

기계학습: 대용량/패널자료와 학습분석학 자료 분석으로의 활용

유진은(한국 교원대학교)

송유경

2019-2학기 iLED 스터디
(191113)

순서



1.

서론



2.

기계학습 개관



3.

학습분석학 연구 실태



4.

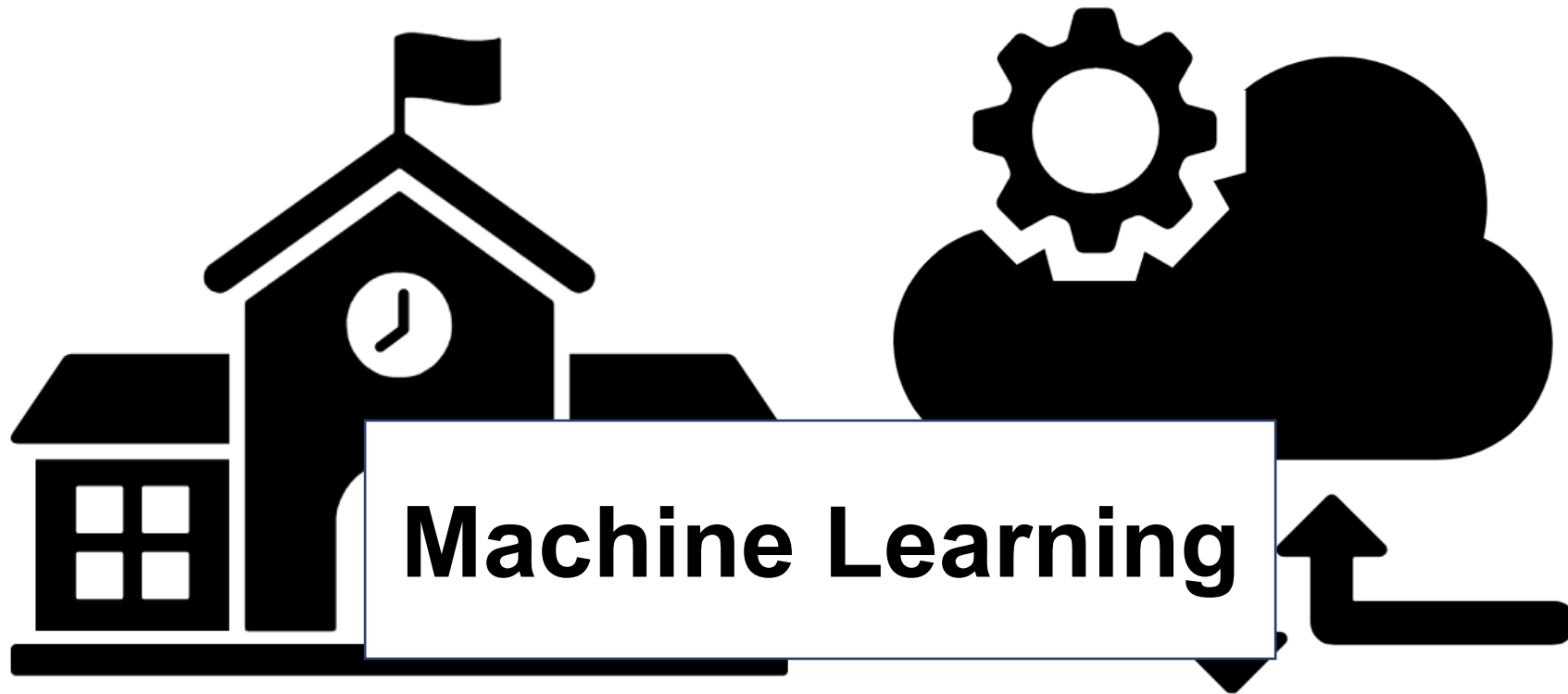
논의

01

서론



MINERVA



연구의 목적



기계학습이란?



기계학습의 교육적 활용 연구

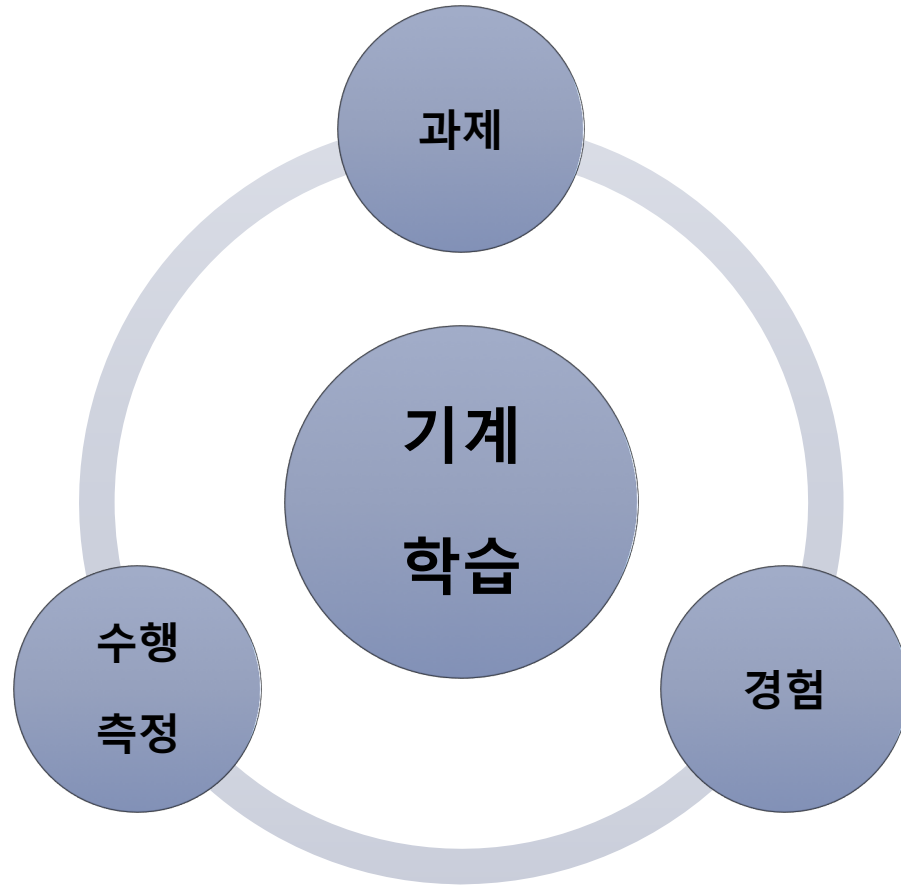


학습분석학 자료 분석과 기계학습

02

기계학습이란?

