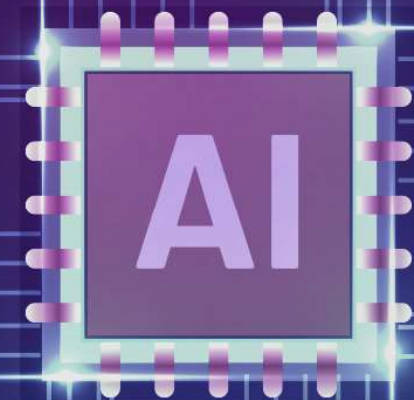


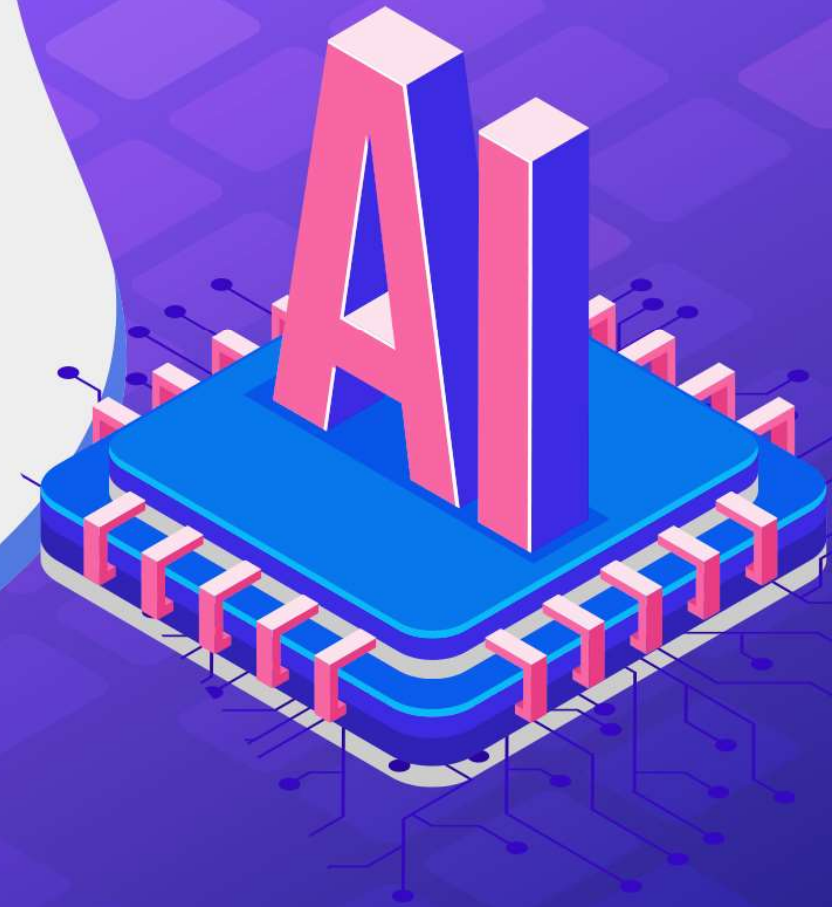
# 머신러닝의 학습과 분류 (1)

김재광 교수 (소프트웨어융합대학 글로벌융합학부)



# Contents

- 머신러닝의 개요
- 머신러닝의 학습 방법



# 머신러닝의 개요

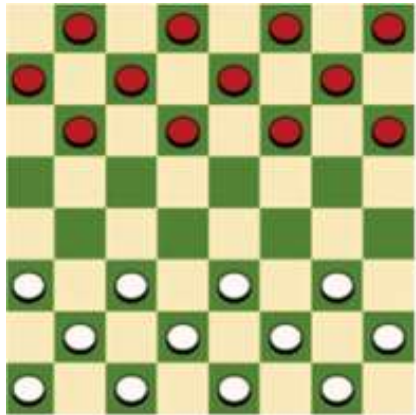
## ● 머신러닝(Machine Learning: ML)의 정의

- ‘기계학습’으로도 불리는 인공지능의 한 분야
- 인간의 학습 능력을 컴퓨터로 실현하려는 기법
- 1959년 아서 새뮤얼(Arthur Samuel)이 최초로 정의
- “프로그램을 명시적으로 작성하지 않고 컴퓨터에 학습할 수 있는 능력을 부여하기 위한 연구 분야이다.”
- 1998년 톰 미첼(Tom M. Mitchell)이 구체적으로 정의
- ‘머신(machine)’은 컴퓨터, ‘러닝(learning)’은 학습, 따라서 머신러닝이란 ‘컴퓨터를 통한 학습’을 나타냄

## 🧠 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 역사적 배경

- 1952년 새무얼(Samuel)이 '체커(Checker)' 개발
- 체커는 최초의 머신러닝 프로그램
- 체커는 당시로는 가장 복잡한 게임 프로그램 중 하나
- 체커는 경험으로부터 학습하는 방법을 사용
- 알파고와 같은 AI 바둑 S/W 작성의 바탕이 됨



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 연도별 주요 개발 모델과 특징

연도	개발자	모델	특징 or 종류
1952년	Arthur Samuel	Checker Program	최초의 머신러닝
1957년	Frank Rosenblatt	Perceptron	최초의 신경망 모델
1986년	Rumelhart 등	Multilayer Perceptron	Back-Propagation 알고리즘
1986년	Quinlan	Decision Tree	ID3
1995년	Vapnik, Cortes	Support Vector Machine	이진 분류기

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝

- 머신러닝은 데이터에서 지식을 추출하는 작업
- 코딩하지 않고 예제를 통해 학습할 수 있는 시스템
- 경험을 통해 데이터 기반으로 학습하고 예측
- 데이터로부터 유용한 규칙 등을 추출하는 기능
- 프로그래밍하기 어려운 작업의 해결에 주로 활용됨



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝에 대한 설명들

- “머신러닝은 명시적으로 프로그래밍하지 않고도 컴퓨터를 작동시키는 과학이다.”
- “머신러닝은 규칙기반 프로그래밍에 의존하지 않고, 데이터로부터 학습할 수 있는 알고리즘을 기반으로 한다.”
- “머신러닝 알고리즘은 예제를 일반화하여 중요한 작업을 수행하는 방법을 파악할 수 있다.”

