

MACHINE LEARNING 기계 학습

1장. 소개

PREVIEW

■ 사람의 학습

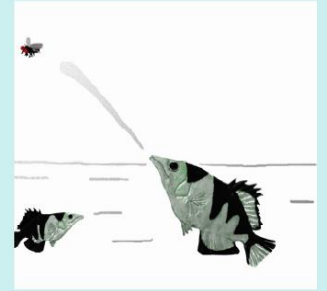
- 수학, 과학, 역사뿐 아니라 수영, 자전거 타기 등

■ 동물의 학습

- 예) 물총물고기의 목표물 맞추기 능력 향상



(a) 자전거 타기 학습

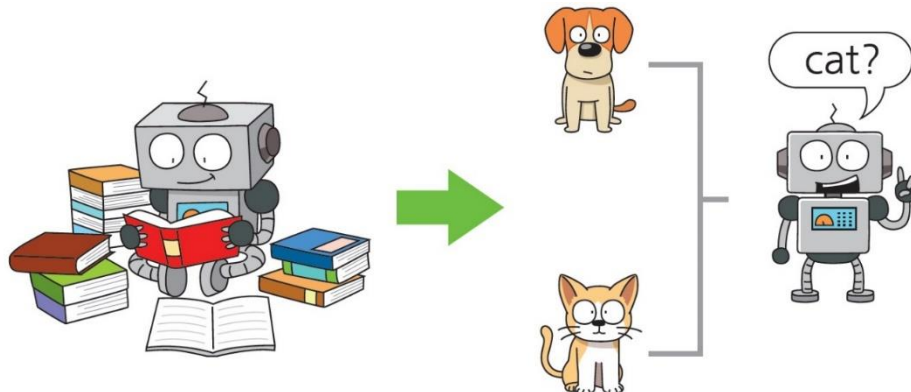


(b) 물총물고기의 사냥

그림 1-1 사람과 동물의 학습

■ 기계 학습

- 그렇다면 기계도 학습할 수 있을까?
- 경험을 통해 점점 성능이 좋아지는 기계를 만들 수 있을까?
- 이 책은 이 질문에 대한 답을 찾아가는 길



1.1 기계 학습의 정의

■ 학습이란? <표준국어대사전>

“경험의 결과로 나타나는, 비교적 지속적인 행동의 변화나 그 잠재력의 변화. 또는 지식을 습득하는 과정[국립국어원2017]”

■ 기계 학습이란?

■ 인공지능 초창기 사무엘의 정의

“Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort. 컴퓨터가 경험을 통해 학습할 수 있도록 프로그래밍할 수 있다면, 세세하게 프로그래밍해야 하는 번거로움에서 벗어날 수 있다[Samuel1959].”

- “데이터를 이용해서 명시적으로 정의되지 않은 패턴을 컴퓨터로 학습하여 결과를 만들어 내는 학문 분야 ”

1.1.2 지식기반 방식에서 기계 학습으로의 대전환

■ 인공지능의 탄생

- 컴퓨터의 뛰어난 능력
 - 사람이 어려워하는 일을 아주 쉽게 함
 - $80932.46789076 \times 0.39001324$ 와 같은 곱셈을 고속으로 수행(현재는 초당 수십억개)
 - 복잡한 함수의 미분과 적분 척척
- 컴퓨터에 대한 기대감 (컴퓨터의 능력 과신)
 - 사람이 쉽게 하는 일, 예를 들어 고양이/개 구별하는 일도 잘 하지 않을까
 - 1950년대에 인공지능이라는 분야 등장

■ 초창기는 지식기반 방식이 주류

- 예) "구멍이 2개이고 중간 부분이 홀쭉하며, 맨 위와 아래가 둥근 모양이라면 8이다"

