

# 1. 머신 러닝

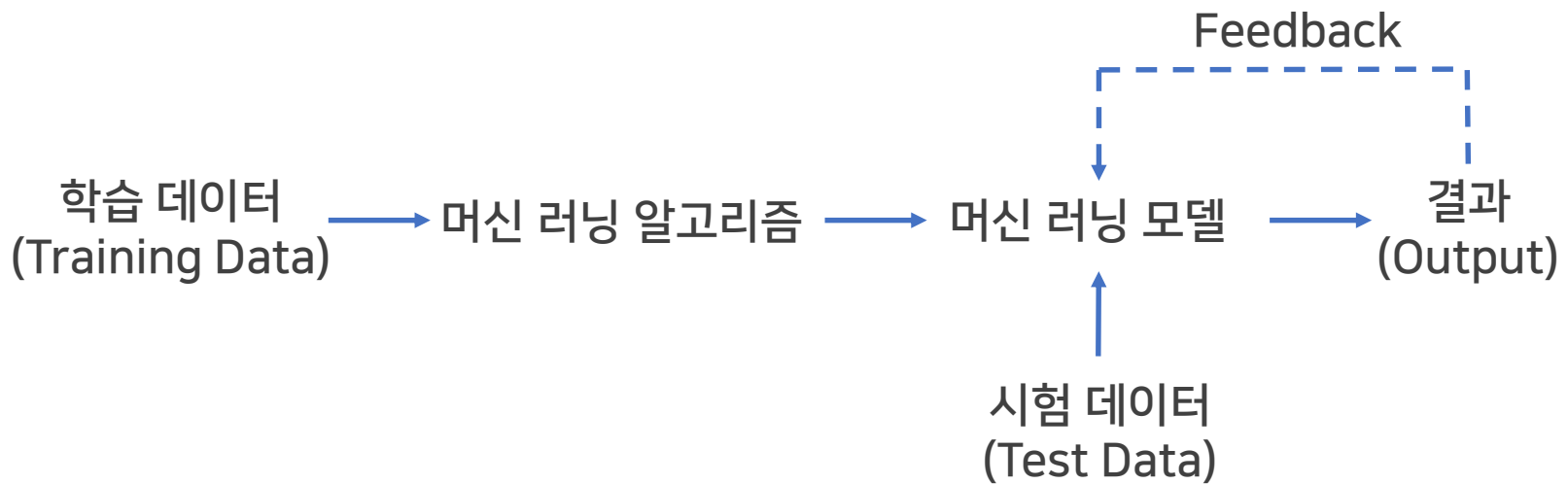
## 머신 러닝의 정의

“컴퓨터에 명시적인 프로그램 없이 배울 수 있는 능력을 부여하는 연구 분야”

- 1959, 아서사무엘(Arthur Samuel) -

“특정한 작업(Task)을 달성하기 위해 경험(Experience)이 축적될수록 과업 수행의 성능(Performance)이 향상되는 컴퓨터 프로그램 또는 에이전트를 개발하는 것”

