

# 03

## 머신러닝의 학습과 분류

Learning and Classification of Machine Learning

### CONTENTS

8.1 머신러닝의 개요

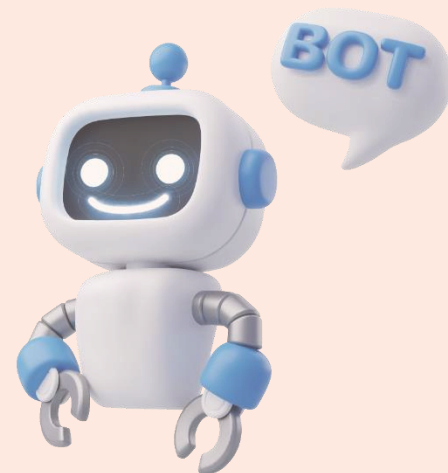
8.2 머신러닝의 학습 방법

8.3 머신러닝의 비지도 학습

8.4 강화 학습

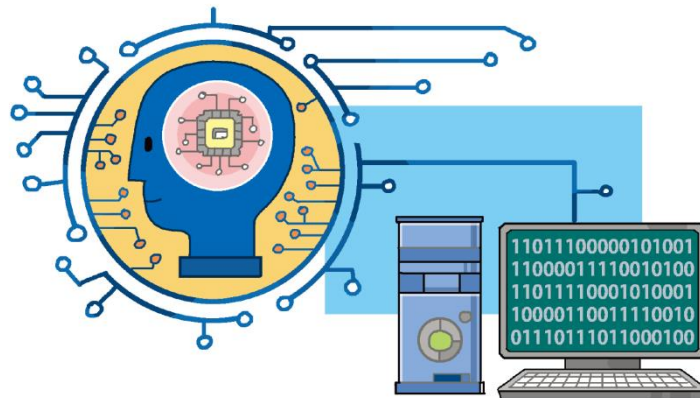
8.5 베이지안 네트워크와 은닉 마르코프 모델

- 가까운 곳에서 인공지능 경험하기
- 생각하고 토론하기
- 인공지능 실습하기
- 연습문제

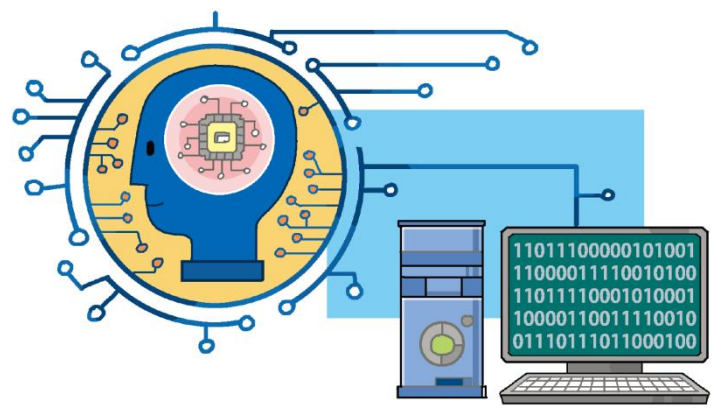


## [학습 목표]

머신러닝의 개념을 이해하고  
여러 가지 학습 방법과  
응용 분야들을 고찰한다.



# [학습 목차]



- 머신러닝의 정의와 종류, 그리고 인공지능과의 관계를 요약해본다.
- 머신러닝의 종류와 다양한 응용 분야를 파악해본다.
- 머신러닝의 3가지 학습 방법과 지도 학습에 따른 분류와 회귀를 알아본다.
- 머신러닝의 비지도 학습 클러스터와 클러스터링의 개념과 응용을 살펴본다.
- 머신러닝의 한 영역인 강화 학습의 개념과 응용 분야를 다룬다.
- 베이지안 네트워크와 은닉 마르코프 모델을 간략히 알아본다.



# 8.1 머신러닝의 개요

## (1) 머신러닝(Machine Learning : ML)의 정의

- '기계학습'으로도 불리는 인공지능의 한 분야
- 인간의 학습 능력을 컴퓨터로 실현하려는 기법
- 1959년 아서 새무얼(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



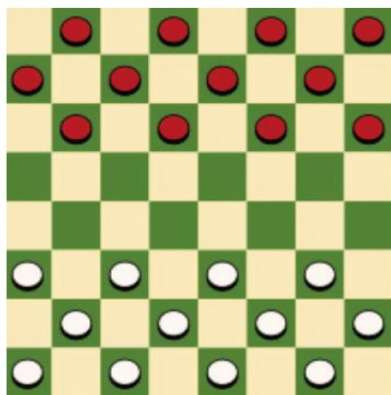
[그림 8.1] 머신러닝의 분야들



# 8.1 머신러닝의 개요

## (2) 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



[그림 8.2] 체커 게임 보드와 체커 프로그램을 개발하는 장면



# 8.1 머신러닝의 개요

## 머신러닝의 연도별 주요 개발 모델과 특징

〈표 8.1〉 머신러닝의 연도별 주요 개발 모델과 특징

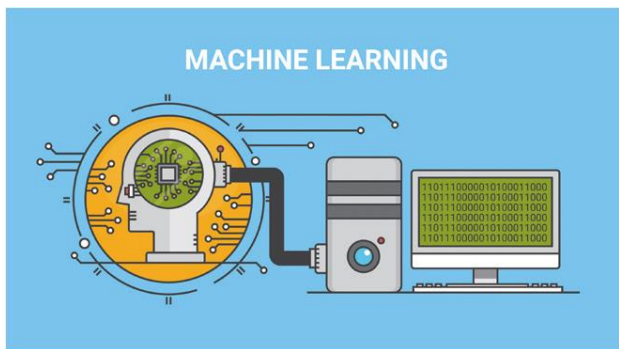
연도	개발자	모델	특징 또는 종류
1952년	Arthur Samuel	Checker Program	최초의 머신러닝
1957년	Frank Rosenblatt	Perceptron	최초의 신경망 모델
1986년	Rumelhart 등	Multilayer Perceptron	Back-propagation 알고리즘
1986년	Quinlan	Decision Tree	ID3
1995년	Vapnik, Cortes	Support Vector Machine	이진 분류기



# 8.1 머신러닝의 개요

## (3) 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



[그림 8.3] 머신러닝



# 8.1 머신러닝의 개요

## 머신러닝의 다양한 정의

- “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.”





# 8.1 머신러닝의 개요

## 머신러닝의 학습 개념

- 머신러닝을 통한 간단한 학습의 예
- 입력과 출력이 여러 개의 데이터 쌍으로 주어짐

$(1, 2), (2, 4), (4, 8), (7, 14), (5, 10), \dots$

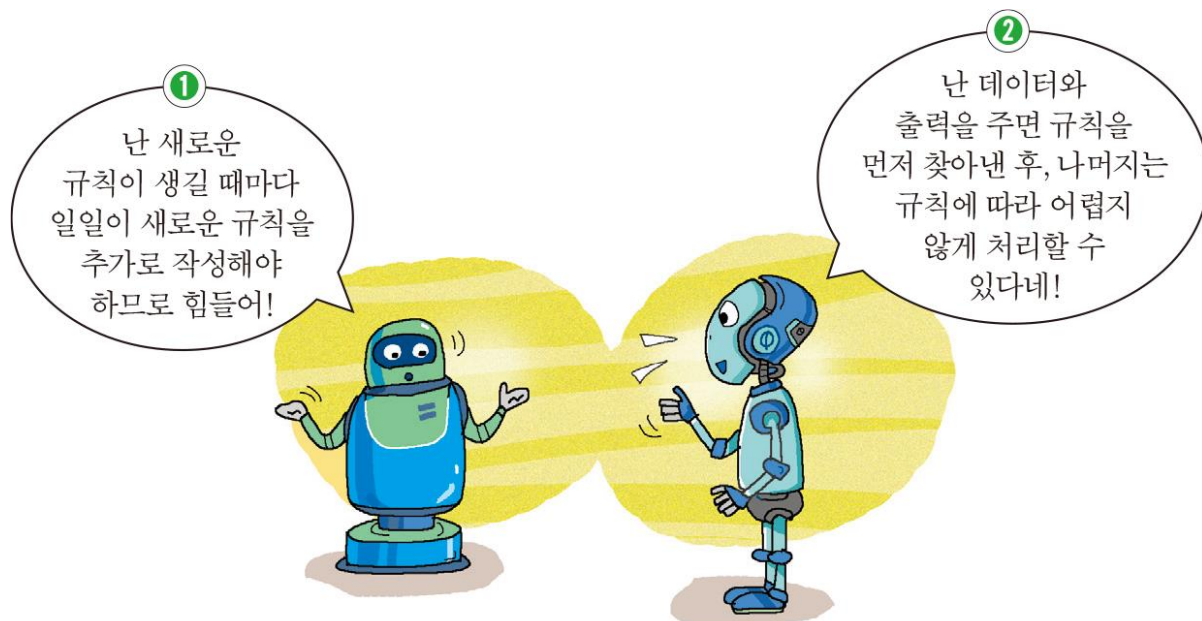
- 학습 후, 출력이 입력의 2배임을 유추
- $(3, ?), (8, ?)$  등의 질문에 6, 16 등으로 답변



