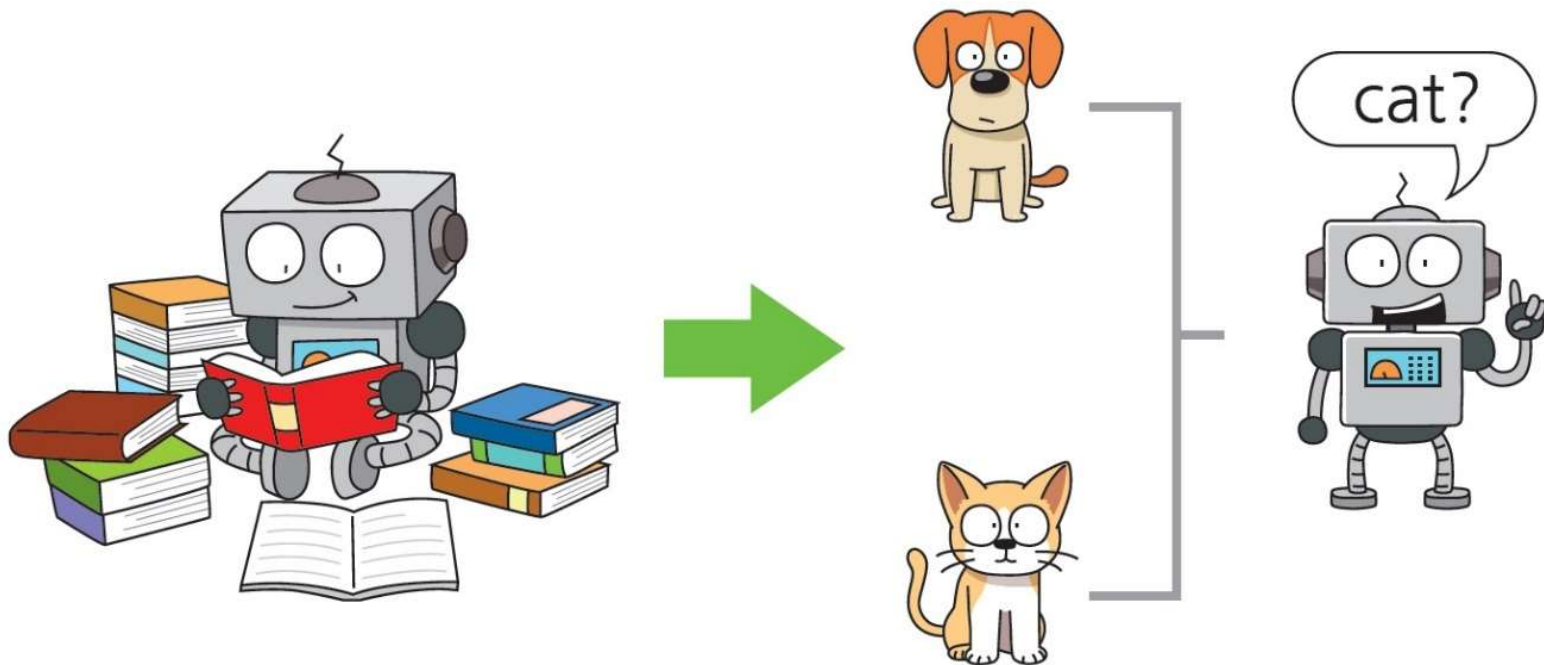


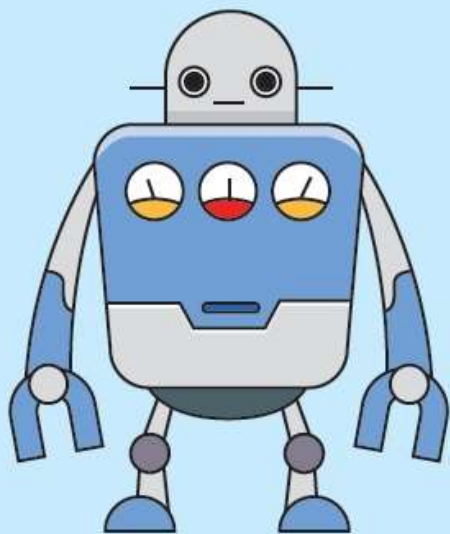
머신러닝 (기계학습)

기계학습이란?

- 현재의 컴퓨터는 스스로 학습할 수 없기 때문에 우리가 컴퓨터에게 어떤 작업을 시키려면 반드시 프로그램을 작성하여 작업을 지시하여야 한다.
- 컴퓨터가 스스로 학습할 수 있다면 컴퓨터는 프로그램 없이도 여러 가지 일을 할 수 있을 것이다.



인공지능, 기계학습, 딥러닝



인공지능
(Artificial Intelligence)



기계학습
(Machine Learning)



딥러닝
(Deep Learning)

1950년 1960년 1970년 1980년 1990년 2000년 2010년

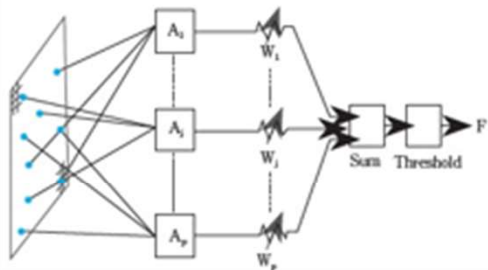
기계 학습은 어디에 이용되는가?