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 학습 개념

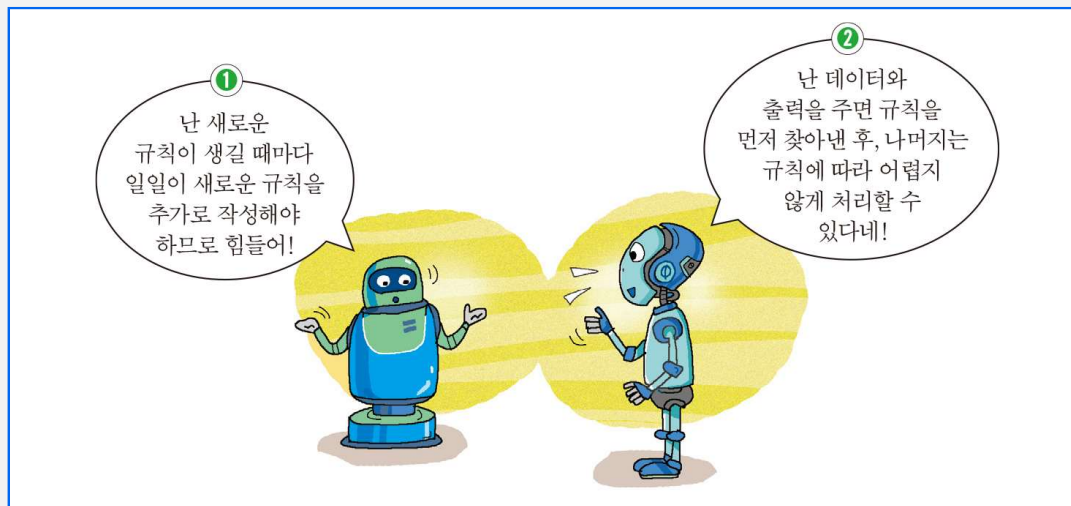
- ex. 입력과 출력의 여러 개의 데이터 쌍  
(1, 2), (2, 4), (4, 8), (7, 14), (5, 10), .....
- 데이터로부터 학습 후, 출력이 입력의 2배임을 유추함  
(3, ?), (8, ?) 등의 질문에 6, 16 등으로 답변함



## AI 머신러닝의 개요 (Cont'd)

### ● 머신러닝과 전통적인 프로그래밍과의 차이점

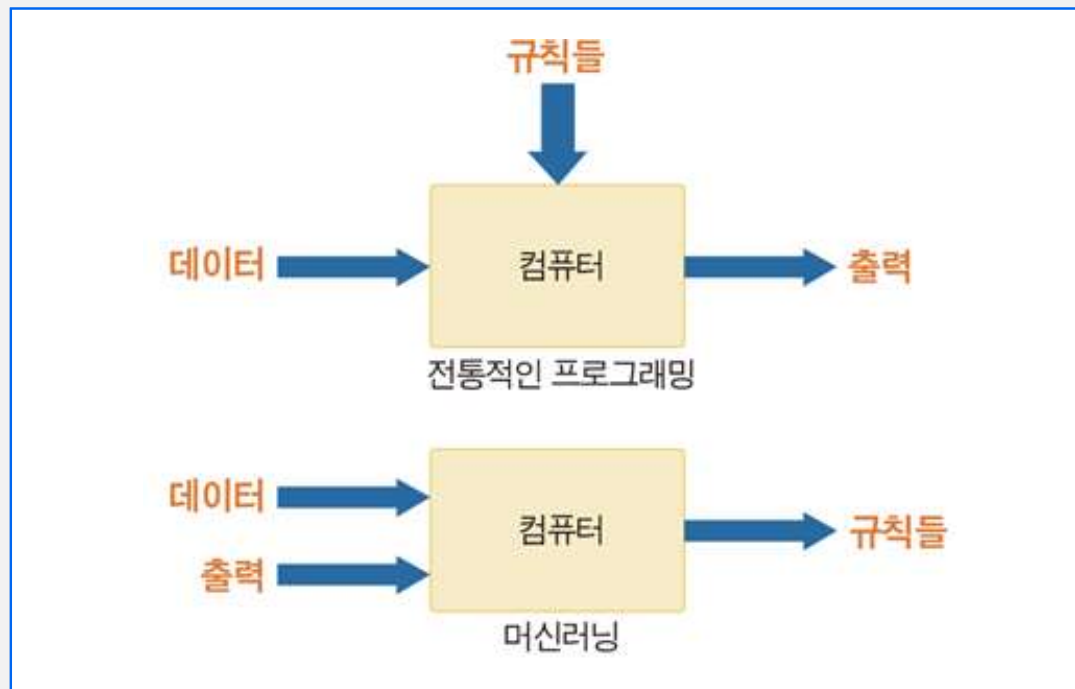
- 전통적인 프로그래밍에서는 모든 규칙들을 작성함  
만약 규칙들이 추가로 작성될 경우 유지 관리가  
어렵다는 문제가 있음
- 그러나 머신러닝은 시간에 따라 점차 효율이 향상됨  
입출력 데이터의 관계를 학습하여 규칙을 생성함



## AI 머신러닝의 개요 (Cont'd)

### ● 머신러닝과 전통적인 프로그래밍과의 차이점 (cont'd)

- 프로그래밍에서는 데이터와 규칙이 결합하여 출력을 생성
- 머신러닝에서는 데이터와 출력이 함께 들어가서 규칙 생성



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝과 인공지능과의 관계

- 머신러닝은 인공지능에 속하는 부분 집합
- 인공지능은 머신러닝을 포함하는 상위 개념
- 따라서 추구하는 개념과 목표가 다소 다름
- 인공지능은 추론, 계획 등과 머신러닝을 포함



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝과 인공지능의 차이점

- 머신러닝은 데이터로부터 학습하여 지식을 획득
- 인공지능은 지식을 획득한 후, 그것을 활용함

	머신러닝	인공지능
주요 활동	학습을 통한 지식의 획득	지식의 획득과 활용
구현과 실현	데이터로부터 학습 실현	복잡한 문제 해결을 위한 지능의 구현
개발 목표	스스로 학습하는 알고리즘 개발	인간을 닮은 지능적 시스템의 개발

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝 과정에서의 고려 사항

- 주어진 데이터로부터 원하는 답을 찾을 수 있을까?
- 문제 해결을 위해 데이터가 충분한가?
- 어떤 머신러닝 기법을 적용하면 좋을까?
- 추출할 데이터의 특성은 무엇인가?
- 머신러닝의 결론은 무엇으로 설정할 것인가?
- 생성된 출력을 실제 응용에 어떻게 사용할 것인가?