인공지능의 역사

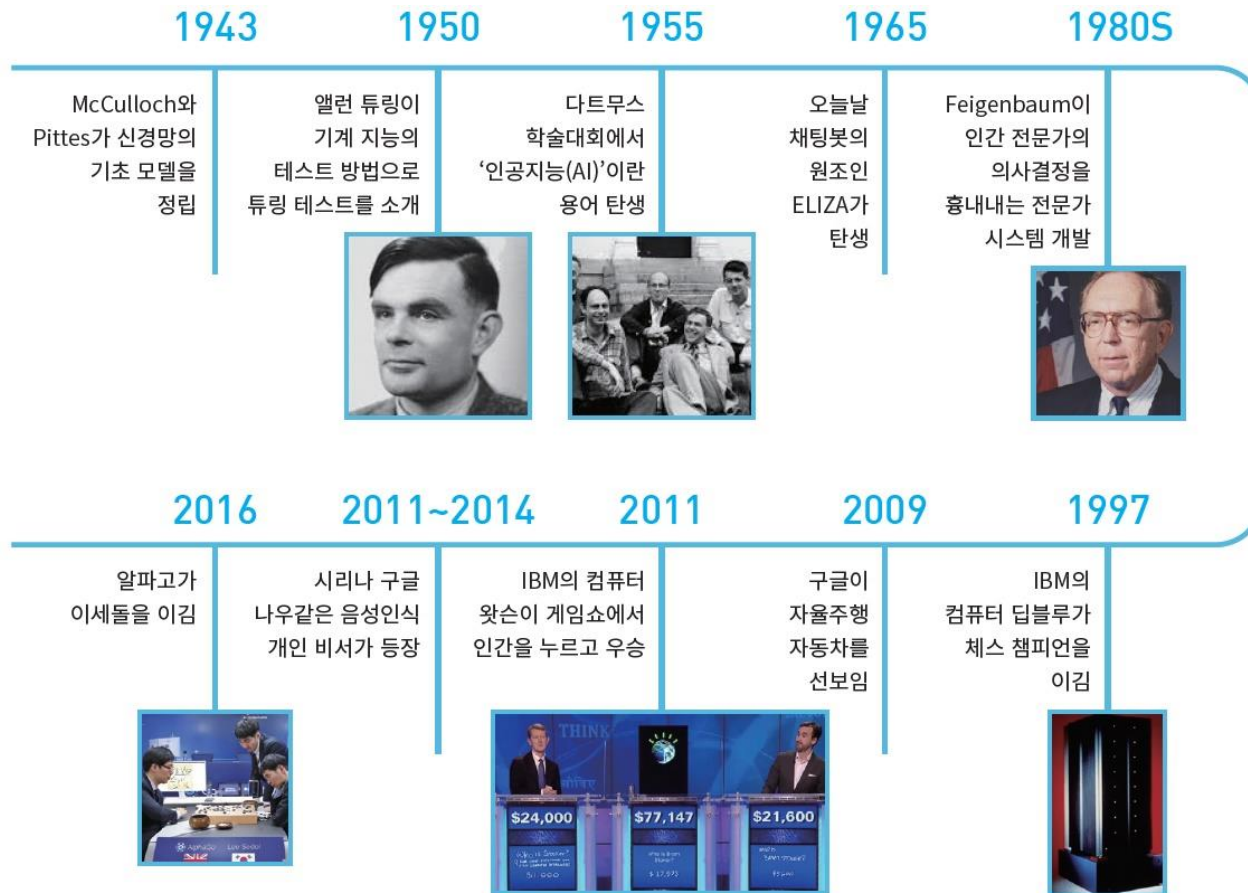


그림 1-17 인공지능의 역사

앨런 튜링

■ 1950년 앨런 튜링(Alan Turing)

- '생각하는 기계'를 만들 가능성을 고찰한 획기적인 논문을 출간
 - "Computing Machinery and Intelligence" 논문
- '생각'은 정의하기 어렵다는 것을 지적하고 유명한 튜링 테스트를 고안



인공지능 vs 기계학습 vs 딥러닝

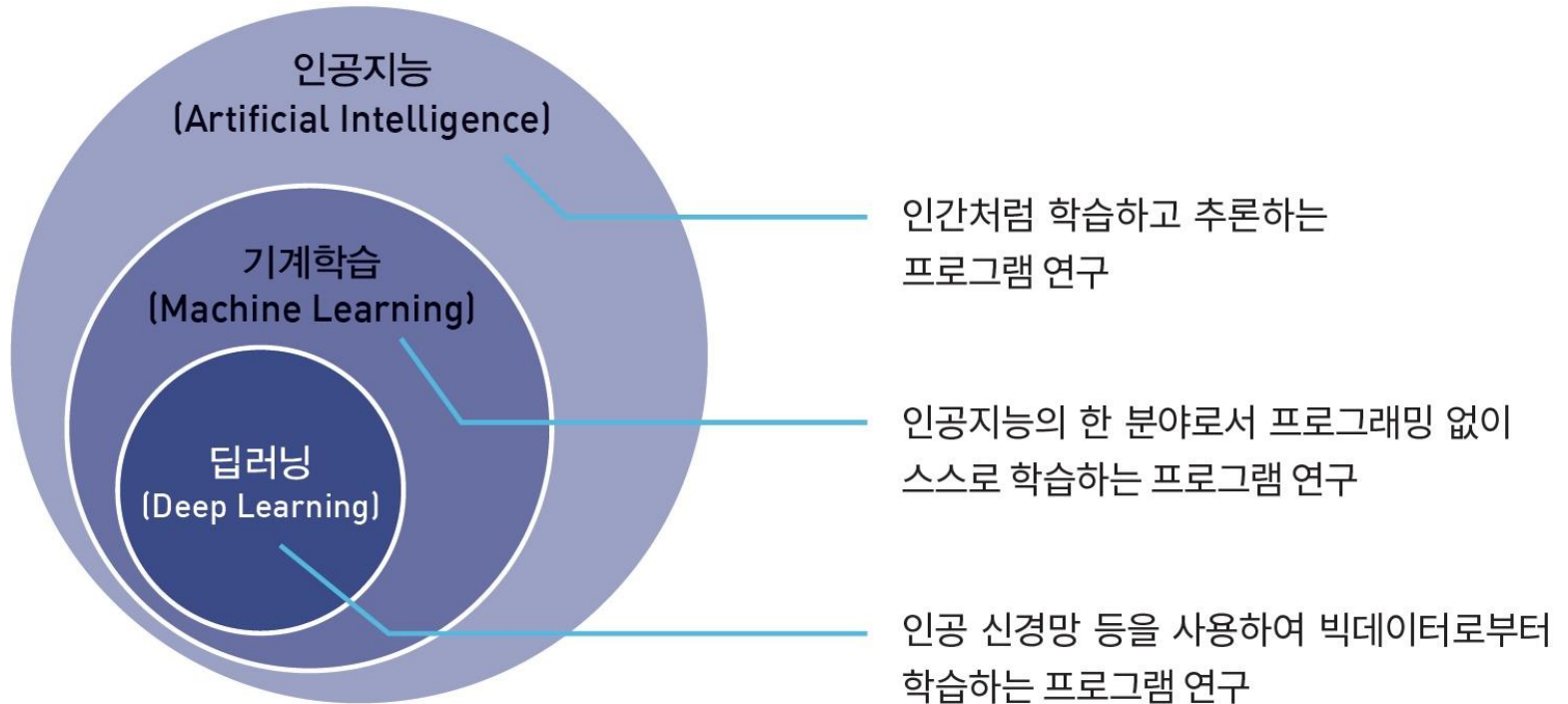


그림 1-9 인공지능, 기계학습, 딥러닝의 관계

머신러닝 : 데이터 + 패턴인식 + 컴퓨터 계산

■ 데이터

- 머신러닝은 항상 데이터를 기반으로 함
- 데이터를 기반으로 한다는 점에서 통계학과 가장 가까움

■ 패턴인식

- 사용자가 일일이 정해놓은 패턴으로 데이터를 분석하는 것이 아니라 데이터를 보고 패턴을 추리는 것이 머신러닝의 핵심
- 통계학: 데이터에서 패턴을 찾아내는 학문(머신러닝의 가장 기본적이며 핵심적인 개념)