기계 학습이란?



- Field of study that gives computers the ability to learn without being explicitly programmed
Arthur Samuel(1959)

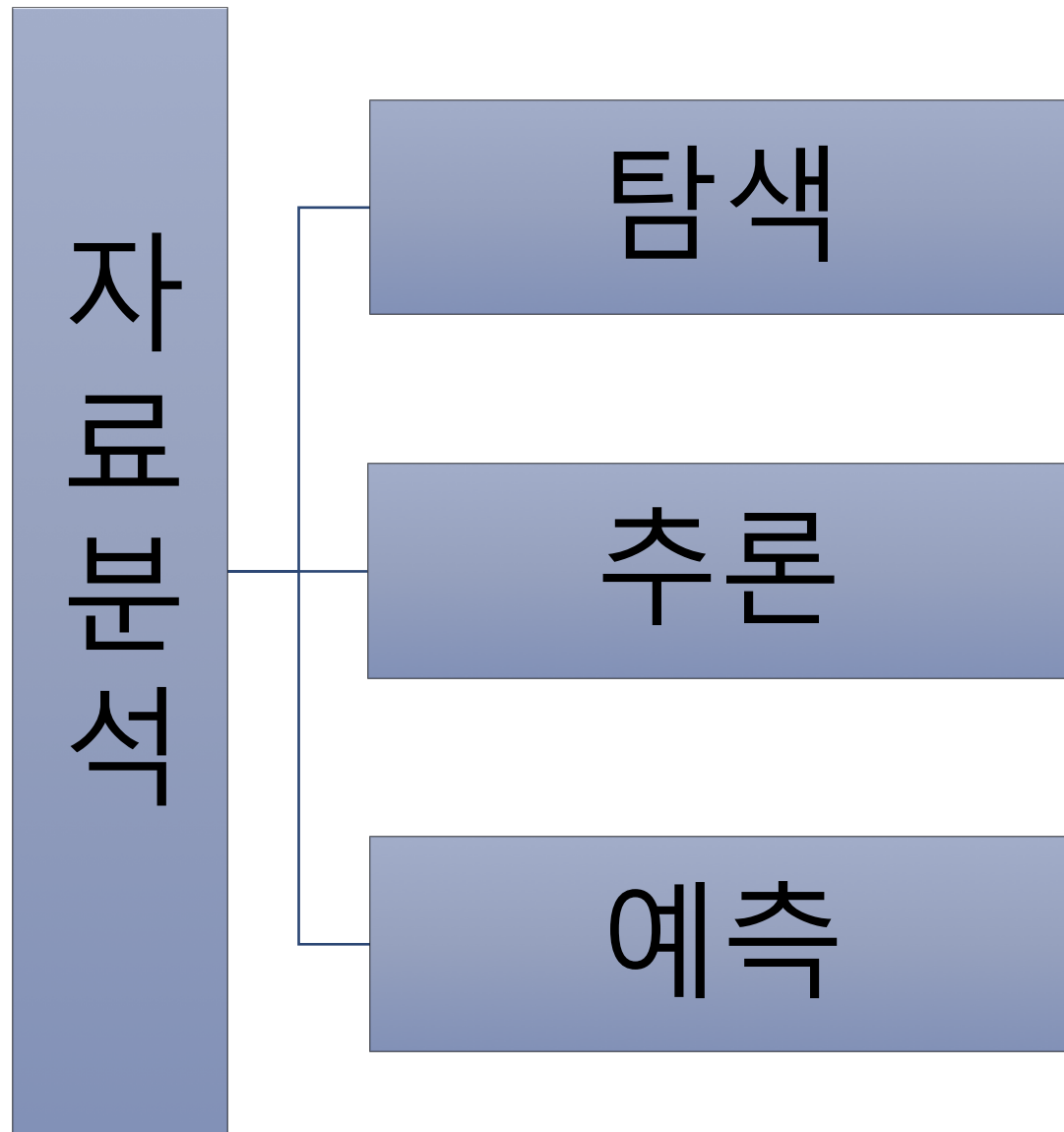
- 어떤 **과제**를 수행하고 그 **수행의 측정**을 통해 **경험**이 쌓이면서 점점 수행이 향상되는 기계
=> 학습이 일어남

Mitchell(1997)

- 기계(또는 컴퓨터)로 하여금 데이터를 이용하여 학습하도록 하는 것이며 그 결과로 최적의 해를 찾아내도록 하는 방법

유진은(2019)