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 종류

- 신경망(Neural Network)
  - ✓ 생물의 신경 네트워크 구조와 기능을 모방한 모델
- 클러스터링(Clustering)
  - ✓ 주어진 데이터를 클러스터라는 부분 집합들로 분리하는 것
- 분류(Classification)
  - ✓ 주어진 데이터를 비슷한 것들끼리 분류하는 것
- 의사결정 트리(Decision Tree)
  - ✓ 트리 구조 형태의 예측 모델로 의사를 결정하는 모델
- 나이브 베이즈(Naive Bayes)
  - ✓ 베이즈 정리를 바탕으로 한 조건부 확률 모델 분류

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 활용 분야

분 야	응 용
영상인식	문자인식, 물체인식
얼굴인식	Facebook에서 얼굴인식
음성인식	Bixby, Siri, Alexa 등
자연어 처리	자동번역, 대화분석
정보검색	스팸 메일 필터링
검색엔진	개인 맞춤 식 추천시스템
로보틱스	자율주행 자동차, 경로탐색

## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 활용 분야 (cont'd)

- 머신러닝은 여러 산업 분야에 다양하게 활용 가능
- 최근 딥러닝이 물체의 인식 등에 획기적으로 성공함
- 미국에서는 머신러닝 기술로 빅데이터를 분석하는 '데이터 과학자(data scientist)'의 수요가 급증
- 자율주행 자동차, 문자인식 등과 같이 알고리즘 개발이 어려운 문제 해결에 활용되고 있음

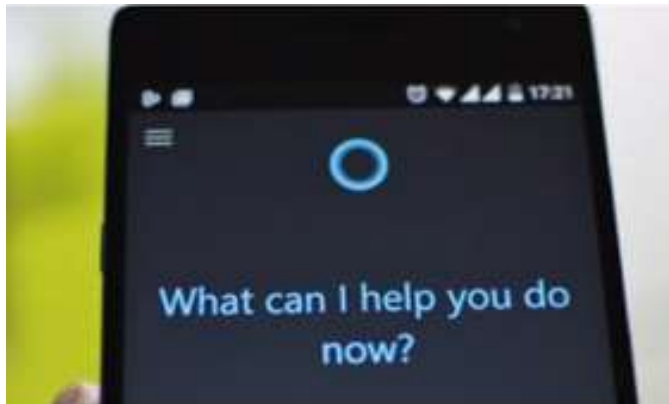




## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 개인비서(personal assistant)에의 활용

- 머신러닝은 스마트폰의 개인 비서에 활용됨
- 음성인식, 언어 습관, 행동 패턴 등을 학습
- 머신러닝 기법을 이용한 지능적인 역할 담당
- 삼성의 Bixby, 애플의 Siri, 구글의 Assistant 등



## 머신러닝의 개요 (Cont'd)

- 헬스케어(health care) 분야에서의 역할

- 건강과 관련된 헬스케어(health care) 분야에 활용
- 센서들을 통해 환자의 건강 예측이나 개선에 활용
- 환자의 심장 박동 등의 건강 정보를 모아 분류와 학습



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 SNS에의 활용

- 머신러닝은 SNS에도 상당히 중요한 역할
- 여러 번 검색해본 책, 영화 등의 습관을 학습
- 적절한 시각에 알려주거나 광고를 보내기도 함
- 페이스북에서는 출신학교나 친구들의 관계를 적용
- 머신러닝이 새로운 친구 제안