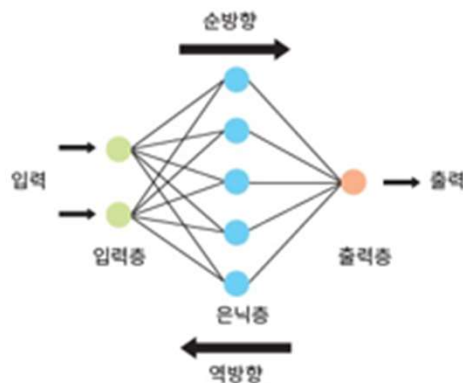
기계 학습은 어디에 이용되는가?

- 이들 분야들은 살펴보면 복잡한 데이터들이 있고, 이들 데이터에 기반하여 결정을 내려야 하는 분야이다.
 - 영상 인식, 음성 인식처럼 프로그램으로 작성하기에는 규칙과 공식이 너무 복잡할 때
 - 주식 거래나 에너지 수요 예측, 쇼핑 추세 예측의 경우처럼 데이터 특징이 계속 바뀌고 프로그램을 계속해서 변경해야 하는 상황일 때
 - 구매자가 클릭할 확률이 가장 높은 광고가 무엇인지를 알아내는 시스템
 - 넷플릭스에서 비디오 추천 시스템
 - 자율 주행자동차

