

# Machine Learning

## Chapter 1 머신러닝 개요

- Machine Learning 개념을 이해 할 수 있다.
- Machine Learning의 종류 및 과정을 알 수 있다.

Data → Model  
(알고리즘)

데이터 특성과 패턴을 학습하여,  
미지의 데이터에 대한 결과를 예측하는 것

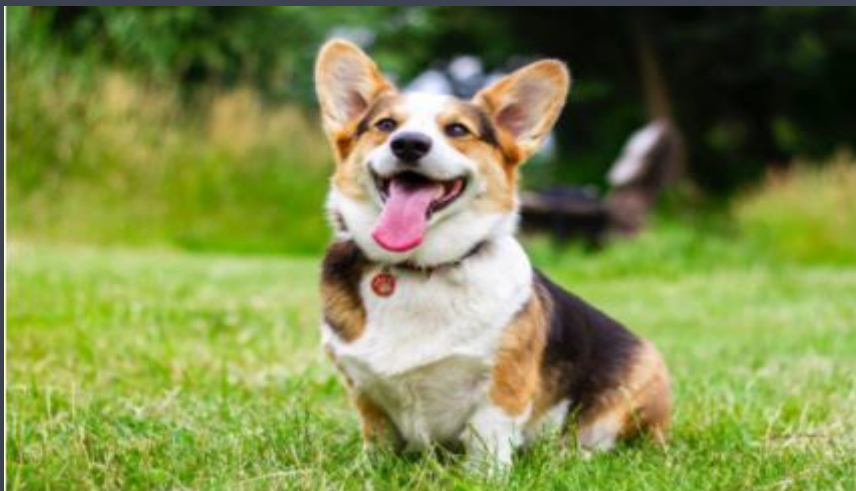
지도학습 (Supervised Learning)

비지도학습 (Unsupervised Learning)

강화학습 (Reinforcement Learning)

## 지도 학습 (Supervised Learning)

- 데이터에 대한 Label(명시적인 답)이 주어진 상태에서 컴퓨터를 학습시키는 방법.



## 분류(classification)

어떠한  $x$ 라는 데이터가 들어왔을 때  
이 데이터가 어떤 class에 속하는지  
예측

Ex)강아지나 고양이사진이 주어졌을때  
강아지나 고양이로 예측

## 회귀(Regression)

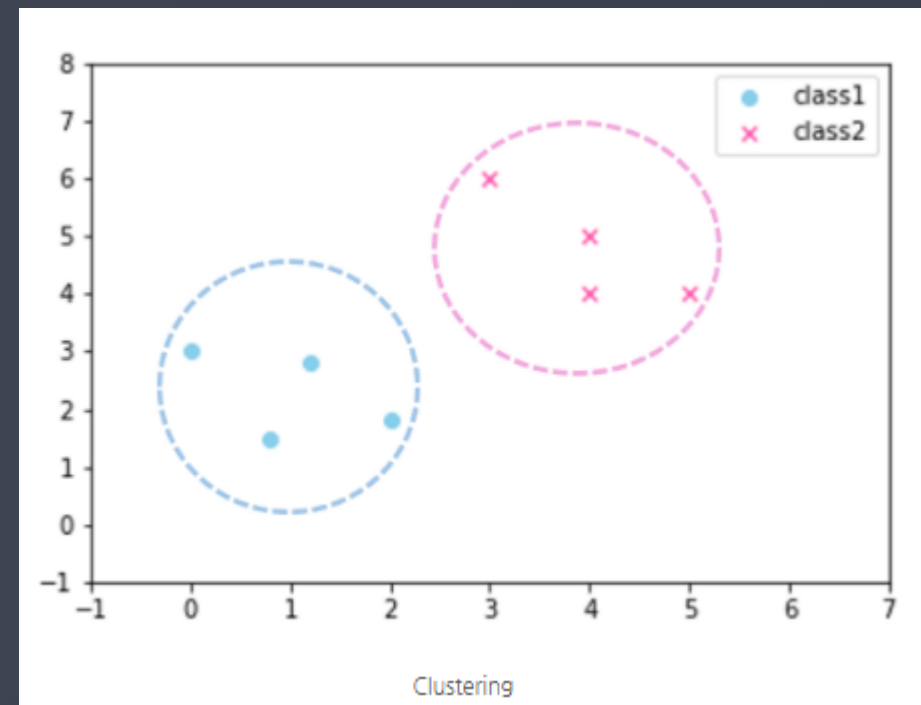
어떠한  $x$ 라는 데이터가 들어왔을 때  
"어떠한 값"을 예측

Ex)몸무게라는  $x$  데이터를 넘겨줬을 때  
키  $y$ 를 예측

## 비지도 학습 (Unsupervised Learning)

- 데이터에 대한 Label(명시적인 답)이 없는 상태에서 컴퓨터를 학습시키는 방법.
- 데이터의 숨겨진 특징, 구조, 패턴을 파악하는데 사용.
- 데이터를 비슷한 특성끼리 묶는 군집(Clustering)과 차원 축소(Dimensionality Reduction) 등이 있다.