■ 컴퓨터 계산

- 계산 속도를 높이고, 또 더 많은 데이터를 효율적으로 다루어야 함
- 분산 처리 등 시스템 구성에 관한 연구
- 단순히 수학적 모델의 구축이나 증명에만 그치는 것이 아니라 실제 데이터에 대해 계산해서 결과를 만들어 낸다는 점에서 전산학의 한 분야로 볼 수 있음

1.2 머신러닝을 이해하는데 필요한 배경 지식

■ 수학(선형대수, 미분)

- 행렬 연산(행렬곱)과 역행렬 개념이 많이 사용
- 깊이 있는 내용을 이해하려면 대학에서 배우는 선형대수(행렬 분해에 대한 지식 및 미분에 대한 지식)

■ 통계학 (통계, 확률)

- 머신러닝의 많은 부분이 통계학과 접근 방법이 비슷
- 분포란 무엇인지, 정상분포 혹은 가우스분포란 무엇인지, 왜 중요한지, 상관관계란 무엇인지, 회귀란 무엇인지 등
- 확률의 정의와 조건부 확률이 중요

■ 프로그래밍

- 머신러닝 라이브러리가 이미 많이 존재: 밑바닥부터 모든 것을 만들지 않아도 머신러닝 시스템을 구축할 수 있음

1.3 머신러닝 발전사

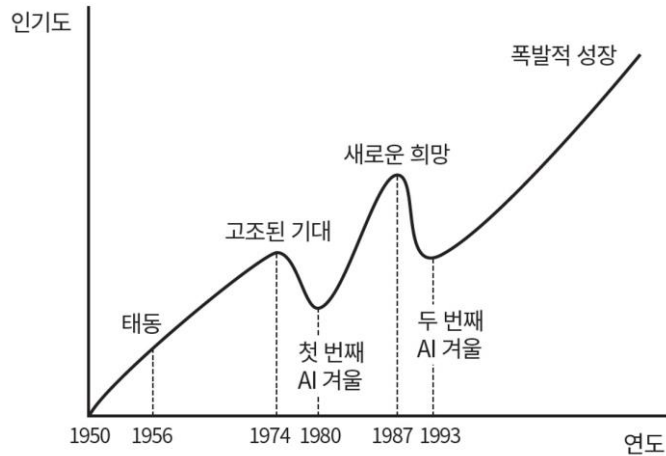
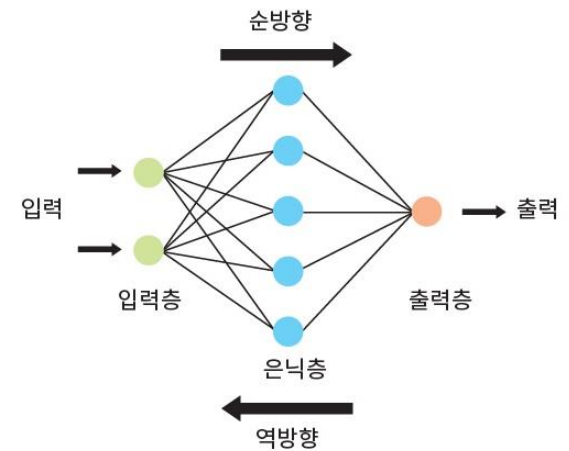
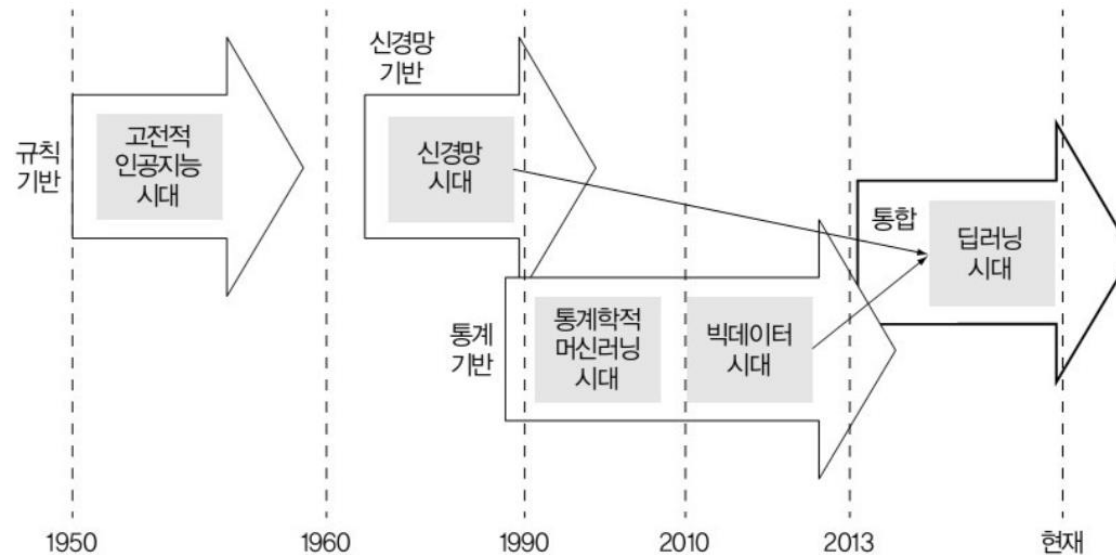