기계 학습의 주요 개념



추론, 예측을 위한 예비 단계

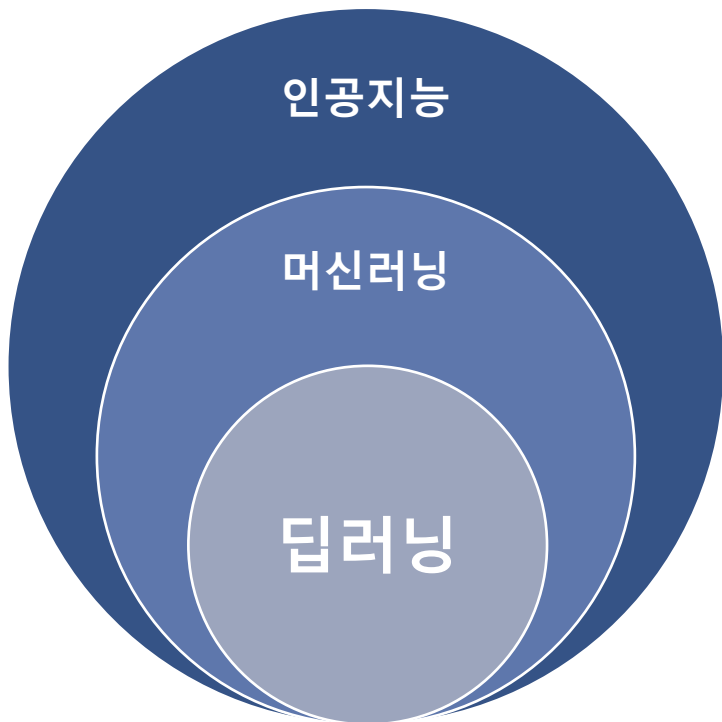
연구자가 설정한 가설을 검증

예) 출석률이 나쁜 학생은 성적이 나쁠 것이다

가설 검증 x, 주요 변수 뽑아서 **예측**

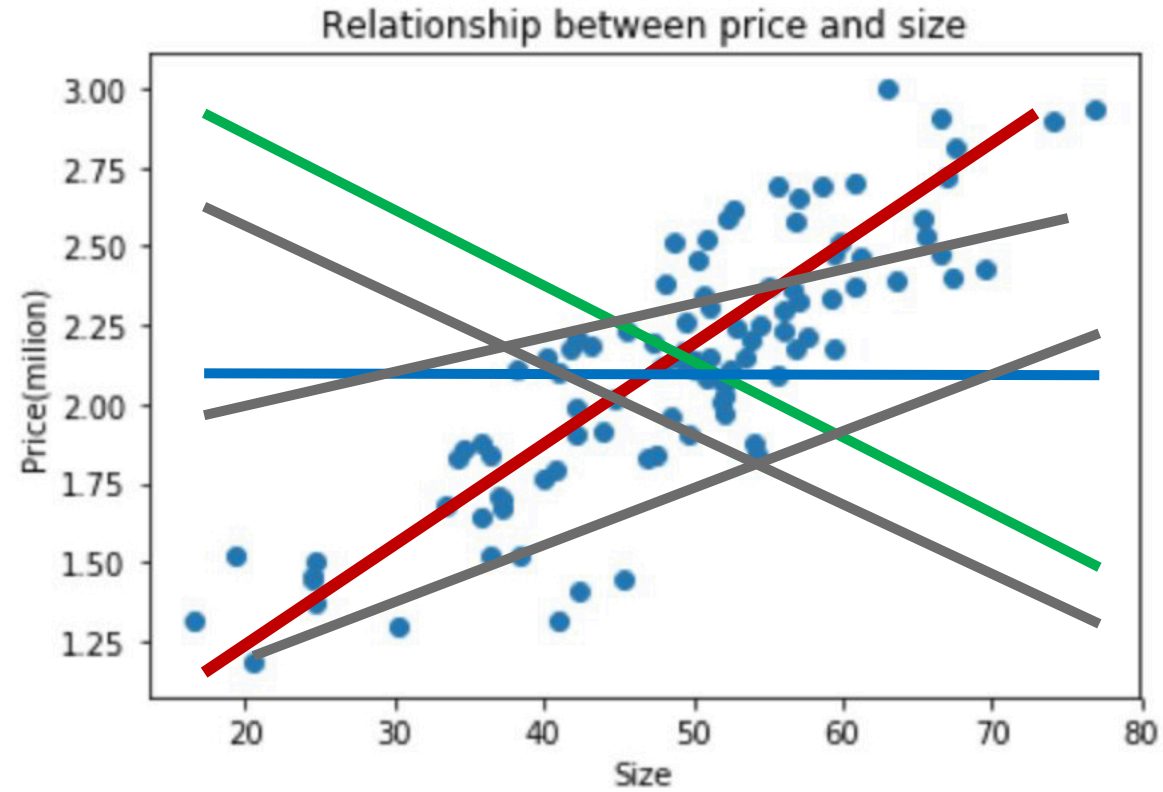
특정 학생이 학습부진인지 아닌지 예측

기계 학습과 인공지능

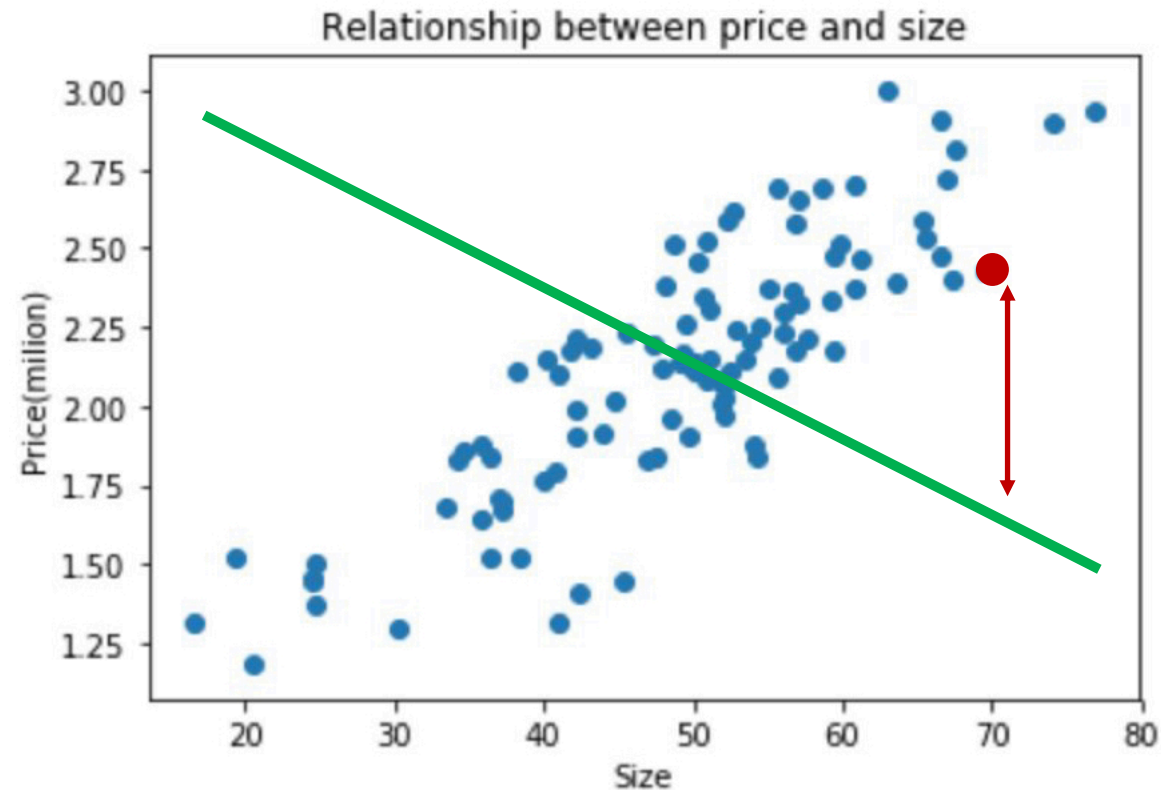


- 인공지능: 머신러닝, 자연어처리, 로봇 공학 등을 포괄하는 개념
- 머신러닝: 빅데이터를 기반으로 스스로 성능을 개선할 수 있도록 하는 알고리즘
- 딥러닝 : 인공 신경망을 통해 학습하는 것

