

LG 전자 DX (AI + BigData) 교류

1주차 - 머신러닝 overview



KNU KYUNGPOOK
NATIONAL UNIVERSITY



LG전자

소 속 : 경북대학교/인공지능학과

이 름 : 김현철 교수

E-Mail : hyunchul_kim@knu.ac.kr

Contents

- 서론
 - 인공지능개념
 - 인공지능역사
- 기계학습 개요
 - 인공지능정의
 - 머신러닝을 사용하는 이유
 - 머신러닝의 분류
 - 머신러닝 알고리즘의 유형
 - 머신러닝의 주요 도전과제

❖ 인공지능 개념

➤ <표준국어대사전> <https://stdict.korean.go.kr/>

✓ 지능: 계산이나 문장 작성 따위의 지적 작업에서, 성취 정도에 따라 정하여지는 적응 능력

- 학습: 경험이 쌓임에 따라 점점 성능이 좋아지는 지적작업
- 인식: 오감을 통해 외부 환경의 상태를 알아내는 지적작업
- 추론: 이미 알고 있는 사실을 바탕으로 새로운 사실 또는 새로운 지식을 알아내는 지적작업
- 창작: 세상에 없던 새로운 것을 만들어 내는 지적작업

✓ 인공지능: 인간의 지능이 가지는 학습, 추리, 적응, 논증 따위의 기능을 갖춘 컴퓨터 시스템

❖ 인공지능 개념

➤ 『Artificial Intelligence: Foundations of Computational Agents』
(Poole 저, 2017)

- ✓ 인공지능: the field that studies the synthesis and analysis of computational agents that act intelligently
- ✓ 인공지능: 지능적으로 행동하는 계산 에이전트를 만들고 분석하는 학문 분야
 - 에이전트: 주어진 환경에서 주어진 목표를 향해 행동하는 것
 - 환경과 자신의 목표에 맞게 적절한 행동을 수행
 - 변화하는 환경과 목표에 유연하게 대처
 - 과거 경험으로부터 학습
 - 인식과 계산 능력의 한계에 적절하게 대처

❖ 인공지능 개념

➤ 인공지능

✓ 인간이 가진 지적 능력을 컴퓨터를 통해 구현하는 기술

✓ 인공지능의 구분

- 강인공지능(Strong AI) : 인간의 능력을 초월한 성능을 가진 AI, 다양한 지능의 복합체(예, 터미네이터에 등장하는 인조인간)
- 약인공지능(Weak AI) : 특정 영역에서 도구로 사용하기 위해 설계된 AI, 한가지 지능에 특화된 인공지능(현재 인공지능 제품들, 인공지능 스피커, 언어 번역기, 영상 인식기, 알파고 등)

❖ 인공지능 개념

➤ 앨런 튜링의 튜링 테스트

- ✓ Can machines think? (1950, Computing machinery and intelligence)
- ✓ 튜링 테스트를 통과한 기계는 생각한다고 간주해도 된다(튜링의 주장)

➤ 철학자 존 설(1980, Minds, brains, and program)

- ✓ 컴퓨터 프로그램은 중국어 질문을 전혀 이해하지 못한 채 튜링 테스트를 통과할 수 있음
 - > 튜링 테스트를 통과해도 생각한다고 말하면 안 됨

1.1 서론

❖ 인공지능 개념

➤ 캡차(CAPTCAH)

- ✓ 2000년 미국 카네기멜론 대학교의 연구원들이 사람과 로봇을 구별하기 위해 만든 튜링 테스트
- ✓ 로봇이 구별하기 힘든 문자를 제시하여 이를 맞추면 사람으로 인정하는 방식



1.1 서론

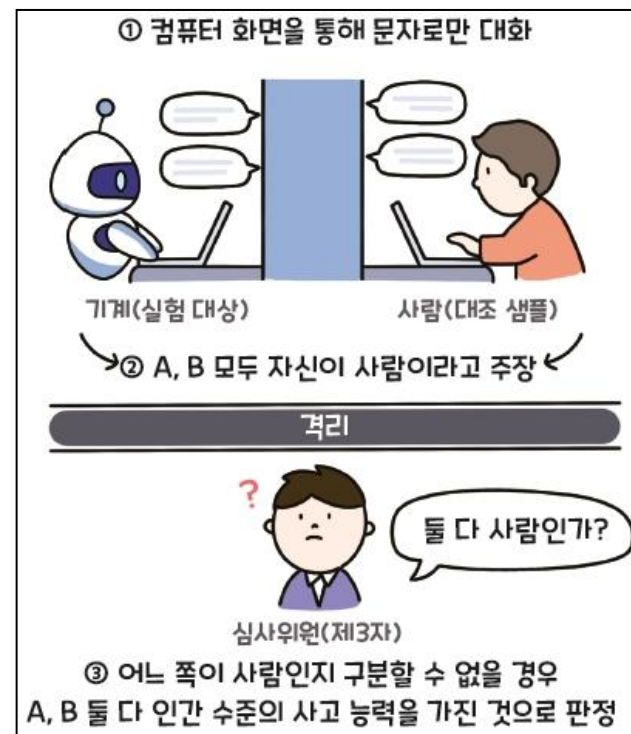
❖ 인공지능 개념

➤ 2014, 유진 구스트만(슈퍼컴퓨터)

- ✓ 튜링테스트 첫 통과 사례
- ✓ 레딩대 '튜링 테스트 2014' 행사
- ✓ 5분 길이의 텍스트 대화에서 심사위원

33% 이상에게 진짜 인간이라는 확신을 줌

- ✓ 우크라이나에 사는 13세 소년, 유진
- ✓ 블라디미르 베셀로프, 유진 뎀첸코,
세르게이 올라센 등이 개발



1843

에이다 “... 해석엔진은 꽤 복잡한 곡을 자곡할 수 도 있다 ” 논문발표



어거스트 에이다 바이런([1815년](#) ~ [1852년](#)) 최초의 프로그래머

1.1

서론

- 1843 에이다 “... 해석엔진은 꽤 복잡한 곡을 자곡할 수 도 있다 ” 논문발표
- 1943 워런 맥클록과 월터 피츠, 최초의 인공신경망 연구
- 1950 앨런튜링 “Can machines think?” 튜링테스트



1.1

서론

- 1843 에이다 “... 해석엔진은 꽤 복잡한 곡을 자곡할 수 도 있다 ” 논문발표
- 1943 워런 맥클록과 월터 피츠, 최초의 인공신경망 연구
- 1950 앨런튜링 “**Can machines think?**” 튜링테스트
- 1956 다트머스 컨퍼런스(존 매카시, 마빈민스키, 클로드새넌, 앤런뉴얼 등)



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

- 1843 에이다 “... 해석엔진은 꽤 복잡한 곡을 작곡할 수 도 있다 ” 논문발표
- 1943 워런 맥클록과 월터 피츠, 최초의 인공신경망 연구
- 1950 앨런튜링 “**Can machines think?**” 튜링테스트
- 1956 다트머스 컨퍼런스(존 매카시, 마빈민스키, 클로드새넨, 앤런뉴얼 등)
- 1958 로젠블랫이 퍼셉트론 제안, 인공지능 언어 **Lisp** 탄생



▲ 퍼셉트론 시연

- 1959 사무엘이 기계학습을 이용한 체커 게임 프로그램 개발



▲ 체커 프로그램 시연

- 1843 에이다 “... 해석엔진은 꽤 복잡한 곡을 자곡할 수 도 있다 ” 논문발표
- 1943 워런 맥클록과 월터 피츠, 최초의 인공신경망 연구
- 1950 앨런튜링 “**Can machines think?**” 튜링테스트
- 1956 다트머스 컨퍼런스(존 매카시, 마빈민스키, 클로드새넨, 앤런뉴얼 등)
- 1958 로젠블랫, 퍼셉트론 제안, 인공지능 언어 **Lisp** 탄생(존 맥카시)
- 1959 사무엘, 기계학습을 이용한 체커 게임 프로그램 개발
- 1965 자데, 퍼지 이론 발표
- 1966 세계 최초의 챗봇 엘리자 공개
- 1969 민스키, 퍼셉트론의 과대포장 지적, 신경망 퇴조 시작
- 1972 인공지능 언어 **Prolog** 탄생, 스탠퍼드 대학교 마이신 전문가시스템 개발시작
- 1973 라이트힐 보고서를 계기로 인공지능 내리막길, 1차 인공지능 겨울 시작
- 1974 웨어보스, 오류 역전파 알고리즘 발표

1966 미국 정부의 지원 중단

1957년 최초 인공위성 스푸트닉호 러시아

미국, 국립연구회를 중심으로 러시아 과학논문 자동번역 프로젝트 지원

보고서: 과학논문의 기계번역은 불가능했으며, 조만간 가능하리라는 전망도 없다.

1969 민스키, 퍼셉트론 한계 증명

로젠블랫의 퍼셉트론은 '**AND**', '**OR**' 같은 선형불리가 가능한 문제에서는 적용할 수 있지만 '**XOR**' 와 같은 문제에는 적용할 수 없다

단순 퍼셉트론 모델의 한계 증명. 인공신경망의 한계로 착각, 학계에서 외면

1971 영국 정부의 지원 중단

제임스 라이트힐 경, 영국 인공지능 연구현황 보고서

현재의 인공지능 기술이 '폭발적 조합의 증가' 문제에 제대로 대처할 수 없다는 점을 들어 AI 연구에서 중요한 결과는 커녕, 의미있는 결과조차 찾지 못해서 '**AI**'라는 별도의 과학분야는 필요없다고 주장

- 1980 존설, 중국인의 방 논문 발표
- 1986 다층 퍼셉트론으로 신경망 부활
- 1987 Lisp 머신의 시장 붕괴로 2차 인공지능 겨울 시작

1차 붐

1956-1974

간단한 문제해결



2차 붐

1980-1987

전문가시스템



3차 붐

1993-

기계학습/딥러닝

모라벡의 역설: 인간에겐 쉬운 것(보거나 걷는 것)이 AI 에겐 어렵다

상식의 저주: 인간에겐 상식 수준의 지식들도 AI 는 모두 배워야만 가능

프레임문제: 주어진 현실 상황에서 모든 것을 다 고려해서 한없이 생각할 수 밖에 없다

프레임문제

인간 : 관계 있는 지식만을 꺼내서 그것을 사용

철학자 대니얼 데닛이 고안한 예

동굴 안에 로봇을 움직이게 하는 배터리가 있고 그 위에 시한폭탄이 장착돼 있다. 로봇은 배터리를 가져 오지 않으면 배터리가 다 떨어져 움직일 수 없게 되므로 동굴에서 배터리를 가져오라는 지시를 받음

연구자들은 이것을 수행시키기 위해 로봇을 설계

로봇 1호: 배터리를 동굴에서 가져올 수 있었지만 배터리 위에 놓여 있는 시한폭탄도 같이 가져와 동굴에서 나온 후에 폭탄이 폭발. 시한폭탄이 실려 있다는 것은 알고 있었지만 배터리를 가져오면 폭탄도 함께 가져오게 된다는 사실을 몰랐던 것

로봇 2호: 배터리를 반출할 때 폭탄도 같이 반출할 것인가, 아닌가를 판단시키기 위해서 ‘자산이 무언가를 하면 그 행동에 따라서 부차적으로 일어나는 것’도 고려할 수 있도록 개량. 로봇 2호는 배터리를 앞에 두고 생각하기 시작. ‘자신이 왜건을 끌어당기면 벽의 색이 바뀔 것인가’, ‘천장이 무너지지는 않을까’ 등등. 하지만 모든 현상이 일어날 것인가 아닌가를 생각한 탓에 타임 오버로 시한폭탄이 폭발

로봇 3호: ‘목적을 수행하기 전에 관계없는 사항은 고려하지 않도록’ 개량. 로봇 3호는 관계있는 것과 없는 것을 분류하는 작업에 몰두하고 무한한 생각을 거듭한 탓에 동굴에 들어가기 전에 동작이 멈춤. 목적과 관계없는 사항도 워낙 많기 때문에 그 모든 것을 고려하는 것 만도 긴 시간이 걸렸기 때문

지식 - > 데이터 형태로 저장

1990

Machine Learning

데이터 - > 지식

학습

직간접적 경험이나 훈련에 의해 지속적으로 지각하고, 인지하며,
변화시키는 행동 변화

데이터를 특징에 따라 분류하는 것

(안 르쿤, 제프리 힌튼, 요수아 벤지오, 앤드류 응)

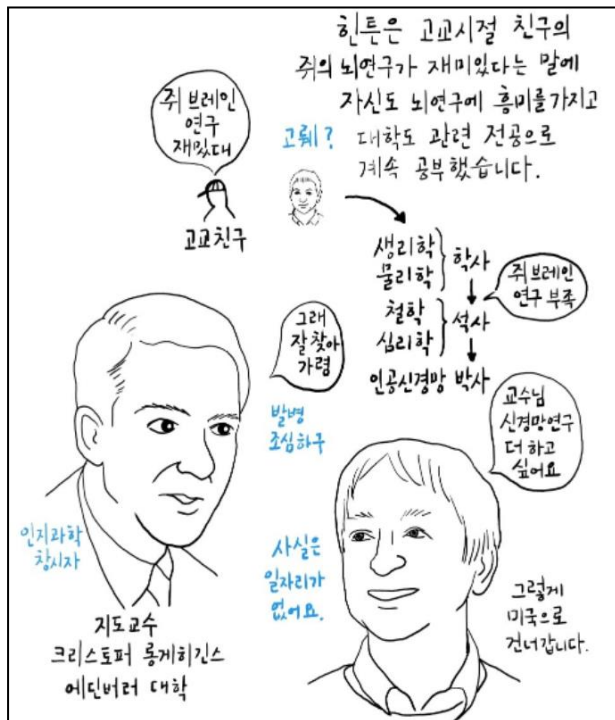


1990

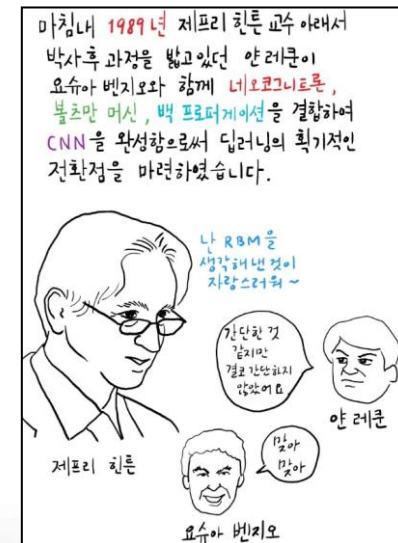
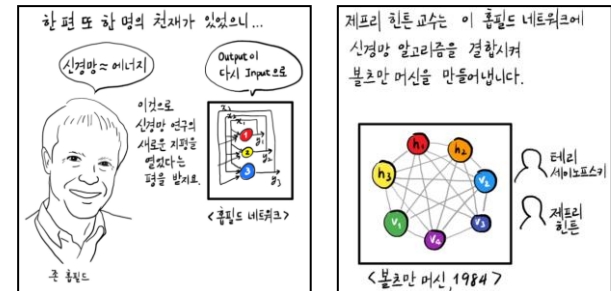
Deep Learning

데이터를 특징에 따라 분류하는 것 (인간, 기계)