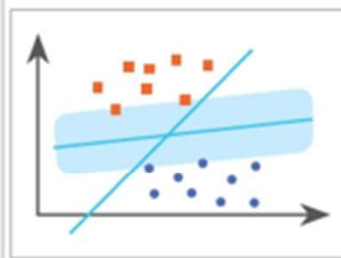
기계 학습의 역사



1957 퍼셉트론



1985 역전파 학습 알고리즘



1995 SVM



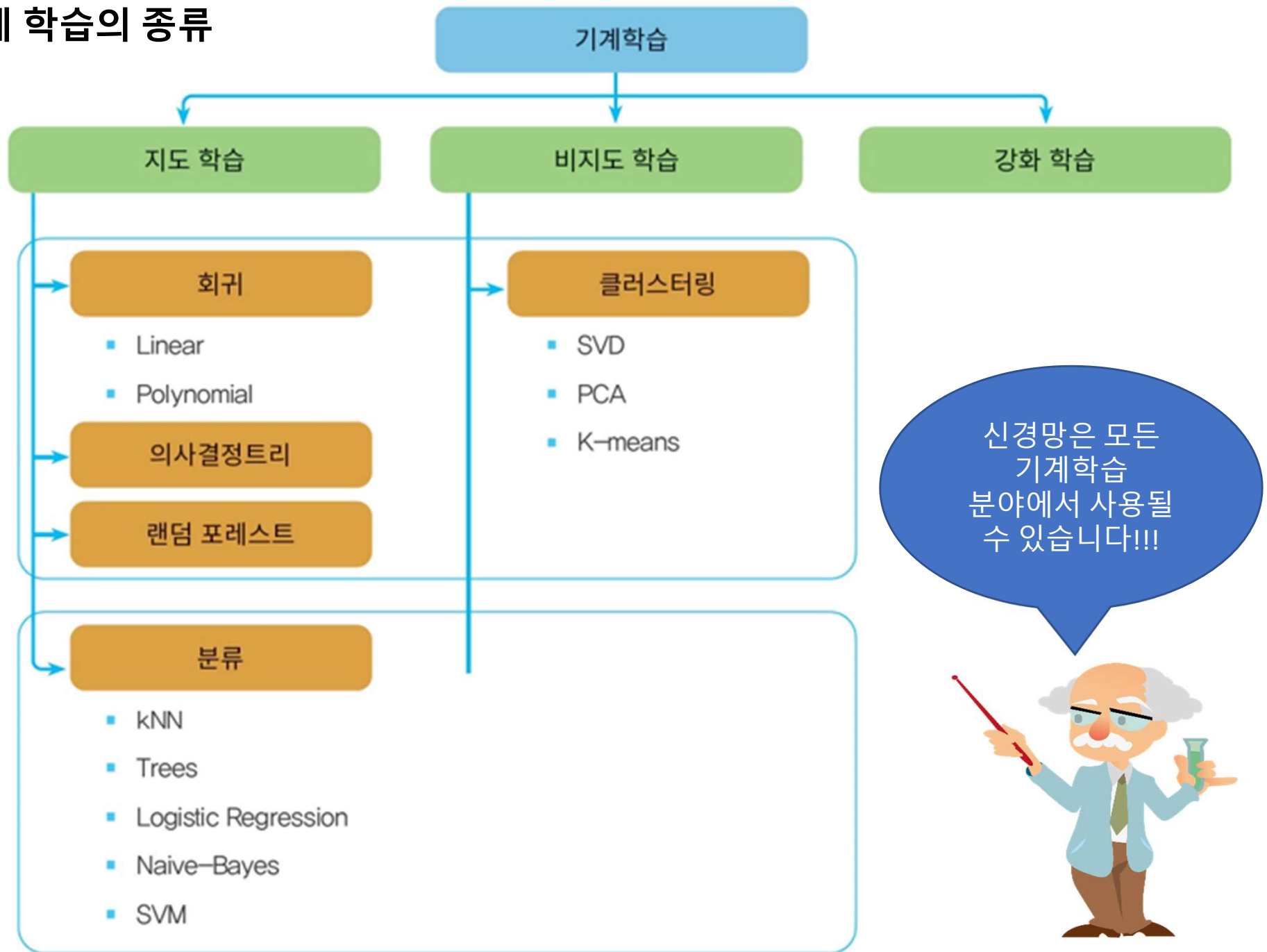
1998 필기체 인식



2012
ImageNet에서
우승



기계 학습의 종류



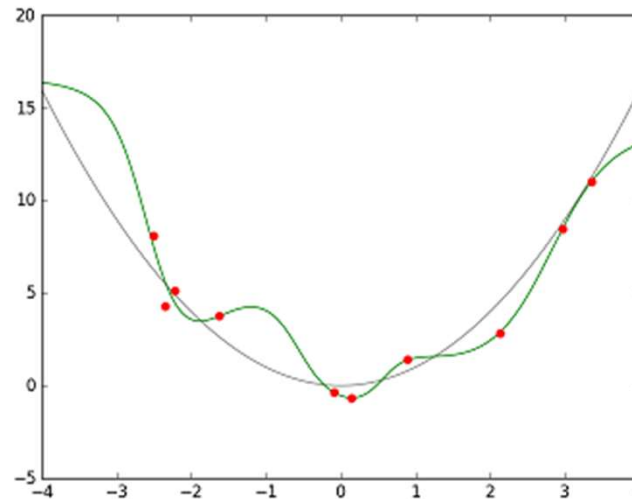
기계 학습

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

우리는 함수를
학습합니다.



입력



출력

기계 학습(machine learning) == 함수 근사(function approximation)

특징 (features)

- 특징이란 우리가 학습 모델에게 공급하는 입력이다. 가장 간단한 경우에는 입력 자체가 특징이 된다.