그렇다면, 기계는 어떻게 학습할까?



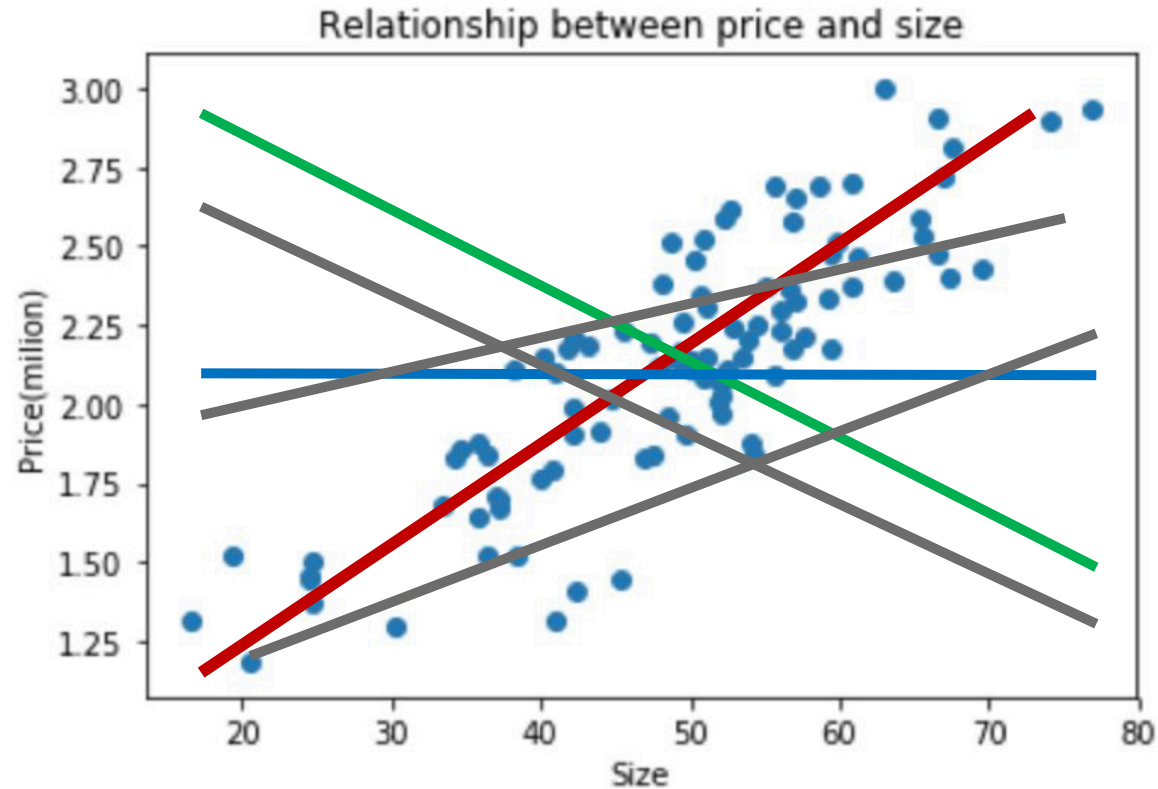
그렇다면, 기계는 어떻게 학습할까?