자율학습



학습



참고: 적송 글, 루나 그림

- 1991 파이썬 언어 탄생
- 1997 **IBM** 딥블루가 세계 체스 챔피언 카스파로프 이김(2승3무1패)
- 1998 르쿤, 컨볼루션 신경망의 실용적인 학습 알고리즘 제안
- 1999 엔비디아, **GPU** 공개
- 2000 컴퓨터 비전 패키지 **OpenCV** 최초 공개
- 2007 파이썬의 기계학습 라이브러리 사이킷 런 최초 공개
- 2009 딥러닝 패키지 씨아노 서비스 시작
- 2010 **ImageNet**(대규모 자연 영상 **DB**) 탄생, 1회 **ILSVRC** 대회 개최
- 2011 **IBM** 왓슨, 미국 **TV** 퀴즈쇼 '제퍼디'에서 우승



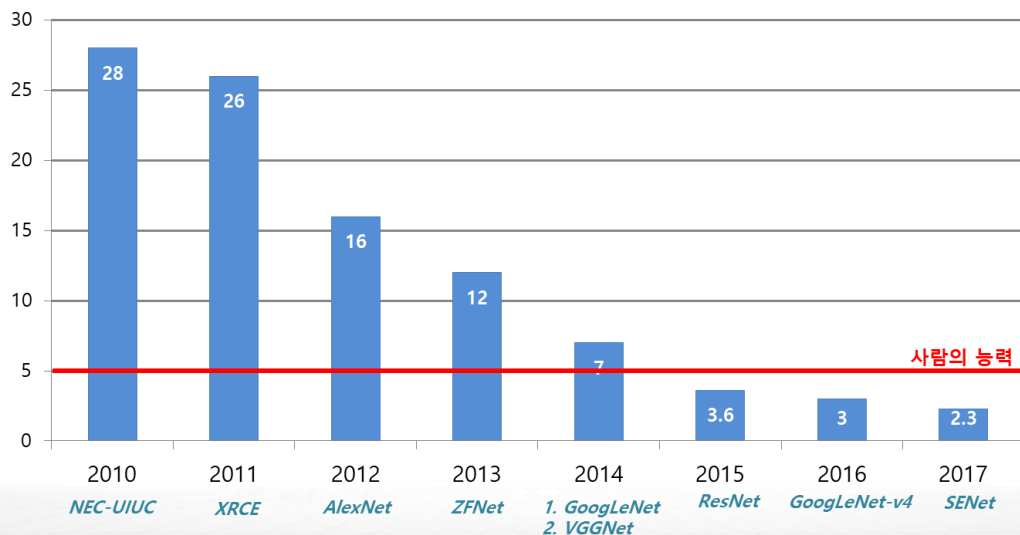
2012

딥러닝으로 **MINIST(필기 숫자 DB) 0.23%** 오류율 달성,
AlexNet 발표(3회 ILSVRC 우승, 컨볼루션 신경망의 가능성을 보여줌)

ImageNet Large Scale Visual Recognition Challenge

- ❖ 대용량의 이미지셋을 주고 이미지 분류 알고리즘의 성능을 평가하는 대회
- ❖ 1,000개가 넘는 카테고리로 분류된 100만개의 이미지

우승 알고리즘의 분류 에러율(%)



- 2014** 딥러닝 패키지 카페(**Caffe**) 서비스 시작
- 2015** 딥러닝 패키지 텐서플로(**TensoFlow**) 서비스 시작
OpenA 창립
클라우드 슈밥이 4차 산업혁명 언급
- 2016** 딥러닝 패키지 파이토치(**PyTorch**) 서비스 시작
딥러닝 패키지 케라스(**Keras**) 서비스 시작
알파고 승리(이세돌과 바둑대국)
- 2017** 알파고 제로, 알파고를 100:0 이김
구글, **Teachable machine** 공개
- 2020** **OpenAI** 재단이 3세대 언어 모델 **GPT-3** 발표

1.1 서론

알파고

- ❖ 구글 **2011년** 알파고를 만들어낸 영국의 인공지능 스타트업 딥마인드 **4억 달러**에 인수
- ❖ 딥마인드의 창업자 데미스 하사비스: 체스신동(**14세 체스 세계2위**)
- ❖ **2016년 이세돌 9단 대결(4:1)**



1.2 기계학습 개요

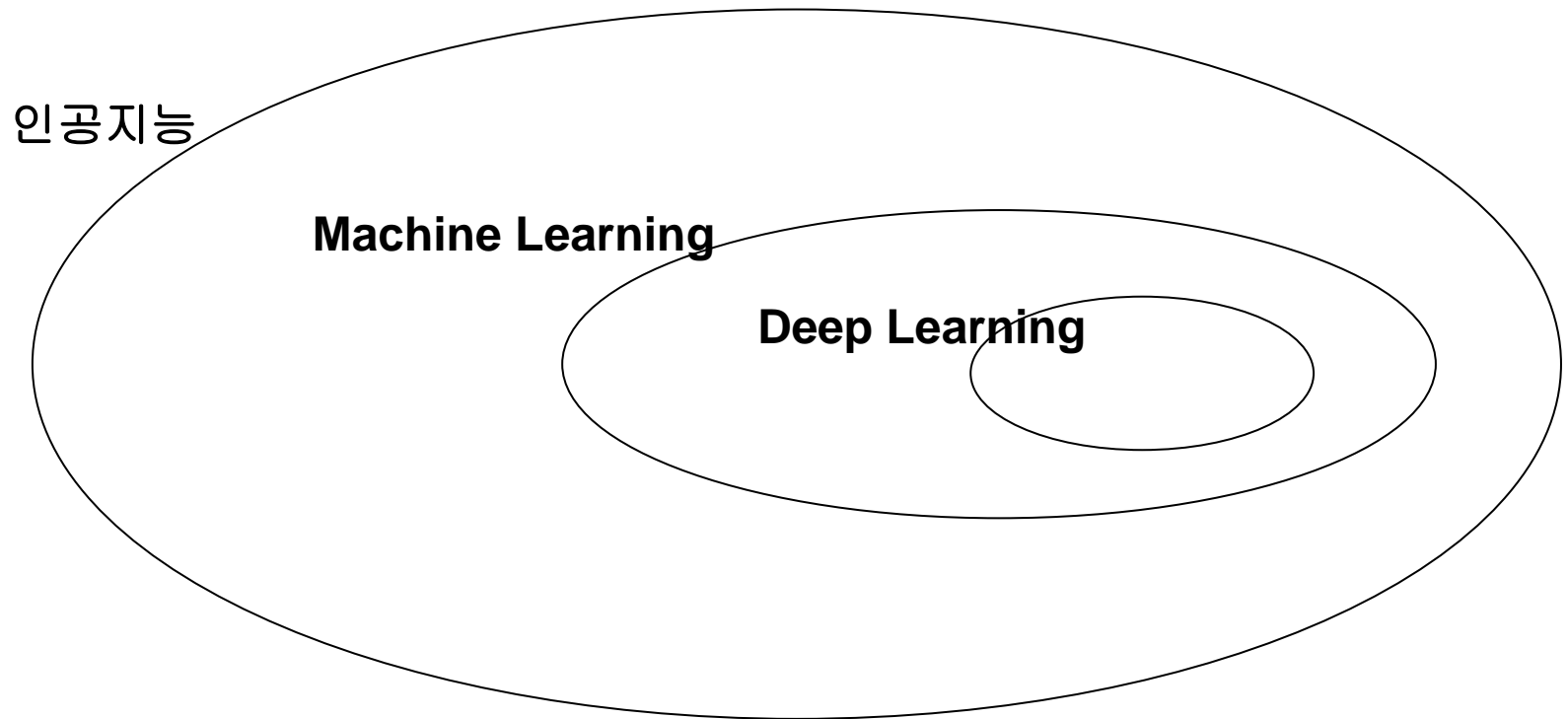
❖ 지능(Intelligence)

- 무언가를 이해하고 배우는 능력
- 오직 인간만 가지고 있는 고유의 성질
- 본능적으로 행동하는 것이 아니라 생각하고 이해함으로써 행동으로 옮기는 능력

❖ 인공지능(Artificial Intelligence)

- 컴퓨터가 학습하고 생각하여 스스로 판단할 수 있도록 만드는 기술
- (사전)지성을 갖춘 존재 또는 시스템에 의해 만들어진 인공적인 지능
- (전통)컴퓨터가 인간의 지능적인 행동을 모방할 수 있도록 하는 SW로, 인간이 가진 지적 능력의 일부 또는 전체를 구현하는 것
- (기술)인간의 지능으로 할 수 있는 사고, 학습, 자기계발 등을 컴퓨터가 할 수 있는 방법을 연구하는 컴퓨터공학 및 정보기술의 한 분야

인공지능(Artificial Intelligence)



1.2 기계학습 개요

□머신러닝

- 컴퓨터를 인간처럼 학습하게 함으로써 인간의 도움 없이도 컴퓨터 스스로가 새로운 규칙을 발견할 수 있도록 하는 기술
- 기본적으로 **알고리즘을 이용해** 데이터를 분석하고, 분석을 통해 학습하며, 학습한 내용을 기반으로 판단이나 예측을 함
- 머신러닝이 스스로 학습하여 데이터를 처리하는 과정
 - 빅데이터를 입력
 - 데이터를 분석하여 모델을 만듦
 - 모델을 이용하여 의사결정 및 예측 등을 수행

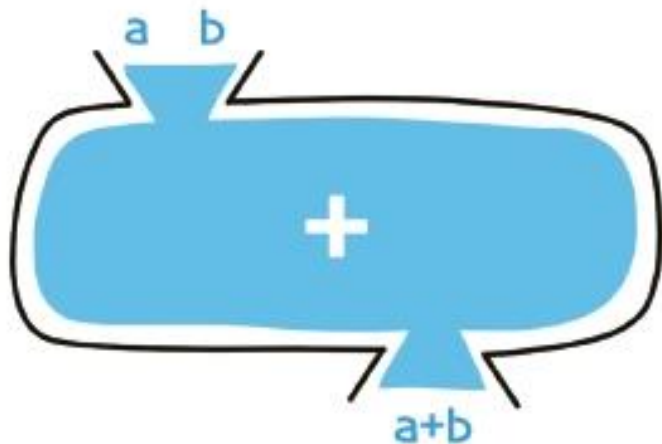
1.2 기계학습 개요

□ 머신러닝

➤ 기존 프로그램과 머신러닝의 차이

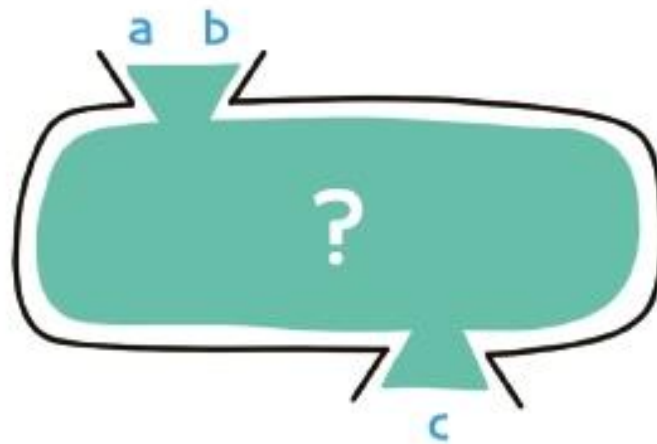
기존 프로그래밍

$a+b$ 의 결과값이 나오는 프로그램 설계



머신러닝

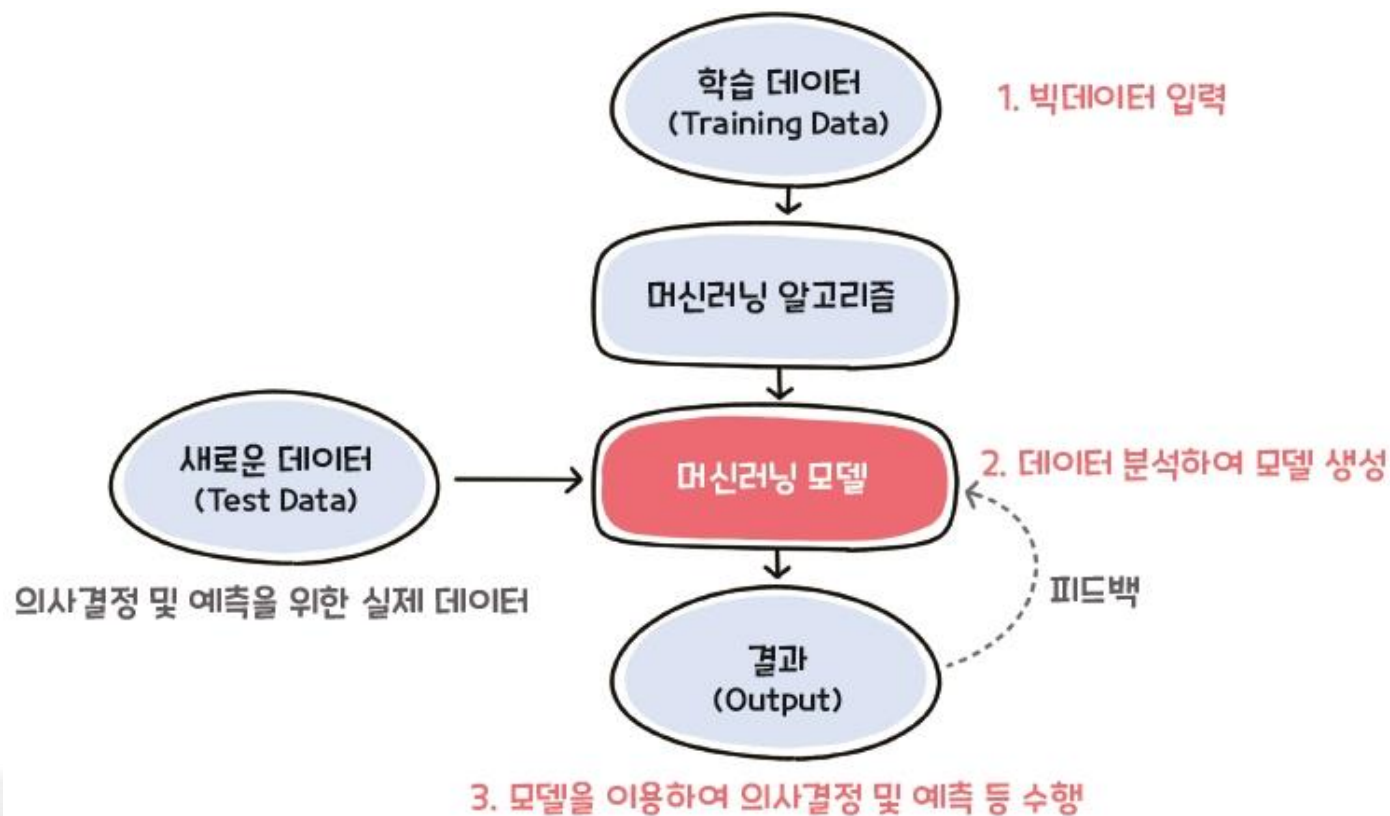
대용량의 데이터와 머신러닝 알고리즘 활용



1.2 기계학습 개요

□ 머신러닝

➤ 머신러닝 학습절차

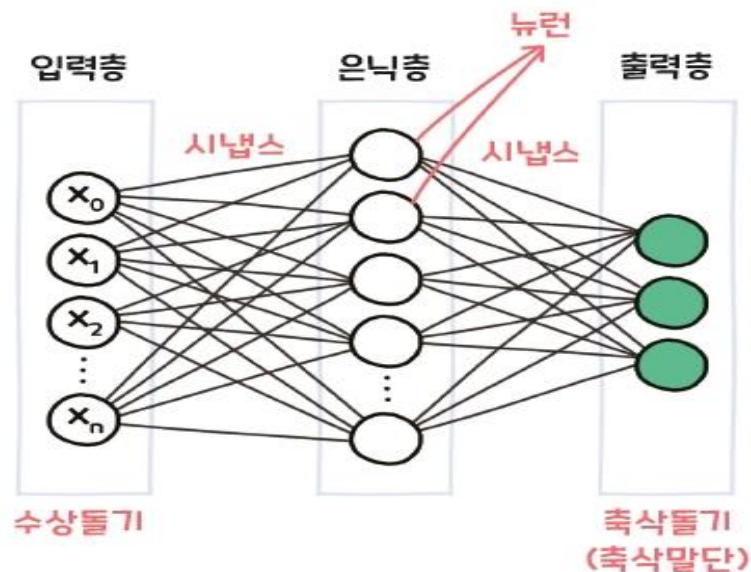
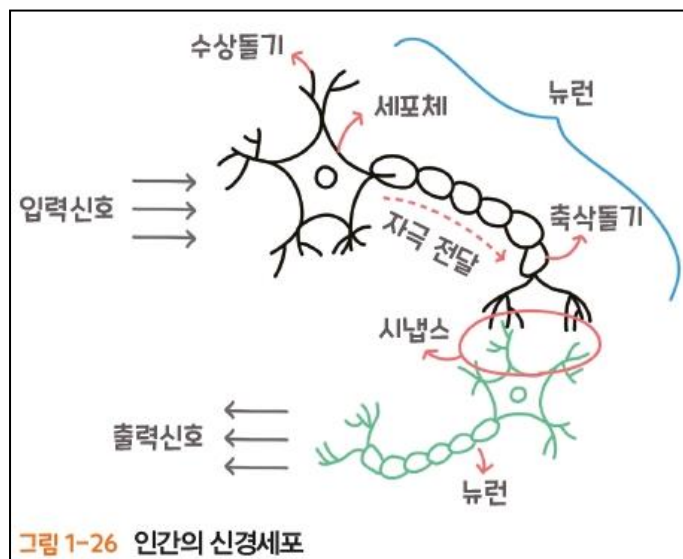