## 비지도 학습 (Unsupervised Learning)



clustering



## 강화 학습 (Reinforcement Learning)

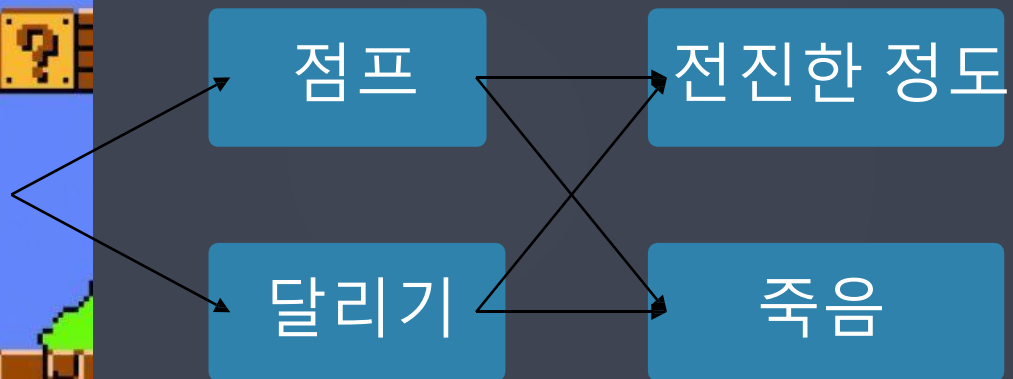
- 문제와 답을 주지 않고 목표와 보상만 제공하여 컴퓨터를 학습시키는 방법.
- 기계는 더 많은 보상을 얻을 수 있는 방향으로 행동을 학습.
- 주로 게임이나 로봇을 학습시키는데 많이 사용.

## 강화 학습 (Reinforcement Learning)

- 보상 시스템에 따라 최적의 액션 시퀀스(action sequence)를 결정하는 것



환경



액션

보상



지도학습



비지도학습



강화학습

XOR 연산을 학습 해보자

## scikit-learn

- 파이썬에서 쉽게 사용할 수 있는 머신러닝 프레임워크, 라이브러리
- 회귀, 분류, 군집, 차원 축소, 특성공학, 전처리, 교차검증, 파이프라인 등 머신러닝에 필요한 기능을 갖춘
- 학습을 위한 샘플 데이터도 제공

## XOR 연산 학습하기

P	Q	P xor Q
0	0	0
0	1	1
1	0	1
1	1	0

문제[Data]

답[Label]

## XOR 연산 학습하기

Model  
(알고리즘)

KNN 분류 모델을 사용

## XOR 연산 학습하기

```
knn=KNeighborsClassifier(n_neighbors=1)
knn.fit(문제 , 답)
knn.predict(새로운 문제)
score=metrics.accuracy_score(실제답, 예측결과)
```



# 머신러닝(Machine Learning) 과정

1. Problem Identification(문제정의)
2. Data Collection(데이터 수집)
3. Data Preprocessing(데이터 전처리)
4. EDA(탐색적 데이터분석)
5. Model Selection(모델 선택)
6. Fit(학습)
7. Evaluation(평가)

## 1. Problem Identification(문제정의)

- Classification 분류
- Regression 회귀

## 2. Data Collection(데이터 수집)

- 공공데이터  
(<https://www.data.go.kr/>)
- 웹크롤링 (뉴스, SNS, 블로그)
- Kaggle

## 3. Data Preprocessing(데이터 전처리)

- 결측치, 이상값 조정
- Encoding  
Categorical Data를 수치 데이터로 변경
- Feature Engineering (특성공학)  
단위 변환, 새로운 속성 추가

#### 4. EDA(탐색적 데이터분석)

- 데이터를 관찰 후 전처리 전략 수립  
→ 시각화(pandas, matplotlib, seaborn)
- 예측 모델에 넣을 Feature(특성) 결정

## 5. Model Selection(모델 선택)

- 목적에 맞는 적절한 모델 선택
- KNN, Decision Tree,  
Linear Model, Ridge, Lasso ...  
HyperParameter tuning  
(하이퍼파라미터 조정)

## 6. Fit(학습)

- Train 데이터와 Test 데이터를  
7:3 정도로 나눔

## 머신러닝(Machine Learning) 과정





## 7. Evaluation(평가)

- Accuracy(정확도)
- Mean squared error(평균제곱오차)