## 머신러닝의 개요 (Cont'd)

### ● 머신러닝의 동영상 추천에의 활용

- 유튜브(YouTube)에 '내 관련 재생 목록'으로 응용됨
- 즐겨 찾던 동영상 리스트가 추천 리스트로 올라옴



[바둑렙 강좌] 두칸벌림 2선침입  
바둑렙 강좌  
조회수 44,926회



프로 9단들도 헛갈리는 함정수!  
[무서운 함정수 5강]  
바둑 도서관  
맞춤 동영상 새 동영상

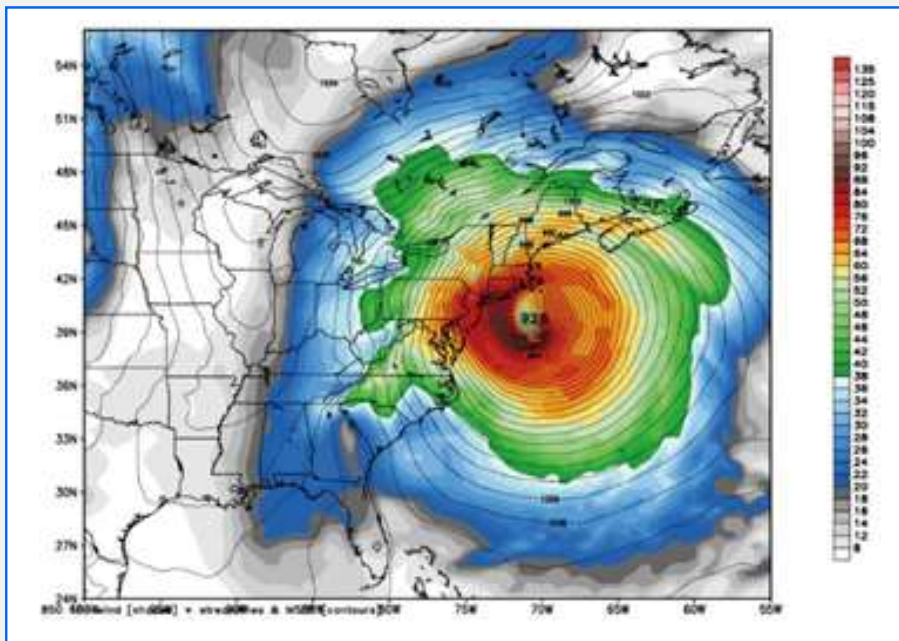


[바둑렙 강좌] 족보에도 없는 끝수를 당하신다면~  
바둑렙 강좌  
조회수 18,382회

## 머신러닝의 개요 (Cont'd)

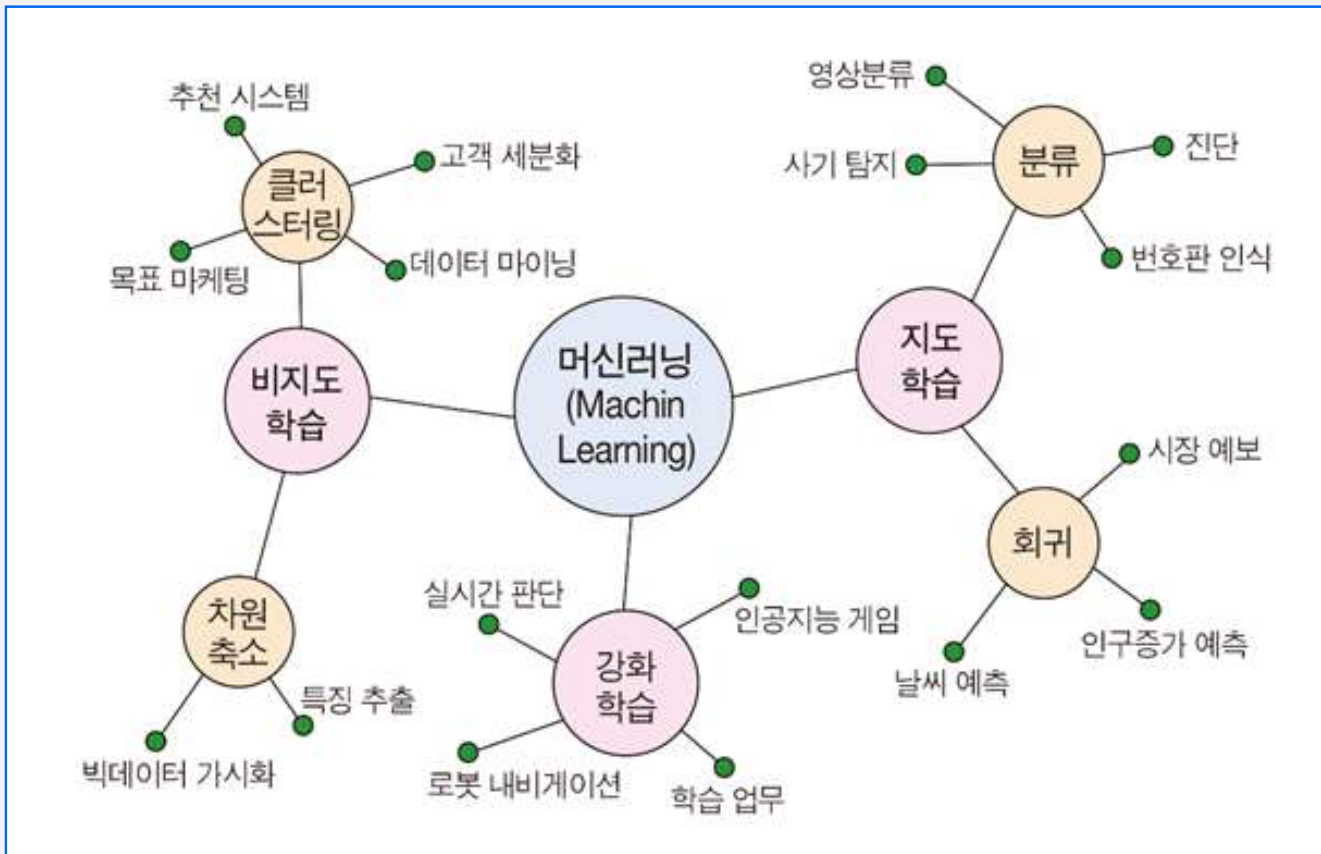
### ● 머신러닝의 기상 예측 등에의 활용

- 데이터와 통계적 도구를 결합하여 결과를 예측
- 결과물은 기상 예측 등에 유용하게 활용됨
- 그 외 사기 탐지, 작업 자동화 등에도 활용



# 머신러닝의 학습 방법

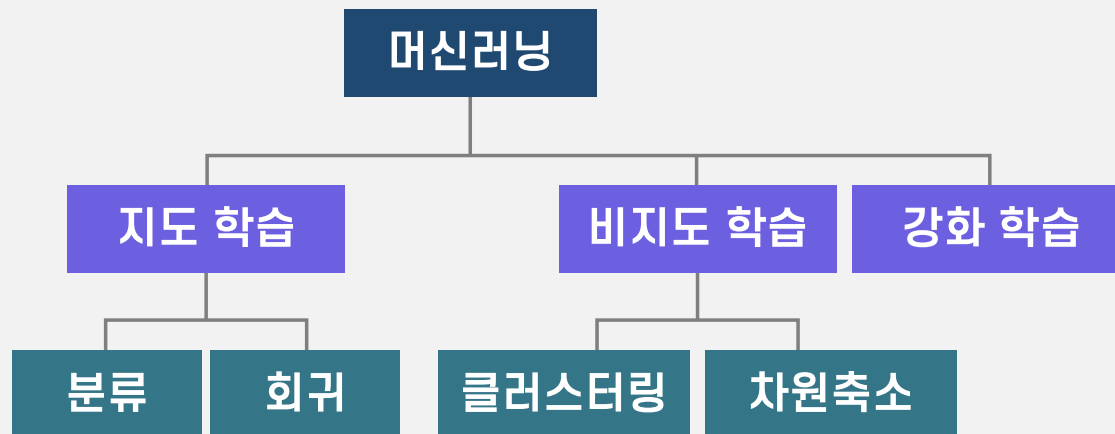
## 머신러닝의 학습 방법과 다양한 활용



## 머신러닝의 학습 방법 (Cont'd)

- 머신러닝에서의 학습 방법

- 학습의 형태에 따라 3가지 학습 방법
- 지도 학습, 비지도 학습, 강화 학습으로 구분



## 머신러닝의 학습 방법 (Cont'd)

- 지도 학습(supervised learning)

- 입력과 미리 알려진 출력을 연관시키는 관계를 학습
- 주어진 입력과 출력 쌍 사이의 대응 관계를 학습
- 자동차 번호판이 오염된 경우 인식하지 못할 수도 있음
- 그러나 오염된 번호판 사례들을 학습시켜 인식률을 높임



## 머신러닝의 학습 방법 (Cont'd)

### ● 비지도 학습(unsupervised learning)

- 출력값을 알려주지 않고 스스로 모델을 구축하여 학습
- 비지도 학습은 입력만 있고 출력 즉 레이블(label)이 없음
- 규칙성을 스스로 찾아내는 것이 학습의 주요 목표
- 결과는 지도 학습의 입력으로 사용 가능