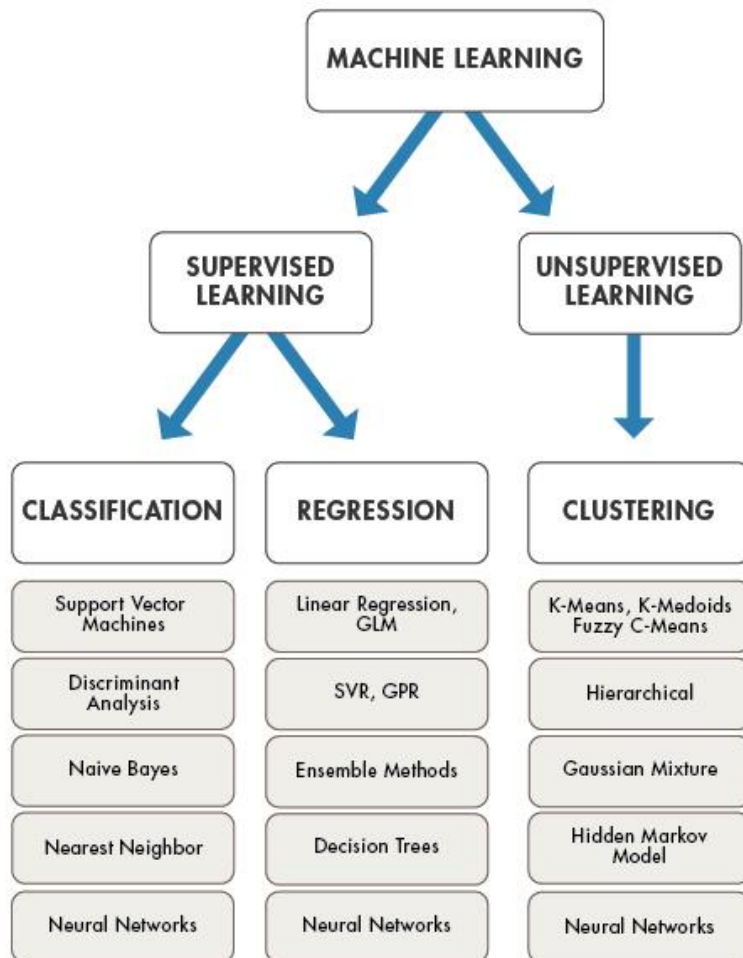
그림 1-2 머신러닝 변천사



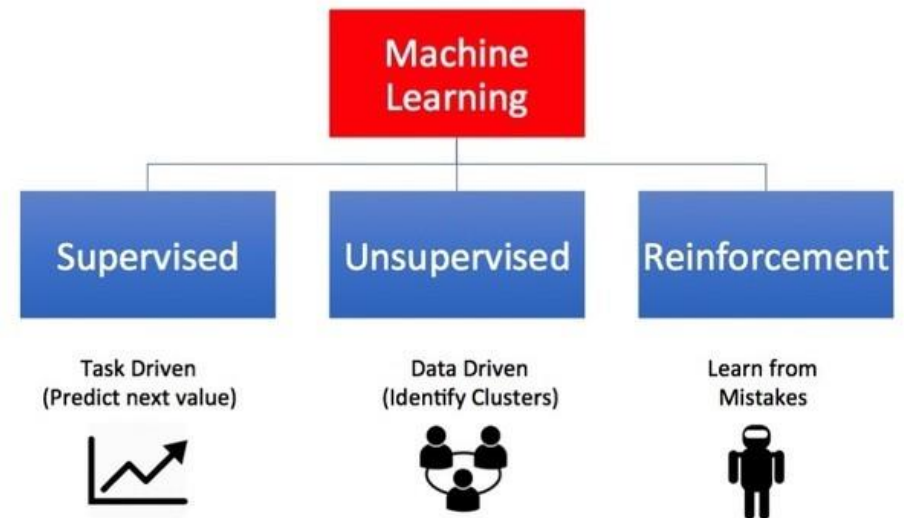
머신러닝의 분류

■ 풀고자 하는 부류에 따라 (강화 학습을 추가하여 3가지로 분류)