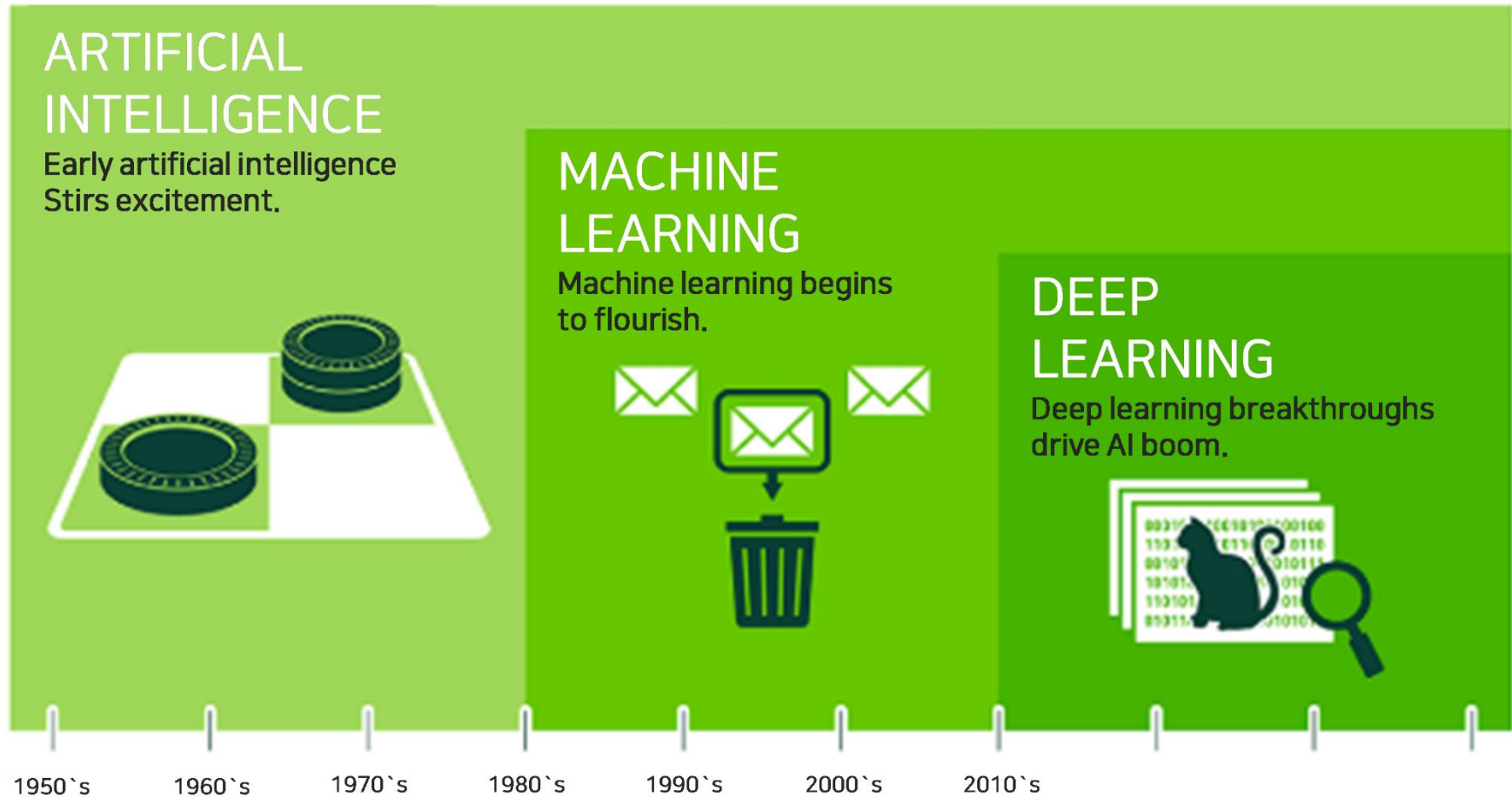
- 1997, 톰 미첼(Tom M. Mitchell) -



# x 머신 러닝과 인공지능



🔍 머신 러닝은 인공지능의 한 분야

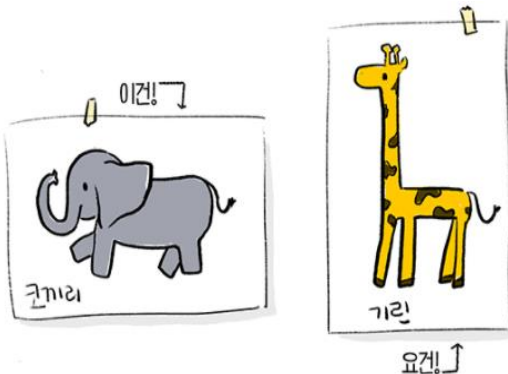


 인공지능을 공부시키는 방법을 머신 러닝이라고 한다.

## 머신 러닝

### 지도 학습 (Supervised Learning)

문제와 정답을 모두 알려주고  
공부시키는 방법



예측, 분류

### 비지도 학습 (Unsupervised Learning)

답을 가르쳐주지 않고  
공부시키는 방법

비지도학습은 답을 가르쳐주지 않고 공부를 시키는거야.



연관 규칙, 군집

### 강화 학습 (Reinforcement Learning)

보상을 통해  
상은 최대화, 벌은 최소화하는  
방향으로 행위를 강화하는 학습

강화학습은 일종의 게임 같이 보상해주는거야



보상

## (참고) 인공지능(AI; Artificial Intelligence)

- 철학적 관점에서는 종종 인공지능을 강인공지능(strong AI)과 약인공지능(weak AI)으로 구분하곤 한다.
- 강인공지능은 쉽게 말하면 인간과 같이(또는 다르게) 지각력이 있고 스스로를 인식하는 것이다.
- 약인공지능은 어떤 문제를 실제로 사고하거나 해결할 수는 없지만 컴퓨터 기반의 인공적인 지능을 만들어 내는 것에 관한 연구다.

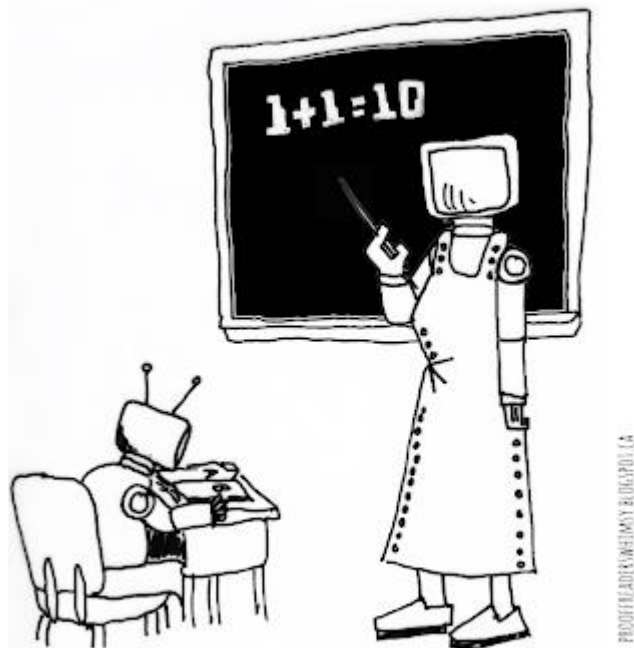
## (참고) 인공지능(AI; Artificial Intelligence)