- 또는 전문가에 의해 해석되어 다른 용도로 활용됨
- 데이터 마이닝(data mining) 기법은 비지도 학습의 예

## 머신러닝의 학습 방법 (Cont'd)

- 강화 학습(reinforcement learning)

- 주어진 입력에 대응하는 행동에 대해 보상(reward)
- 이러한 보상을 이용하여 학습하는 방법
- 주어진 입력에 대한 출력, 즉 정답 행동이 주어지지 않음
- 주요 응용 분야로는 로봇, 게임, 내비게이션 등

## 머신러닝의 학습 방법 (Cont'd)

### 지도 학습의 예

- 주어진 입력과 정해진 출력 간의 관계를 학습
- 각 데이터에 레이블(label) 또는 태그(tag) 표시 붙임
- 데이터에 'P'(pass) 또는 'F'(fail)라는 레이블 활용
- 예를 들어 '이들은 사과'라는 레이블로 학습한 후, 새로운 사과 하나를 제시하면 그것을 '사과'라고 예측하는 방법



## 머신러닝의 개요 (Cont'd)

### ● 지도 학습의 장단점

#### - 장점

- ✓ 이전의 경험으로부터 데이터 출력을 생성
- ✓ 경험을 사용하여 성능 기준을 최적화
- ✓ 다양한 유형의 문제 해결에 도움이 됨

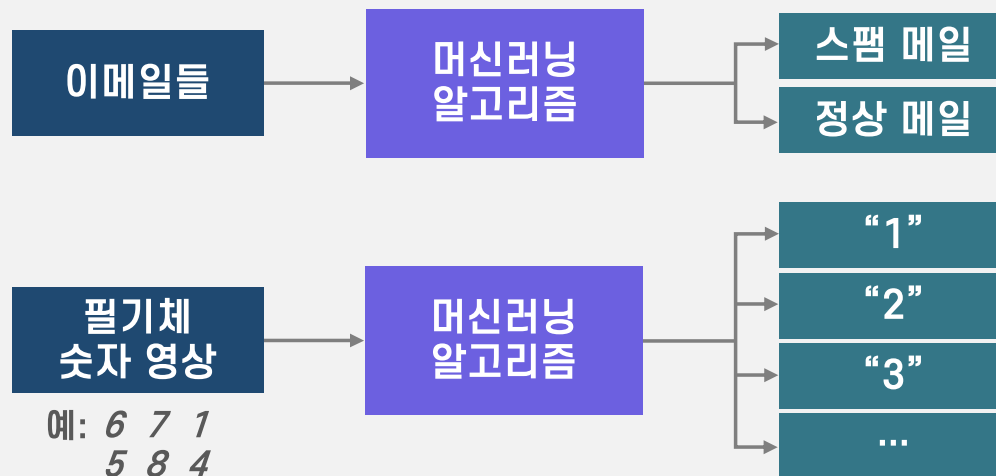
#### - 단점

- ✓ 출력에 반드시 레이블이 있는 데이터들을 사용해야 함
- ✓ 일반적으로 많은 시간이 걸림
- ✓ 빅데이터의 경우 엄청난 시간이 걸릴 수도 있음

## 머신러닝의 학습 방법 (Cont'd)

### ● 분류 (Classification)

- 유사한 특성을 가진 데이터들끼리 묶어서 나누는 것
- 2개로 분류하는 이항 분류, 그 이상의 다항 분류
- 이항 분류는 합격/불합격, 스팸 메일/정상 메일 등 분류 방법
- 0에서 9까지의 아라비아 숫자 인식은 다항 분류



## AI 머신러닝의 학습 방법 (Cont'd)

### ● 분류의 다양한 예

- 일상생활에서 수많은 패턴들을 분류
- 일반 버스/마을 버스/광역 버스 등의 구별
- 많은 남자와 여자 사진을 레이블을 붙여놓고 학습
- 학습 후 새로운 사진에 대해 남자/여자 분류