- (예)
 - 이메일에 "검찰"이라는 문자 포함 여부(yes 또는 no)
 - 이메일에 "광고", "선물 교환권"이나 "이벤트 당첨" 문자열 포함 여부(yes 또는 no)
 - 이메일의 제목이나 본문에 있는 '★'과 같은 특수 기호의 개수(정수)

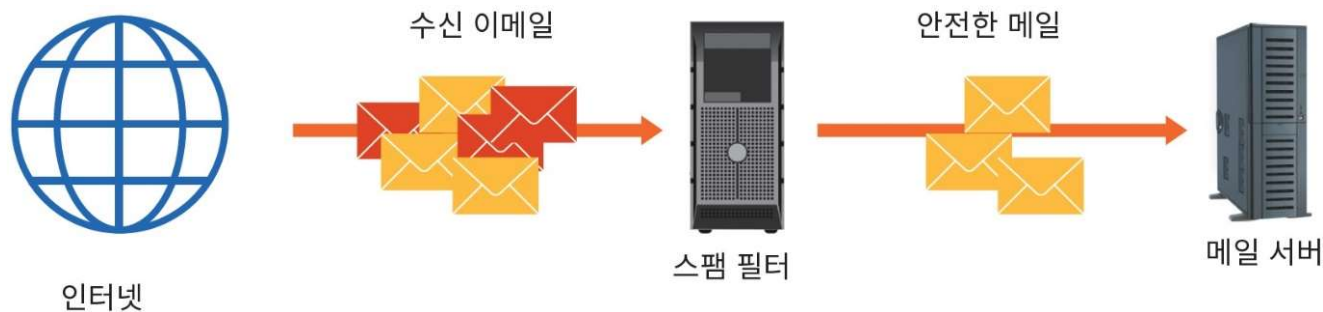


그림 9-6 스팸 필터링 시스템

레이블과 샘플

- 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.

학습 데이터와 텍스트 데이터



그림 9-7 학습 단계

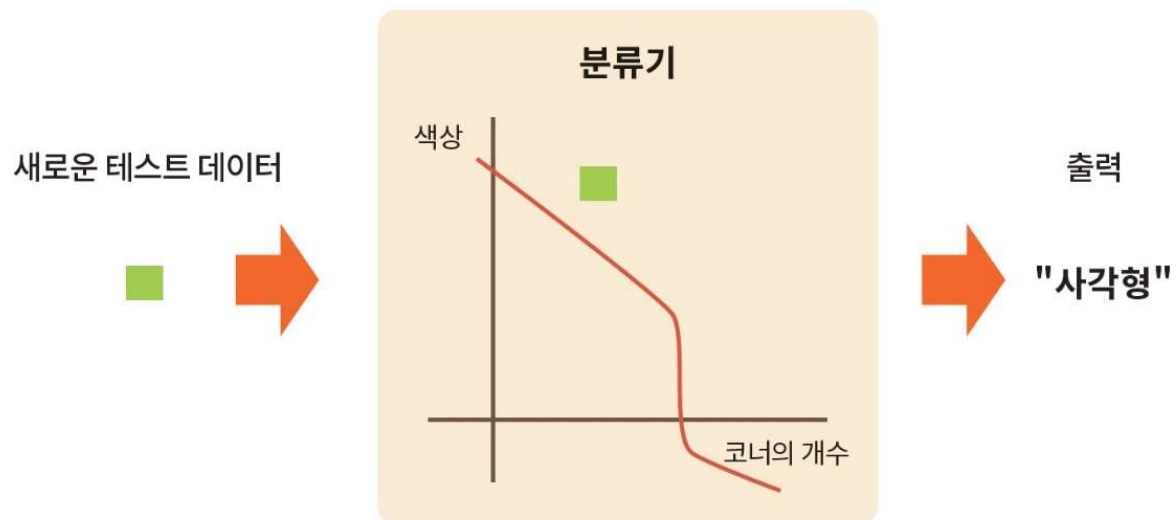


그림 9-8 테스트 단계

기계 학습

- 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.

