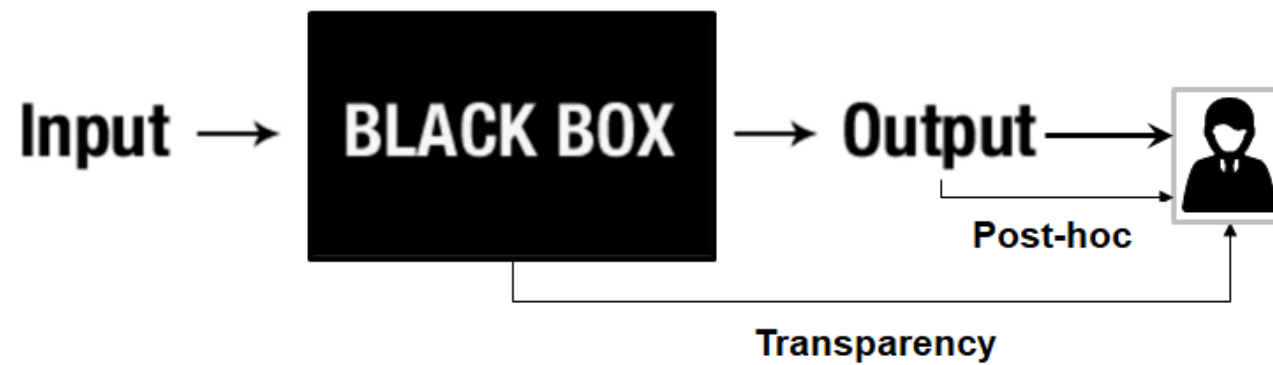
- 둘 중 performance가 좋은 것은 -> ???

XAI, Black-box



- 'Explainable' 한가?
 - 간단할 수록 해석하기 쉽지만, 결과는 좋지 않음.
 - (usually) 복잡할 수록 해석하기 어렵지만, 결과가 좋음

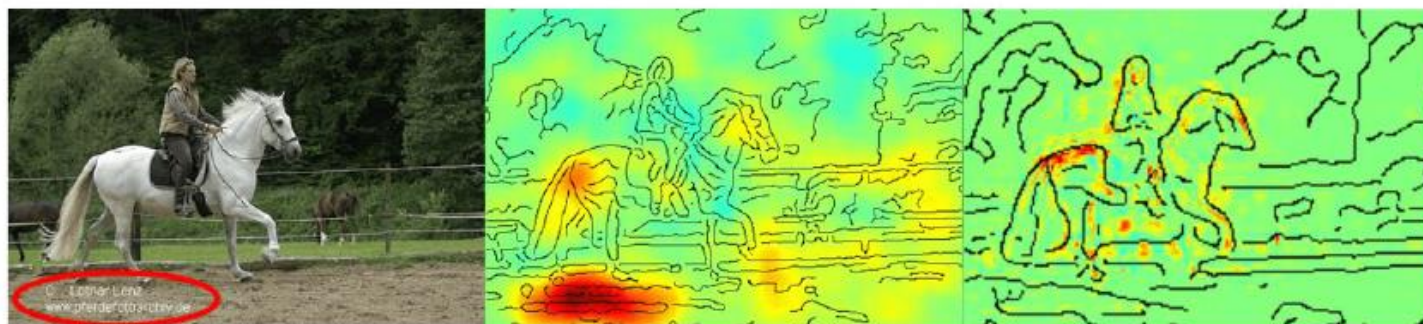
XAI, Black-box



- Black Box
 - 딥러닝 신경망의 구조가 너무 복잡함.
 - 왜 그런 결과가 나왔는지 설명할 수 없음.
 - 알 수 없으니까 == Black Box (<-> White Box)