## 머신러닝의 학습 방법 (Cont'd)

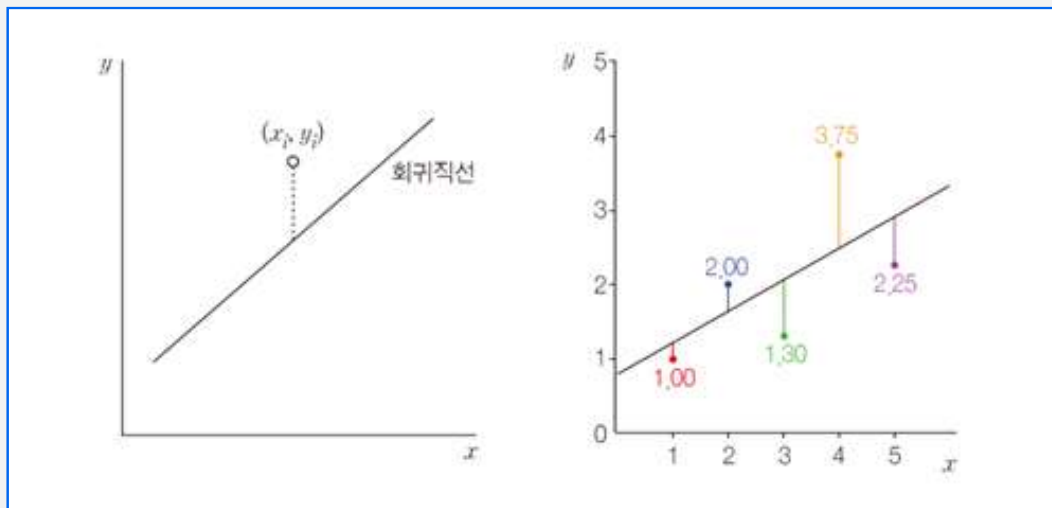
### ● 분류의 다양한 예 (cont'd)

- 사진으로 남자와 여자의 구별
- 개와 고양이의 구분
- 스팸 메일과 정상 메일 구분
- 0에서 9까지의 숫자의 구분
- 알파벳과 한글 문자 등의 구분
- 편지봉투의 손으로 쓴 주소 판별
- 카드 부정 사용 감지
- 의료 영상에서 종양의 존재 여부 판단

## AI 머신러닝의 학습 방법 (Cont'd)

### ● 회귀 (Regression)

- 회귀란 변수들 사이의 관계를 결정하는 통계적 측정
- 하나의 독립 변수를 사용하는 직선 형태의 '선형 회귀'
- 선형 회귀는 각 점에서 회귀 직선까지의 y축 방향의 거리 제곱의 총합을 최소로 해서 얻어지는 직선
- 직선  $y = a + bx$ 를 x에 대한 y의 회귀 직선이라 함