1.2 기계학습 개요

□ 딥러닝

➤ 인공신경망(ANN, Artificial Neural Network)

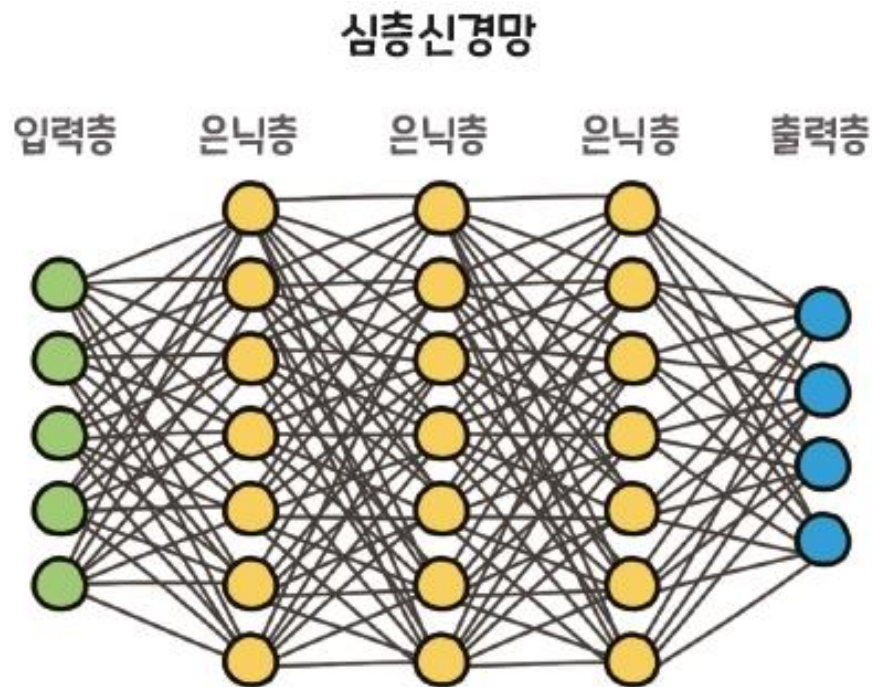
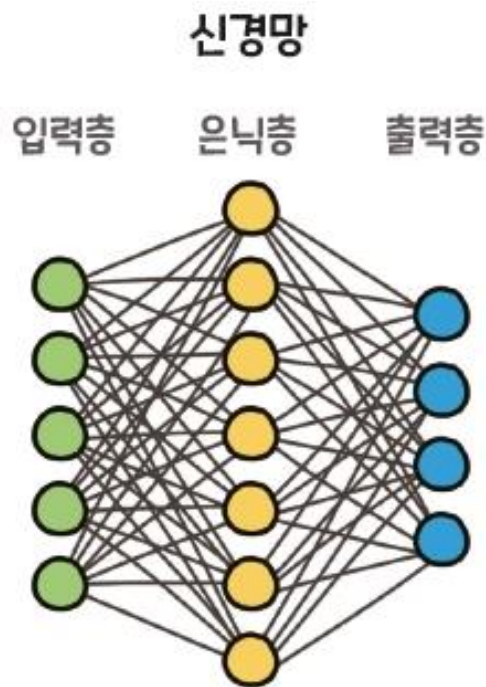
- ✓ 여러 뉴런이 서로 연결되어 있는 구조의 네트워크



1.2 기계학습 개요

□ 딥러닝

➤ 심층신경망(DNN, Deep Neural Network)

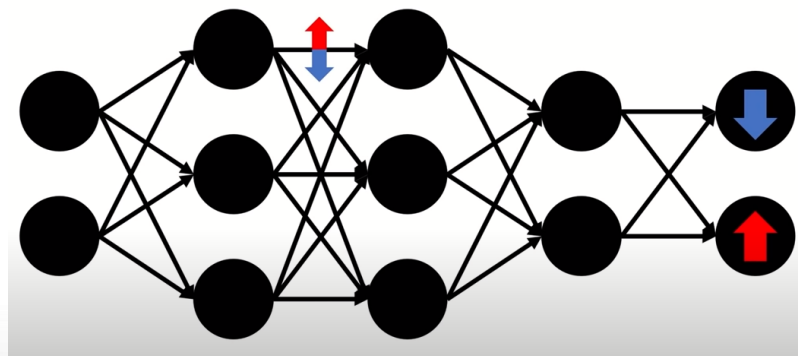
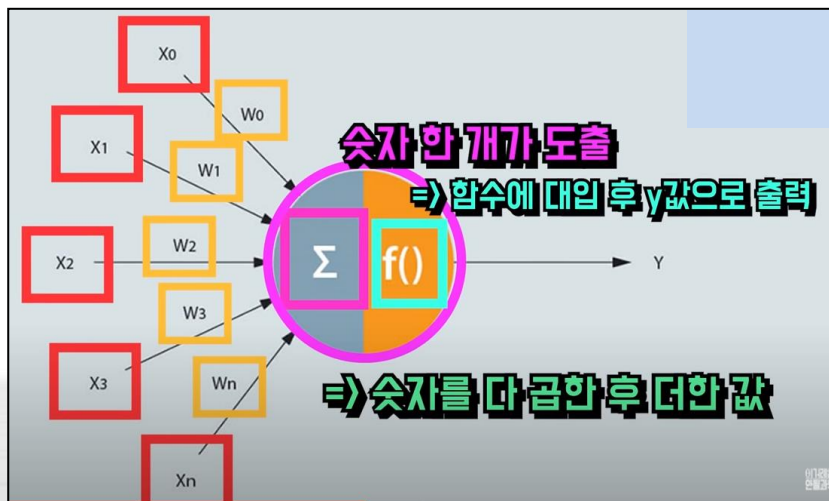


1.2 기계학습 개요

□ 딥러닝

➤ 심층신경망(DNN, Deep Neural Network)

- ✓ 여러 은닉층을 가진 인공신경망을 사용하여 머신러닝 학습을 수행하는 기술
- ✓ 딥러닝의 '딥(Deep)'은 연속된 신경망 층(layer)을 깊게(deep) 쌓는다는 의미
- ✓ 이 신경망이 깊어질수록 성능이 향상됨



1.2 기계학습 개요

□ 딥러닝

- 합성곱신경망(CNN, Convolutional Neural Network)
 - ✓ 시신경 모방: 추출과 분류의 기능 특화
 - ✓ 문장분류, 얼굴인식
- 순환신경망(RNN, Recurrent Neural Network)
 - ✓ 반복적이고 순차적인 데이터 학습 특화
 - ✓ 음성인식, 번역분야
- GAN, Generative Adversarial Networks)
 - ✓ 10년간 나온 아이디어 중 가장 흥미로운 것
 - ✓ 2개의 모델이 서로 대결하며 학습하는 것
 - ✓ 위조지폐와 경찰 모델
 - ✓ 사진

1.2 기계학습 개요

□ 머신러닝과 딥러닝의 차이점

➤ 인간의 개입 유무

- ✓ 특징 추출(Feature Extraction) : 머신러닝에서 컴퓨터가 스스로 학습하려면 사람이 인지하는 데이터를 컴퓨터가 인지할 수 있는 데이터로 변환해야 하는데, 이 작업을 위해 데이터별로 어떤 특징을 가지는지 찾아내고 그것을 토대로 데이터를 벡터로 변환하는 것

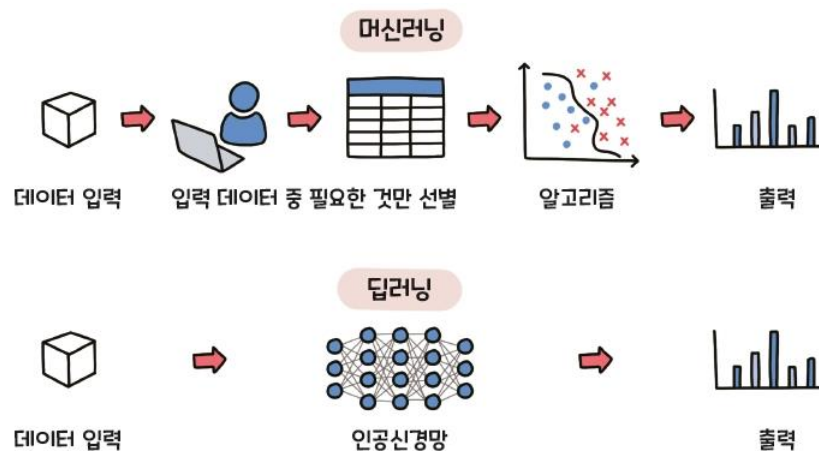
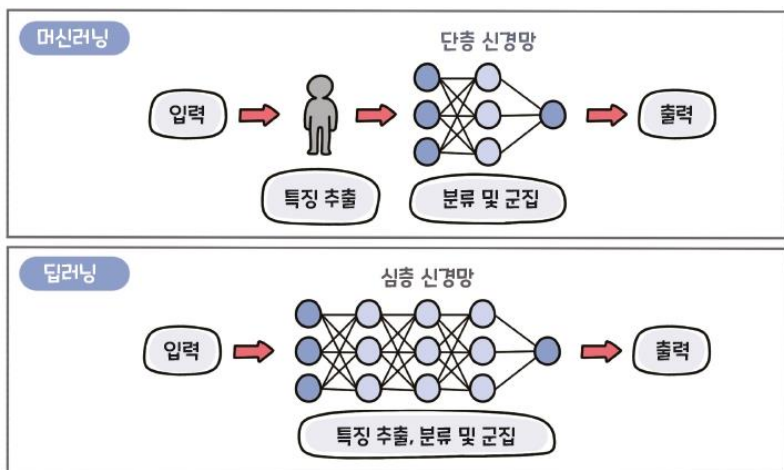


1.2 기계학습 개요

❑ 머신러닝과 딥러닝의 차이점

➤ 인간의 개입 유무

- ✓ 머신러닝은 사람이 학습 데이터에 레이블(정답)을 알려주거나 데이터의 특징을 추출하는 등 어느 정도 개입
- ✓ 딥러닝은 인간의 개입 없이 컴퓨터 스스로 학습함

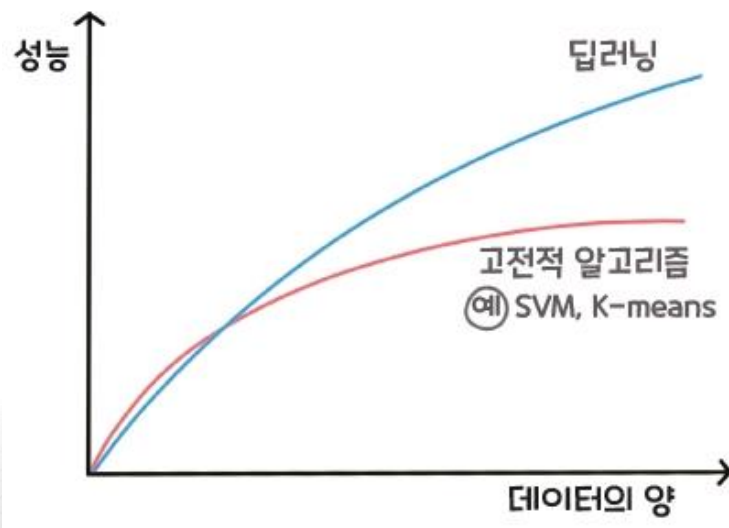


1.2 기계학습 개요

□ 머신러닝과 딥러닝의 차이점

➤ 데이터 의존도(Data Dependencies)

- ✓ 딥러닝은 주어진 문제를 해결하기 위해 중요한 특징을 직접 추출
- ✓ 그래서 데이터의 양이 충분하지 않으면 정확한 특징을 추출할 수 없음
- ✓ 반면, 충분한 양의 데이터가 주어진다면 사람이 인지하지 못한 중요한 특징들까지 찾아낼 수 있을 정도로 좋은 성능 발휘

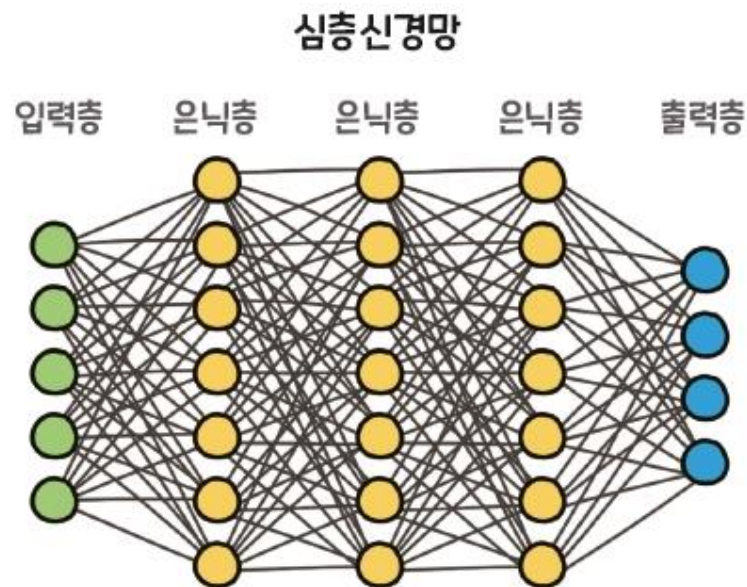
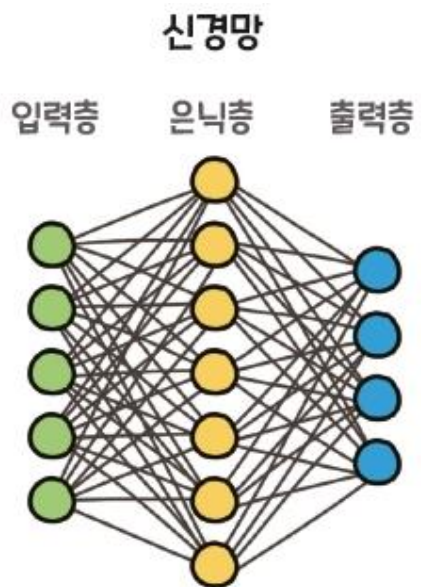