“특정한 작업(Task)을 달성하기 위해 경험(Experience)이 축적될수록 과업 수행의 성능(Performance)이 향상되는 컴퓨터 프로그램 또는 에이전트를 개발하는 것”  
- 1997, 톰 미첼(Tom M. Mitchell) -

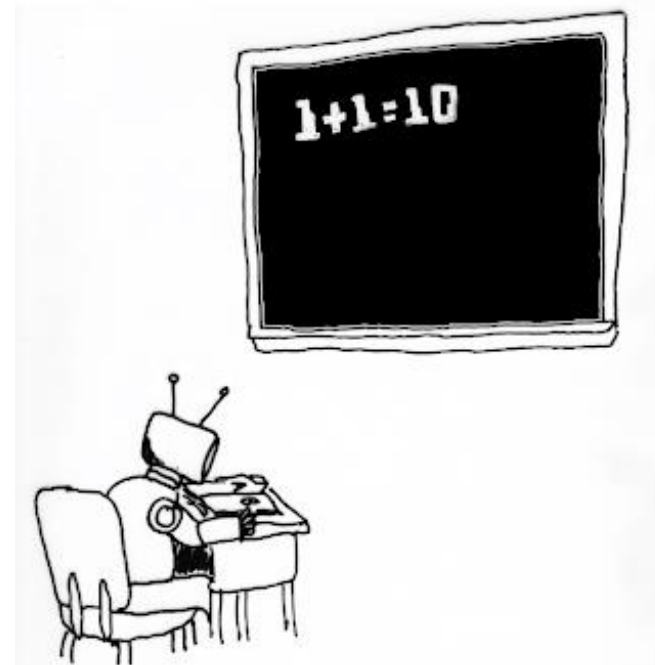
- 톰 미첼의 정의를 다시 한 번 보면 알 수 있듯이, 여기서 다룰 인공지능이란 SF 영화에서 자주 등장하는 ‘강인공지능’이 아닌 ‘약인공지능’이다.
- 결국 주어진 데이터에 맞는 분석 방법, 알고리즘 등은 데이터를 보고 ‘사람’이 설계해야 한다. 기계는 그 이후에 데이터를 보고 ‘학습’하는 것이다.

🔍 지도 학습(supervised learning) vs 비지도 학습(unsupervised learning)

SUPERVISED MACHINE LEARNING



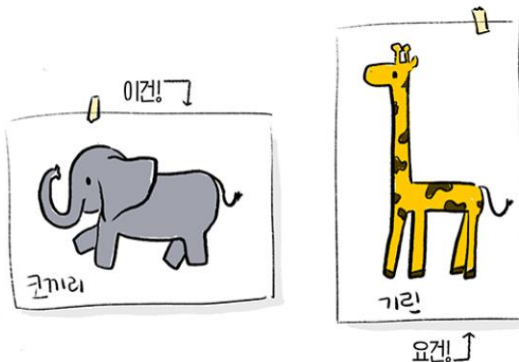
UNSUPERVISED MACHINE LEARNING



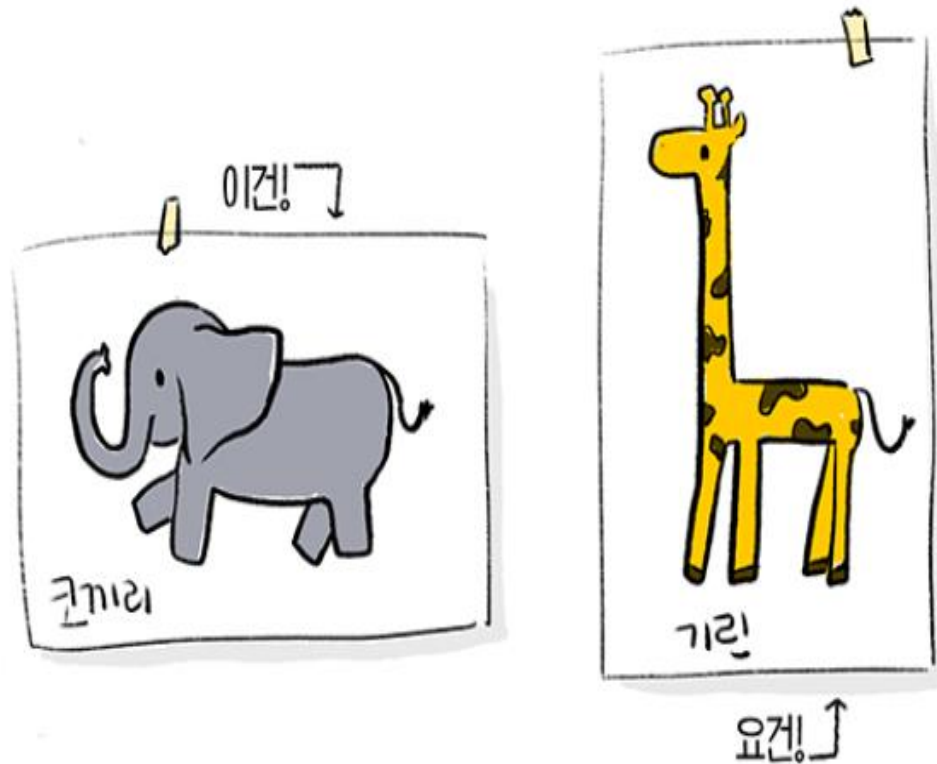
## 🔍 지도 학습이란 무엇인가?