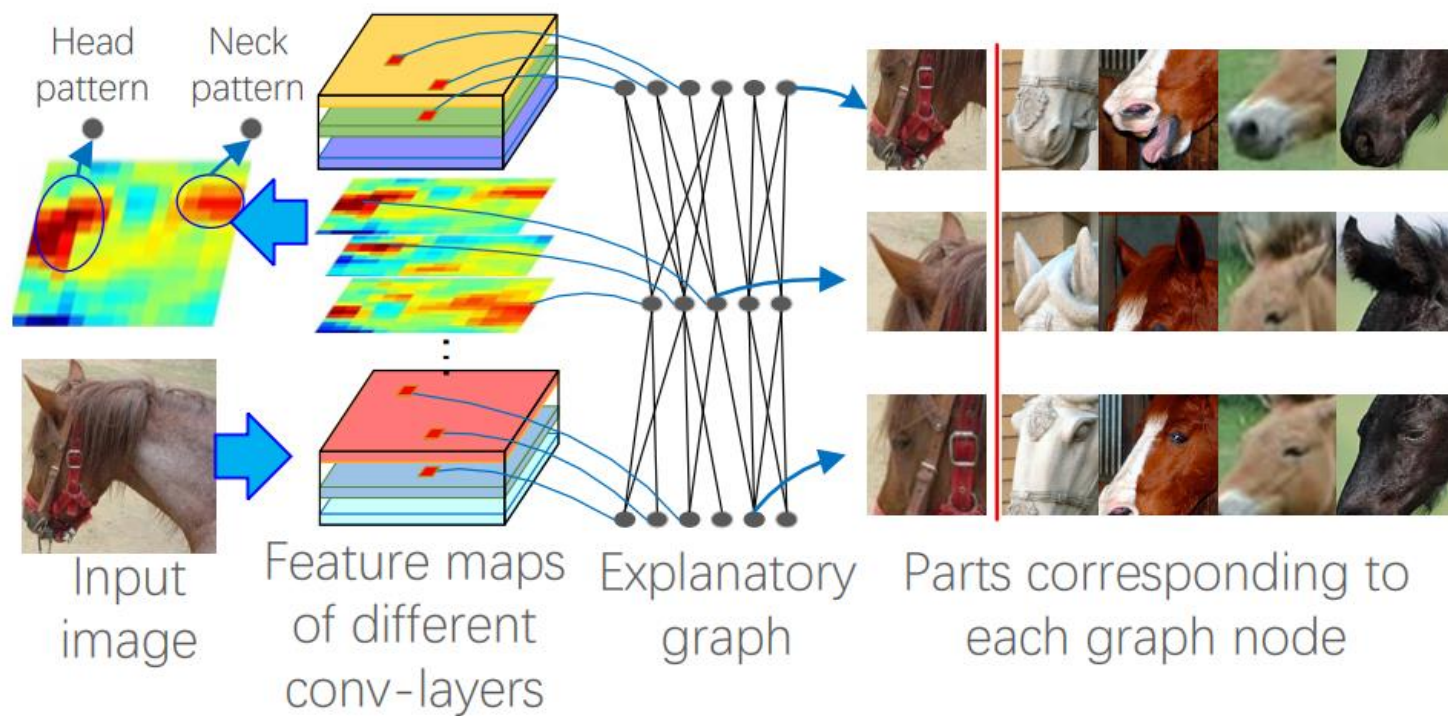
XAI, Black-box

	aeroplane	bicycle	bird	boat	bottle	bus	car
Fisher	79.08%	66.44%	45.90%	70.88%	27.64%	69.67%	80.96%
DeepNet	88.08%	79.69%	80.77%	77.20%	35.48%	72.71%	86.30%
	cat	chair	cow	diningtable	dog	horse	motorbike
Fisher	59.92%	51.92%	47.60%	58.06%	42.28%	80.45%	69.34%
DeepNet	81.10%	51.04%	61.10%	64.62%	76.17%	81.60%	79.33%
	person	pottedplant	sheep	sofa	train	tvmonitor	mAP
Fisher	85.10%	28.62%	49.58%	49.31%	82.71%	54.33%	59.99%
DeepNet	92.43%	49.99%	74.04%	49.48%	87.07%	67.08%	72.12%