# 8.1 머신러닝의 개요

## (4) 머신러닝과 전통적인 프로그래밍과의 차이점

- 전통적인 프로그래밍에서는 모든 규칙들을 작성함  
만약 규칙들이 추가로 작성될 경우 유지 관리가 어려운 문제점
- 그러나 머신러닝은 시간에 따라 점차 효율이 향상됨  
입출력 데이터의 관계를 학습하여 규칙을 생성

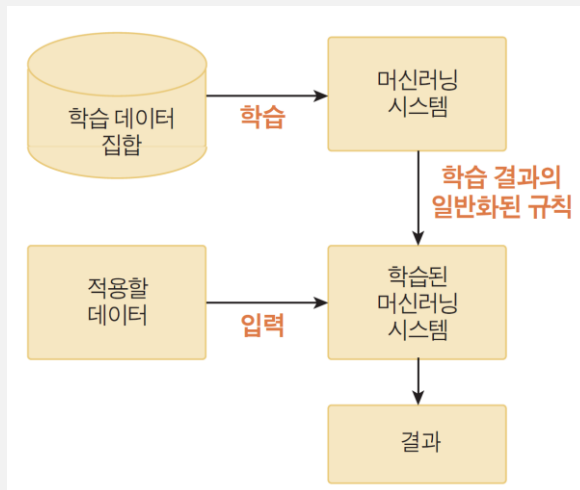




# 8.1 머신러닝의 개요

## 여기서 잠깐! 머신러닝 시스템과 학습

머신러닝에서는 학습 데이터 집합을 대상으로 학습시키는데, 그 결과 머신러닝 시스템은 학습 결과의 일반화된 규칙을 생성한다. 학습된 머신러닝 시스템에다 적용할 데이터를 입력시키면 적절한 결과를 얻을 수 있다.

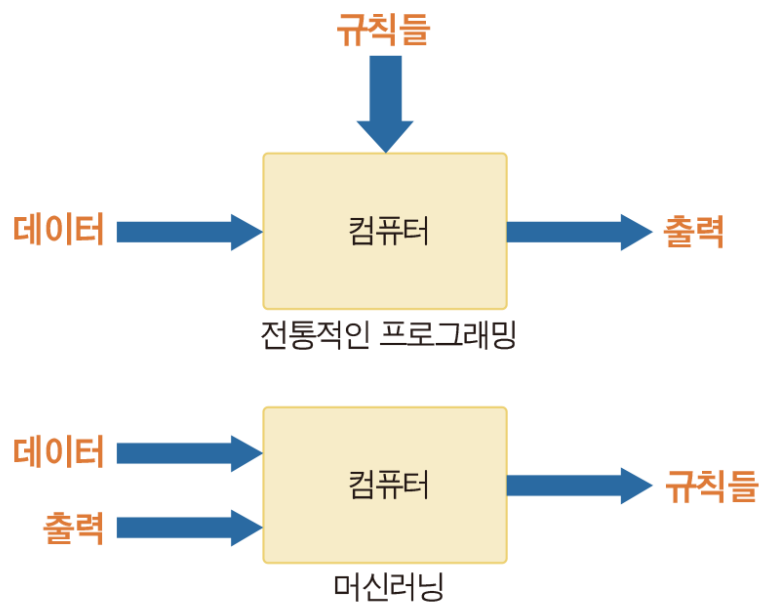




## 8.1 머신러닝의 개요

### 전통적인 프로그래밍과 머신러닝의 차이점

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



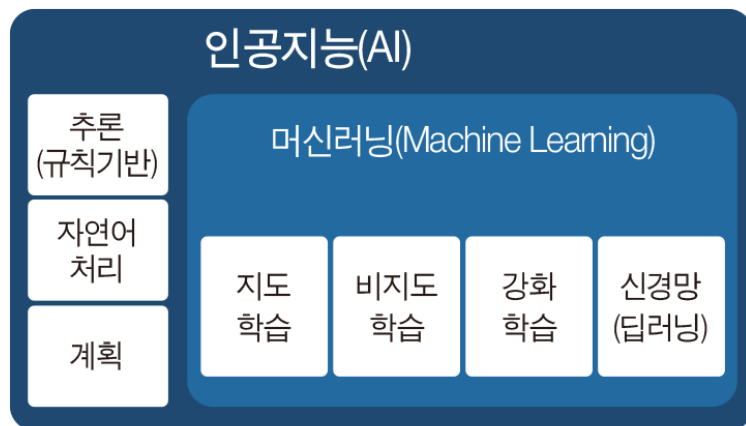
[그림 8.4] 전통적인 프로그래밍과 머신러닝의 차이점



# 8.1 머신러닝의 개요

## (5) 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



[그림 8.5] 머신러닝과 인공지능



# 8.1 머신러닝의 개요

## 머신러닝과 인공지능의 차이점

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

〈표 8.2〉 머신러닝과 인공지능의 차이점

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



# 8.1 머신러닝의 개요

## (6) 머신러닝 과정에서의 고려 사항

### 머신러닝 구상과정에서의 고려 사항들

- 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?



# 8.1 머신러닝의 개요

## (7) 머신러닝의 종류

### 머신러닝의 주요 종류들

- 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





# 8.1 머신러닝의 개요

## (8) 머신러닝의 활용 분야

〈표 8.3〉 머신러닝의 활용 분야

활용 분야	응용
영상인식	문자인식, 물체인식
얼굴인식	Facebook에서의 얼굴인식
음성인식	Bixby, Siri, Alexa 등
자연어 처리	자동 번역, 대화 분석
정보 검색	스팸 메일 필터링
검색 엔진	개인 맞춤식 추천 시스템
로보틱스	자율주행 자동차, 경로 탐색



# 8.1 머신러닝의 개요

## 머신러닝의 활용 분야

- 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



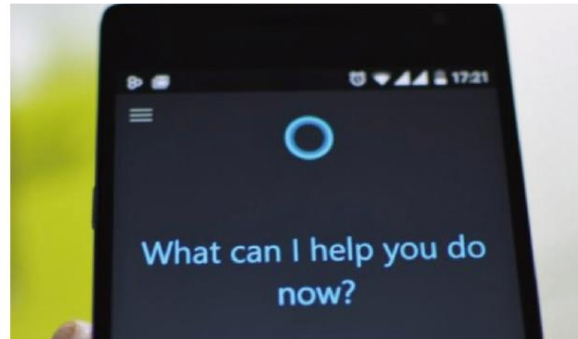
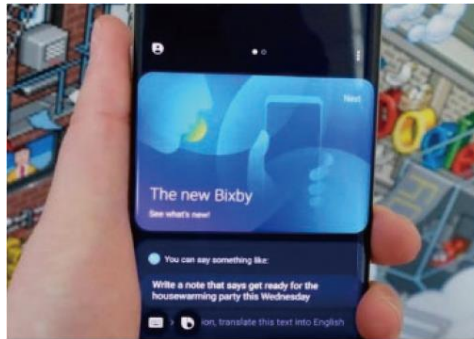
[그림 8.6] 자율주행 자동차