### 지도 학습


라벨된 학습 데이터셋으로  
학습을 수행하는 알고리즘



? 답을 가르쳐주고 공부를 시킨다 ?





 지도 학습: 반응변수를 예측해내는 것이 목적이다.

- 회귀 예측분석(regression): 수량형 반응변수를 예측한다.
- 분류분석(classification): 범주형 반응변수를 예측한다.

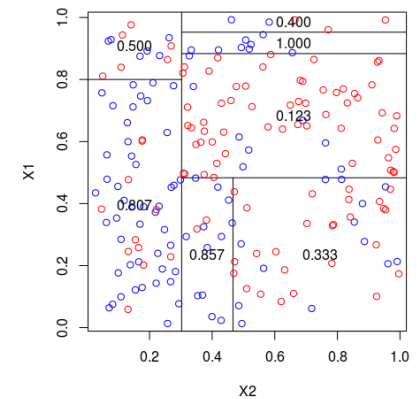
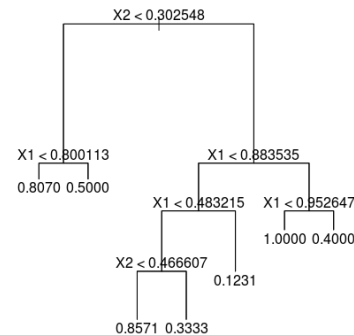
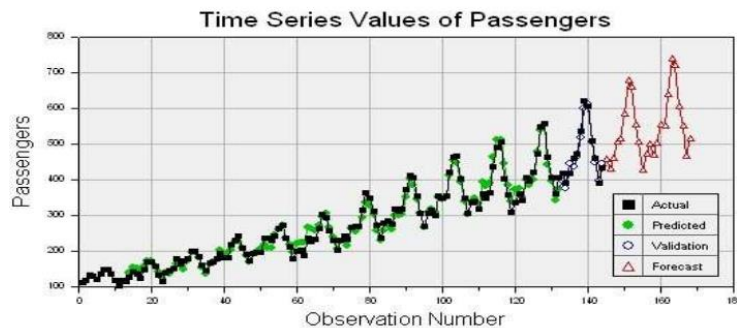
## 🔍 지도 학습에는 어떤 기법들이 있는가? (1/2)

### 회귀 : Regression

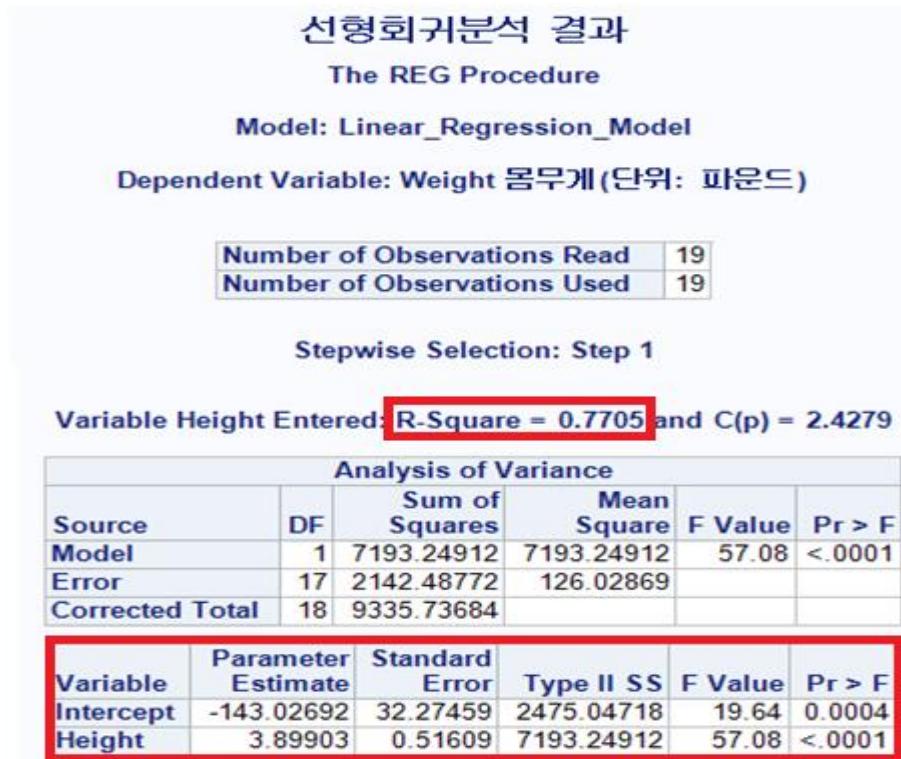
#### ▪ 목적 : 수치형 반응 변수를 예측

- ✓ 얼마나 많은가?와 같은 양적인 문제 해결에 도움되는 알고리즘으로  
분류 문제에 비해 모형의 복잡도가 높음  
예) 구매 금액, 매출액, 이용 고객 수 예측 등
- ✓ 선형 회귀, 회귀나무 등