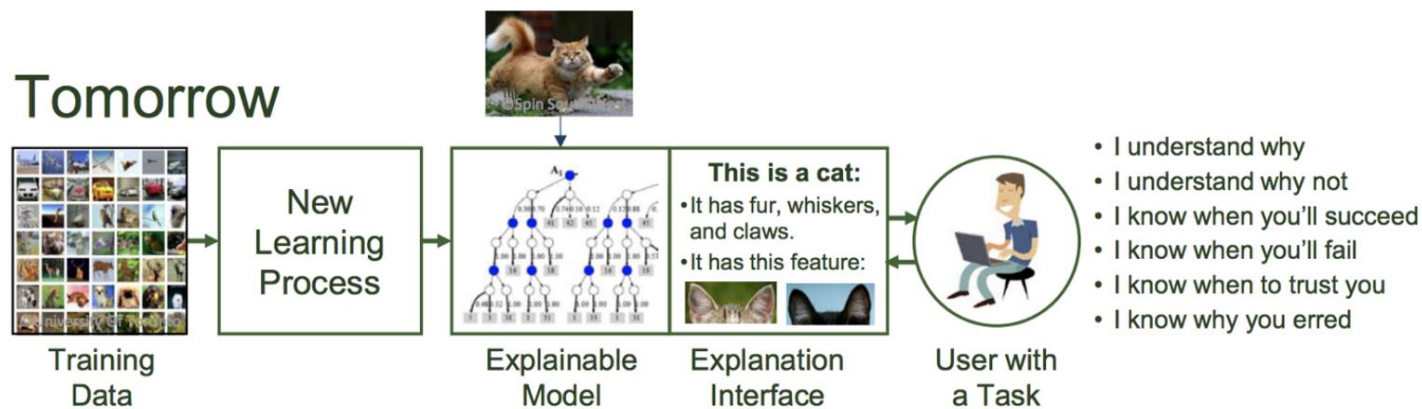
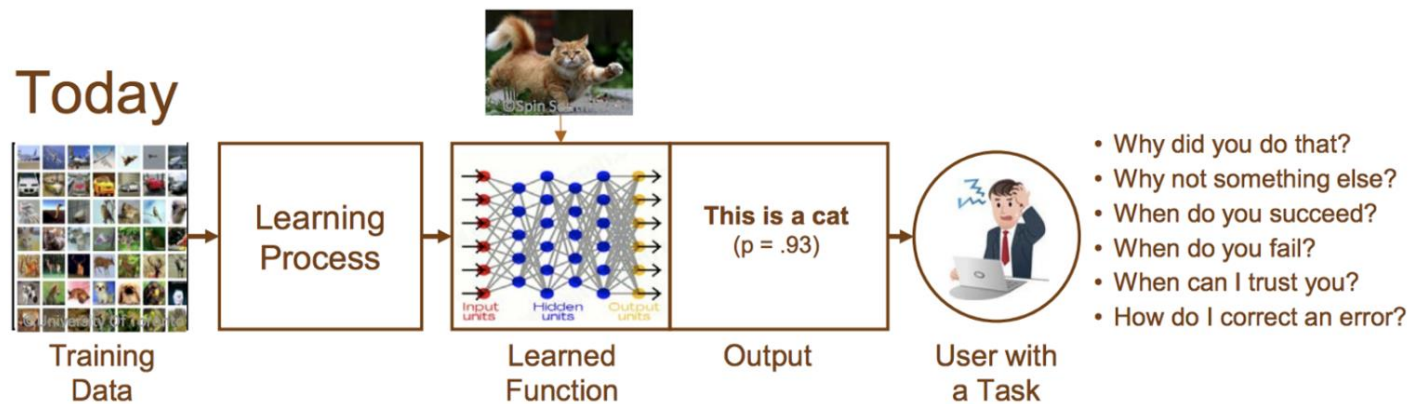
- CNN (Computer Vision)
 - 딥러닝 알고리즘의 내부를 뜯어 본다.
 - 어떤 부분을 보고 알고리즘이 말이라고 판단했는가?

XAI, Black-box



- CNN (Computer Vision)
 - 딥러닝 알고리즘의 내부를 뜯어 본다.
 - 어떤 부분을 보고 알고리즘이 말이라고 판단했는가?
 - 주로 Heat map을 사용하여 시각화.

XAI, Black-box



- 매우 높은 신뢰성이 보장되어야 하는 영역에 사용되는 DNN

못 다한 이야기

- Overfitting
- Forward propagation
- Back propagation
- Linear/Nonlinear
- Dimension Reduction (DR)
- Overfit되어도 좋은 data와 그렇지 않은 데이터
- 왜 Activation function은 비선형이어야 하는가
- Hyper-plane
- Learning-rate
- XOR Problem

등등..