## 8.1 머신러닝의 개요

### 머신러닝의 개인비서(personal assistant)에의 활용

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



[그림 8.7] 스마트폰의 개인 비서



# 8.1 머신러닝의 개요

## 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
- 최근 큰 관심을 끌고 있음



[그림 8.8] 머신러닝의 헬스케어에서의 활용



# 8.1 머신러닝의 개요

## 머신러닝의 SNS에의 활용

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



[그림 8.9] 다양한 소셜 미디어 서비스



## 8.1 머신러닝의 개요

### 머신러닝의 동영상에의 활용

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



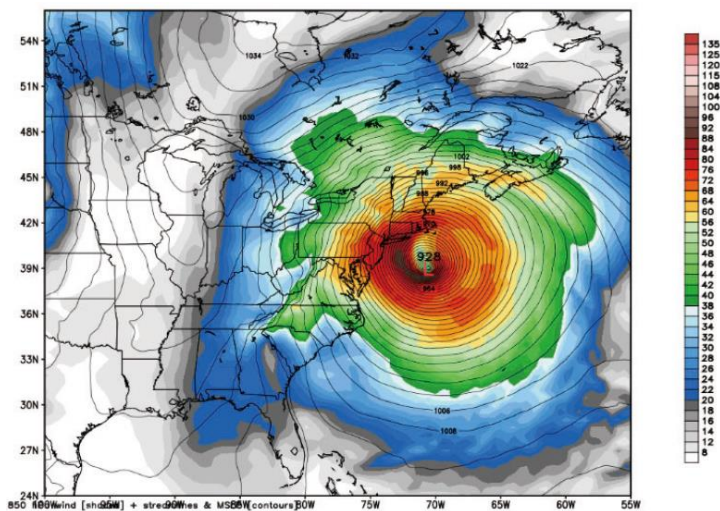
[그림 8.10] 동영상 추천 리스트



## 8.1 머신러닝의 개요

### 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



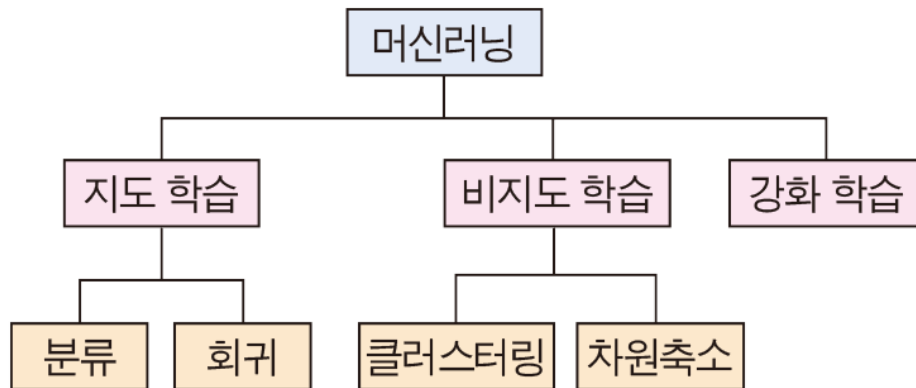
[그림 8.11] 기상 예측



## 8.2 머신러닝의 학습 방법

### (1) 머신러닝에서의 학습 방법

- 학습의 형태에 따라 3가지 학습 방법
- 지도 학습, 비지도 학습, 강화 학습으로 구분
- 지도학습은 분류와 회귀, 비지도 학습은 클러스터링과 차원 축소로 다시 나누어짐