1.2 기계학습 개요

□ 머신러닝과 딥러닝의 차이점

➤ 심층신경망의 사용 여부

- 딥러닝은 심층신경망을 이용하여 입력 데이터에서 특징을 추출하고 스스로 결과(예측 혹은 분류)를 도출
- 심층신경망을 사용하는 것은 딥러닝만의 뚜렷한 특징

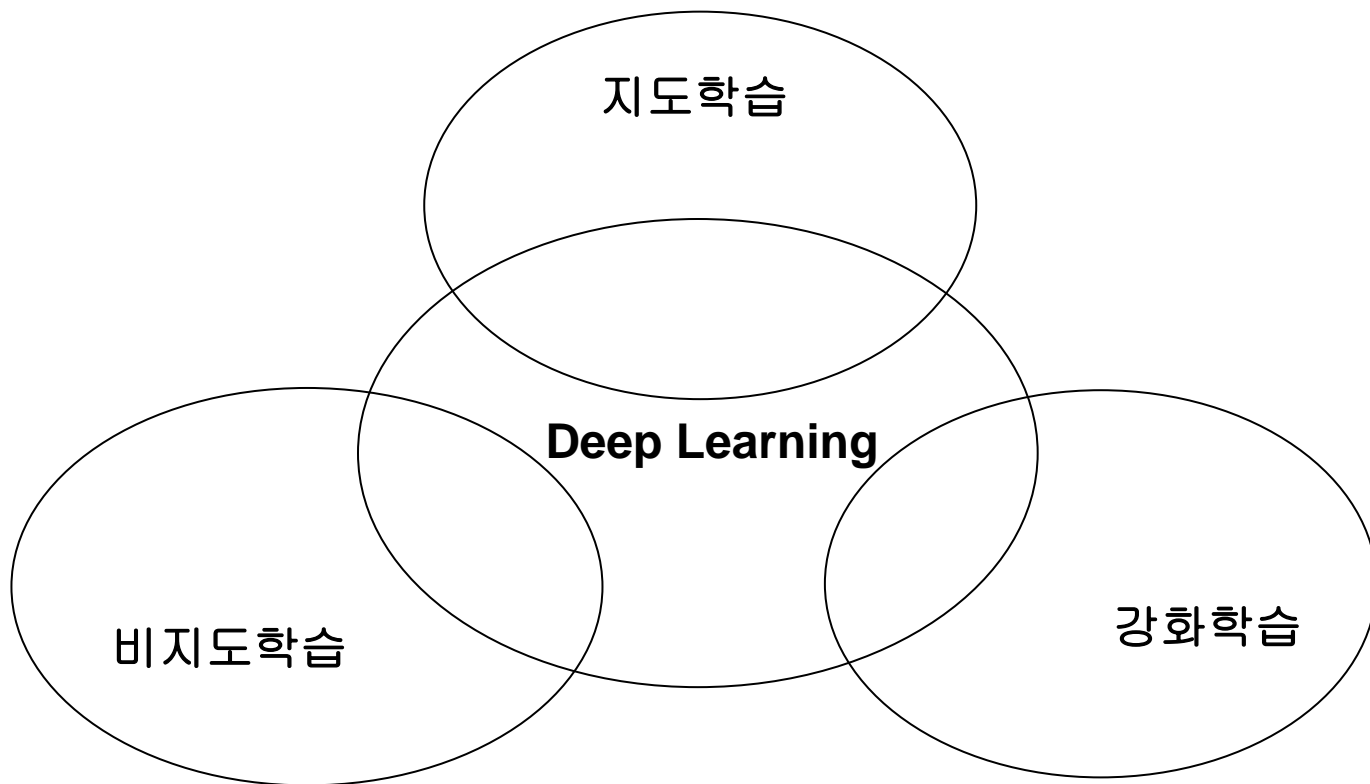


1.2 기계학습 개요

□ 머신러닝과 딥러닝의 차이점

구분	머신러닝	딥러닝
필요한 데이터의 양	적은 양의 데이터도 가능	빅데이터
정확도	낮음	높음
훈련 시간	짧은 시간 안에 가능	오래 걸림
하드웨어	CPU만으로도 가능	GPU
하이퍼파라미터 튜닝	제한적	다양한 방법으로 튜닝 가능

□ 인공지능(Artificial Intelligence)

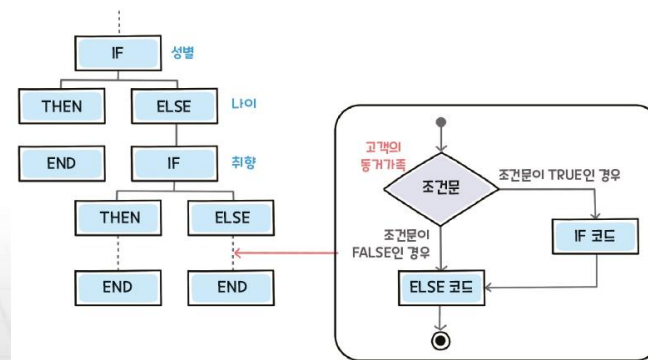


1.2 기계학습 개요

❑ 머신러닝을 사용하는 이유

➤ 기존 프로그래밍의 한계

- ✓ 고객의 성별 · 나이 · 취향 등이 반영된 프로그램에 ‘고객의 동거가족’ 변수를 추가하고 싶다면, 만들어져 있는 기존 프로그램을 변경해야 함
- ✓ 뿐만 아니라 기존 변수들과의 관련성까지 고려해 프로그램 전체를 수정해야 함
- ✓ ‘고객의 동거가족’에 대한 데이터를 데이터베이스에 저장하기 위한 수정 필요
- ✓ 즉, 변수가 하나 더 추가되었을 뿐인데도 프로그램에서 수정해야 할 부분들이 상당히 많다는 것을 알 수 있음

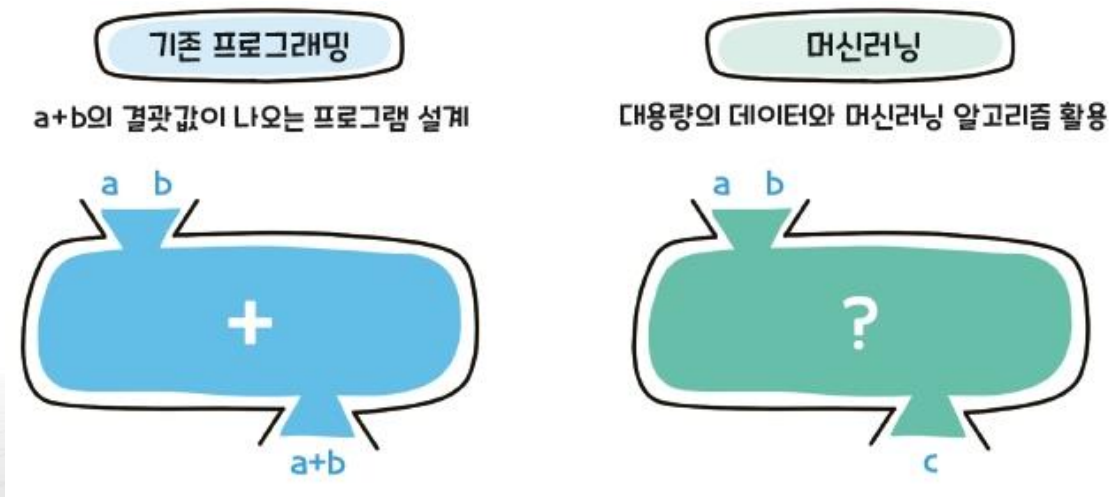


1.2 기계학습 개요

□머신러닝을 사용하는 이유

➤ 머신러닝의 유용성

- ✓ 하지만 빠른 의사결정이 필요한 시기에 기존 프로그래밍 방식이 적절하지 않음
- ✓ 그래서 이를 해결하기 위해 머신러닝 방식을 사용하는 것
- ✓ 대용량의 데이터와 많은 변수가 관련되어 있고 기존에 사용했던 규칙의 프로그램으로는 복잡한 작업이나 문제를 해결 할 수 없을 때 머신러닝은 아주 유용한 해결책임



1.2 기계학습 개요

□ 머신러닝을 사용하는 이유

➤ 머신러닝의 유용성

- ✓ 얼굴 인식이나 음성 인식과 같이 규칙 기반 프로그램으로 답을 낼 수 없는 복잡한 상황
- ✓ 거래 기록에서 사기를 감지하는 경우와 같이 규칙이 지속적으로 바뀌는 상황
- ✓ 주식 거래, 에너지 수요 예측, 쇼핑 추세 예측의 경우처럼 데이터 특징이 계속 바뀌는 상황

1.2 기계학습 개요

□머신러닝의 분류

- 지도학습 : 예측이나 분류를 위해 사용
- 비지도학습 : 군집을 위해 사용
- 강화학습 : 환경에서 취하는 행동에 대한 보상을 이용하여 학습을 진행

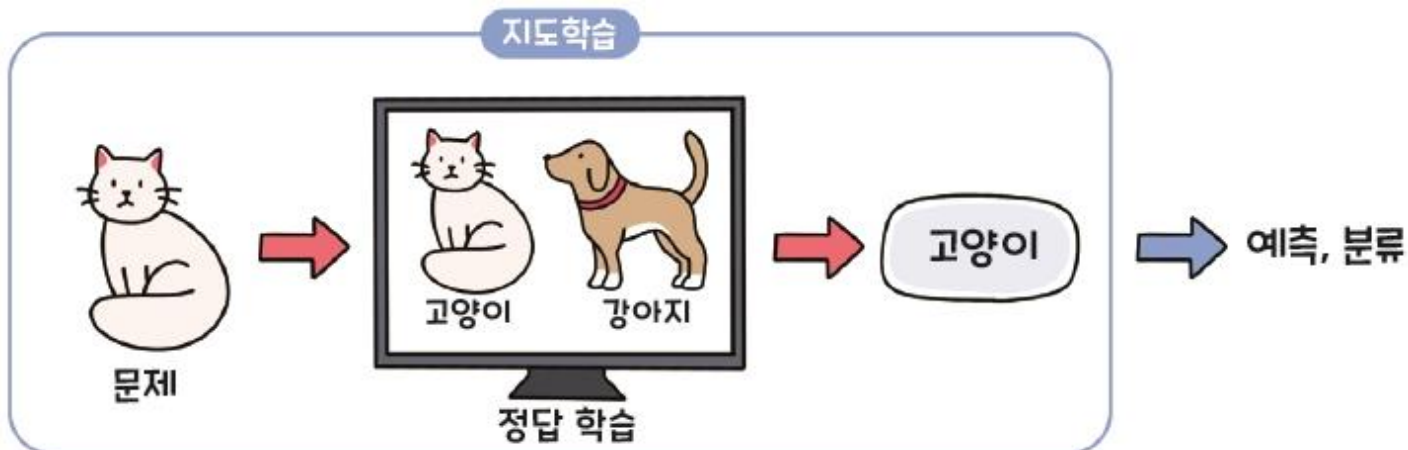


1.2 기계학습 개요

□머신러닝의 분류

➤ 지도학습 : 예측이나 분류를 위해 사용

- ✓ 문제와 답을 함께 학습함으로써 미지의 문제에 대한 올바른 답을 예측하는 학습
- ✓ 지도학습에서 사용하는 모델로는 크게 예측과 분류가 있음

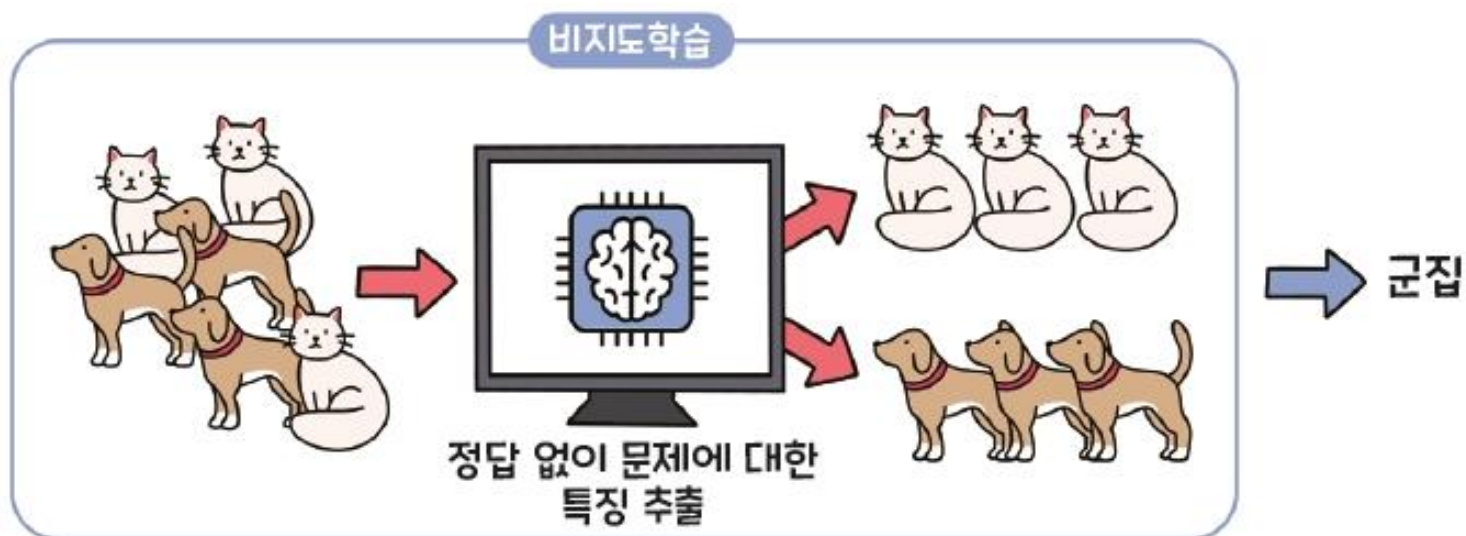


1.2 기계학습 개요

□머신러닝의 분류

➤ 비지도학습(Unsupervised Learning)

- ✓ 지도학습과 다르게 조력자의 도움 없이 컴퓨터 스스로 학습하는 형태
- ✓ 컴퓨터가 훈련 데이터를 이용하여 데이터들 간의 규칙성을 찾음



1.2 기계학습 개요

□머신러닝의 분류

➤ 비지도학습(Unsupervised Learning)

- ✓ X (입력 데이터)와 y (지도학습에서 레이블)의 관계를 파악했던 지도학습과는 달리, 비지도학습은 x 간의 관계를 스스로 파악함
- ✓ 즉, 지도학습과 다른 점은 y (레이블)의 차이
- ✓ 비지도학습에서 사용하는 모델로는 군집(Clustering)이 있음

구분	지도학습	비지도학습
필요한 데이터 종류	x (학습 데이터), y (레이블)	x (학습 데이터)

1.2 기계학습 개요

□ 강화학습(Reinforcement Learning)