예측값 - 실제값 = 오차

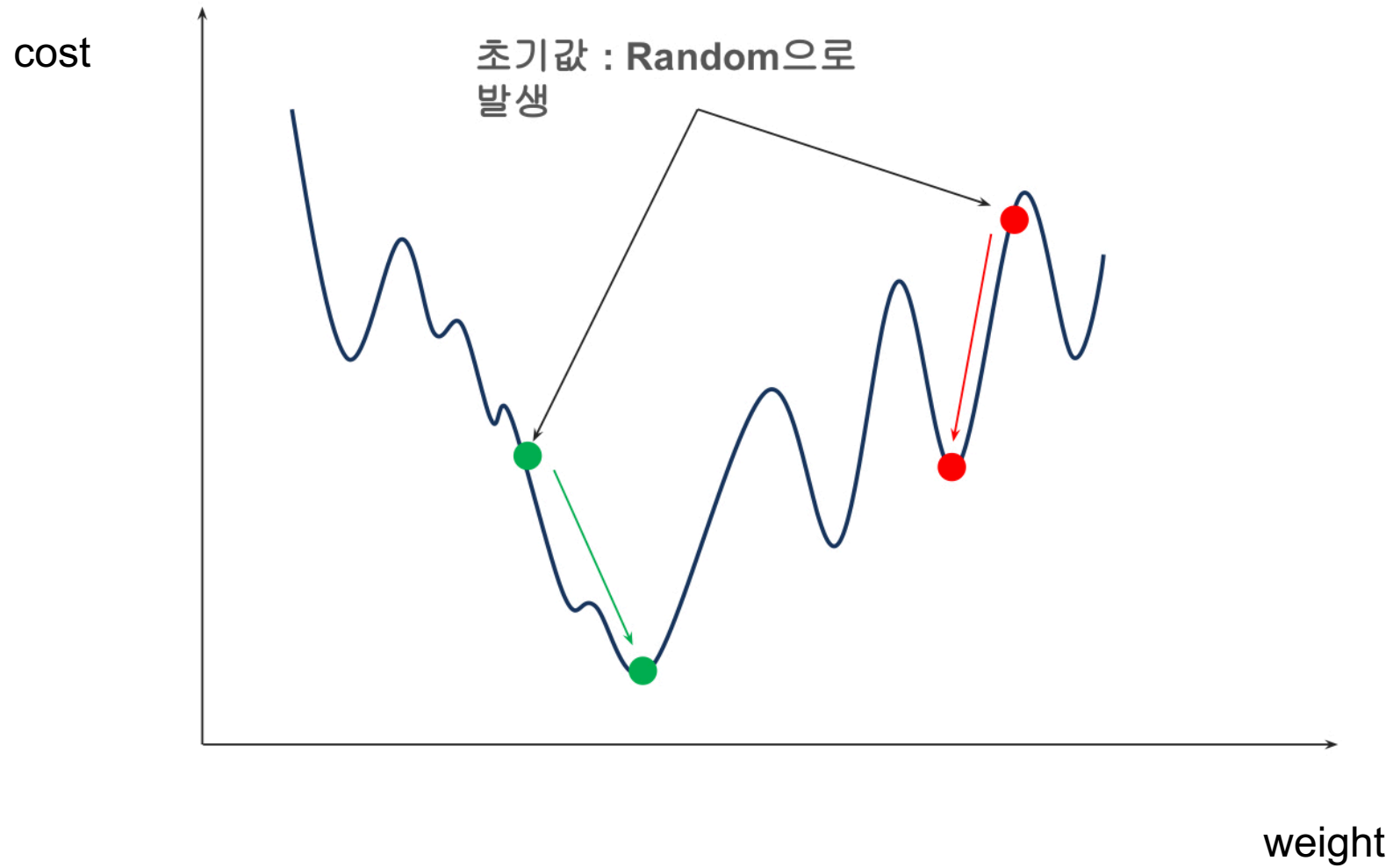
오차들의 평균 = cost

그렇다면, 기계는 어떻게 학습할까?



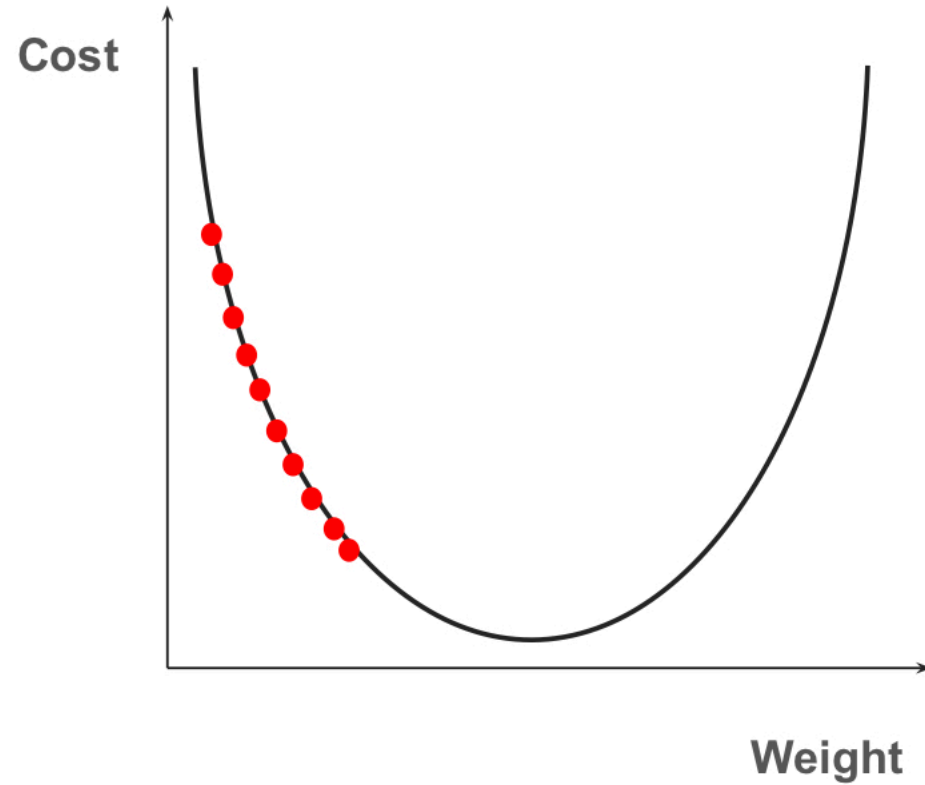
각 값의 모든 Cost 를 비교하여
Cost가 (최소)가 되는 직선을 찾기!

그렇다면, 기계는 어떻게 학습할까?