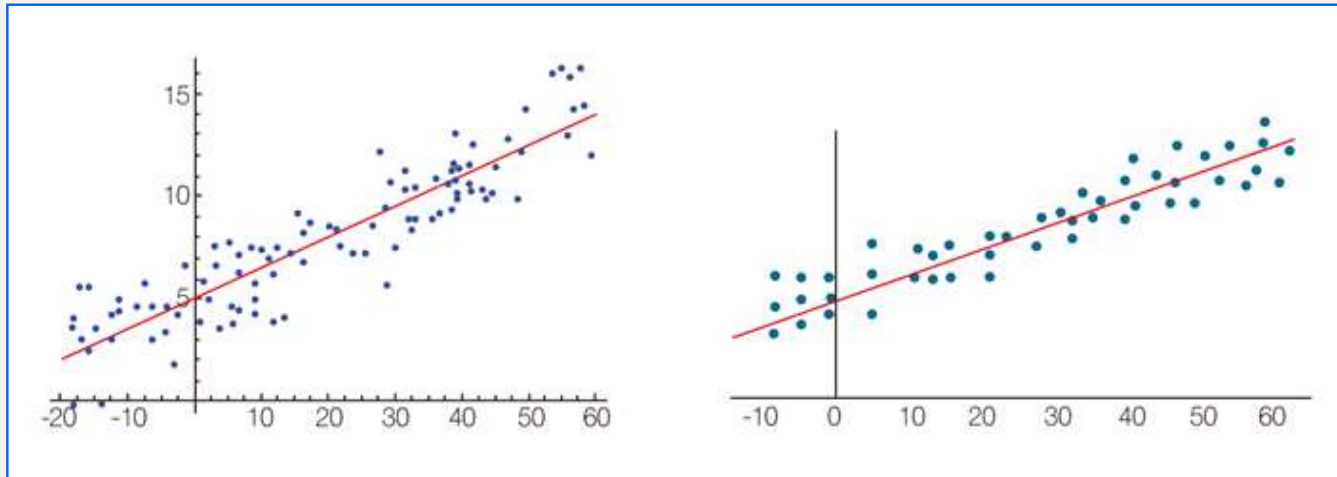




## AI 머신러닝의 학습 방법 (Cont'd)

### ● 회귀 분석 (Regression analysis)

- 변수 사이의 회귀에 대해 검정이나 추정을 하는 것
- 회귀 분석은 학습 데이터를 사용하여 출력값 예측
- 산출물은 항상 확률론적 의미를 내포



## 머신러닝의 학습 방법 (Cont'd)

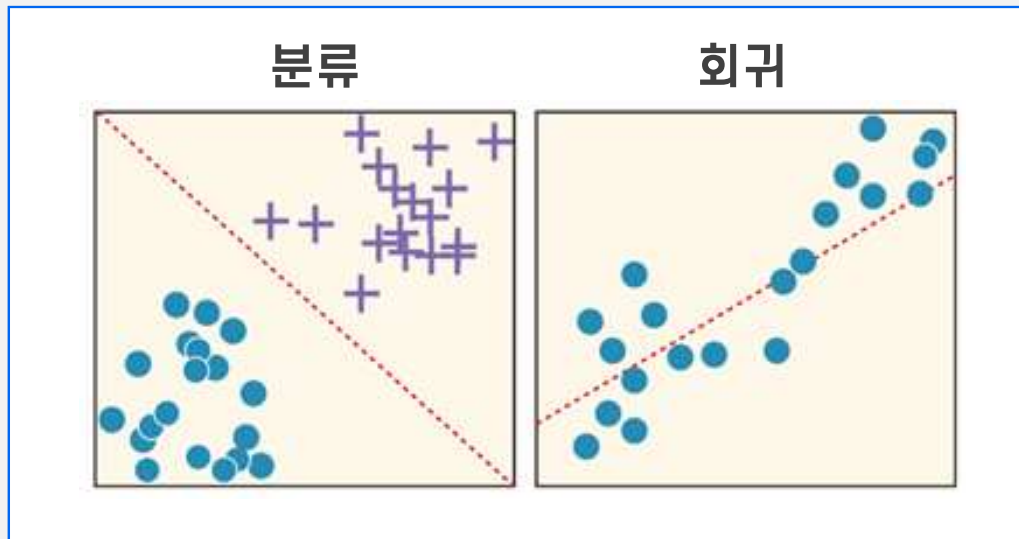
### ● 회귀와 회귀 분석의 예측에의 활용

- 날씨에 대한 예측
- 월별 판매액을 보고 다음 달 판매액 예측
- 금융, 투자, 비즈니스적 가격 판단
- 금값이나 원유 가격 예측
- 주택 가격의 예측
- 장단기 주가 예측
- 원유 가격 추정 등

## 🧠 머신러닝의 학습 방법 (Cont'd)

### ● 분류와 회귀의 차이점

- 분류는 일정한 기준에 따라 명백하게 구분 짓는 것
- 회귀는 오차 제곱의 합을 최소화하는 직선을 긋는 작업  
따라서 명확히 직선으로 구별되는 것이 아님



## 머신러닝의 학습 방법 (Cont'd)

### ● 분류와 회귀의 차이점 (cont'd)

- 분류의 출력은 남자/여자 등과 같은 선택식 출력
- “내일 날씨는 더울 것이다.”와 같은 이분법적 선택
- 회귀의 출력은 연속값으로 나타냄
- “내일 기온?”에 대해 “18.3도로 추정된다.” 등의 형태



## 머신러닝의 학습 방법 (Cont'd)

- 몇 가지 유형의 주요 분류 방법
  - Naive Bayes 분류기
  - 의사결정 트리
  - SVM
  - K-Nearest Neighbor(K-NN) 등

## 머신러닝의 학습 방법 (Cont'd)

### ● Naive Bayes 분류기

- 나이브 베이즈 분류기는 머신러닝의 한 분야
- 자료의 분류를 베이즈 정리를 활용하여 판단
- 나이브 베이즈 분류기는 조건부 확률 모델
- 모든 특성값은 서로 독립이라고 가정

**Naive Bayes Classifier**

In Machine learning

$$P(A / B) = (P(B/A) * P(A) / P(B))$$

## 머신러닝의 학습 방법 (Cont'd)

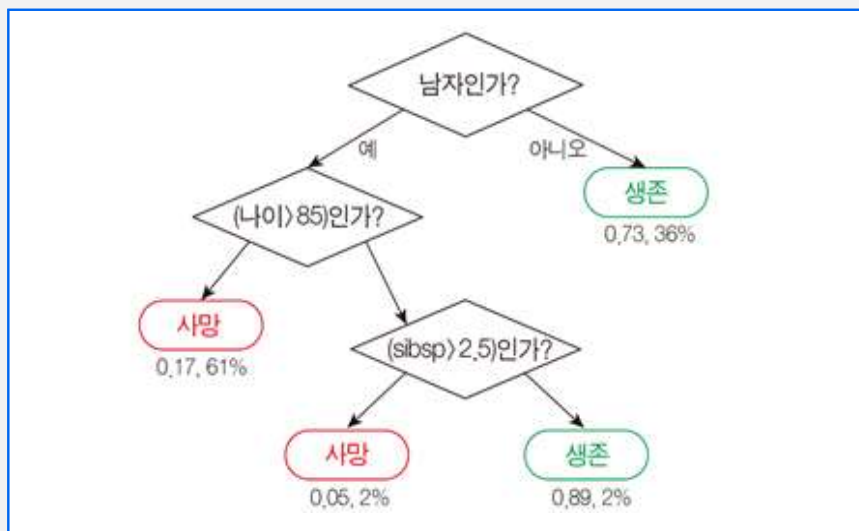
### ● Naive Bayes 분류기 (cont'd)

- 구축하기 쉽고, 대규모 데이터 세트에 유용함
- 지도 학습 환경에서 효율적으로 훈련될 수 있음
- 복잡한 실제 상황에서 비교적 잘 작동
- 주가의 상승이나 하락이 예상되는 종목들을 분류
- 문서의 내용에 따라 문서 분류 가능
- 이메일 내용에 따라 스팸/정상 메일로 분류

## 머신러닝의 학습 방법 (Cont'd)

### ● 의사결정 트리 (Decision Tree)

- 관측값과 목표값을 연결하는 예측 모델
- 최대 2가지의 판단을 하는 이진 트리 사용
- '스무고개' 문답처럼 선택 방법으로 진행
- 주택이나 자동차 구입비용 등의 추정에 활용
- 타이타닉호 탑승객의 생존 여부를 나타내는 결정 트리