- 자신이 한 행동에 대해 보상(Reward)을 받으며 학습하는 것
- 컴퓨터가 주어진 상태에 대해 최적의 행동을 선택하도록 학습하는 방법



1.2 기계학습 개요

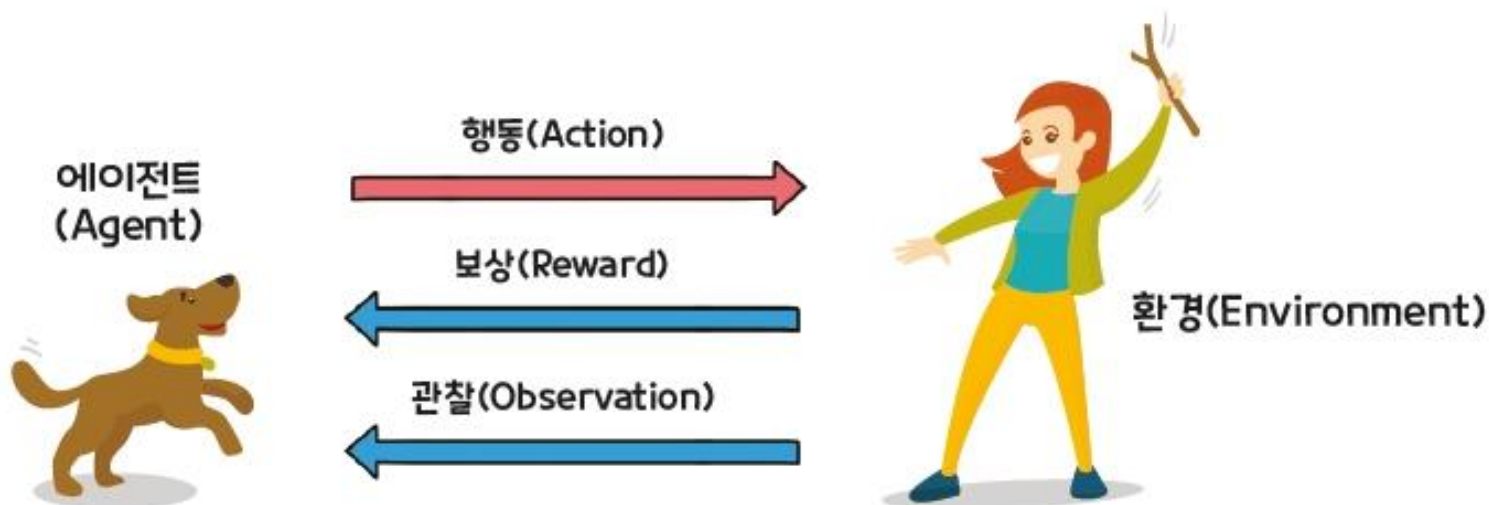
□ 강화학습(Reinforcement Learning)

- 강화학습을 이해하기 위해 알아야 할 개념들
- 에이전트(Agent) : 주어진 문제 상황에서 행동하는 주체
- 상태(State) : 현재 시점에서의 상황
- 행동(Action) : 플레이어가 취할 수 있는 선택지
- 보상(Reward) : 플레이어가 어떤 행동을 했을 때 따라오는 이득
- 환경(Environment) : 문제 그 자체를 의미
- 관찰(Observation) : 에이전트가 수집한(보고 듣는) 환경에 대한 정보

1.2 기계학습 개요

□ 강화학습(Reinforcement Learning)

- 주어진 환경에서 에이전트가 선택한 행동에 따라 그 행동이 옳은 선택이면 상을 받고, 잘못된 선택이면 벌을 받음
- 강화학습은 에이전트가 상태를 계속 주시하면서 보상이 높은 쪽으로 학습(행동)하게 됨



1.2 기계학습 개요

□ 인공지능 발전

Y. LeCun

How Much Information is the Machine Given during Learning?

▶ “Pure” Reinforcement Learning (cherry)

- ▶ The machine predicts a scalar reward given once in a while.

▶ A few bits for some samples

▶ Supervised Learning (icing)

- ▶ The machine predicts a category or a few numbers for each input
- ▶ Predicting human-supplied data
- ▶ 10→10,000 bits per sample

▶ Self-Supervised Learning (cake génoise)

- ▶ The machine predicts any part of its input for any observed part.
- ▶ Predicts future frames in videos
- ▶ Millions of bits per sample

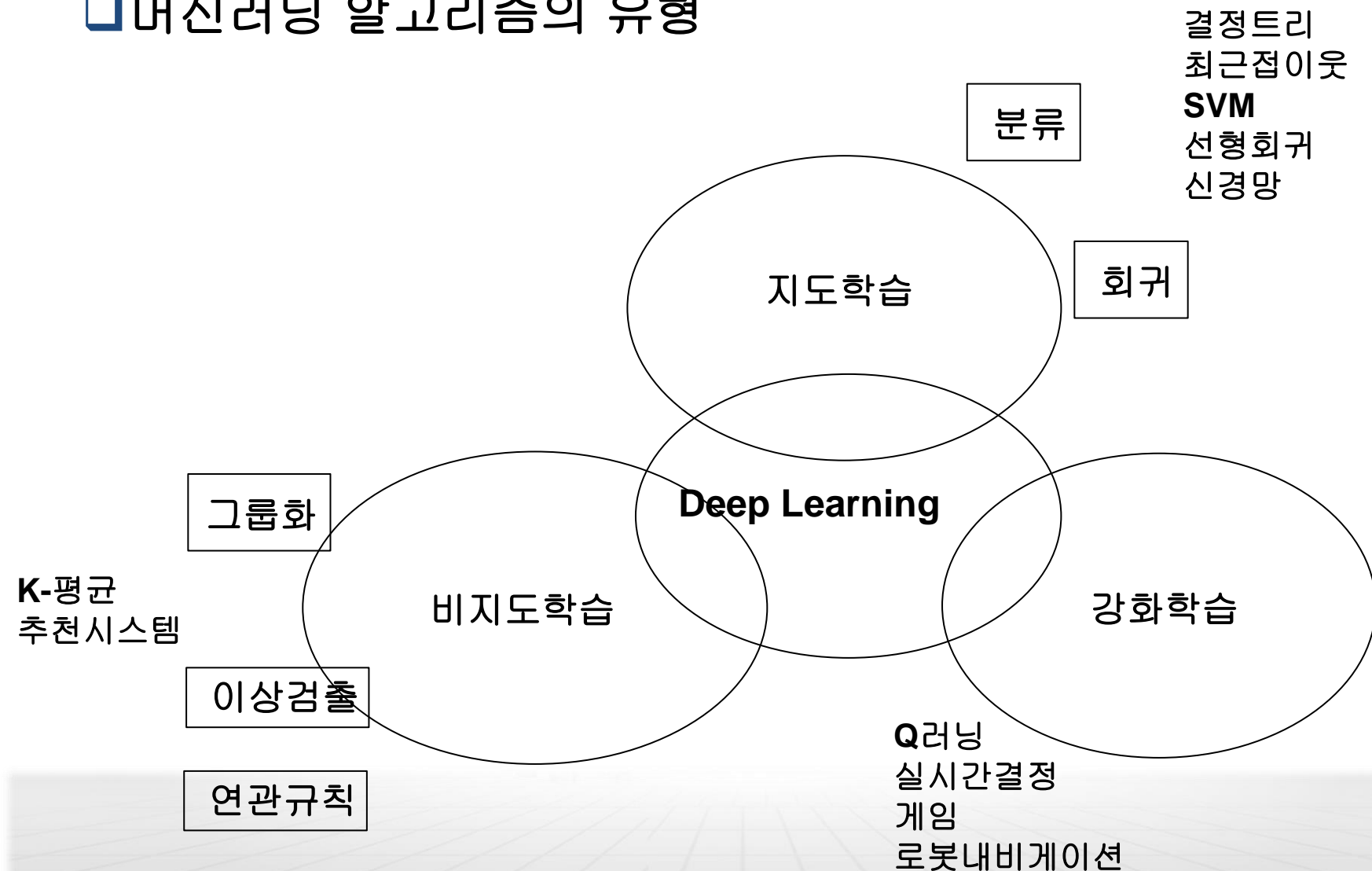


□ 머신러닝 알고리즘의 유형

- 지도학습 : 분류와 예측
- 비지도학습 : 군집
- 강화학습 : 큐러닝과 딥큐러닝

1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

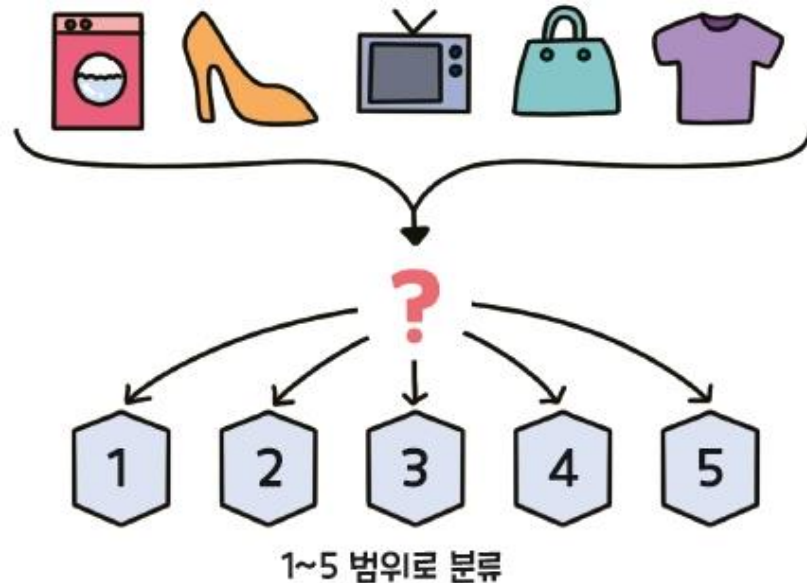


1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 분류(Classification)

- ✓ 레이블이 포함된 데이터를 학습하고 유사한 성질을 갖는 데이터끼리 분류한 후, 새로 입력된 데이터가 어느 그룹에 속하는지를 찾아내는 기법



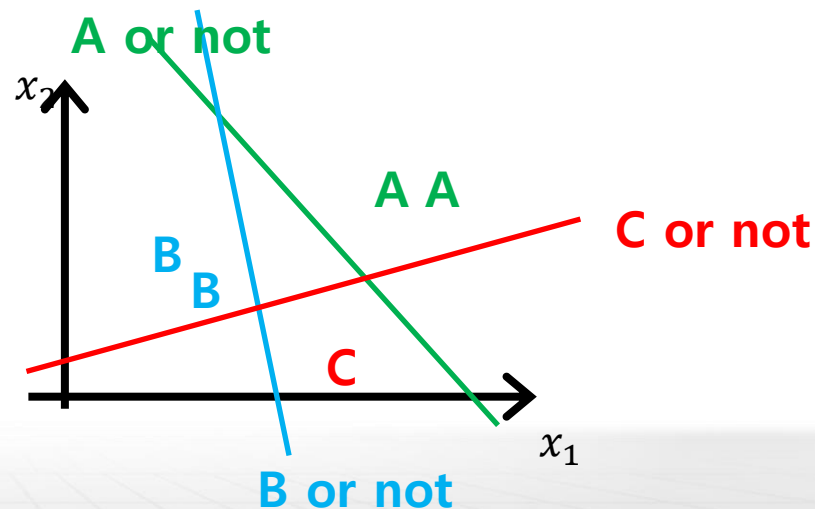
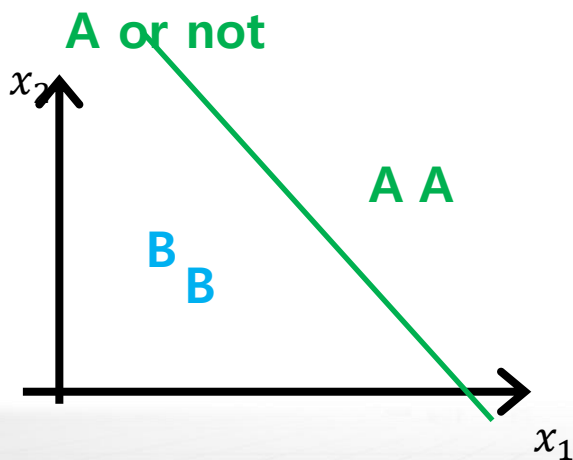
1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 분류(Classification)

✓ 분류의 종류

- 이진 분류(Binary Classification) : 데이터를 2개의 그룹으로 분류
- 다중 분류(Multiclass Classification) : 데이터를 3개의 그룹 이상으로 분류



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 분류(Classification)

✓ 분류에 해당하는 알고리즘

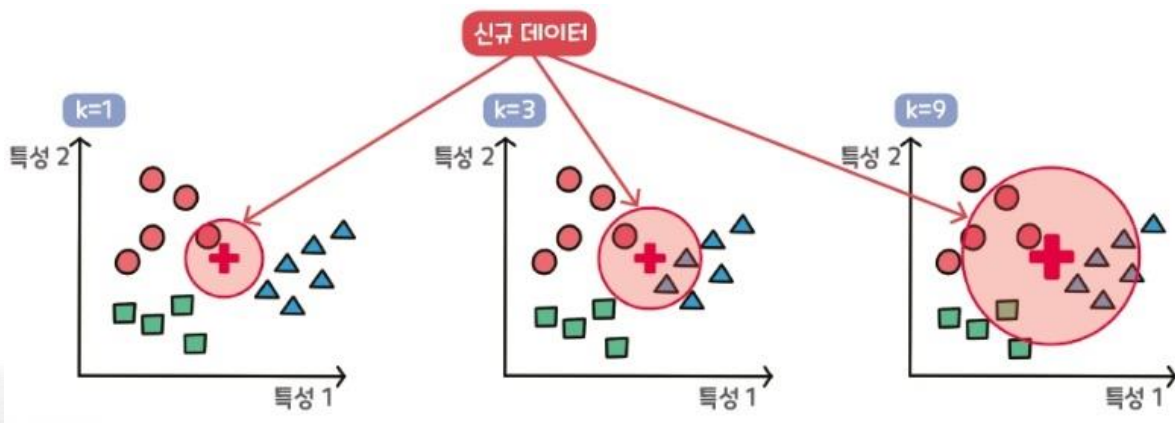
- K-최근접 이웃(KNN)
- 서포트 벡터 머신(SVM)
- 의사결정나무 (Decision Tree)
- 로지스틱 회귀(Logistic Regression)
- 신경망

1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ K-최근접 이웃(KNN, K-Nearest Neighbors)

- ✓ 새로운 데이터가 들어왔을 때 기존 데이터의 그룹(K개의 그룹) 중 어떤 그룹에 속하는지 분류하는 알고리즘
- ✓ (예) K=1일 때 신규 데이터가 입력되면 빨간 원으로 분류, K=3일 때 신규 데이터는 파란 삼각형으로 분류, K=9일 때도 파란 삼각형으로 분류 됨

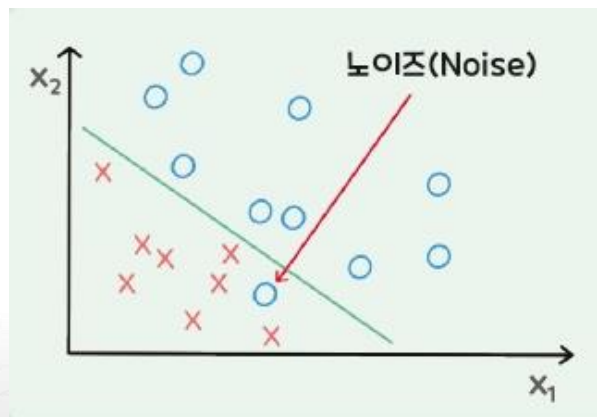


1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ K-최근접 이웃(KNN, K-Nearest Neighbors)