### TIME SERIES ANALYSIS



## 지도 학습 : 선형 회귀, 시계열 분석 사례



< 선형 회귀 분석 예시 >  
Y=몸무게, X1=성별, X2=나이, X3=키로 하는 다  
중선형회귀분석



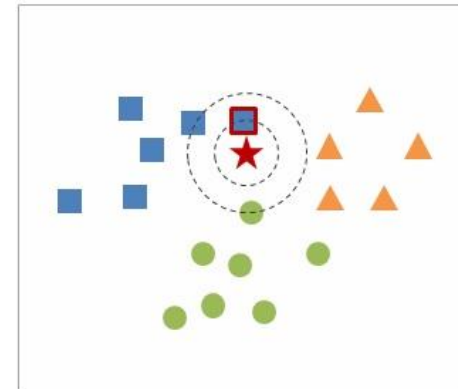
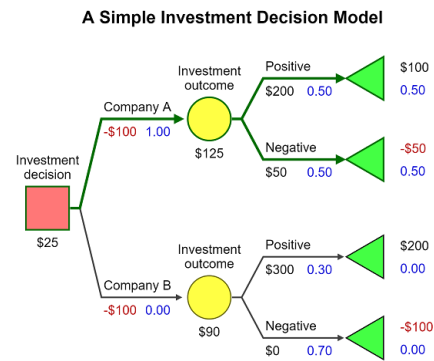
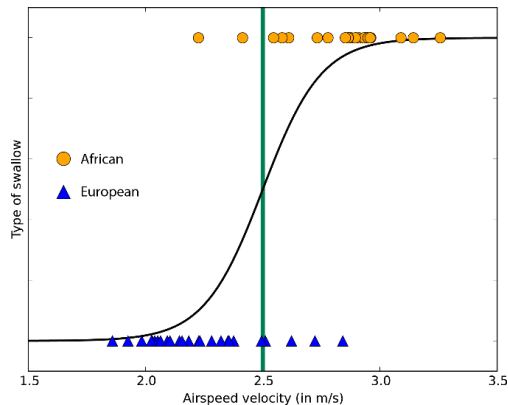
< 시계열 분석 예시 >  
2003~2015년 교통사고 건수의  
추세 분석

## 지도 학습에는 어떤 기법들이 있는가? (2/2)

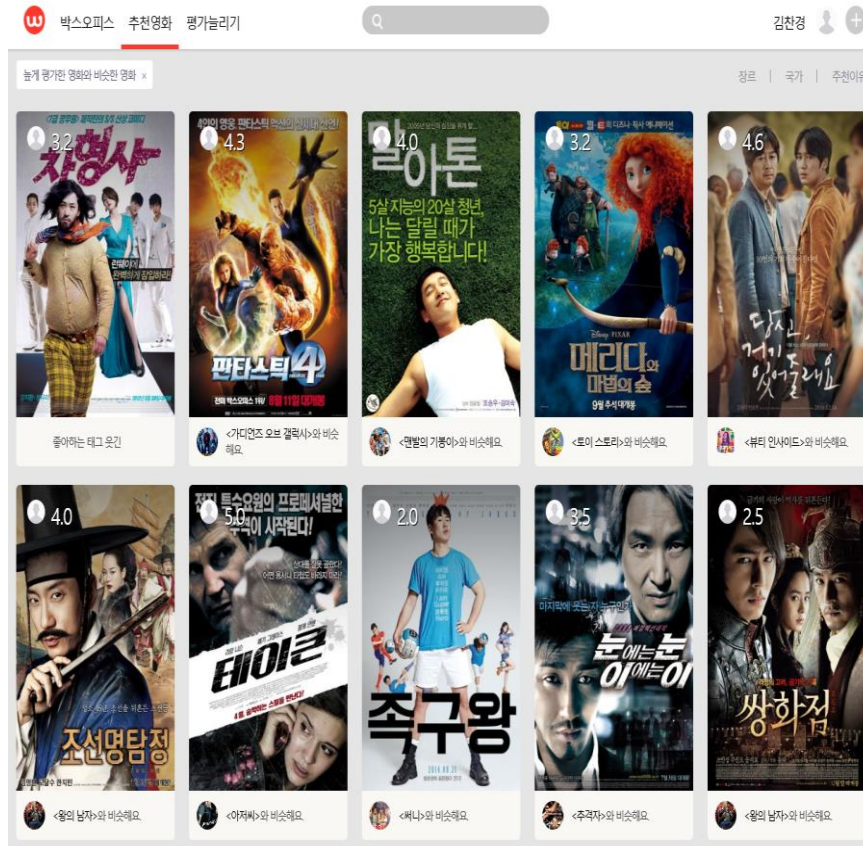
### 분류 : Classification

#### ■ 목적 : 범주형 반응 변수를 예측

- ✓ 새로운 데이터와 형성된 모형으로부터 보지못한 데이터 분류
- ✓ 이진 분류(binary classification) 문제가 대다수
  - 예) 구매/비구매, 사기거래/정상거래, 양성반응/음성반응 등
- ✓ 로지스틱 회귀, 의사결정나무, k-NN 인접 이웃 등