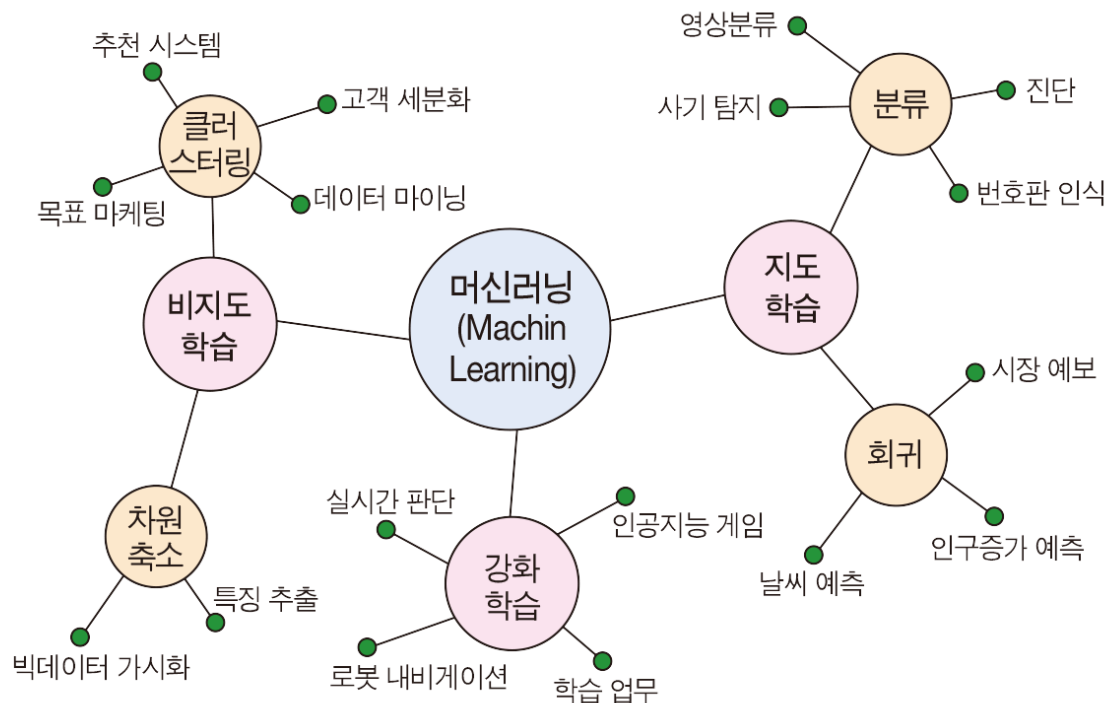
[그림 8.12] 머신러닝 학습의 분류





## 8.2 머신러닝의 학습 방법

### 머신러닝의 학습 방법과 활용 분야 체계



[그림 8.13] 머신러닝의 학습 방법과 활용 분야 체계

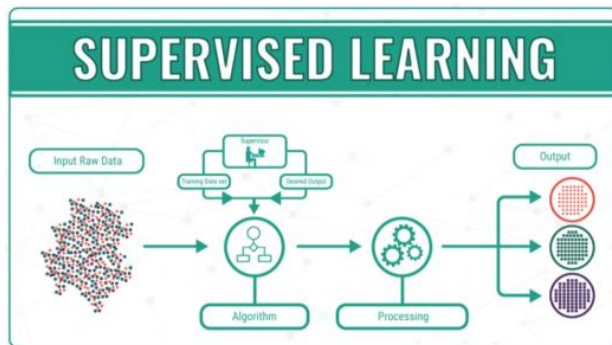


## 8.2 머신러닝의 학습 방법

### (1) 머신러닝에서의 학습 방법

#### ① 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

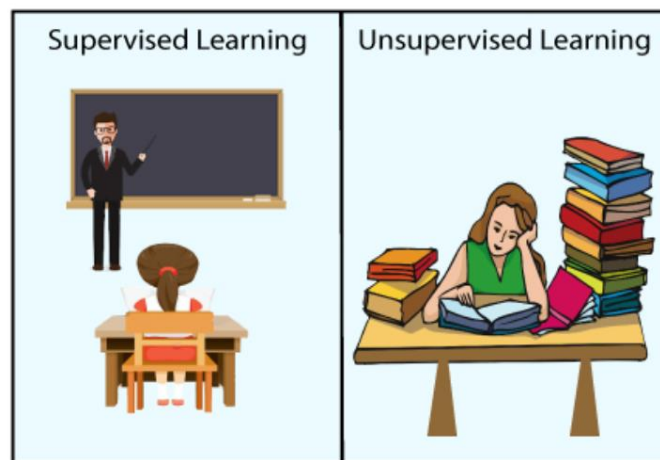




## 8.2 머신러닝의 학습 방법

### ② 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





## 8.2 머신러닝의 학습 방법

### ③ 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

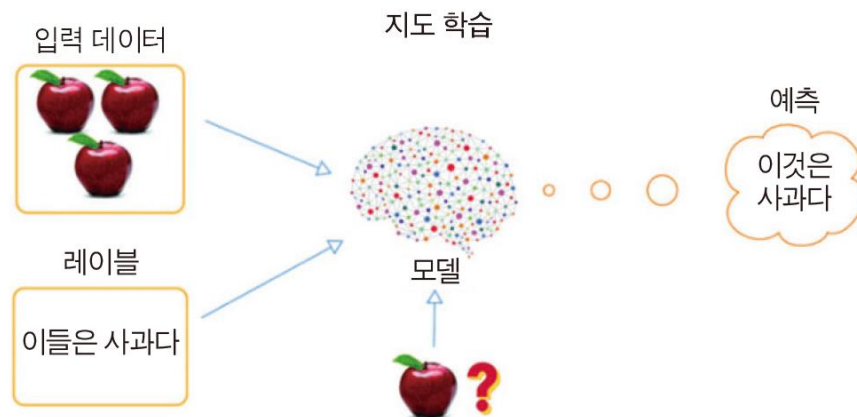




## 8.2 머신러닝의 학습 방법

### (2) 지도 학습(supervised learning)

- 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



[그림 8.14] 레이블이 있는 지도 학습



## 8.2 머신러닝의 학습 방법

### 지도 학습의 장단점

#### ① 장점

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

#### ② 단점

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



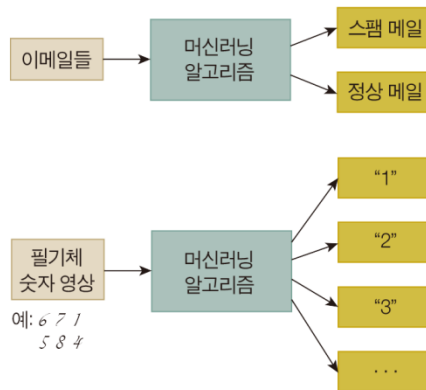
## 8.2 머신러닝의 학습 방법

### (3) 지도 학습에서의 분류와 회귀

지도 학습은 분류와 회귀로 나누어짐

#### ① 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



[그림 8.15] 여러 가지 분류



## 8.2 머신러닝의 학습 방법

### 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







## 8.2 머신러닝의 학습 방법

### 분류의 다양한 응용 예

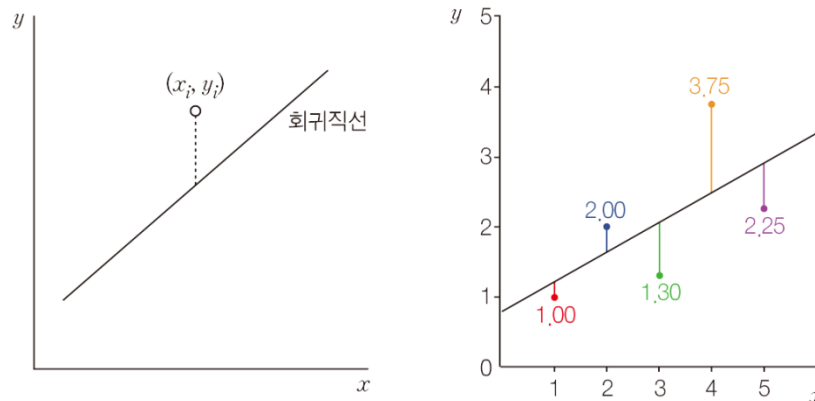
- 사진으로 남자와 여자의 구별
- 개와 고양이의 구분
- 스팸 메일과 정상 메일 구분
- 0에서 9까지의 숫자의 구분
- 알파벳과 한글 문자 등의 구분
- 편지봉투의 손으로 쓴 주소 판별
- 카드 부정 사용 감지
- 의료 영상에서 종양의 존재 여부 판단



## 8.2 머신러닝의 학습 방법

### ② 회귀(regression)

- 회기란 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  $x$ 에 대한  $y$ 의 회귀 직선이라 함



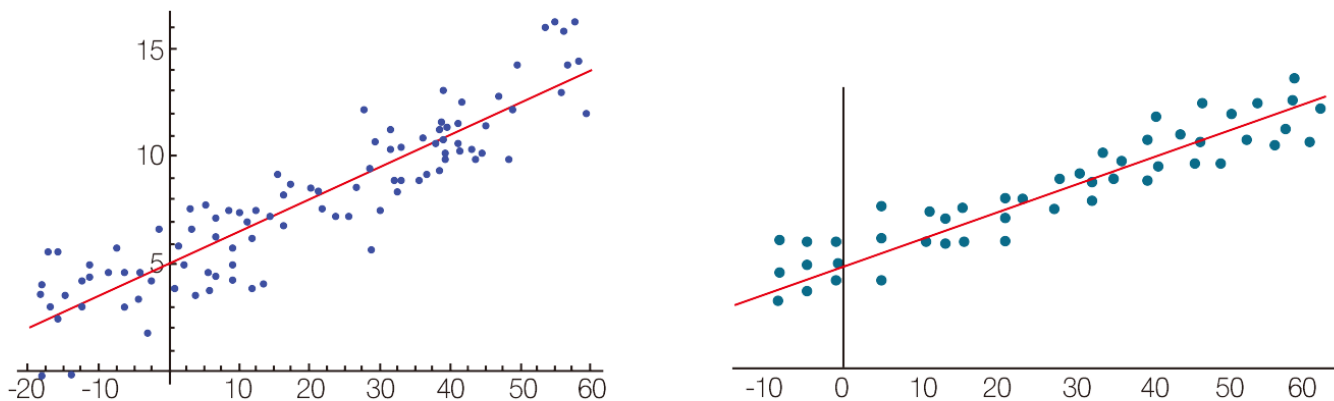
[그림 8.16] 회귀직선



## 8.2 머신러닝의 학습 방법

### 회귀 분석(regression analysis)

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



[그림 8.17] 선형 회귀



## 8.2 머신러닝의 학습 방법

### 회귀와 회귀 분석의 예측에의 활용

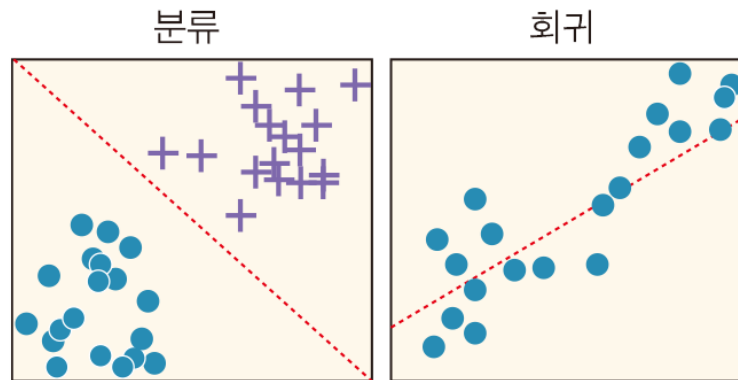
- 날씨에 대한 예측
- 월별 판매액을 보고 다음 달 판매액 예측
- 금융, 투자, 비즈니스적 가격 판단
- 금값이나 원유 가격 예측
- 주택 가격의 예측
- 장단기 주가 예측
- 원유 가격 추정 등



## 8.2 머신러닝의 학습 방법

### ③ 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



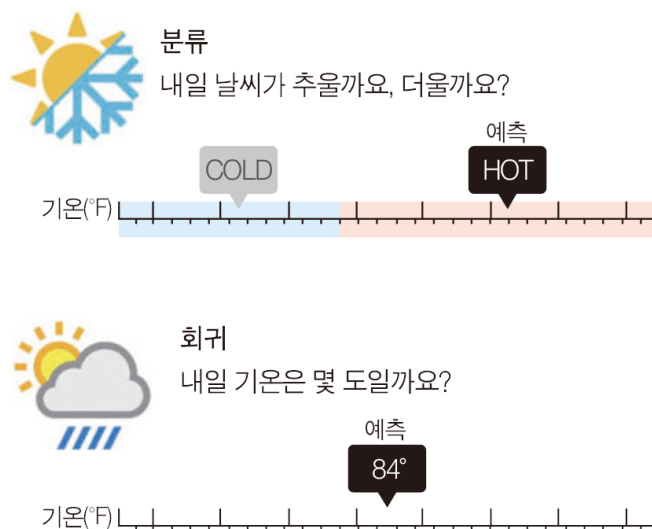
[그림 8.18] 분류와 회귀의 비교



## 8.2 머신러닝의 학습 방법

### 분류와 회귀의 차이 구분

- 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



[그림 8.19] 분류와 회귀의 차이점



## 8.2 머신러닝의 학습 방법

### (3) 분류의 방법

#### 몇 가지 유형의 주요 분류 방법

- Naive Bayes 분류기
- 의사결정 트리
- SVM
- K-Nearest Neighbor(K-NN) 등



## 8.2 머신러닝의 학습 방법

### ① 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

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

[그림 8.20] 나이브 베이즈 분류





## 8.2 머신러닝의 학습 방법

### 나이브 베이즈 분류의 장점

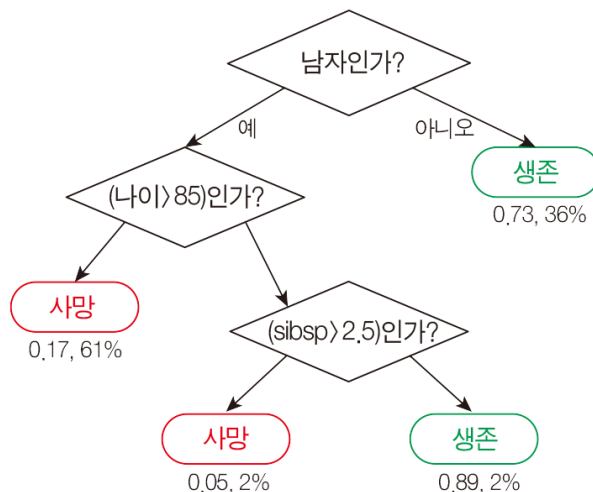
- 구축하기 쉽고, 대규모 데이터 세트에 유용함
- 지도 학습 환경에서 효율적으로 훈련될 수 있음
- 복잡한 실제 상황에서 비교적 잘 작동
- 주가의 상승이나 하락이 예상되는 종목들을 분류
- 문서의 내용에 따라 문서 분류 가능
- 이메일 내용에 따라 스팸/정상 메일로 분류