- ✓ KNN은 학습 데이터 내에 존재하는 노이즈의 영향을 크게 받지 않으며, 학습 데이터 수가 많을 때 꽤 효과적인 알고리즘
- ✓ 하지만 어떤 하이퍼파라미터가 분석에 적합한지는 불분명해, 데이터 각각의 특성에 맞게 연구자가 임의로 선정해야 한다는 단점이 있음
- ✓ 노이즈(Noise) : 데이터에 무작위의 오류(Random Error) 또는 분산(Variance)이 존재하는 것임



1.2 기계학습 개요

□머신러닝 알고리즘의 유형

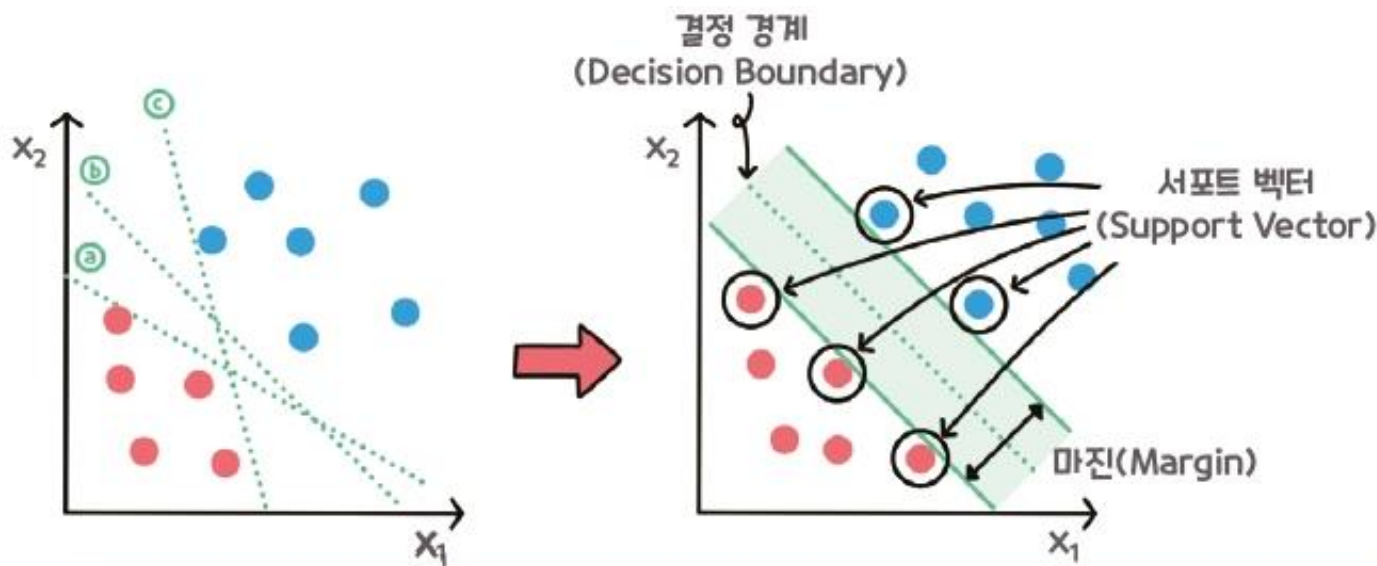
➤ 서포트 벡터 머신(SVM, Support Vector Machine)

- ✓ 주어진 데이터가 어느 그룹에 속하는지 분류하는 모델
- ✓ 두 분류 사이의 여백을 의미하는 마진을 최대화하는 방향으로 데이터를 분류
- ✓ SVM은 마진을 극대화하는 선을 찾아 분류하므로 마진이 크면 클수록 새로운 데이터가 들어오더라도 잘 분류할 가능성이 높아짐
- ✓ SVM은 사용 방법이 쉽고 예측 정확도가 높다는 장점
- ✓ 하지만 모델 구축에 시간이 오래 걸리고 결과에 대한 설명력이 떨어지는 단점

1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 서포트 벡터 머신(SVM, Support Vector Machine)



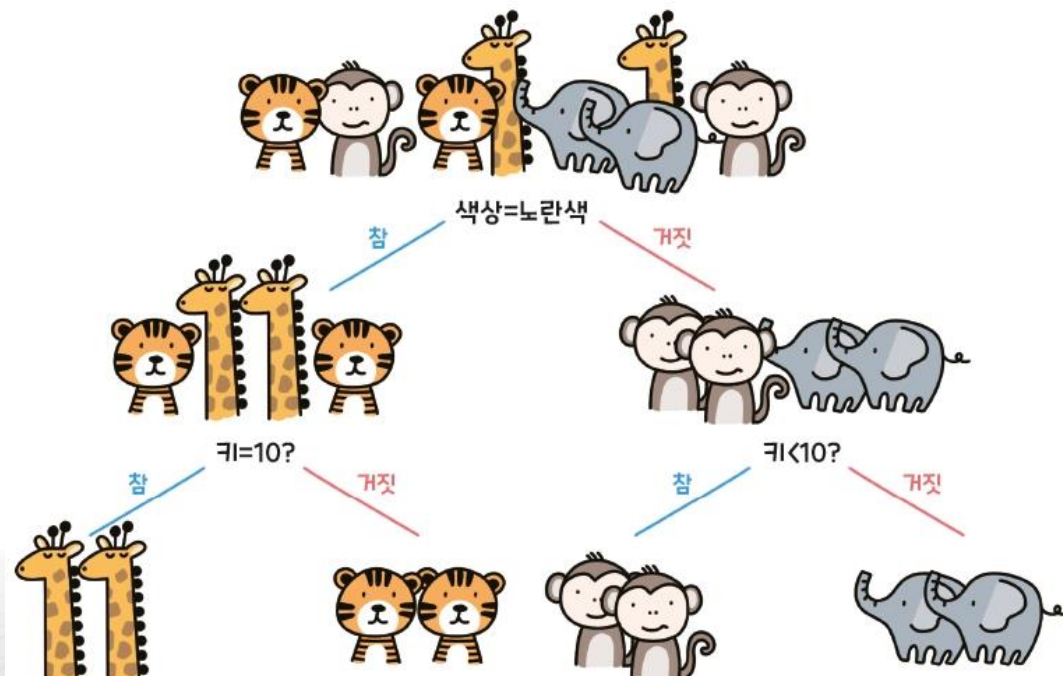
- 결정 경계(Decision Boundary) : 분류를 위한 기준선
- 서포트 벡터(Support Vector) : 결정 경계와 가장 가까운 위치에 있는 데이터
- 마진(Margin) : 결정 경계와 서포트 벡터 사이의 거리

1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 의사결정나무(Decision Tree)

- ✓ 의사결정 규칙을 나무 형태로 분류하는 분석 방법
- ✓ 상위 노드에서 시작하여 분류 기준값에 따라 하위 노드로 확장하는 방식이 '나무'를 닮았다고 하여 '의사결정나무'라고 불림



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 의사결정나무(Decision Tree)

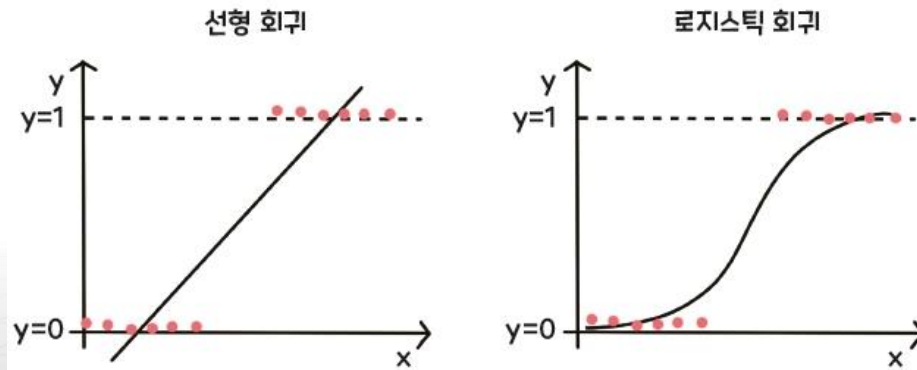
- ✓ 의사결정나무는 분석 과정이 직관적이고 이해하기 쉬움
- ✓ 인공신경망의 경우 분석 결과에 대한 설명이 어려운 블랙박스 모델인 반면, 의사결정나무는 분석 과정을 눈으로도 관측할 수 있음
- ✓ 그래서 결과에 대한 명확한 설명이 필요할 때 많이 사용함

1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 로지스틱 회귀

- ✓ 로지스틱 회귀는 선형 회귀와는 다르게 종속변수가 범주형 데이터
- ✓ 즉, 입력 데이터가 주어졌을 때 해당 데이터의 결과가 0과 1 사이의 값을 가짐
- ✓ 결괏값이 정해진 범주 내에서 나오므로 확률적인 의미에서 사건 발생 가능성을 예측하는 데 사용할 수 있음
- ✓ 선형 회귀는 종속변수로 올 수 있는 값에 대한 제약이 없는 반면, 로지스틱 회귀의 종속변수는 값이 제한적이라는 것에 주목해야 함

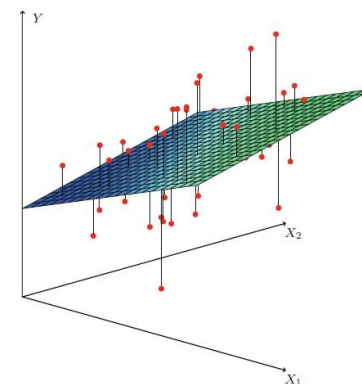
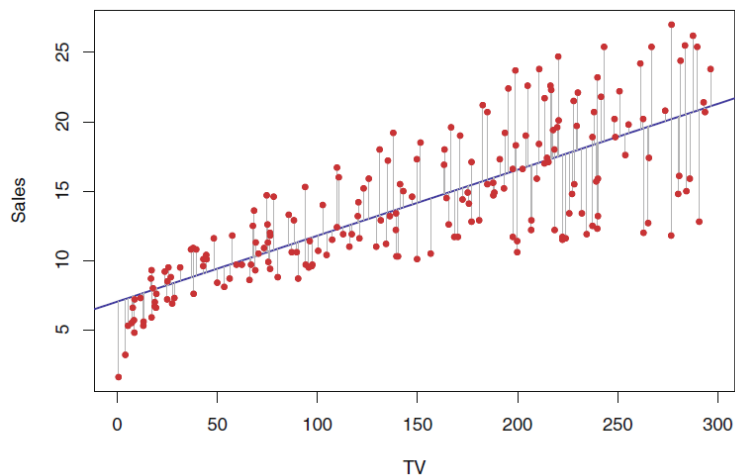
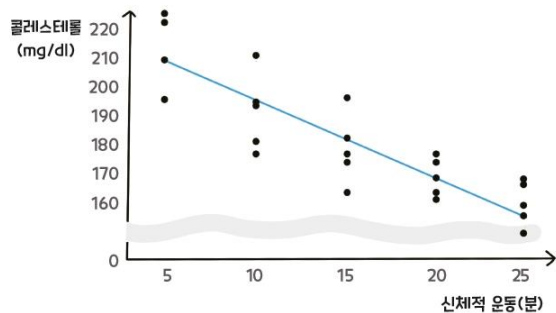


1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 회귀 (Regression)

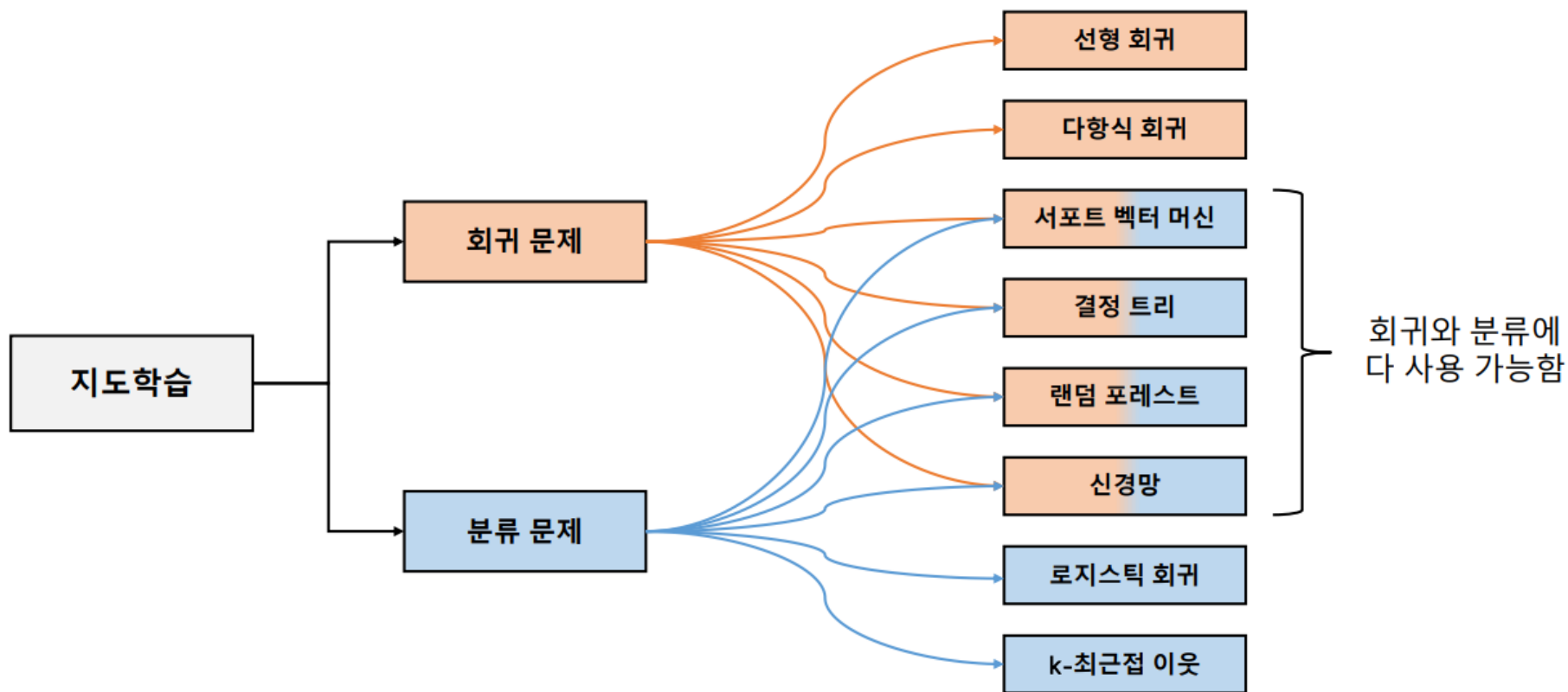
- ✓ 주어진 입력-출력 쌍을 학습한 후에 새로운 입력값이 들어왔을 때, 합리적인 출력값을 예측
- ✓ 연속형 변수들에 대해 변수 간 관계를 추정하는 분석 방법이며, 선형 회귀는 독립변수와 종속변수가 직선의 형태를 취하는 관계