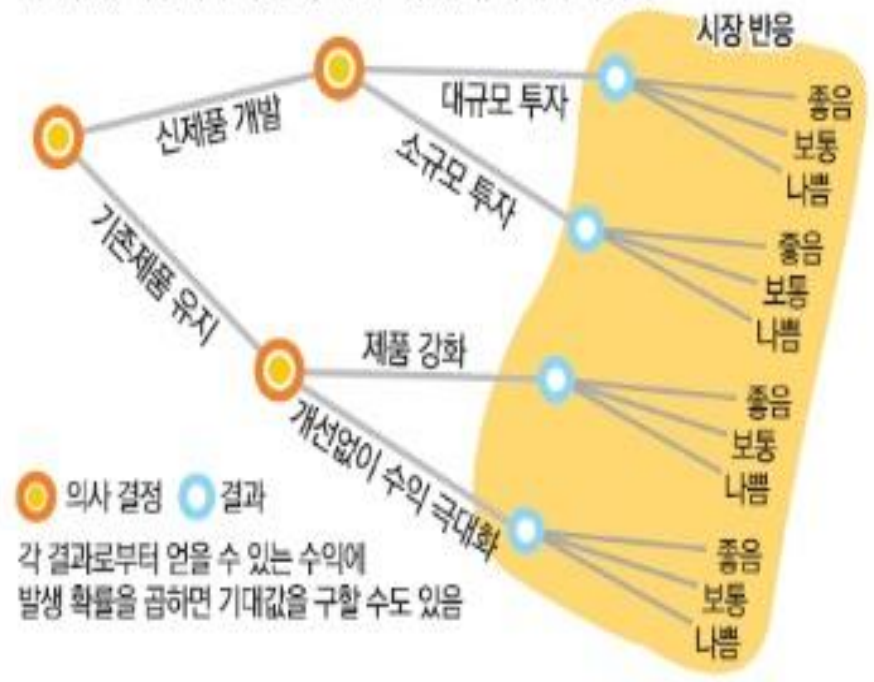
## 지도 학습 : k-NN 인접이웃, 의사결정나무 사례



< k-NN 인접이웃 예시 >  
영화추천시스템

[그림3] 의사결정 나무의 사례

신제품을 개발해야 하는가, 기존 제품을 유지해야 하는가



< 의사결정나무 예시 >  
신제품 개발에 대한 의사결정


## 🔍 비지도 학습이란 무엇인가?

### 비지도 학습

라벨된 학습데이터셋 없이  
데이터의 패턴과 관계를  
발견하는 알고리즘

비지도학습은 답을 가르쳐주지 않고 공부를 시키는거야.



 비지도 학습: 변수들 간의 혹은 관측치 간의 관계를 밝혀내는 것이 목적이다.

- 군집분석(clustering): 관측치들을 변수들 간의 유사성으로 그룹핑한다.
- 차원 감소(dimensionality reduction): 변수들을 관측치들 간의 유사성을 이용하여 적은 수로 줄여준다.
- 피처 가공(feature engineering 혹은 피처 추출(feature extraction): 주어진 변수로부터 지도학습의 입력변수로 사용할 수 있는 특징값을 변환하거나 생성한다.



## 🔍 비지도 학습에는 어떤 기법들이 있는가? (1/2)

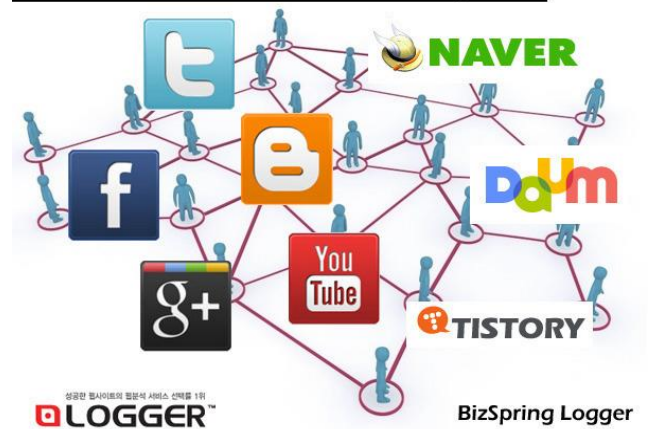
### 연관 규칙 : Association Rule Mining

- 목적 : 연관성이 높은 아이템들로 구성된 규칙 집합을 생성
  - ✓ 추천 시스템에 주로 사용되며 “장바구니 분석” 이라고도 불림
  - 예) 아이폰을 구매하는 고객들은 범퍼를 함께 구매한다
  - ✓ 연관성 분석, 소셜 네트워크 분석 등

고객의 장바구니에는  
맥주와 기저귀가 함께 있다!