## 8.2 머신러닝의 학습 방법

### ② 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



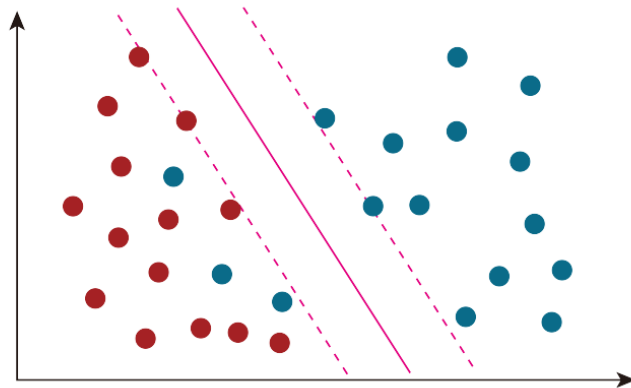
[그림 8.21] 결정 트리의 한 예



## 8.2 머신러닝의 학습 방법

### ③ 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



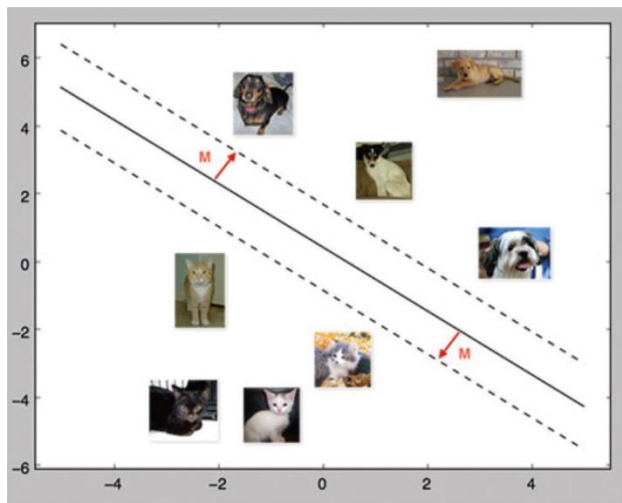
[그림 8.22] 두 클래스를 분류하는 SVM



## 8.2 머신러닝의 학습 방법

### SVM 분류의 활용

- 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



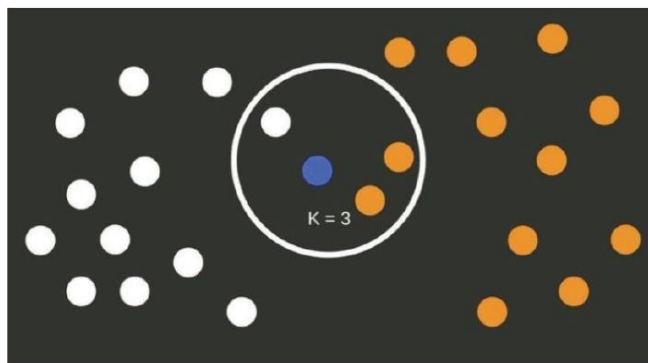
[그림 8.23] SVM에 의한 개와 고양이의 특성 분류



## 8.2 머신러닝의 학습 방법

### ④ 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



[그림 8.24]  $k = 3$ 인 경우의 분류 판단



## 8.2 머신러닝의 학습 방법

### K-NN의 장단점과 활용 분야

- 장점은 매우 간단하며 빠르고 효과적인 알고리즘  
또 어떤 데이터라도 유사성 측정 가능
- 단점으로는 적절한  $k$ 를 선택해야 한다는 점  
새로운 데이터에 대해 일일이 거리를 계산한 후 분류

### K-NN의 활용 분야

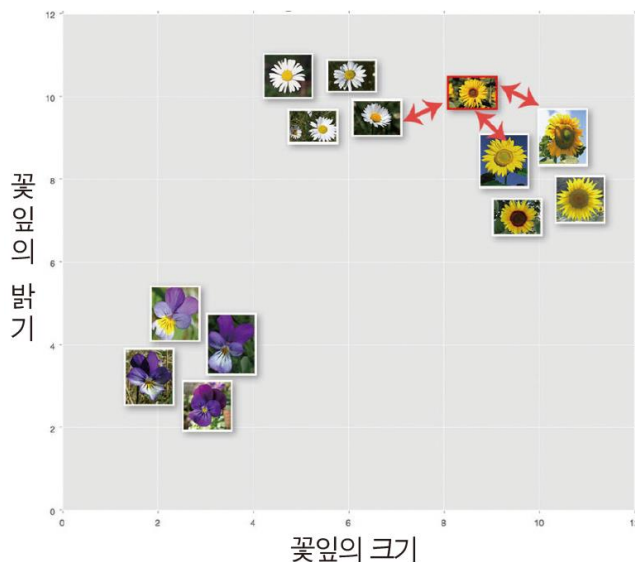
- 영화나 음악 추천에 대한 개인별 선호 예측
- 수표에 적힌 광학 숫자와 글자인식
- 얼굴인식과 같은 컴퓨터 비전
- 유방암 등 질병의 진단과 유전자 데이터 인식
- 재정적인 위험성의 파악과 관리, 주식 시장 예측



## 8.2 머신러닝의 학습 방법

### K-NN의 꽃잎 분류에의 적용 예

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



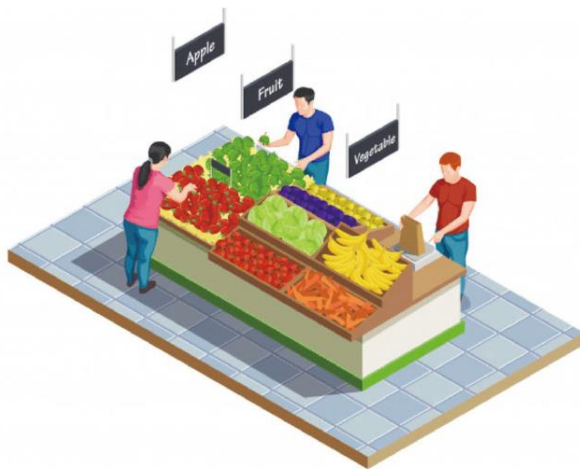
[그림 8.25] 꽃잎의 크기와 밝기에 따른 K-NN 분류



## 8.3 머신러닝의 비지도 학습

### (1) 클러스터(cluster)와 클러스터링(clustering)

- 클러스터는 유사한 여러 개의 클래스로 나뉘어진 데이터
- 클러스터링은 유사한 특성을 가진 그룹들로 묶는 작업
- 같은 클러스터의 것은 다른 클러스터의 것보다 더 유사
- 이와 같은 유사한 것들끼리의 집합이 바로 클러스터