지도 학습

- 회귀(regression) : 회귀에서는 입력과 출력이 모두 실수이다.
 - “사용자가 이 광고를 클릭할 확률이 얼마인가요?”

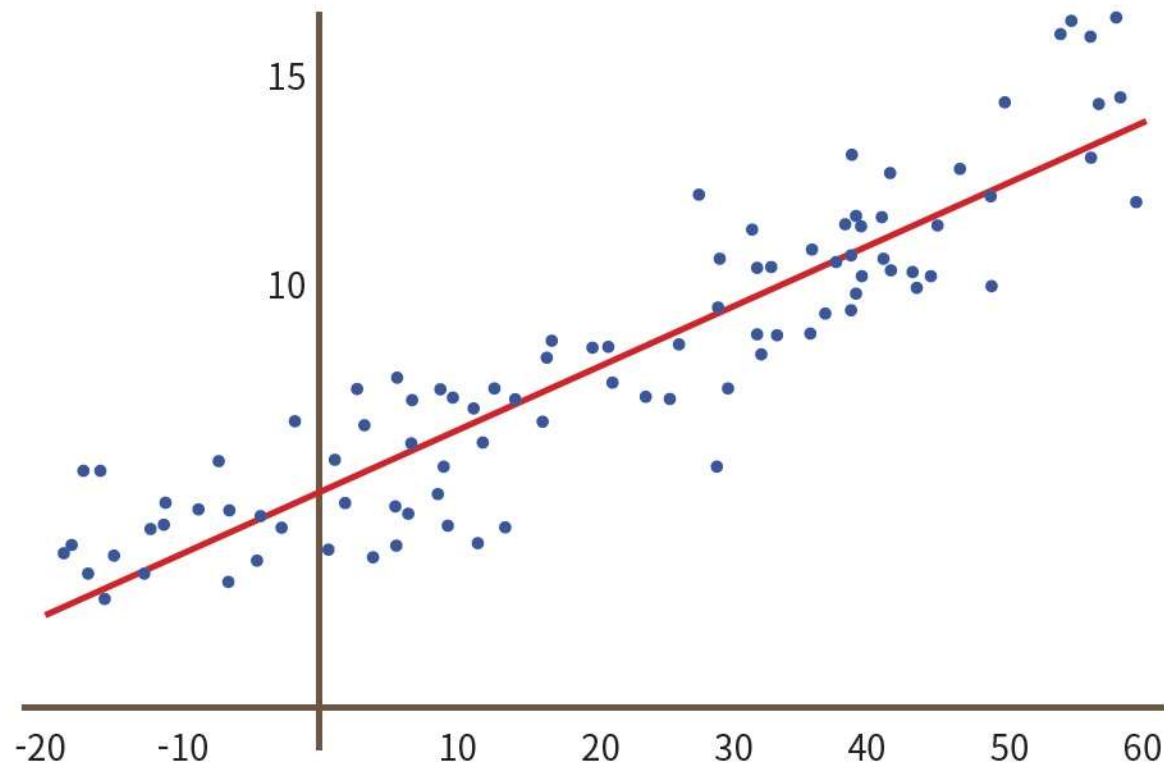
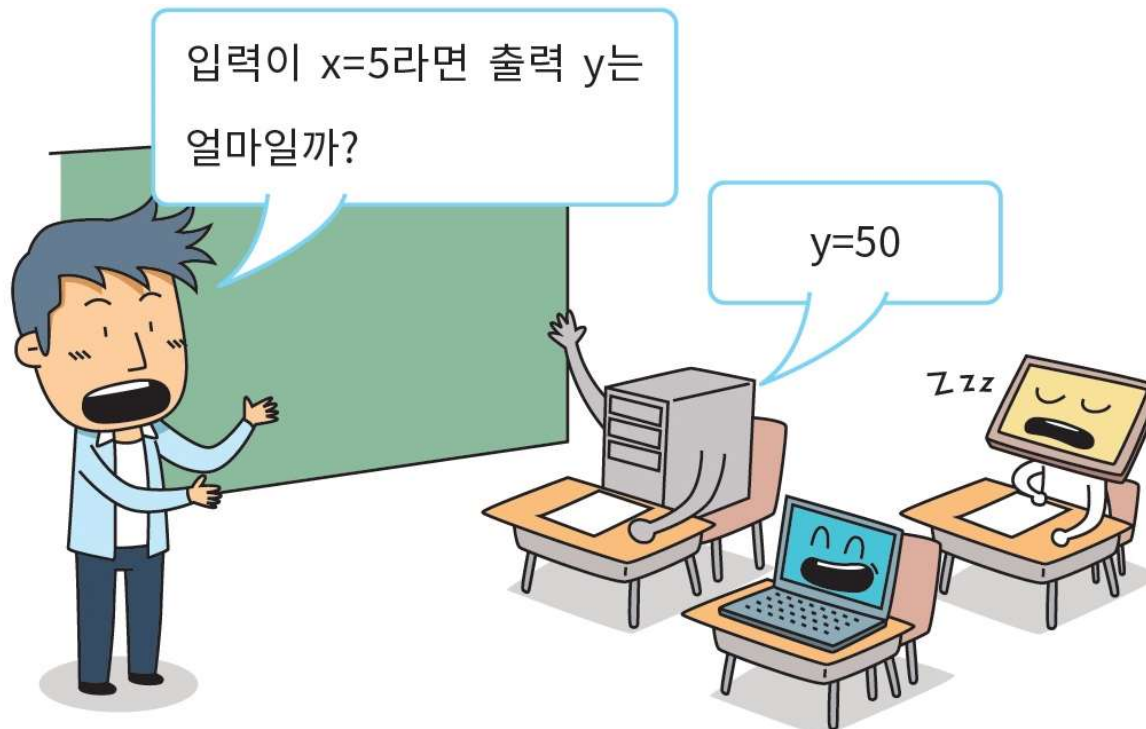