- 지도 학습: 모든 훈련 샘플이 레이블 정보를 가짐
- 비지도 학습: 모든 훈련 샘플이 레이블 정보를 가지지 않음

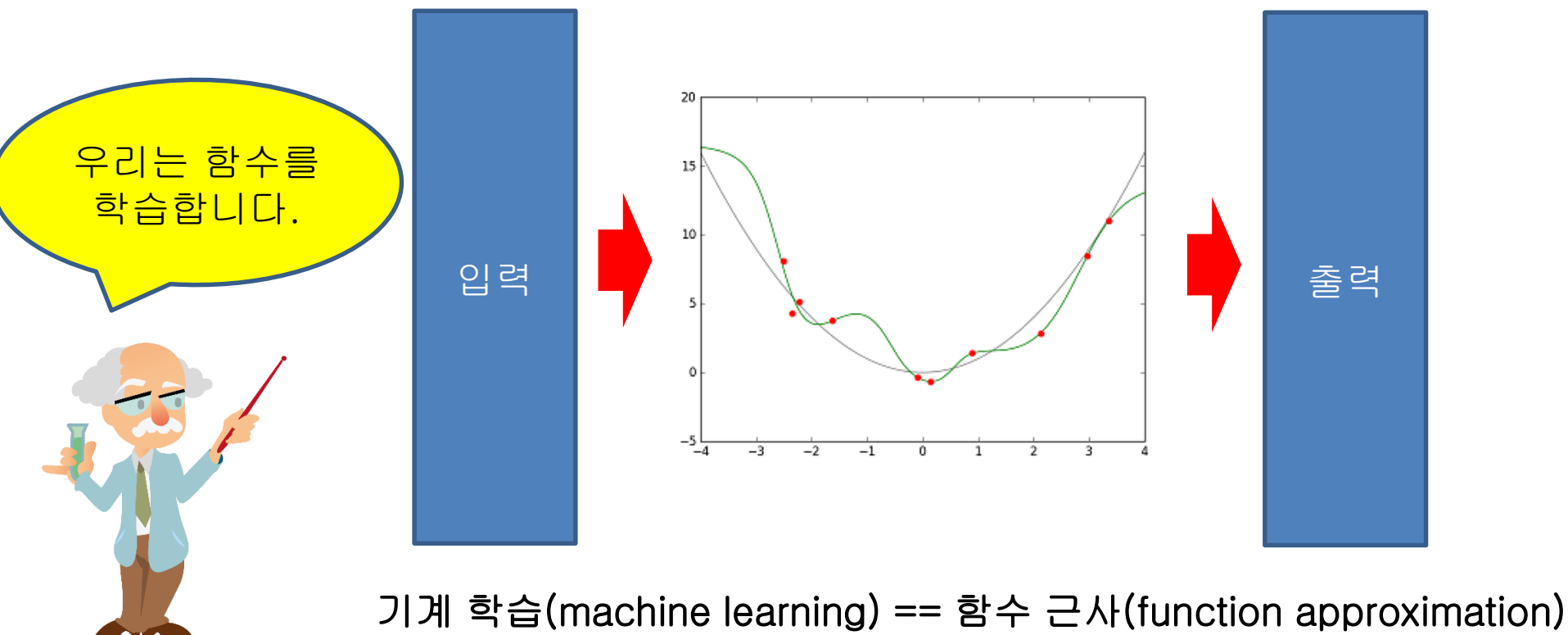


Types of Machine Learning



기계학습 기본 프레임

- 기계 학습은 항상 입력을 받아서 출력하는 함수 $y=f(x)$ 를 학습한다고 생각할 수 있다. (함수 근사)



기계학습 기본 용어

■ 특징(Feature)

- 특징이란 우리가 학습 모델에게 공급하는 입력이다. 가장 간단한 경우에는 입력 자체가 특징이 된다.
- 예를 들어서 꽃 이미지의 경우 색상, 크기 등의 입력이 특징이 된다

■ 레이블(label)

- $Y = f(X)$ 에서 y 변수에 해당한다.
- 예를 들어서 농작물의 향후 가격, 사진에 표시되는 동물의 종류, 동영상의 의미 등 무엇이든지 레이블이 될 수 있다.

■ 샘플, 또는 예제

- 샘플은 기계 학습에 주어지는 특정한 예이다. $y = f(X)$ 에서 x 에 해당한다. 레이블이 있는 샘플도 있고 레이블이 없는 샘플도 있다. 지도 학습을 시키려면 레이블이 있어야 한다.

학습과 예측

- 학습(learning)은 모델을 만들거나 배우는 것을 의미한다.
- 예측(prediction)은 학습된 모델을 레이블이 없는 샘플에 적용하는 것을 의미한다. 즉 학습된 모델을 사용하여 유용한 예측(y')을 해내는 것이다.

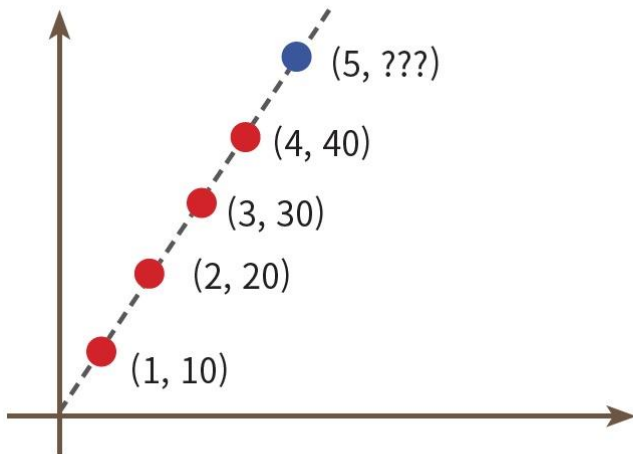
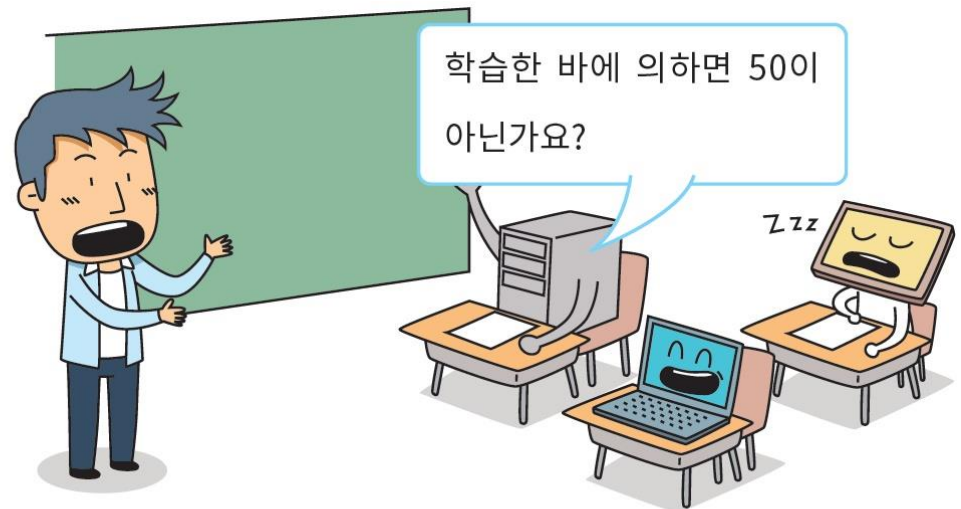