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

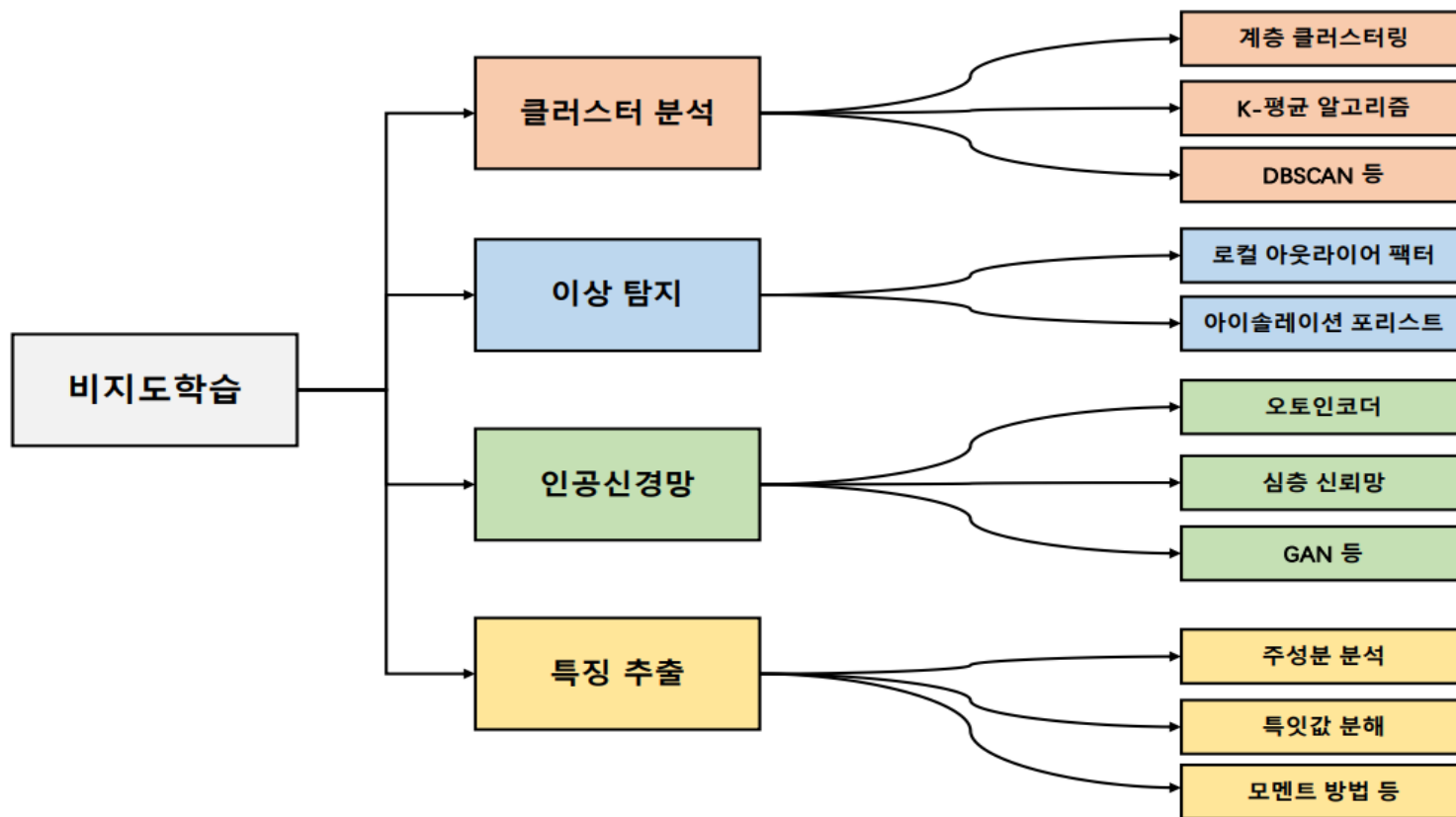
➤ 지도학습



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 비지도학습



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 군집

- ✓ k-평균, DBSCAN, 계층 군집 분석
- ✓ 이상치 탐지와 특이치 탐지
- ✓ 원-클래스 SVM, 아이솔레이션 포레스트

➤ 시각화와 차원 축소

- ✓ 주성분 분석(PCA), 커널 PCA
- ✓ 지역적 선형 임베딩, t-SNE

➤ 연관 규칙 학습

- ✓ 어프라이어리(Apriori)
- ✓ 이클렛(Eclat)

1.2 기계학습 개요

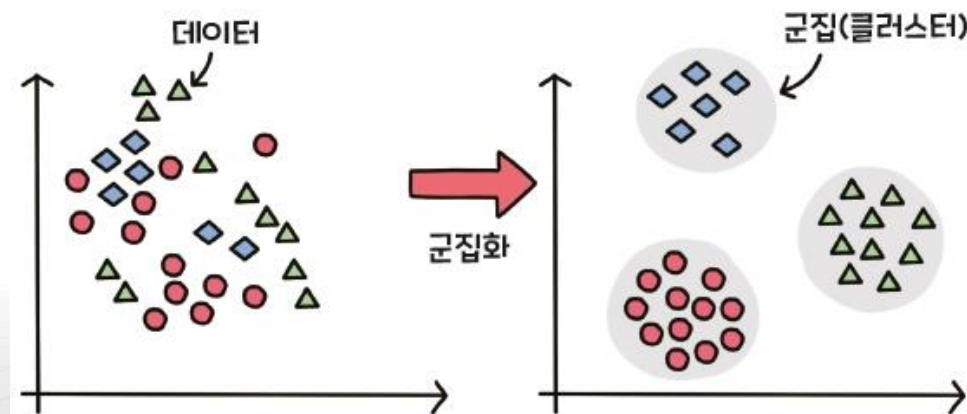
□머신러닝 알고리즘의 유형

➤ 군집(Cluster, 클러스터)

- ✓ 비슷한 특징을 가진 데이터들의 집단

➤ 군집화(Clustering, 클러스터링)

- ✓ 데이터가 주어졌을 때 그 데이터들을 유사한 정도에 따라 군집으로 분류하는 것
- ✓ 왼쪽 그래프를 보면 다양한 데이터들이 서로 섞여 있지만, 군집화 과정을 진행하면 오른쪽 그래프와 같이 비슷한 데이터끼리 군집으로 묶임



1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 군집화(Clustering, 클러스터링)

- ✓ k-평균 군집화(K-Means Clustering)
- ✓ 'K'는 주어진 데이터로부터 묶여질 그룹(군집의 수)
- ✓ 'Means'는 각 군집의 중심과 데이터들의 평균 거리를 의미
- ✓ 클러스터의 중심을 중심점(Centroids)이라고 함

1.2 기계학습 개요

□ 머신러닝 알고리즘의 유형

➤ 군집화(Clustering, 클러스터링)

- (a) : 일반적인 데이터 분포입니다.
- (b) : 데이터셋에서 K개의 중심점을 임의로 지정하는데, 여기에서는 $K=2$ 의 값으로 중심점 2개를 설정했습니다.
- (c) : 데이터들을 가장 가까운 중심점에 할당합니다.
- (d) : (c)에서 할당된 결과를 바탕으로 중심점을 새롭게 지정합니다.
- (e) : 중심점이 더 이상 변하지 않을 때까지 (c)~(d) 과정을 반복합니다.
- (f) : 최종적인 군집이 형성됩니다.

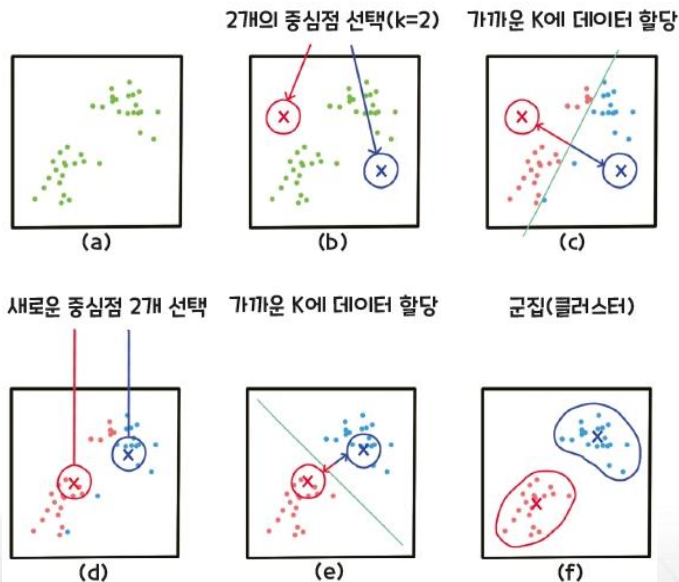


그림 8-23 K-평균 군집화 과정

□ 머신러닝 알고리즘의 유형

➤ 밀도기반 클러스터링(DBSCAN)

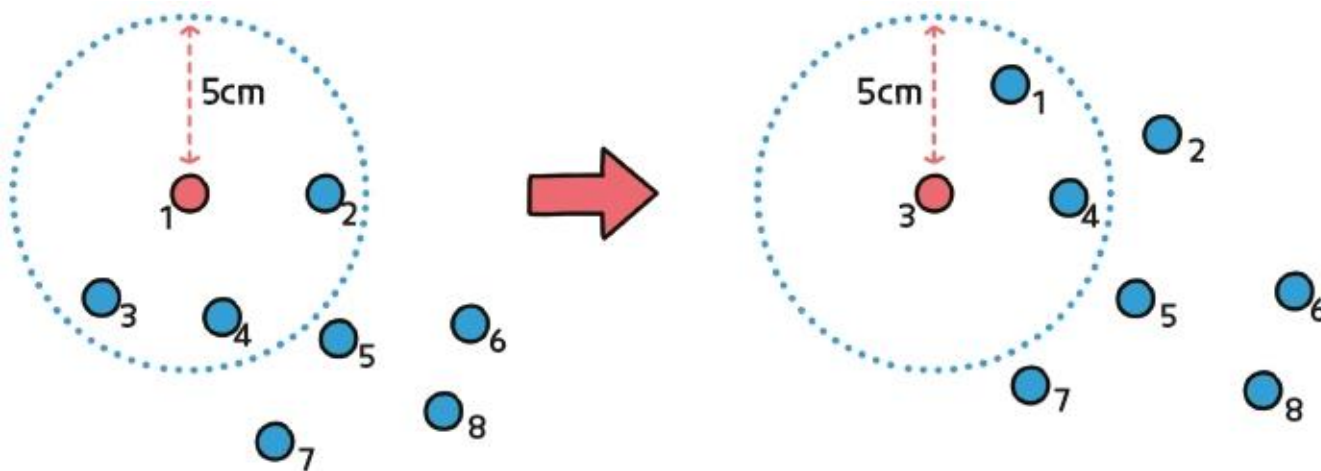
- ✓ 밀도를 기반으로 군집화하는 매우 유용한 군집 알고리즘
- ✓ 밀도기반 클러스터링은 데이터들의 분포
- ✓ 밀도기반 클러스터링을 이해하기 위한 관련 용어
 - ϵ (Epsilon, 거리) : 하나의 점으로부터의 반경
 - minPth(Minimum Points, 최소점) : 군집을 이루기 위한 최소한의 데이터 수

1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 밀도기반 클러스터링(DBSCAN)

- ✓ 밀도기반 클러스터링의 진행 과정($\epsilon=5\text{cm}$, $\text{minPth}=4$ 라고 가정)
- ✓ 1단계 : 한 점을 중심으로 반경 5cm 거리에 4개의 데이터가 있는지 ($\text{minPth}=4$ 를 만족하는지) 확인

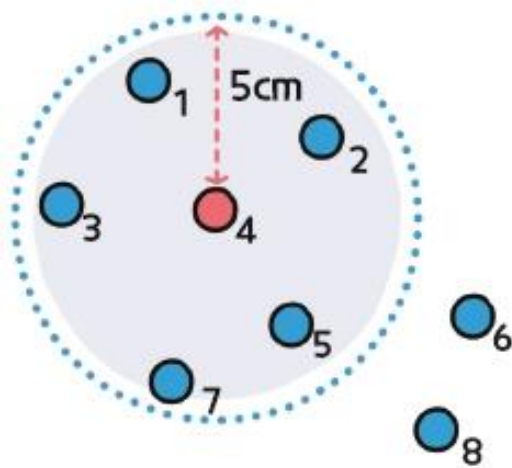


1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 밀도기반 클러스터링(DBSCAN)

- ✓ 2단계 : 이동한 중심점 3을 기준으로 1단계를 반복하는데, 3을 기준으로 반경 5cm 이내에 데이터가 4개 초과 있는지 확인하면, 역시 데이터의 수가 4보다 작으므로 이번에는 중심점을 4로 지정함
- ✓ 3단계 : 4를 중심점으로 했을 때 데이터의 수가 4를 초과하므로 군집이 생성



1.2 기계학습 개요

□머신러닝 알고리즘의 유형

➤ 밀도기반 클러스터링(DBSCAN)

- ✓ K-평균 군집화와 달리 밀도기반 클러스터링은 클러스터 수를 지정할 필요가 없음
- ✓ 더 중요한 것은 밀도기반 클러스터링은 K-평균 군집화가 찾을 수 없는 임의의 모양들을 가질 수 있다는 점임
- ✓ 예를 들어, 밀도기반 클러스터링은 [그림 8-26]의 첫 번째 그림과 같이 다른 군집으로 둘러 싸인 상태에서 또 다른 군집을 가질 수 있음

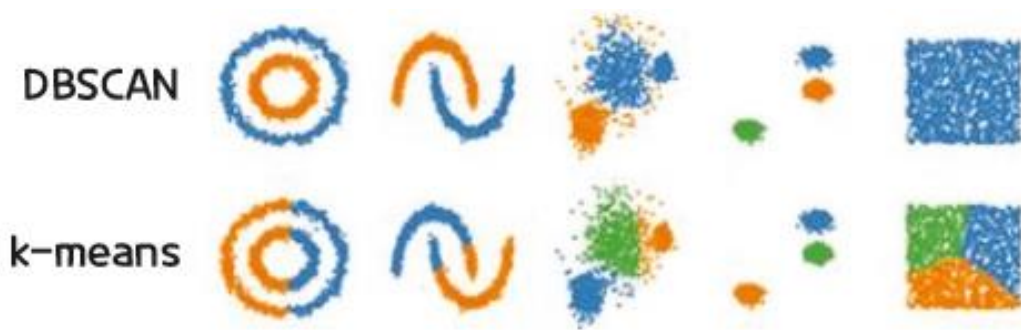
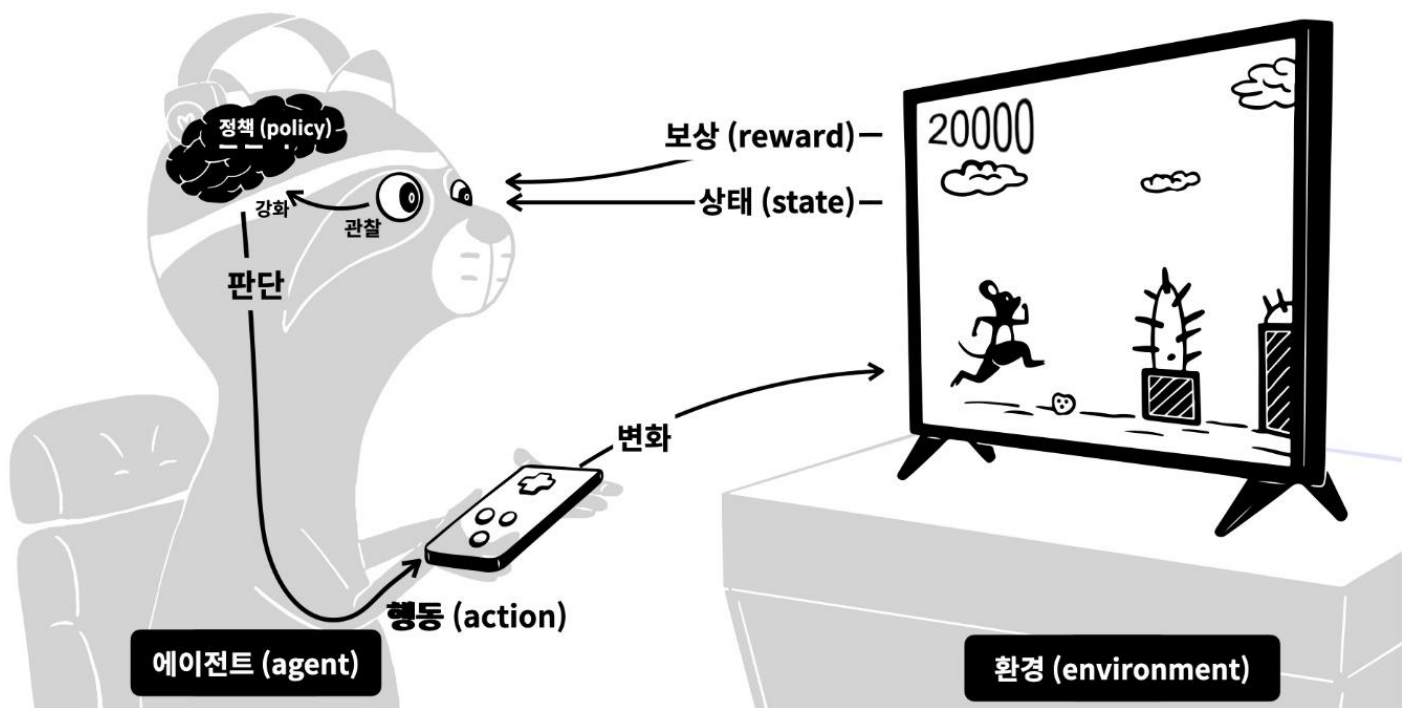