그림 9-10 회귀

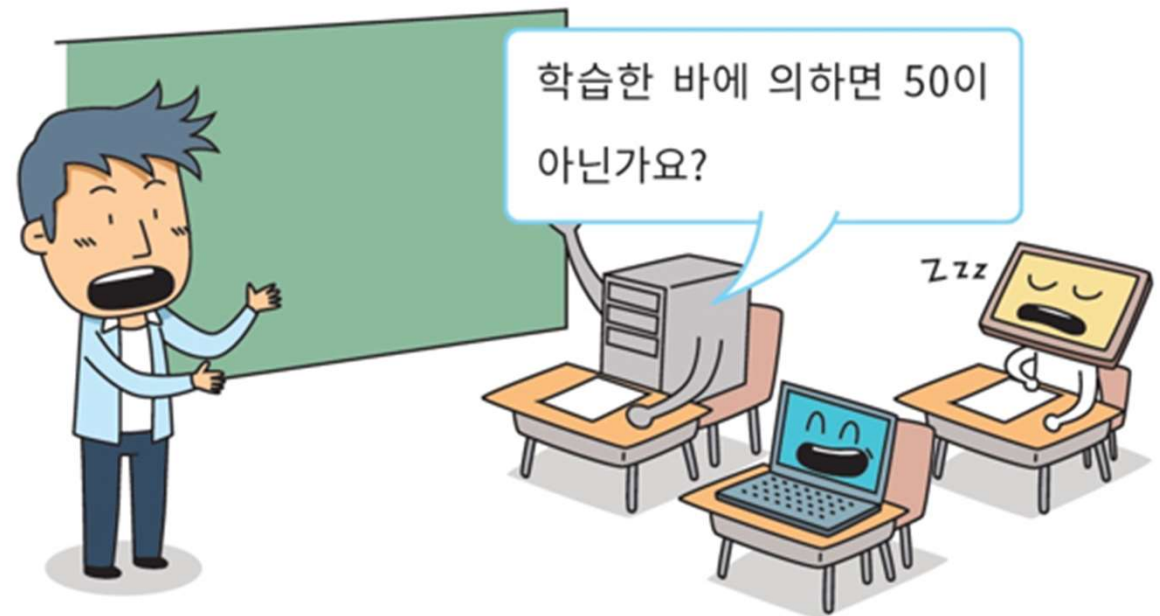
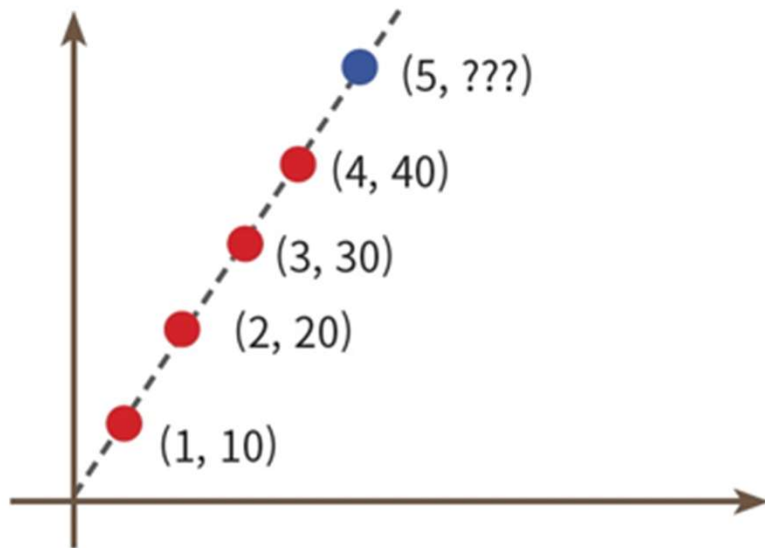
회귀

- 회귀는 실수 입력(x)과 실수 출력(y)이 주어질 때, 입력에서 출력으로의 매핑 함수를 학습하는 것이라 할 수 있다.