그림 9-13 회귀의 예



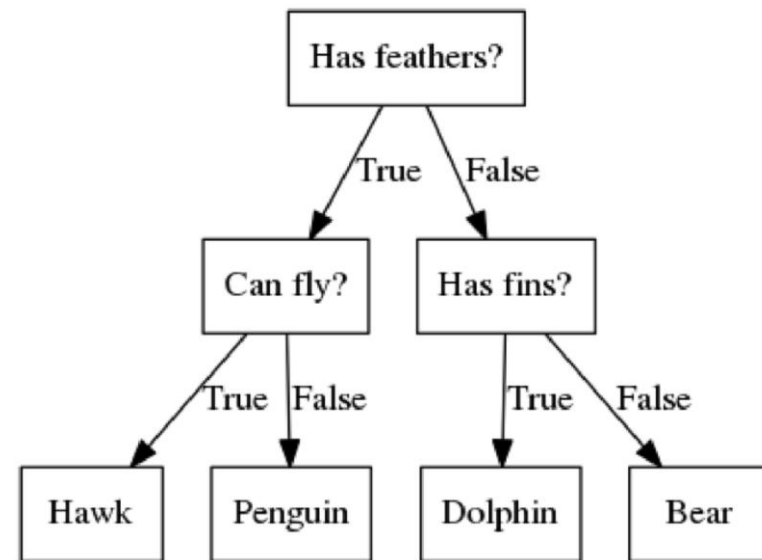
지도 학습의 세부 부류

■ 지도 학습의 3가지 유형

- 회귀: 값 예측
- 분류: 항목 선택
- 랭킹/추천: 순서 배열



회귀



분류

분류의 예

- 분류(classification): 입력을 두 개 이상의 레이블(유형)으로 분할하는 것
- 해당 모델을 학습시킬 때 우리는 레이블을 제공해야 한다.