### Social Network Analysis

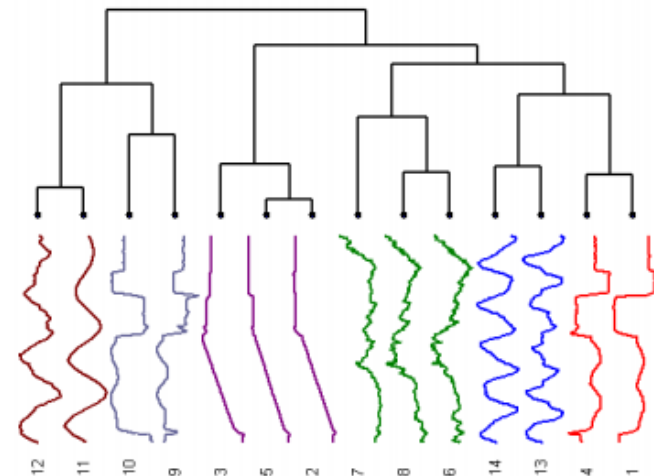
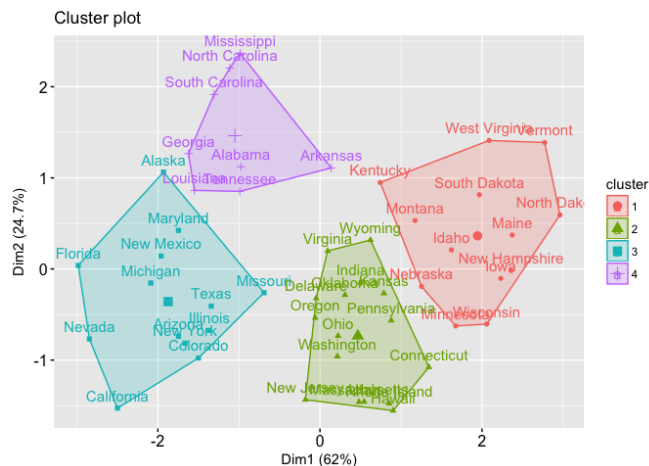




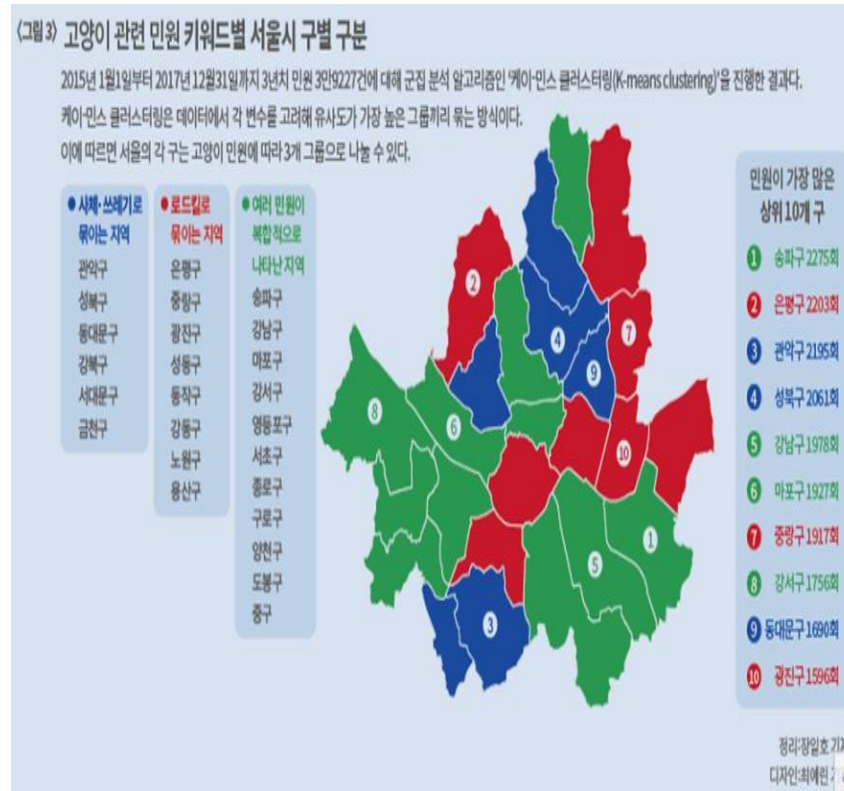
## 비지도 학습에는 어떤 기법들이 있는가? (2/2)