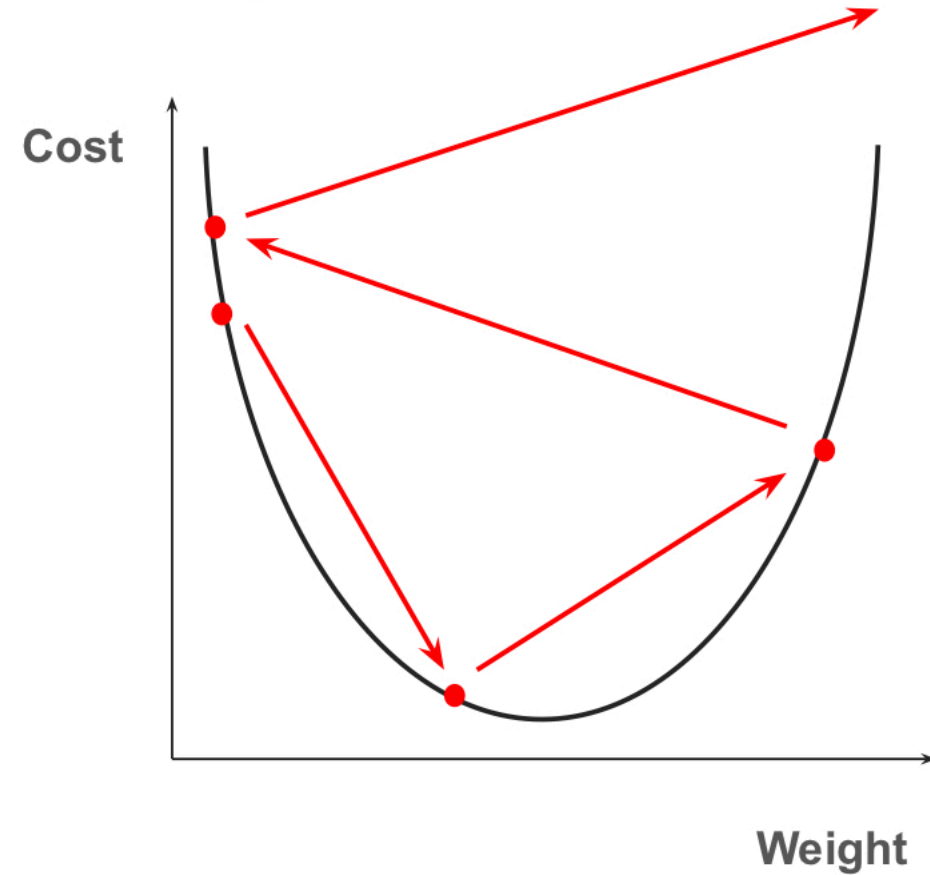


그렇다면, 기계는 어떻게 학습할까?