[그림 8.26] 사과, 과일, 야채의 클러스터

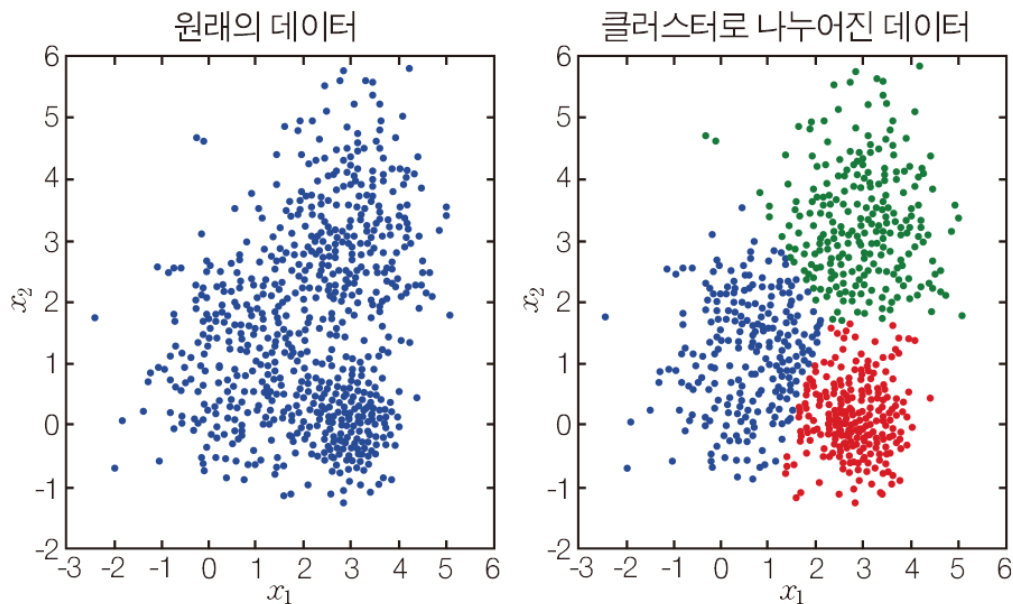




## 8.3 머신러닝의 비지도 학습

### 클러스터의 분류 예

- 원래 데이터를 3개의 클러스터로 분류해 놓은 예
- 빨간색, 파란색, 녹색



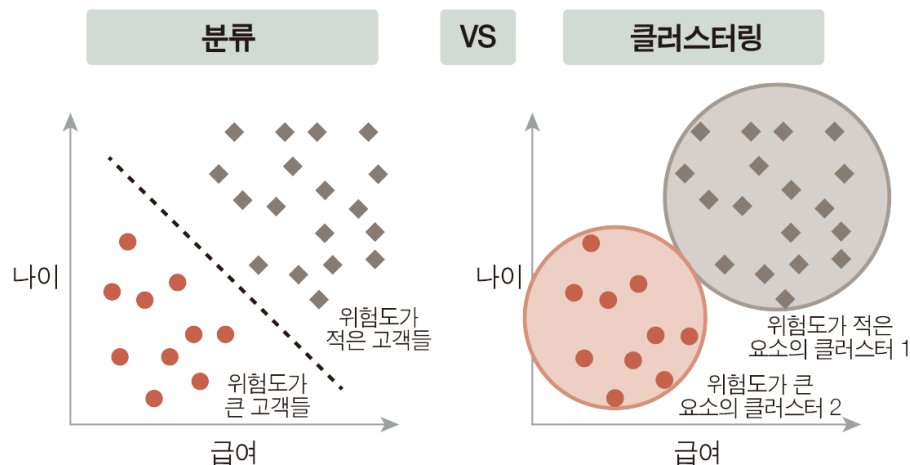
[그림 8.27] 원래 데이터와 3개의 클러스터



## 8.3 머신러닝의 비지도 학습

### 분류와 클러스터링의 차이점

- 분류는 지도 학습 영역, 클러스터링은 비지도 학습 영역
- 분류는 데이터를 기준에 따라 직선으로 분류하는 것
- 클러스터링은 유사성에 따라 몇 개의 클러스터들로 묶는 것
- 급여, 나이, 위험도 상관관계의 예에서의 차이점의 예



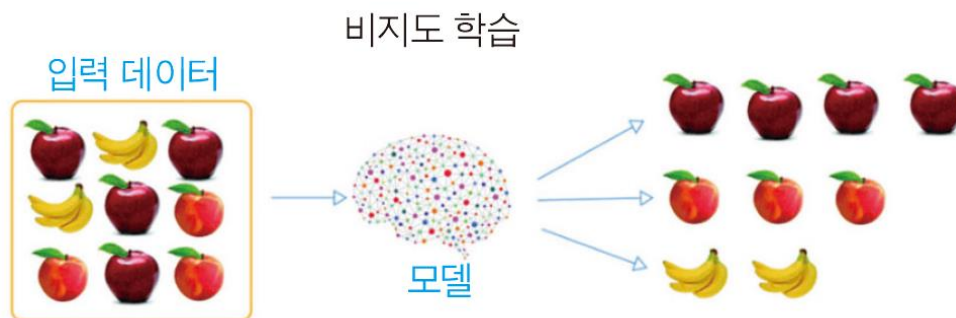
[그림 8.28] 분류와 클러스터링의 차이점



## 8.3 머신러닝의 비지도 학습

### (2) 비지도 학습

- 주어진 입력에 대응하는 출력 정보 없이 학습
- 데이터 분류에 대한 정보가 전혀 없이 패턴을 찾거나 데이터를 분류하려고 할 때 사용하는 학습 방법
- 데이터에 레이블을 전혀 사용하지 않음
- 관계를 스스로 학습한 후, 과일들을 각 그룹으로 알아서 묶기



[그림 8.29] 비지도 학습



## 8.3 머신러닝의 비지도 학습

### 비지도 학습의 주요 응용 분야

- 비슷한 성향의 고객을 그룹으로 묶기
- 블로그에서 주제별로 구분하기
- 유사한 꽃이나 동물들끼리 묶기
- 네트워크상에서의 비정상적인 접근의 탐지



## 8.3 머신러닝의 비지도 학습

### (3) 비지도 학습을 통한 클러스터링과 추천 시스템

#### 머신러닝에서의 주요 비지도 학습 방법

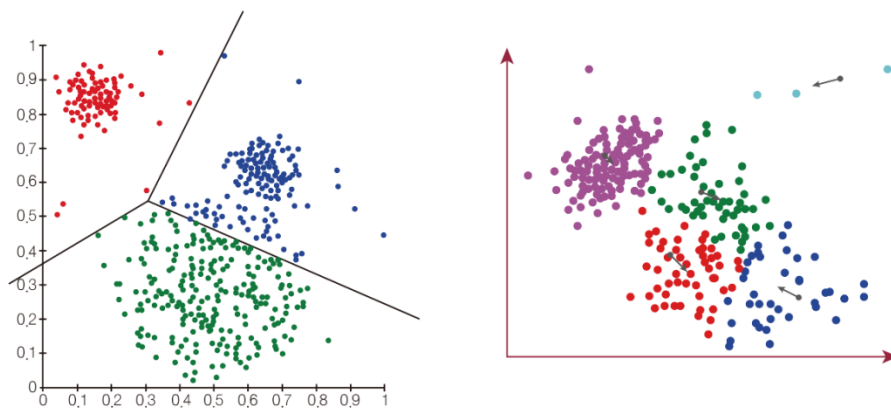
- K-means 클러스터링
- 가우스 혼합 모델
- 계층적 클러스터링
- 추천 시스템 등



## 8.3 머신러닝의 비지도 학습

### ① K-means 클러스터링

- 비지도 학습 알고리즘 중 대표적인 클러스터링 방법
- 우리말로 'K-평균 군집화'라고 함
- 간단하면서도 많이 쓰이는 클러스터링 방법 중 하나
- 유사한 특성을 가진 k개의 데이터 그룹으로 묶는 방법
- 예로 주어진 데이터 집합에서 3개와 4개의 클러스터들



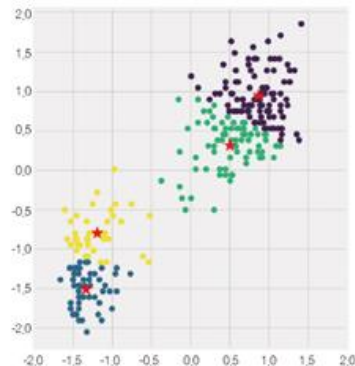
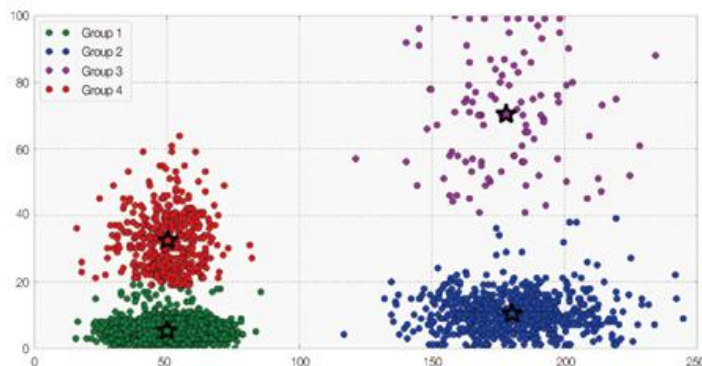
[그림 8.30] 클러스터



## 8.3 머신러닝의 비지도 학습

### 클러스터와 클러스터 중심점(centroid)

- 주어진 데이터 집합에 대해 k개의 클러스터 중심점 찾기
- 각 클러스터에는 클러스터 중심이 있음
- 각 점은 다른 중심점보다 지정된 클러스터 중심점에 더 가까움
- 4개의 클러스터로 구성된 2가지 예
- 별표는 각 클러스터의 중심점을 나타냄





## 8.3 머신러닝의 비지도 학습

### K-means 클러스터링의 장단점

- 장점은 알고리즘이 비교적 간단하고, 수행 속도가 빠르다는 점  
주어진 데이터에 대한 사전 정보 없이 클러스터링을 함  
데이터를 분류하는 머신러닝과 데이터 마이닝의 도구
- 단점은 클러스터링의 개수  $k$ 와 최초로 지정하는 중심점들에  
따라 결과가 다소 달라질 수 있는 점





