

## Chapter. 01

머신러닝의 개요

# I 머신러닝의 정의, 용어

FAST CAMPUS  
ONLINE

직장인을 위한 파이썬 데이터분석

강사. 이경록

## Chapter. 01

# 머신러닝의 정의와 용어

# I 머신러닝이란?

알파고? 딥러닝? 머신러닝?  
인공지능?

# I 인공지능, 머신러닝, 딥러닝

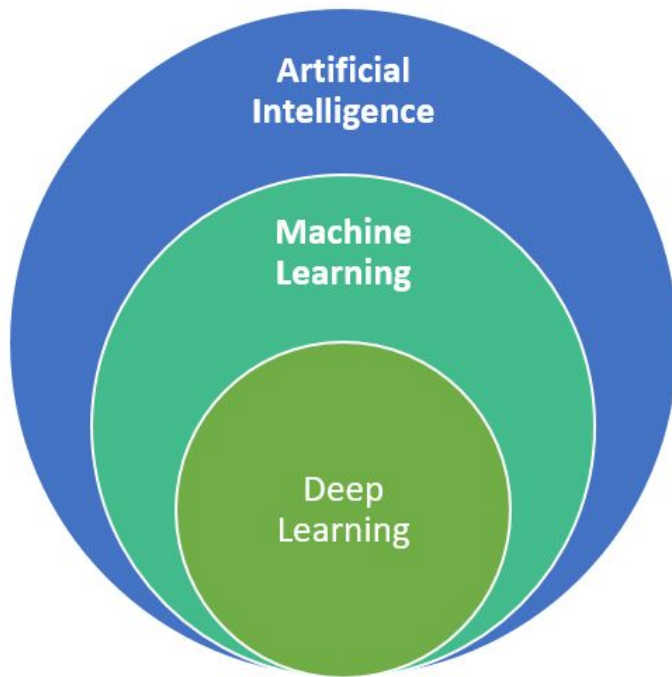


Figure 1: artificial intelligence, machine leaning and deep learning Source: Nadia BERCHANE (M2 IESCI, 2018)

**인공지능:** 사람의 지능을 모방하여, 사람이 하는 것과 같이 복잡한 일을 할 수 있게 기계를 만드는 것

**머신러닝:** 기본적으로 알고리즘을 이용해 데이터를 분석 및 학습하며, 학습한 내용을 기반으로 판단이나 예측

**딥러닝:** 인공신경망에서 발전한 형태의 인공 지능. 머신러닝 중 하나의 방법론

# I 머신러닝이란?

데이터를 기반으로 패턴을 학습하여 결과를  
추론하는 것

Data

Model

Prediction

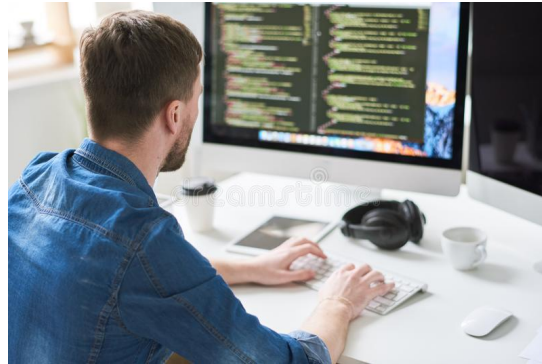
# I 과거에는...

데이터를 기반으로 패턴을 학습하여 결과를  
추론하는 것

Data

Model

Prediction



# I 머신러닝

데이터를 기반으로 패턴을 학습하여 결과를  
추론하는 것

**Data**

**Model**

**Prediction**



# I 지도학습, 비지도학습

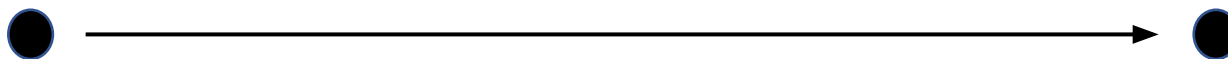
데이터를 기반으로 패턴을 학습하여 결과를  
추론하는 것

Data

Model

Prediction

지도학습  
(Supervised Learning)