- Too Small

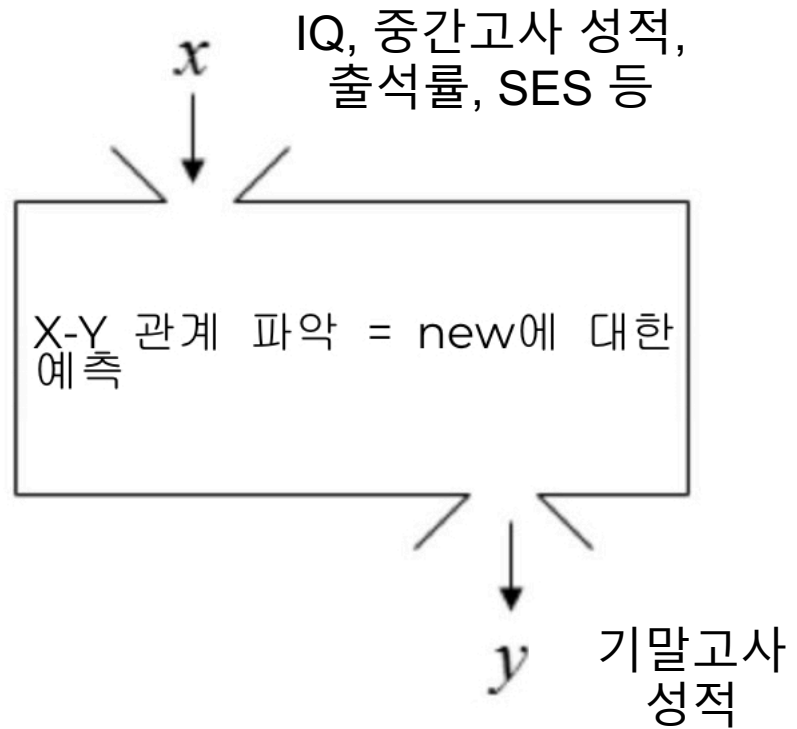


- Too Large

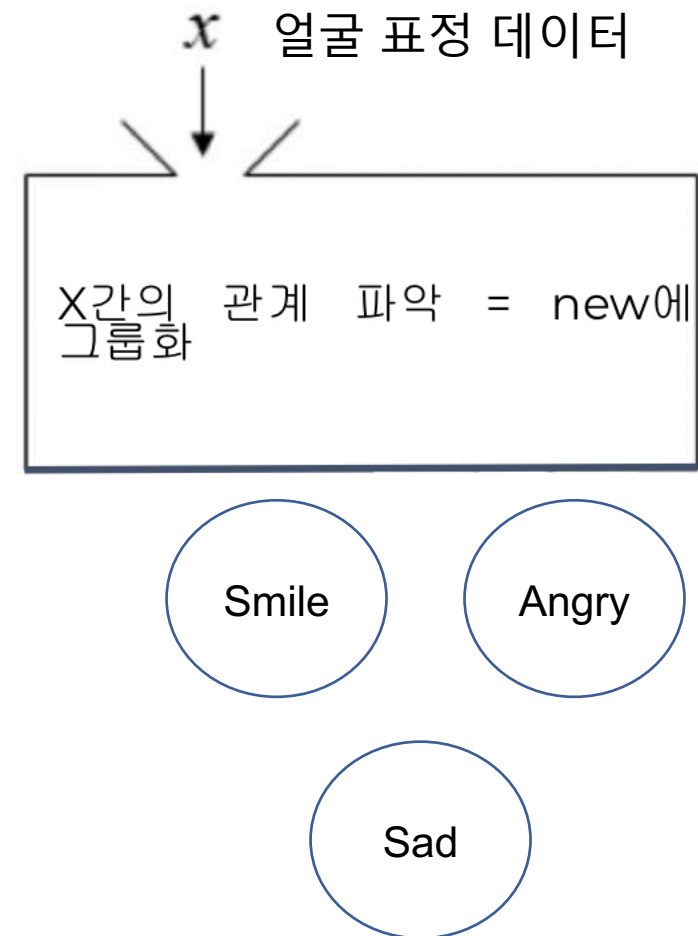


지도 학습과 비지도 학습

지도 학습



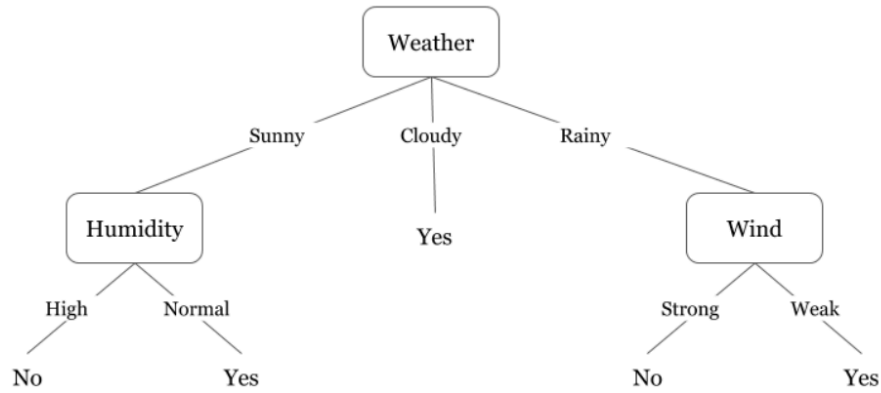
비지도 학습



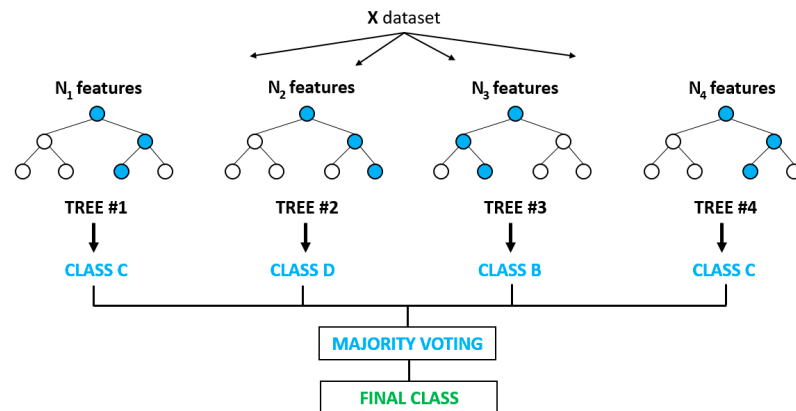
지도 학습과 비지도 학습

지도 학습

- 의사결정나무

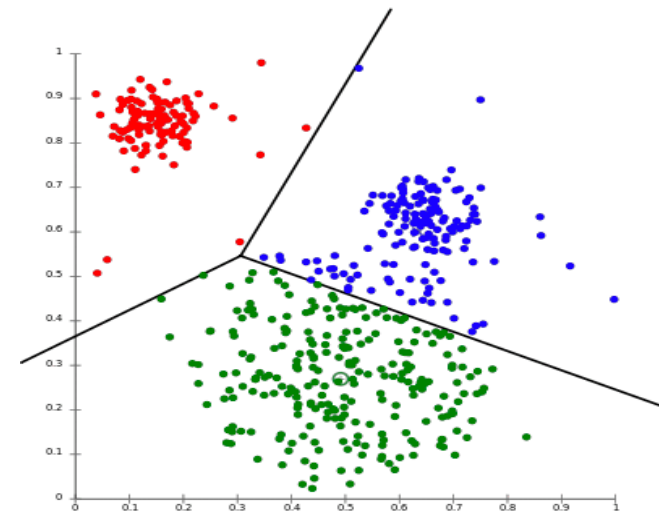


- 랜덤 포레스트



비지도 학습

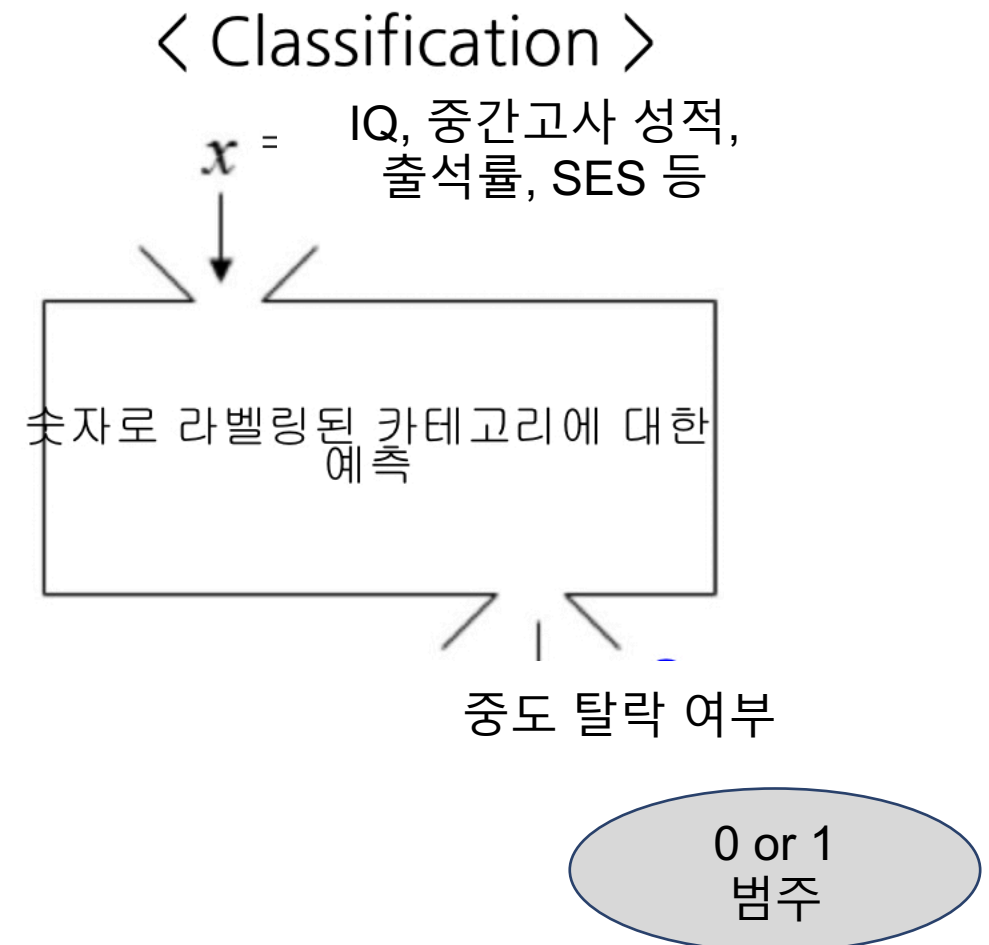
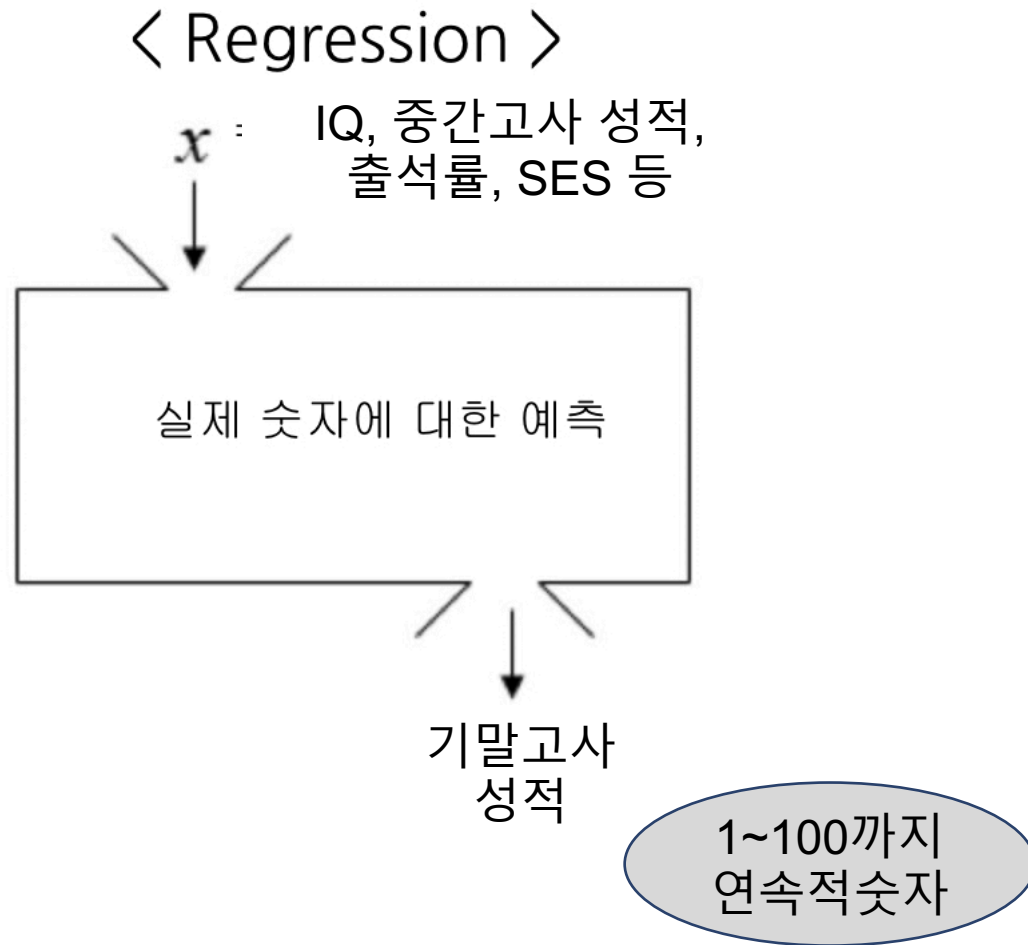
- 군집 분석, 연관성 분석, 텍스트 마이닝