## 8.3 머신러닝의 비지도 학습

### K-means 클러스터링의 활용 분야

- 통계 : 주어진 데이터의 분류나 성향 분석
- 전자상거래 : 고객의 구매 이력으로 고객 분류
- 건강 관리 : 질병과 치료를 위한 패턴 탐지
- 패턴 : 유사한 이미지를 그룹화
- 재무 : 신용카드 사기 탐지
- 회사 : 매출 등을 토대로 회사의 등급 분류
- 기술 : 네트워크 침입과 악의적 활동 탐지
- 기상 예보 : 폭풍 예측



## ② 추천 시스템(Recommender System)

- 추천을 위해 연관 데이터 정의에 도움 주는 클러스터링 방법
- 사용자의 '선호도'를 예측하는 정보 필터링의 일종
- 네이버나 구글 등에서 상업적으로 활용 중
- 현재 검색해본 책이나 동영상 등의 추천
- 또 인기 있는 식당, 연구 관련 기사, 금융 서비스 등 추천



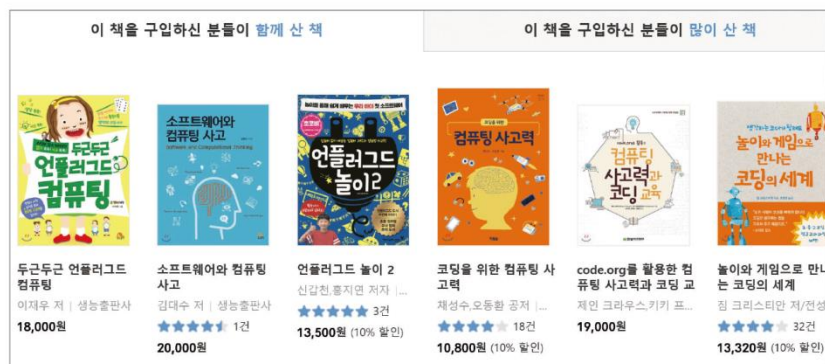
**[그림 8.33] 추천 시스템**



## 8.3 머신러닝의 비지도 학습

### 추천 시스템의 활용과 기타 비지도 학습 방법

- 가령 교보문고에서 책을 검색하면 그 사람이 이전에 검색했던 도서나 관련 도서를 알려줌
- 사용자의 검색 경험 정보 파악, 적절한 광고 내보내기



[그림 8.34] 도서 검색의 경험 활용

- 그 외 비지도 학습 방법에는 가우스 혼합 모델, 계층적 클러스터링, PCA/T-SNE 등이 있음



## 8.3 머신러닝의 비지도 학습

### (4) 지도 학습과 비지도 학습의 특징 비교

〈표 8.4〉 지도 학습과 비지도 학습의 특징 비교

기반	지도 학습	비지도 학습
입력 데이터	입력과 출력(값 또는 레이블)이 지정된 데이터를 사용하여 학습함	출력값이나 레이블이 전혀 없는 데이터를 사용하여 학습함
주요 기능	분류, 회귀	클러스터링, 추천 시스템
계산의 복잡성	비교적 간단함	상당히 복잡함
정확성	매우 정확함	다소 덜 정확함



## 8.3 머신러닝의 비지도 학습

### (4) 지도 학습과 비지도 학습의 특징 비교

- 지도 학습과 비지도 학습의 차이는 지도하는 주체의 유무 여부
- 지도 학습에서는 입력 데이터가 사과임을 모델에게 학습시킴
- 비지도 학습에서는 모델이 스스로 사과와 바나나 등을 분류함



[그림 8.33] 지도 학습과 비지도 학습



## 8.4 강화 학습

### (1) 강화 학습이란?

- 강화 학습은 시행착오를 통해 보상하는 행동 학습
- 최적의 값을 추구하기 위해 당근과 채찍을 사용
- 로봇이 미로에서 옳은 방향으로 진입하면 +2점, 막힌 길로 들어가면 -3점 등
- 입출력이 쌍으로 된 훈련 집합으로 제시되지 않는다는 점에서 일반적인 지도 학습과는 다름



## 8.4 강화 학습

### (2) 강화 학습의 구조와 응용 분야

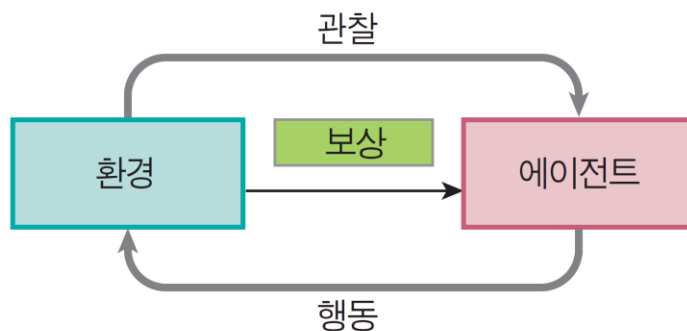
- 보상(reward)이 주어지는 문제 해결에 매우 효과적
- 통신망, 로봇 제어, 엘리베이터 제어, 그리고 체스와 바둑 같은 게임에 주로 응용됨
- 알파고도 강화 학습을 통해 실력 향상
- 게임에서는 거의 필수적으로 강화 학습이 사용됨
- 최근 강화 학습은 스타크래프트와 같은 전략게임, 금융 최적화, 에너지 관리 등에서 사용됨



## 8.4 강화 학습

### 강화 학습의 4가지 기본 구성요소

- ① 에이전트(Agent): 환경에서 어떤 행동을 취할지 결정하는 학습 주체
- ② 환경(Environment): 에이전트의 행동에 반응하여 보상을 제공하는 공간
- ③ 행동(Action): 에이전트가 환경에서 선택할 수 있는 행동
- ④ 보상(Reward): 환경이 에이전트의 행동에 따라 제공하는 피드백으로, 보상을 최대화하는 것이 에이전트의 주된 목표이다.



[그림 8.34] 강화 학습 구조도





## 8.4 강화 학습

### (3) 강화 학습과 큐러닝

- 강화 학습은 에이전트(agent)가 환경과 상호작용하며 보상을 최대화하는 행동을 학습하는 머신러닝의 학습 방법
- 예를 들어, 아이가 뜨거운 물에 손을 데이면 다음부터는 뜨거운 물을 피하게 되는 것과 같은 원리
- 큐함수(Q-function)는 에이전트가 어떤 상태에서 어떤 행동을 취했을 때 얻을 수 있는 예상 보상을 나타냄
- 큐러닝(Q-learning)은 강화 학습의 대표적인 알고리즘으로 큐함수를 학습하는 방법을 사용함
- 큐러닝의 핵심은 벨만 방정식으로 큐함수를 업데이트하는 것
- 에이전트는 시간이 지남에 따라 최적의 행동 전략을 학습함



## 8.5 베이지안 네트워크와 은닉 마르코프 모델

### (1) 베이즈의 정리(Bayesian theorem)

- 과거의 데이터들을 기반으로 미래를 예측하는 모델
- 머신러닝, 통계학, 경제학에 널리 적용되고 있음
- 검색 엔진, 스팸 메일 차단, 금융 이론, 승부 예측, 기상 예측, 의료 분야, 인공지능 등에 폭넓게 활용됨
- 베이즈(Thomas Bayes)는 확률에 대한 연구로 유명
- 베이즈의 정리는 확률적 추론에 이용되는 정리

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$



## 8.5 베이지안 네트워크와 은닉 마르코프 모델

### ‘베이지의 정리’의 응용 예

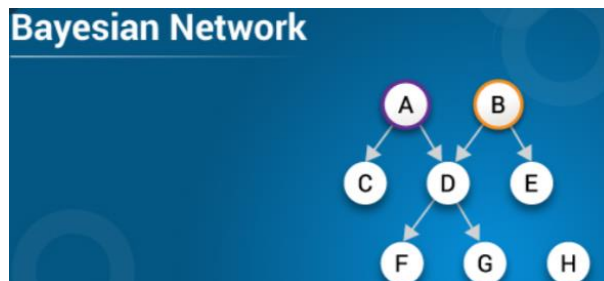
- $P(Y|X)$ 는 ‘X가 주어졌을 때 Y가 발생할 조건부 확률’
- 비교적 구하기 쉬운 확률을 통해 어려운 확률을 추정
- 나이브 베이지안과 은닉 마르코프 모델 등에 적용
  
- 증상과 의학 진단에 활용
- X가 ‘열이 많이 난다’,  $P(X)$ 는 열이 많이 나는 환자가 있을 확률
- Y가 ‘독감’,  $P(Y)$ 는 환자 중에 독감에 걸린 환자가 있을 확률
- $P(Y|X)$ 는 열이 많이 나는 환자가 독감 환자일 확률
- $P(X|Y)$ 는 독감 환자가 열이 많이 나는 확률