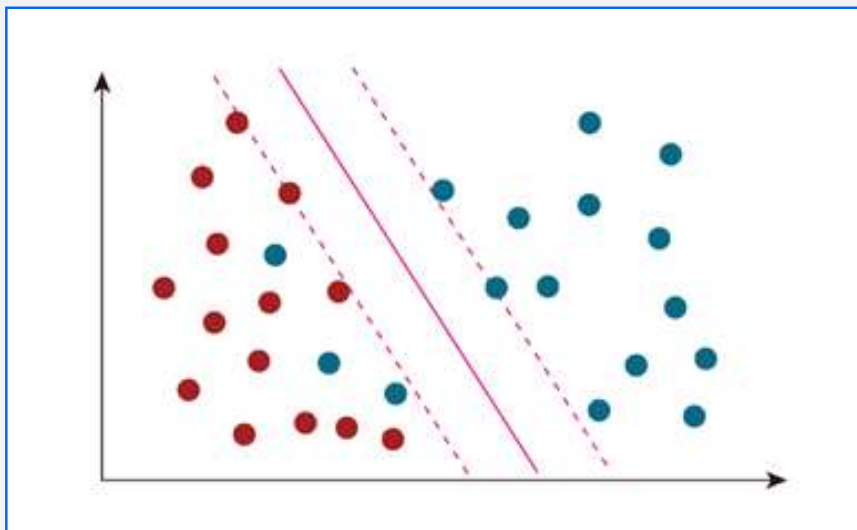




## 🧠 머신러닝의 학습 방법 (Cont'd)

### ● SVM (Support Vector Machine : SVM)

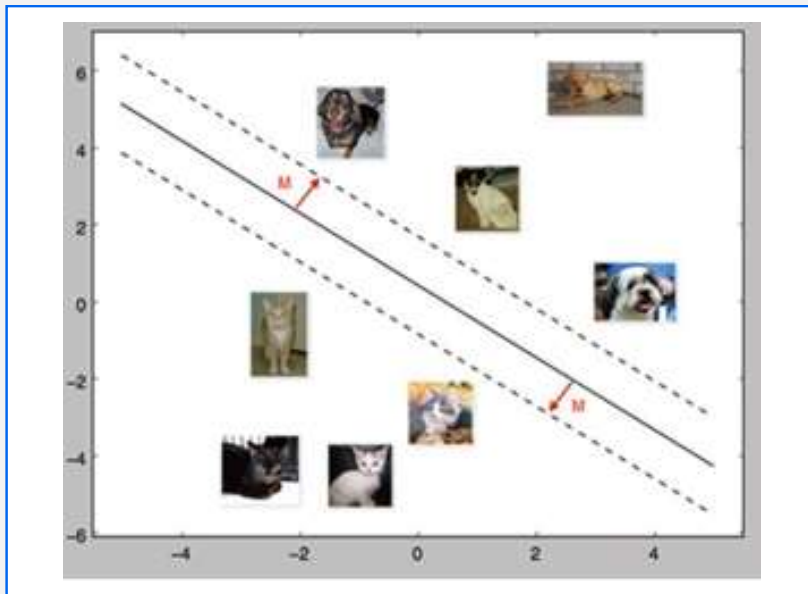
- 1990년에 개발, 통계 학습 이론의 결과 기반
- 데이터를 2개의 영역으로 분류하는 이진 분류기
- 새로운 데이터가 어느 영역에 속하는지를 판단
- 가장 큰 폭을 가진 하나의 경계선을 찾는 알고리즘
- 영역의 여백(margin, gap)이 최대가 되는 중심선 찾기



## AI 머신러닝의 학습 방법 (Cont'd)

### ● SVM (cont'd)

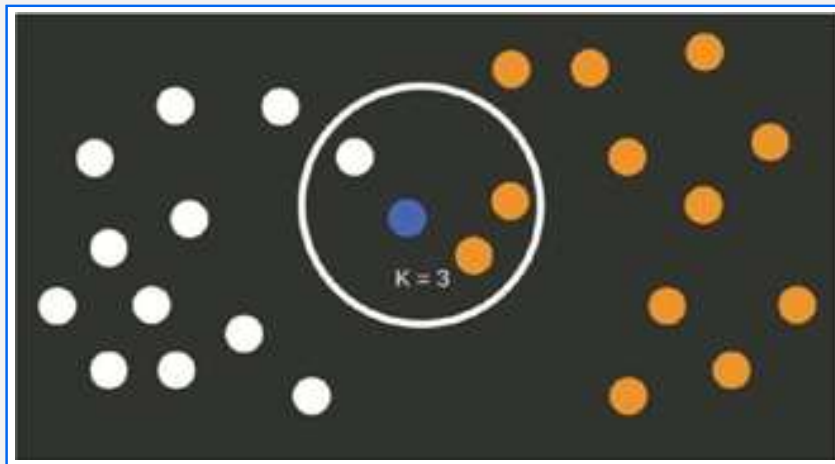
- SVM은 패턴인식과 자료 분석을 위한 지도 학습 모델임
- 분류, 회귀 분석, 멀티미디어 정보 검색 등에 사용
- 두 영역 사이의 여백을 최대로 하는 직선으로 분류
- SVM으로 개와 고양이의 특성을 분류에 활용하는 예



## 🧠 머신러닝의 학습 방법 (Cont'd)

### ● K-Nearest Neighbor (K-NN)

- 1950년대에 개발된 지도 학습 모델의 분류 기법
- 간단한 분류 기법, '최근접 이웃 분류'라고도 불림
- 가장 가까운 것들과의 거리 계산으로 클래스를 분류
- 새로운 입력 데이터와 가장 가까운 k개의 이웃 데이터 선택
- 이웃 데이터들의 클래스 중 다수결로 데이터의 클래스 결정
- 다수결에서 결과가 나오기 위해 k는 반드시 홀수여야 함



## 머신러닝의 학습 방법 (Cont'd)

### ● K-NN (cont'd)

- 영화나 음악 추천에 대한 개인별 선호 예측
- 수표에 적힌 광학 숫자와 글자인식
- 얼굴인식과 같은 컴퓨터 비전
- 유방암 등 질병의 진단과 유전자 데이터 인식
- 재정적인 위험성의 파악과 관리, 주식 시장 예측

## 머신러닝의 학습 방법 (Cont'd)

### ● K-NN (cont'd)

- 장점은 매우 간단하며 빠르고 효과적인 알고리즘
- 어떤 데이터라도 유사성 측정 가능
- 그러나 적절한  $k$ 를 선택하는 문제가 있고, 새로운 데이터가 추가되면 모든 데이터에 대해 거리를 계산한 후 분류함

## 🧠 머신러닝의 학습 방법 (Cont'd)

### ● K-NN (cont'd)

- 꽃잎의 크기와 밝기에 따른 K-NN 분류
- 오른쪽 위에 새로운 꽃잎이 입력으로 들어왔을 때 빨간 화살표의 3가지를 비교한 후 분류하는 것을 보여줌