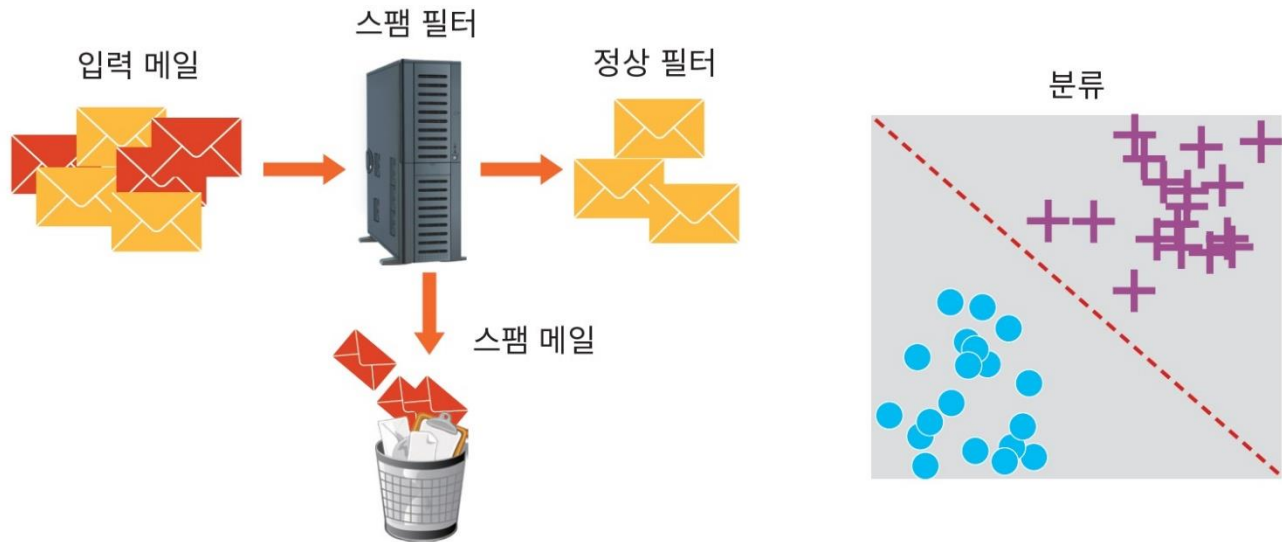


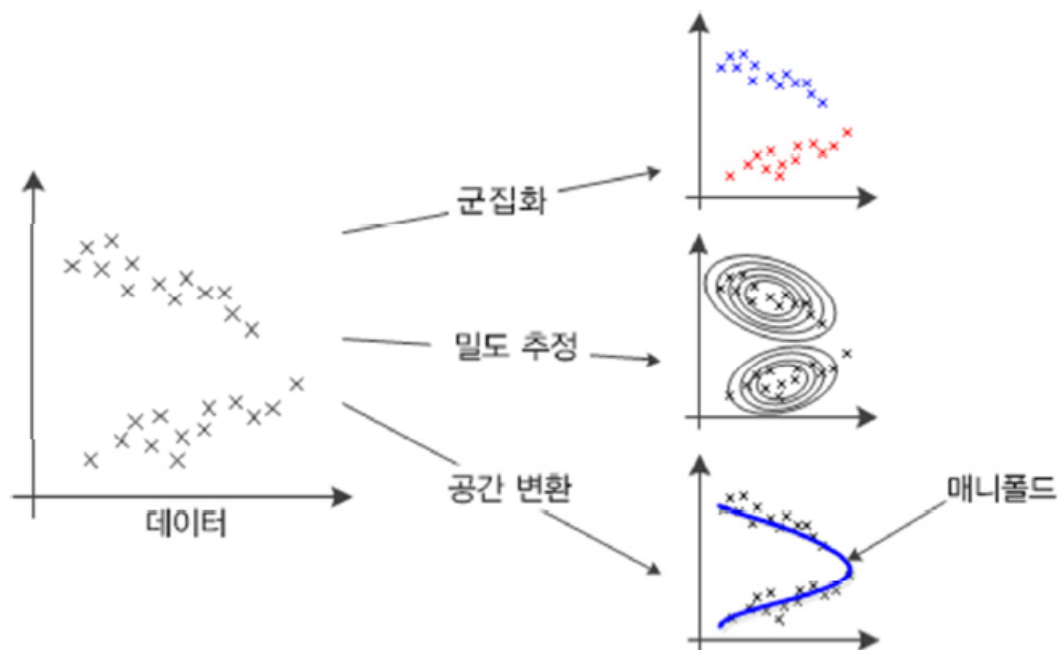
그림 9-11 분류

비지도학습의 세부 부류

■ 세 가지 일반 과업

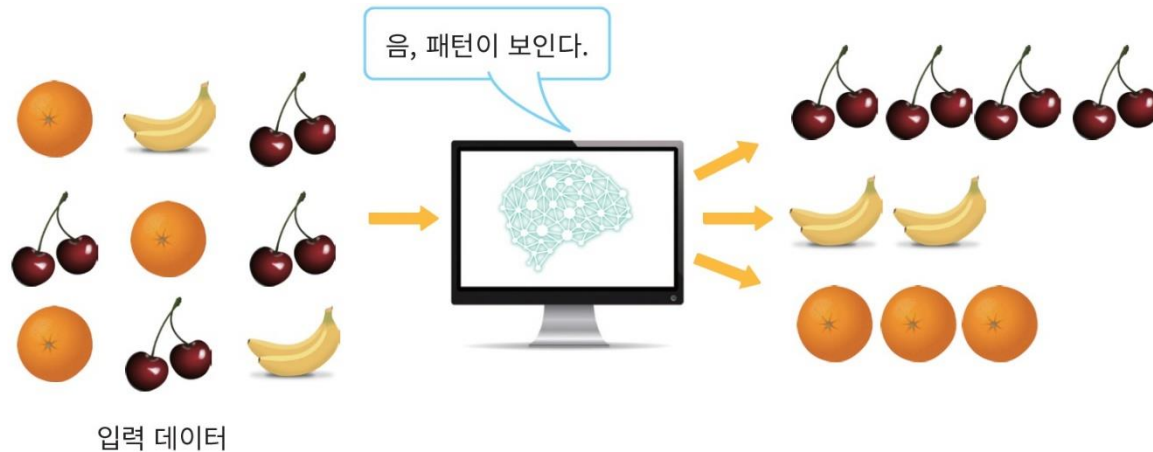
- 군집화: 유사한 샘플을 모아 같은 그룹으로 묶는 일
- 밀도 추정: 데이터로부터 확률분포를 추정하는 일
- 공간 변환: 원래 특징 공간을 저차원 또는 고차원 공간으로 변환하는 일

■ 데이터에 내재한 구조를 잘 파악하여 새로운 정보를 발견해야 함



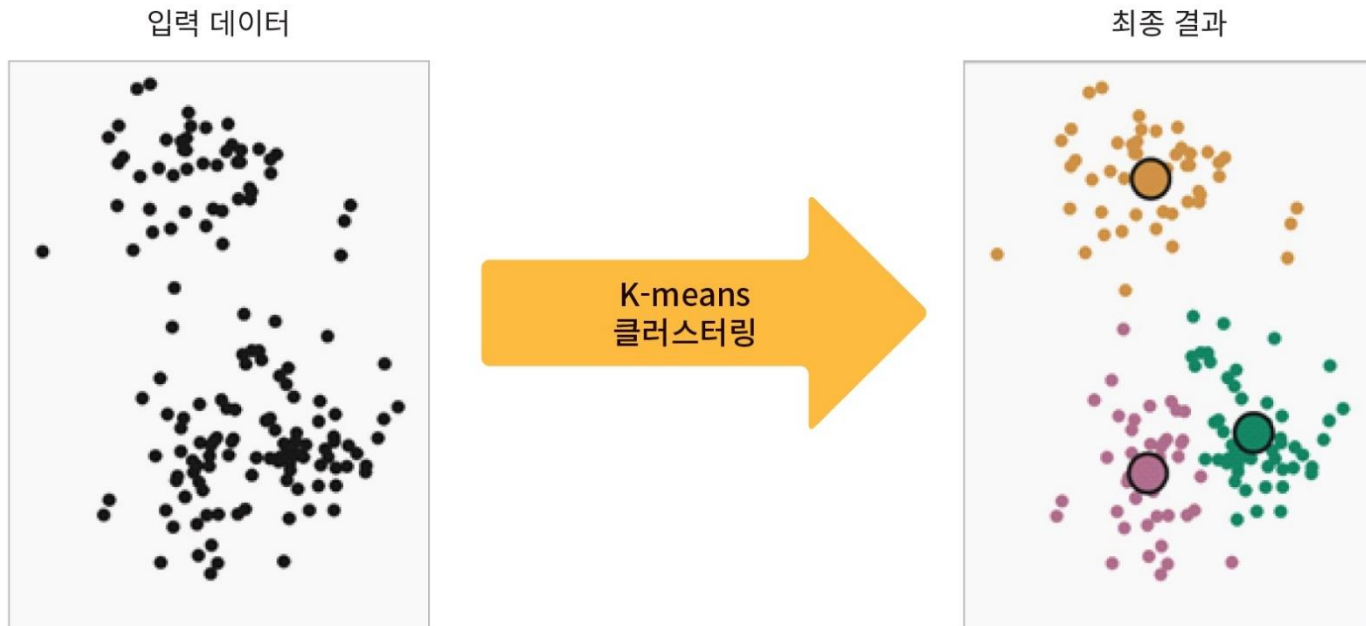
비지도 학습

- 비지도 학습(unsupervised Learning)은 “교사” 없이 컴퓨터가 스스로 입력들을 분류하는 것을 의미한다. 식 $y = f(x)$ 에서 레이블 y 가 주어지지 않는 것이다.
- 데이터들의 상관도를 분석하여 유사한 데이터들을 모을 수는 있다.