그림 8-26 밀도기반 클러스터링과 K-평균 군집화 비교

1.2 기계학습 개요

□ 강화학습 기술

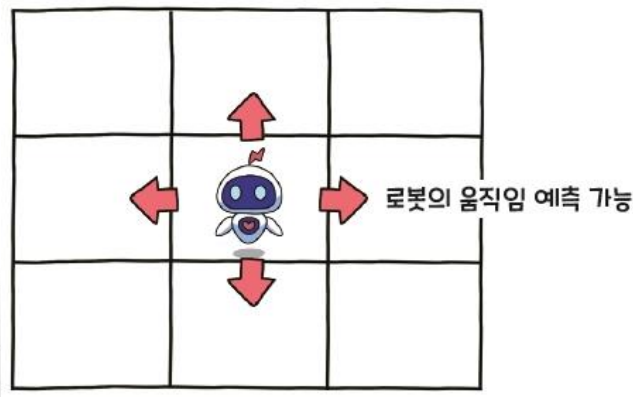


1.2 기계학습 개요

□ 강화학습 기술

➤ 모델이 없는 알고리즘

- ✓ 모델기반 알고리즘은 현재의 상태에서 어떤 행동을 했을 때 다음의 상태가 될 확률을 의미함
- ✓ 예를 들어, 격자 공간에서 로봇이 상하좌우로 이동할 때 로봇의 다음 상태에 대해 직관적으로 파악할 수 있음
- ✓ 모델기반 알고리즘은 이처럼 행동에 따른 상태의 변화를 예측할 수 있어 최적의 솔루션을 얻을 수 있음

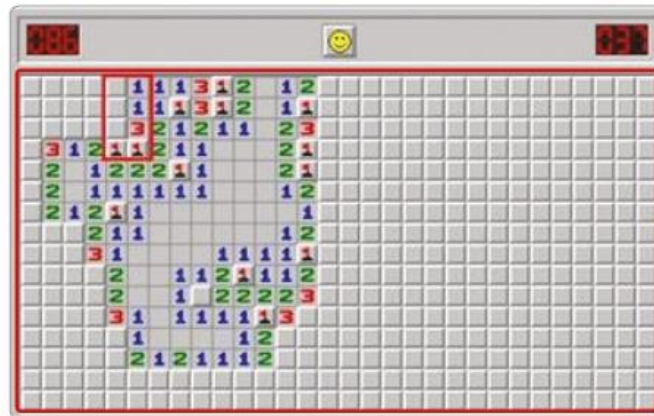


1.2 기계학습 개요

□ 강화학습 기술

➤ 모델이 없는 알고리즘

- ✓ 에이전트가 행동을 통해 받게 되는 보상을 최대로 하는 정책 (Policy)을 찾는 것
- ✓ (예) 지뢰찾기 게임

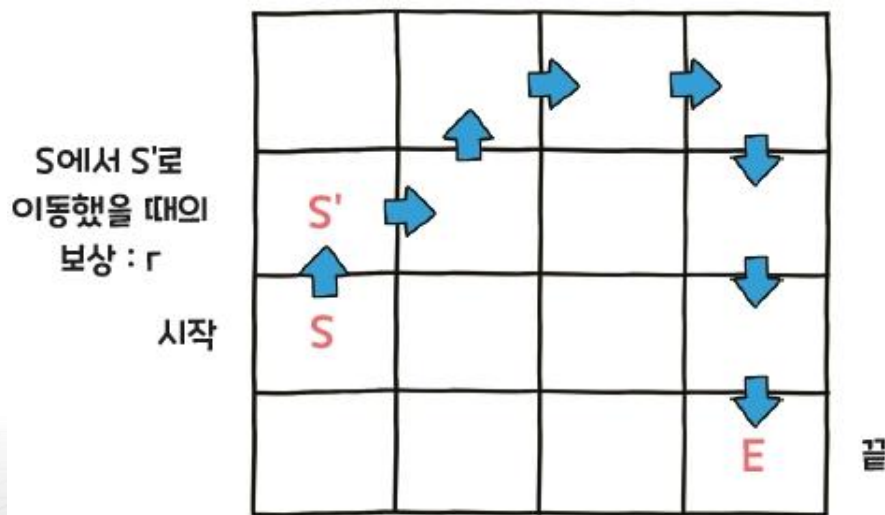


1.2 기계학습 개요

□ 강화학습 기술

➤ 큐러닝(Q-Learning)

- ✓ 특정 상태에서 어떤 결정을 내려야 미래 보상이 극대화될 것인지에 대한 정책을 지속적으로 업데이트하는 것
- ✓ 모델 없이 학습하는 대표적인 강화학습 알고리즘
- ✓ S에서 시작하여 E로 끝나는 미로게임이 있다고 가정



1.2 기계학습 개요

□ 강화학습 기술

➤ 큐러닝(Q-Learning)

- ✓ S에서 시작한 미로게임은 우왕좌왕하면서 우연히 E에 도달
- ✓ 에이전트는 S에서 S'로 이동했을 때 첫 번째 보상값(r)을 받음
- ✓ 이후로는 E에 도달하기 위한 움직임을 계속하면서 보상값을 업데이트해 나갈 것
 - 모든 환경 데이터값(상태, 행동)을 초기화함
 - 현재 상태(S)를 확인함
 - 다음의 작업을 반복함
 - S'로 이동
 - 행동에 따른 보상값(r)을 받음
 - 목적지에 도착할 때까지 이동과 보상값을 [그림 8-30]처럼 테이블에 기록

1.2 기계학습 개요

□ 강화학습 기술

➤ 큐러닝(Q-Learning)

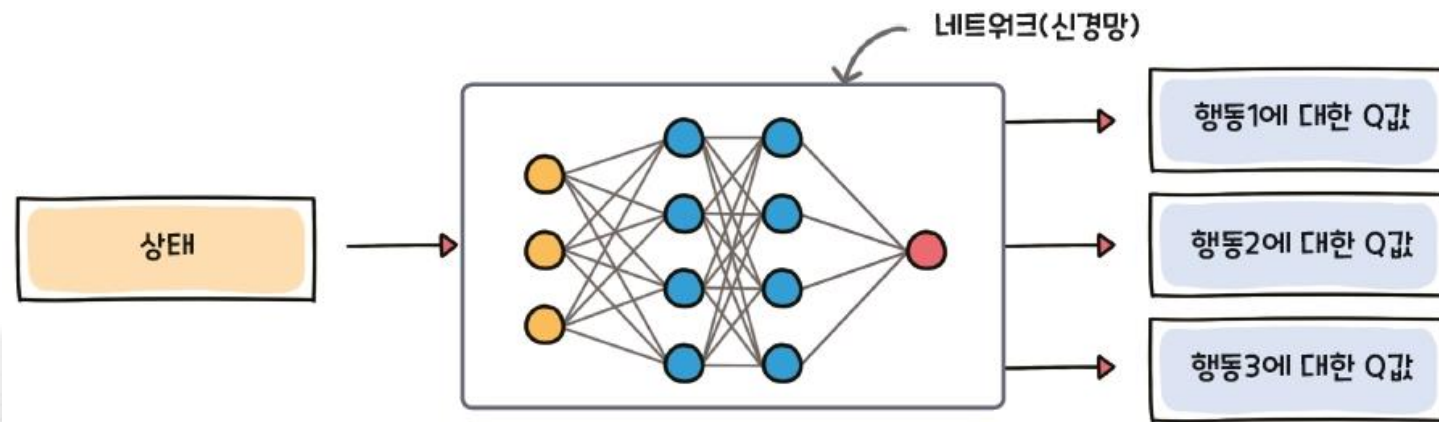


1.2 기계학습 개요

□ 강화학습 기술

➤ 딥큐러닝(DQN: Deep Q Learning)

- ✓ 큐러닝에 신경망을 결합한 알고리즘
- ✓ 큐러닝에서는 보상값(r)을 업데이트하기 위해 테이블을 이용했다면, DQN에서는 네트워크(신경망)를 이용
- ✓ Q값은 전략에 따라 행동했을 때 미래의 보상들에 대한 기댓값의 총합
- ✓ 결국 큐러닝과 DQN 모두 Q값이 높은 쪽으로 행동하는 것을 목표로 함



1.2 기계학습 개요

머신러닝 알고리즘

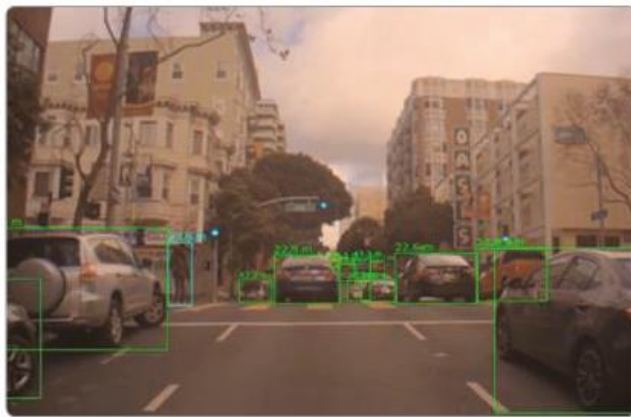


1.2 기계학습 개요

□ 머신러닝의 주요 도전 과제

➤ 많은 데이터 확보

- ✓ 컴퓨터는 인간과 다르게 0, 1이라는 숫자만 인식할 수 있음
- ✓ 즉, 이미지 안의 객체를 숫자로 표현해야 하고 객체 중에서도 자동차만을 인식할 수 있도록 별도의 처리(머신러닝의 가중치 (weight)) 를 해주어야 한다는 뜻
- ✓ 이렇게 복잡한 과정을 거치면서 다양한 유형의 자동차를 정확하게 인식하기 위해서는 수많은 데이터가 필요함



1.2 기계학습 개요

□ 머신러닝의 주요 도전 과제

➤ 많은 데이터 확보

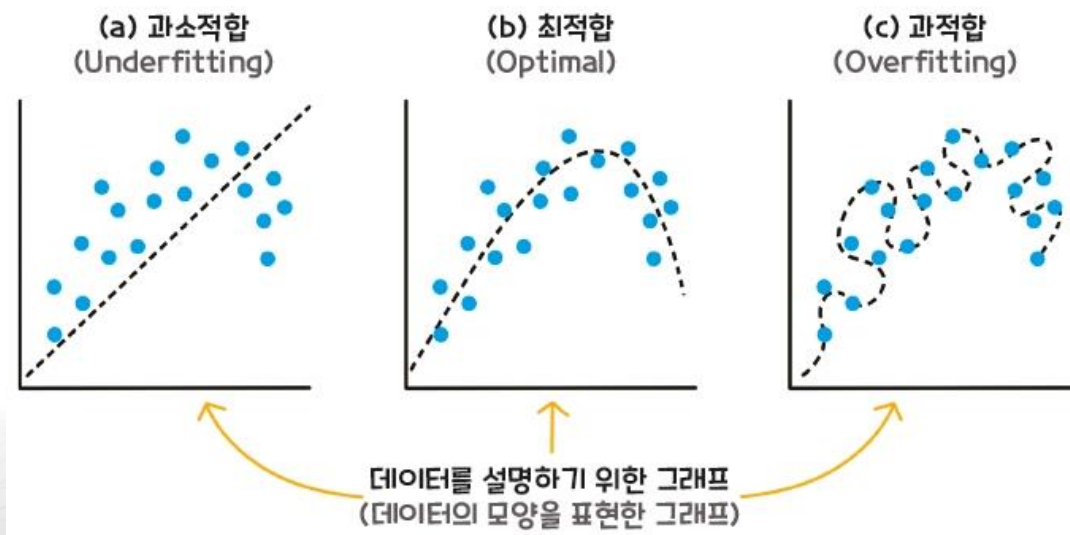
- ✓ 가중치(Weight) : 입력 신호가 출력에 미치는 중요도를 조절하는 매개변수
 - 앞의 이미지에는 건물과 자동차들이 있음
 - 각각의 객체에 단순히 숫자를 부여한다면 건물1=00001, 자동차1=00002, 건물2=00003, 자동차2=00004와 같이 부여되어 컴퓨터는 건물과 자동차를 구별할 수 없을 것
 - 그래서 건물1=00001, 자동차1=11110과 같이 자동차에 더 높은 숫자를 부여하여 건물과 자동차를 구별할 수 있도록 한 것이 가중치임
- ✓ 실제 객체 인식 과정에서 사용되는 가중치는 더 복잡한 과정을 거치지만, 간단하게는 중요한 객체를 부각시키기 위한 값이라고 이해해도 좋음

1.2 기계학습 개요

□ 머신러닝의 주요 도전 과제

➤ 과적합(Overfitting)

- ✓ 훈련 데이터를 너무 과하게 학습하여 실제 데이터를 분석할 때는 성능이 좋지 못한 것을 의미
- ✓ 문제의 복잡도에 비해 데이터가 현저히 부족한 경우, 즉 문제가 정의된 전체 공간을 학습 데이터가 아우르지 못하고 일부 경우에만 집중했을 때 발생함



1.2 기계학습 개요

□ 머신러닝의 주요 도전 과제

➤ 유연성 부족

- ✓ 머신러닝은 유연성이 부족함
- ✓ 머신러닝은 데이터로 시작해서 데이터로 끝나는 기술
- ✓ 다른 사람이 만들어 놓은 모델은 재활용이 가능하더라도 데이터는 공유 어려움
- ✓ 실제로 공유 된 데이터를 사용할 수도 있지만 분석하고자 하는 변수 중 일부가 누락된 경우가 많기 때문에 공유 된 데이터를 이용한 분석은 그 목적에서 벗어난 경우가 많음
- ✓ 따라서 제대로 학습을 하려면 원하는 결과를 위한, 목적에 맞는 ‘나만의 데이터’가 필요함
- ✓ 결국 데이터가 없다면 머신러닝 알고리즘도, 딥러닝 알고리즘도 적용 어려움

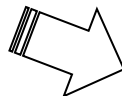
1.2 기계학습 개요

□ 머신러닝의 주요 도전 과제

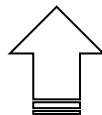
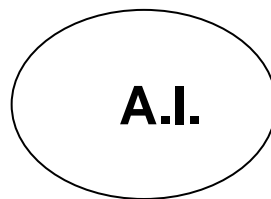
➤ 인공지능 발전

꽃을 피우게 될 학문, 미완성 학문

Data, data, and lots of data



Improved algorithms



Tremendous computing resource

Thank you

Q & A