## 8.5 베이지안 네트워크와 은닉 마르코프 모델

### (2) 베이지안 네트워크(Bayesian network)

- '빌리프 네트워크(Belief network)'라고도 불림
- 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델
- 추론과 학습을 수행하기 위한 효과적인 알고리즘이 존재
- 예를 들어, 질환과 증상 사이의 확률 관계를 나타낼 수 있음
- 증상이 주어지면 다양한 질병의 존재 확률 계산 가능





## 8.5 베이지안 네트워크와 은닉 마르코프 모델

### (3) 은닉 마르코프 모델(Hidden Markov Model, HMM)

- HMM은 마르코프(Markov) 모델의 일종
  - 은닉된 상태와 관찰 가능한 결과로 이루어진 확률형 모델
  - 동적 베이지안 네트워크로 간단히 나타낼 수 있음
  - 대량의 데이터를 통계적으로 분석하여 추론에 응용
  - 음성인식, 자연어 처리 등에 활용
- 머신러닝과 관련된 KAIST 교수님의 강의 동영상(1시간)  
<https://www.youtube.com/watch?v=Wf6lIJZgKg>

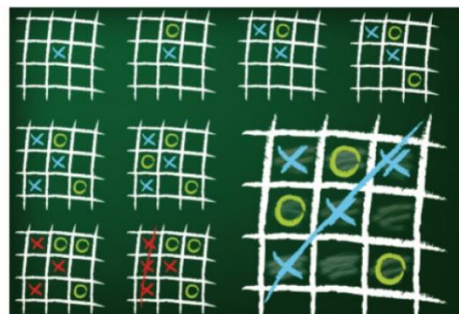
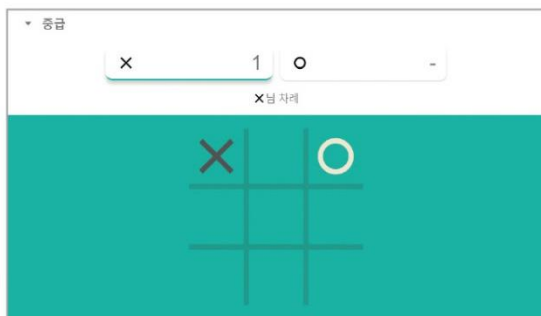
# 가까운 곳에서 인공지능 경험하기

## 틱택토 게임 경험하기

제6장에서 소개한 바와 같이 틱택토(Tic-Tac-Toe)란 미국 어린이들이 즐기는 게임인데, 가로, 세로, 대각선으로 3개가 연속으로 표시된 경우에 이기는 게임이다. 이 게임에서 이길 확률이 많은 전략 즉 휴리스틱을 경험해보자.

Google 검색에서 'tic-tac-toe 게임'을 치면 이 게임이 바로 나오는데, 그냥 첫수 위치를 선택하여 누르면서 게임을 시작하면 된다.

✓ **생각의 관점 Hint** 가장 중심부, 네 귀퉁이로 시작하는 것 등



# 생각하고 토론하기

1. 머신러닝은 현재 패턴을 인식하는 영역에서 큰 역할을 담당하고 있다. 머신러닝을 통한 패턴인식의 가능성과 미래를 생각해보고 의견을 나누어보자.

✓ **아이디어 포인트** 문자인식, 음성인식, 얼굴인식, 동영상인식 기술 등

2. 머신러닝에서 지도 학습이 태그를 하나씩 붙여 가르치며 학습하는 것은 이해가 될 것이다. 그러나 비지도 학습에서 주어진 입력에 대응하는 출력 정보 없이 학습한다고 한다. 그것이 어떤 방법으로 가능할지 그 이유를 생각해보고 의견을 나누어보자.

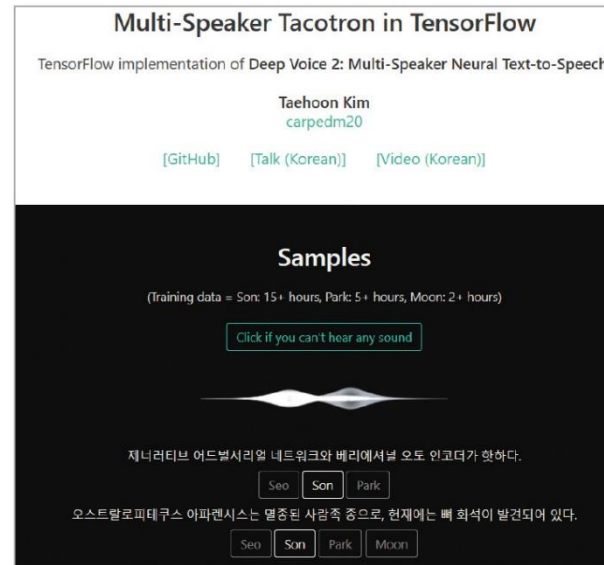
✓ **아이디어 포인트** 각자 전공과의 융합. 예를 들면 음대에서는 인공지능 작곡이나 편곡, 전자공학과의 서의 인공지능 로봇의 구현 등

# 인공지능 실습하기

## 실습 8 텐서플로로 구현된 신경망 특정화자 목소리 시연

특정 화자(손석희)의 목소리를 흉내내어 문장을 읽어준다.

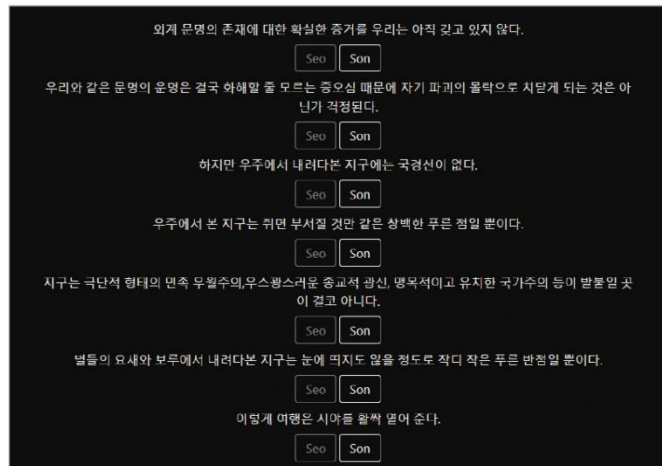
- <https://carpedm20.github.io/tacotron/>





# 인공지능 실습하기

**| 실행방법 |** 해당 사이트에서 Seo, Son, Park을 눌러 음성합성 결과물을 직접 들어보면 된다. 그런데 현재는 Son만 작동되므로 해당 칸을 누르면 손석희 아나운서의 톤으로 음성이 들린다. 여러 가지 경우를 눌러보고 경험하며 유사성을 알아본다.



**| 실행결과 |** 많은 문장을 특정 화자의 목소리로 들을 수 있게 발음한다.

## 퀴즈...

1. 머신러닝은 인간의 학습 능력과 같은 기능을 실현하려는 기법과 분야이다. ○

## 퀴즈...

2. 머신러닝을 흔히 컴퓨터 프로그램이 데이터로부터의 학습하는 과정이라고 말한다. ○

## 퀴즈...

3. 지도 학습을 통해서는 자동차 번호판의 인식률을 높일 수 없다. **X**

## 퀴즈...

4. 비지도 학습에서는 데이터 학습을 위해 레이블이나 태그 표식을 붙인다. **X**

## 퀴즈...

5. 회귀 분석 기술은 학습 데이터를 사용하여 하나의 출력값을 예측하는데, 산출물은 항상 확률론적 의미를 내포하고 있다. ○

퀴즈...

6. 분류는 일정한 기준에 따라 명백하게 구분 짓는 것이고, 회귀는 오차 제곱의 합을 최소화 하는 직선을 긋는 작업이다. ○

퀴즈...

7. 머신러닝에서의 학습 방법 중 강화 학습은 포함되지 않는다. **X**



## 퀴즈...

8. 머신러닝은 건강과 관련된 헬스케어 분야에서도 중요한 역할을 담당한다. ○

## 퀴즈...

9. 의사결정 트리는 이진 트리를 사용하는 관측값과 목표값을 연결하는 예측 모델이다. ○

## 퀴즈...

10. 강화 학습은 보상이 주어지는 문제 해결에 효과적인 학습 방법이 아니다. **X**

# 퀴즈...

본 수업에 사용된 일부 자료 및 영상물은 강의 내용을 보충하기 위해 교육 목적으로 활용하였습니다.  
자료 및 영상물의 불법적 이용, 무단 전재·배포는 법적으로 금지되어 있으니, 학생 여러분께서는 학습 외 용도의 사용을 삼가 바랍니다.