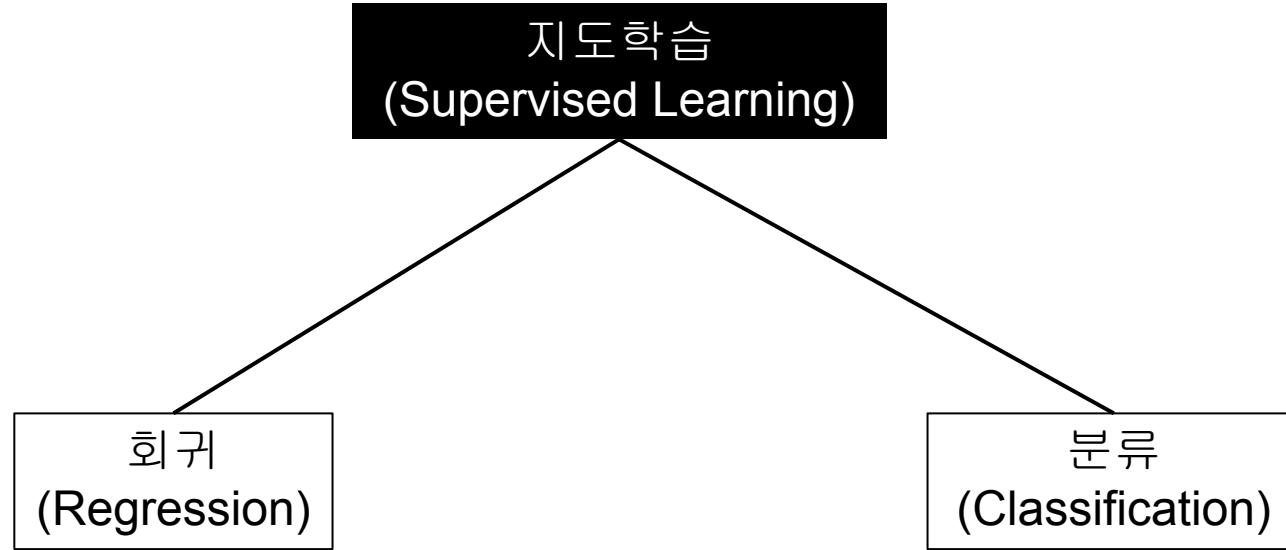
비지도학습  
(Unsupervised Learning)



강화학습  
(Reinforcement Learning)



# I 지도학습



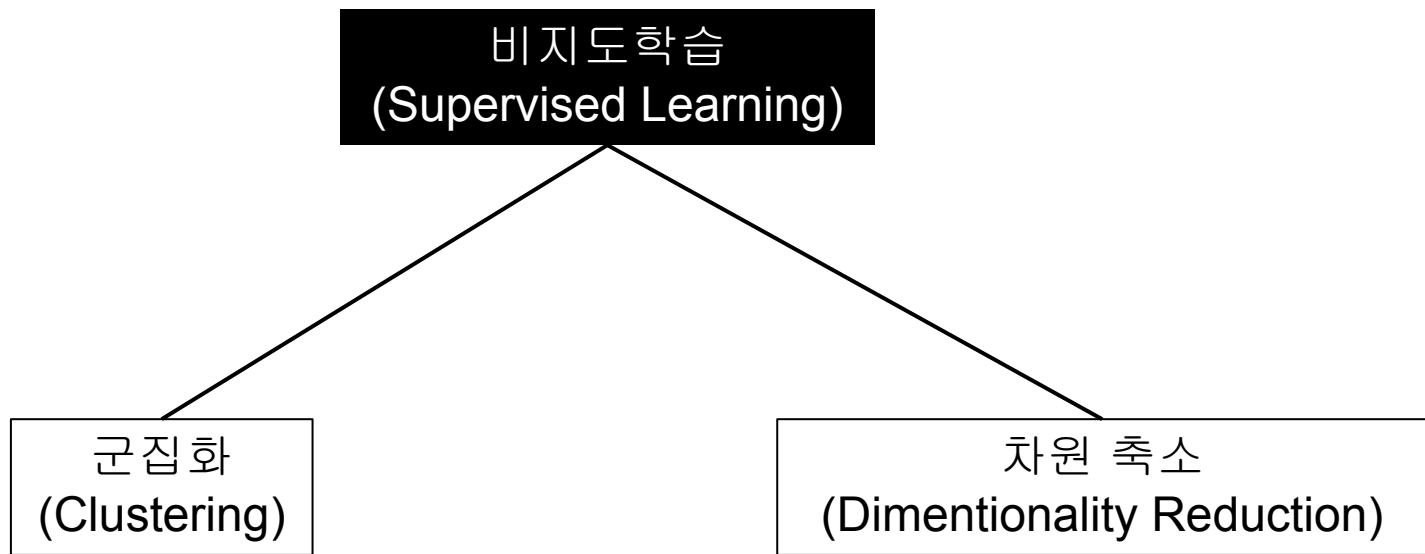
## 수치형 (numeric value)

- 집 값: 3억, 5억 3천 ...
- 가격: 1,000원, 3,500원...
- 온도: 12도, 20도...

## 분류형 (categorical value)

- 스팸메일: 스팸/정상
- 종류 판별: 개/고양이
- 암 진단: 암/정상

# I 비지도학습



## 그룹핑

- 뉴스 분류: 연예, 정치, 교육...
- 사용자 관심사: 쇼핑, 여행, 영화...

# I 머신러닝 장점

1. 복잡한 패턴을 인지할 수 있다
2. 적절한 알고리즘, 다양한 양질의 데이터가 있다면, 좋은 성능
3. 도메인 영역에 대한 지식이 상대적으로 부족해도 가능하다

# I 머신러닝 문제점

1. 데이터의 의존성이 크다 (Garbage in, Garbage out)
2. 과적합의 오류에 빠질 수 있다 (일반화 오류, 데이터 다양성 요구)
3. 풍부한 데이터가 기본적으로 요구된다

# I 머신러닝, 데이터

좋은 성능 = 좋은 데이터 (quality,  
quantity)

데이터 가공 = 전처리  
(pre-processing)

# I 머신러닝, 알고리즘

데이터, 예측해야 할 값에 맞는 **알고리즘**  
사용