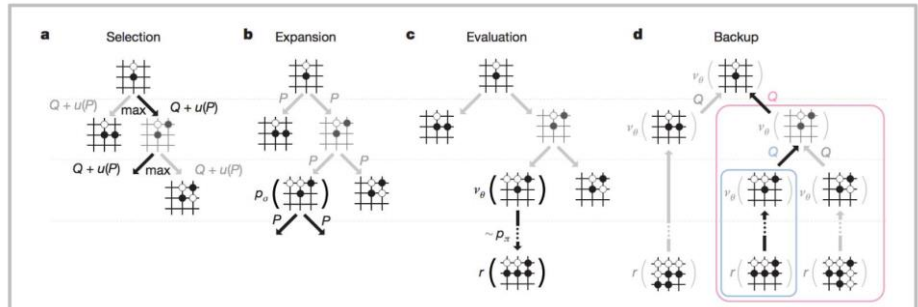
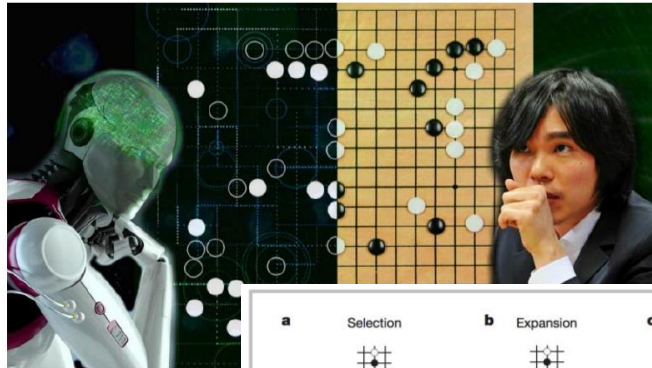
비지도 학습

- 가장 대표적인 비지도 학습이 클러스터링(군집화, clustering)이다.
- 클러스터링이란 데이터간 거리를 계산하여서 입력을 몇 개의 그룹으로 나누는 방법이다.



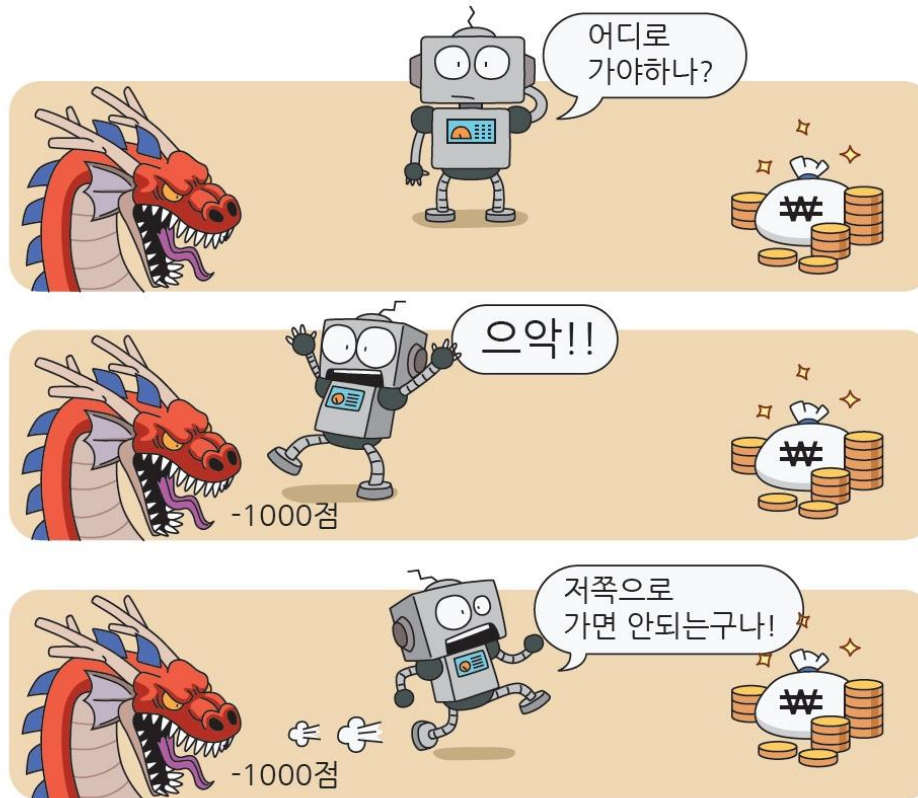
강화학습 (딥러닝)

- 강화 학습 : 목표값을 주어 지도하지만, 목표값의 형태가 지도 학습과 많이 다름
 - 장기 게임, 알파고(지도학습+강화학습) : 현재 판에서 다음 수를 선택하는 것을 학습하는 과정은 지도학습, 다음 수만을 고려하는 게 아니라 승패까지의 전체 수를 고려하여 게임에서 이길 경우 점수를 받고 그렇지 않은 경우에 점수를 받지 못하도록 하여 일련의 이기는 수를 학습하는 것이 강화 학습
 - 화성 탐사 로봇의 경로 탐색 : 로봇은 실제로 환경이 어떤지에 대한 정보가 거의 없음. 따라서 어디로 갈지를 판단하고자 할 때, '현재 가진 정보를 이용해 다음 행동 정하기' 와 '새로운 정보를 얻기 위해 가보지 않은 곳 가기' 중에서 하나를 선택



강화학습

- 강화학습에서는 컴퓨터가 어떤 행동을 취할 때마다 외부에서 처벌이나 보상이 주어진다



강화학습

- 알파고 최종 버전도 강화 학습 사용
- 게임에서 많이 사용된다