### 군집 : Clustering

- 목적 : 전체 데이터를 보다 일관성과 응집성이 높은 세부 그룹으로 나눔
  - ✓ Market/Customer segmentation 등에 주로 사용되는 방법으로  
동일 그룹 내의 개체들이 유사할수록, 다른 그룹 내의 개체들과는 상이할수록  
좋은 군집화 결과임
  - ✓ k-Means Clustering, Hierarchical Clustering 등



## 비지도 학습 : 군집 사례



< k-means Clustering 예시 >  
서울시 고양이 관련 민원 군집화

## 🔍 강화 학습이란 무엇인가?

### 강화 학습

보상을 통해  
상은 최대화, 벌은 최소화하는  
방향으로 행위를 강화하는 학습

강화학습은 일종의 게임 같이 보상해주는거야



## ? 보상을 통해 공부를 시킨다 ?

강화학습은 일종의 게임 같이 보상해주는거야



기계한테 여러 장의 코끼리와 기린 사진들을 보여주고  
코끼리 사진을 고를 때마다 1점씩 주고 기린 사진을 고를 때마다  
1점씩 뺏아서 기계가 높은 점수를 얻기 위해 학습함

## + 강화 학습 사례

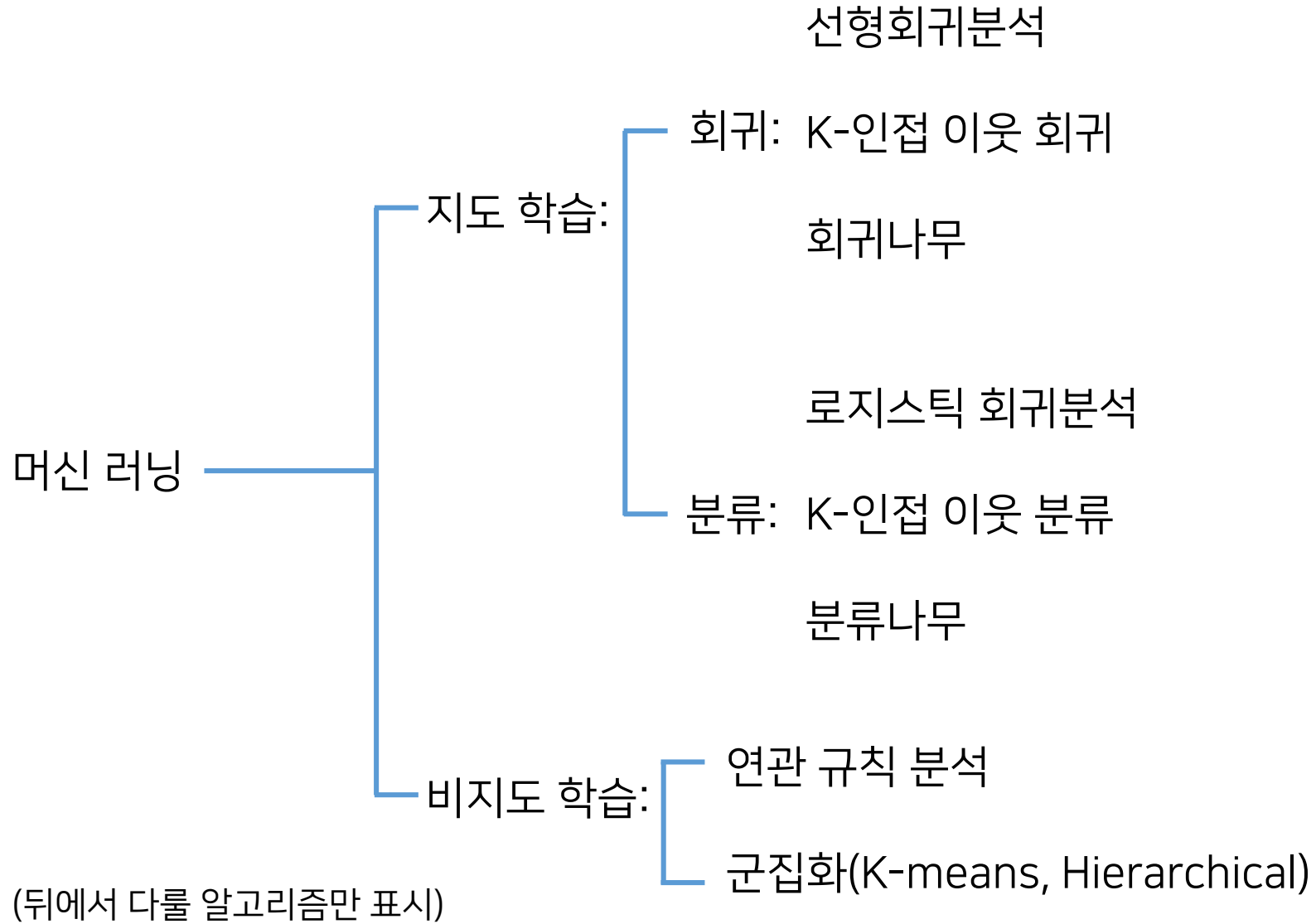
### ✓ 사례1. 쿠키런



## 강화 학습 사례

### 사례2. 벽돌깨기







## scikit-learn algorithm cheat-sheet