$$y = f(x)$$

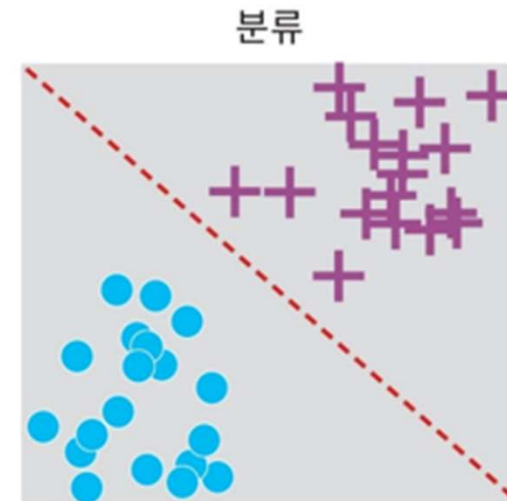


회귀의 예



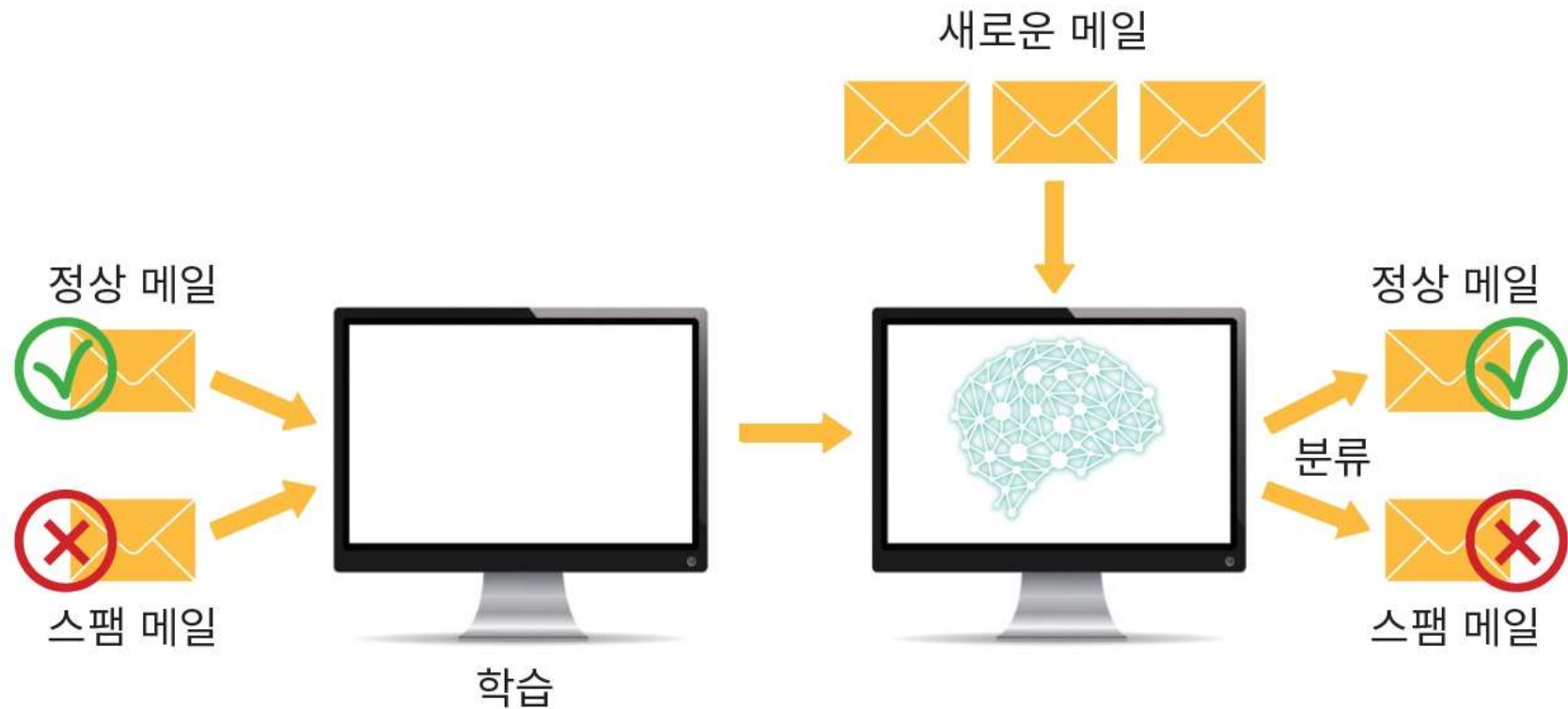
지도 학습

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



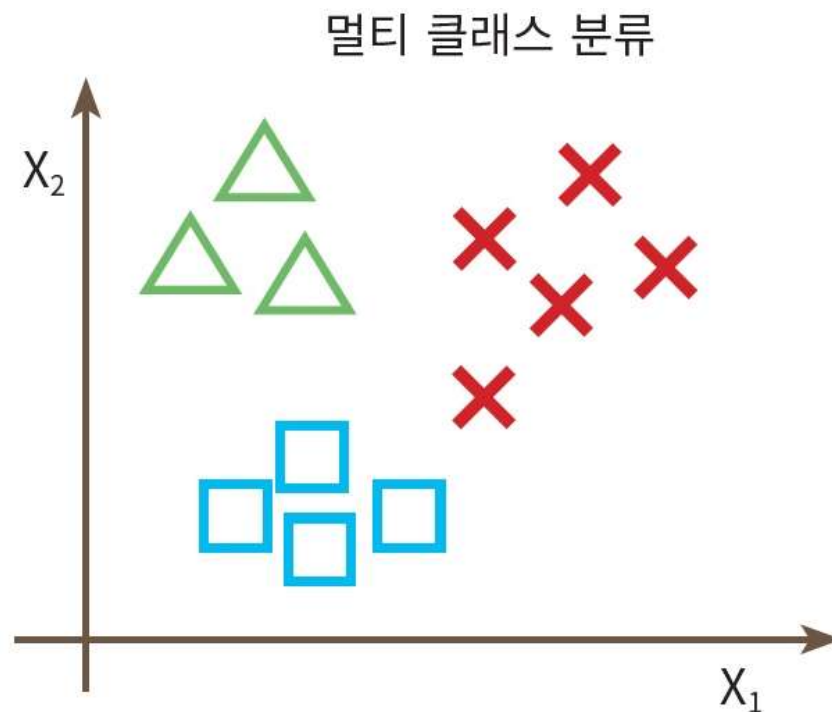
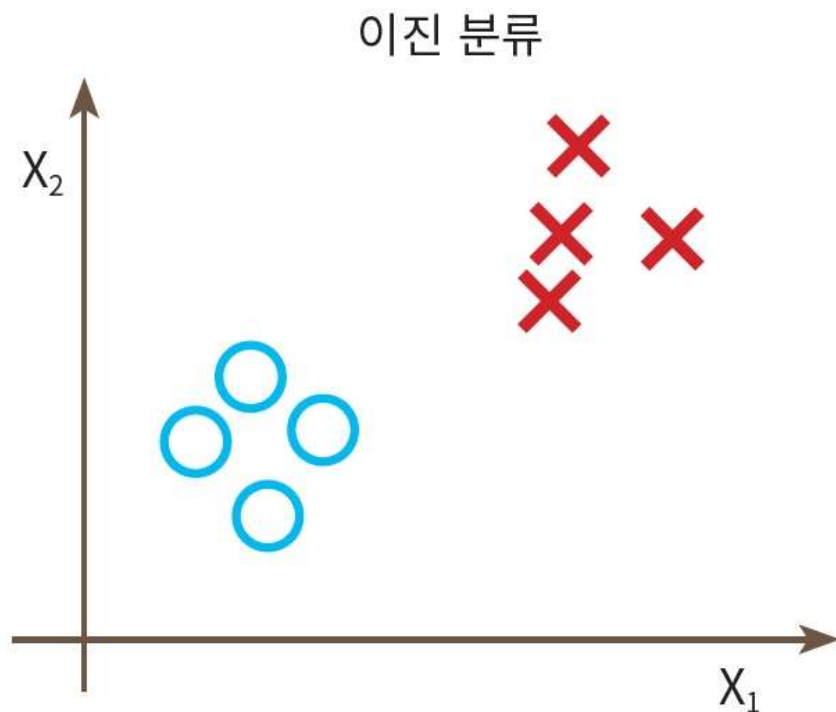
분류

- 앞에 나왔던 식 $y = f(x)$ 에서 출력 y 가 이산적(discrete)인 경우에 이것을 분류 문제(또는 인식 문제)라고 부른다.
- 분류에서는 입력을 2개 이상의 클래스로 나누는 것이다.



분류

- 분류는 지도 학습의 형태로 이루어지는 것이 일반적이다.
- 분류를 수행하기 위한 일반적인 알고리즘에는 신경망, kNN(k-nearest neighbor), SVM(Support Vector Machine), 의사 결정 트리 등이 포함된다.



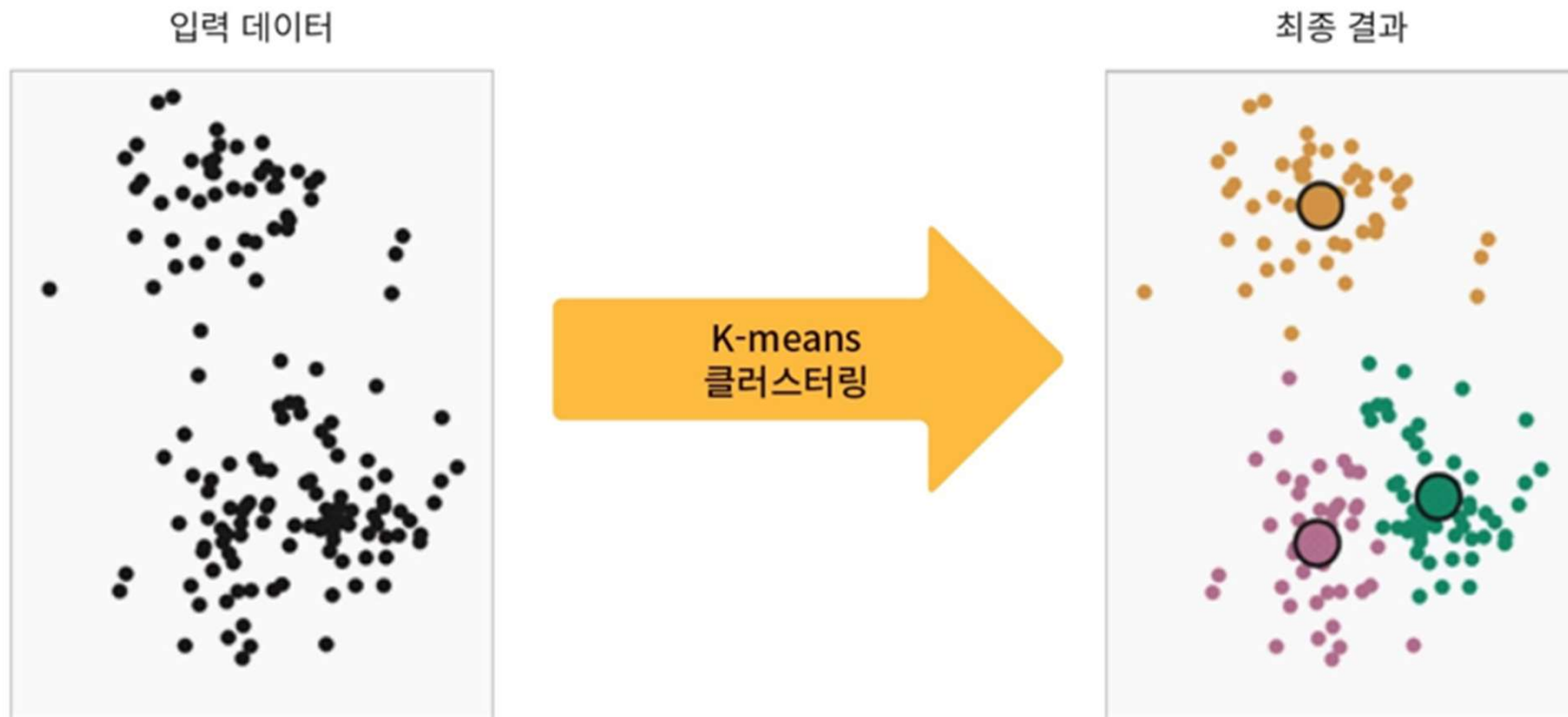
비지도 학습

- 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.



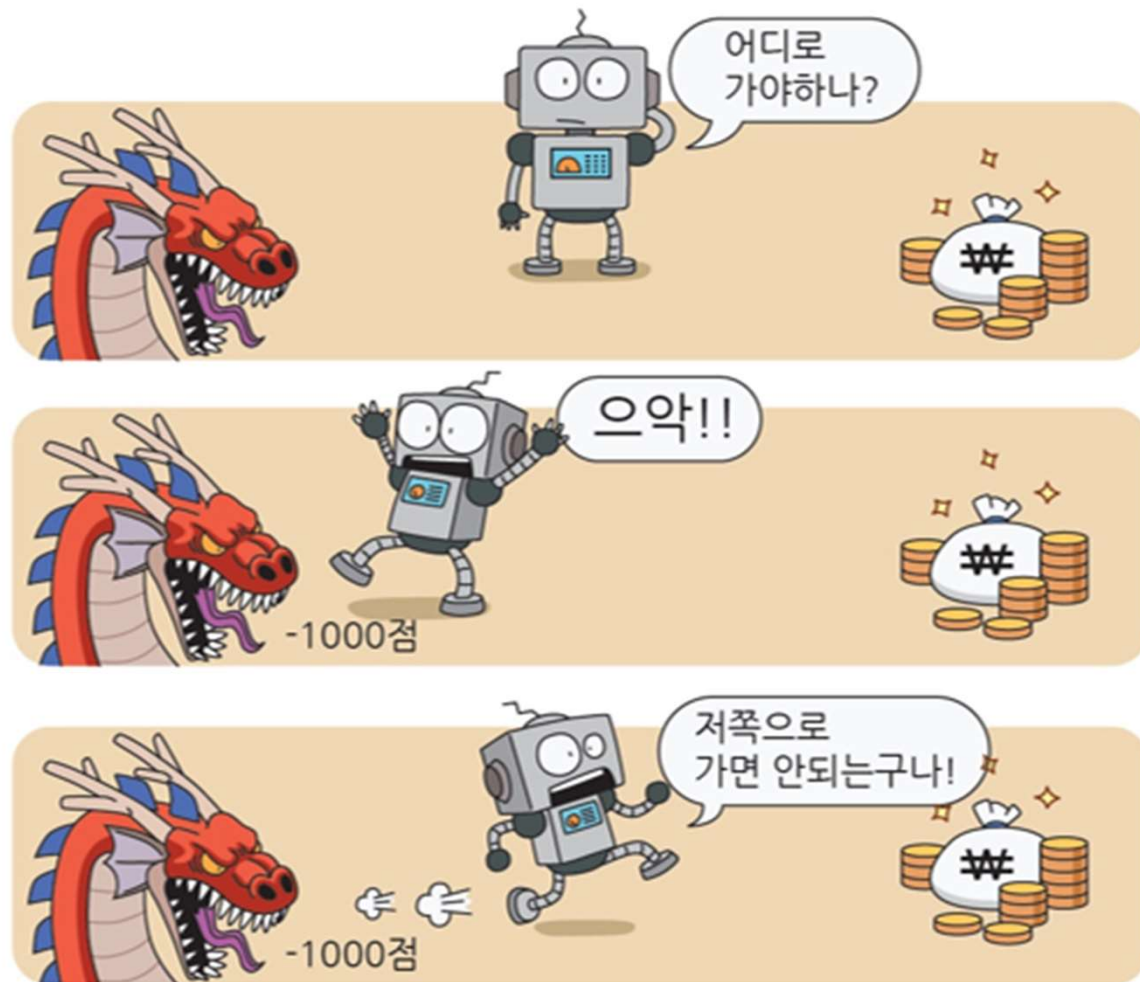
비지도 학습

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



강화 학습

- 강화 학습(Reinforcement Learning)에서는 컴퓨터가 어떤 행동을 취할 때마다 외부에서 처벌이나 보상이 주어진다.



강화 학습

- 알파고 최종 버전도 강화 학습 사용
- 게임에서 많이 사용된다(예: Frozen Lake).