- 연구자의 주관 개입 가능성
but 일일이 변수값을 매기는 데 드는 자원 절약

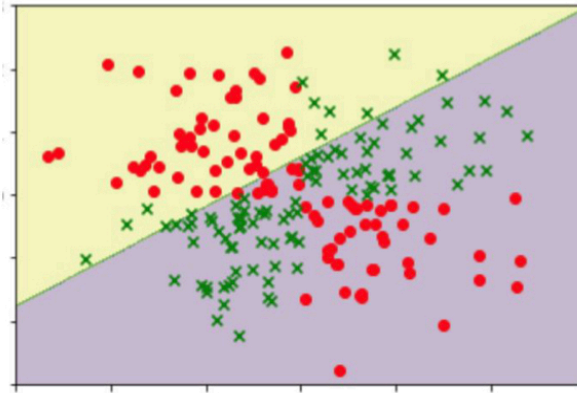
Regression vs Classification

지도 학습



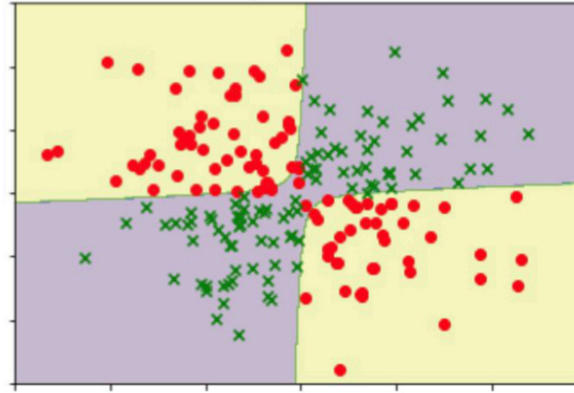
과소적합, 적합, 과적합

Classification



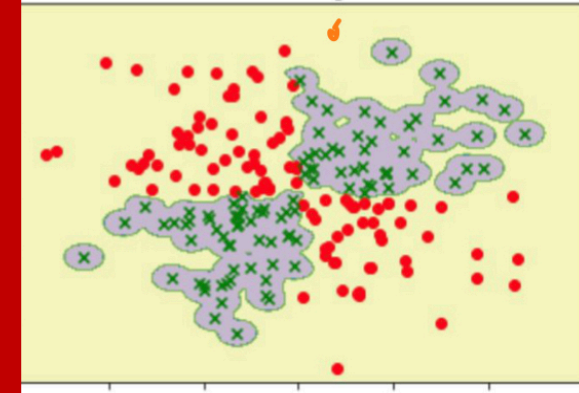
Underfitting

New data에 대한 예측을 잘 못한다.



Fitting

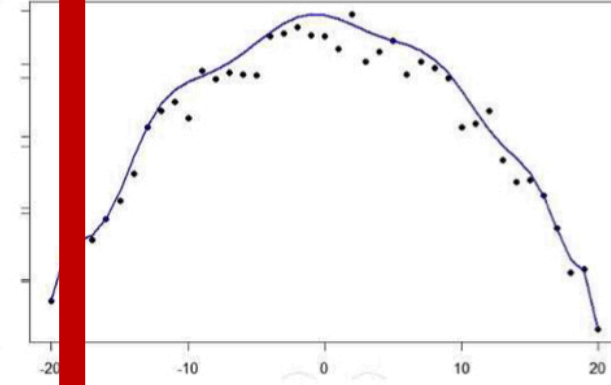
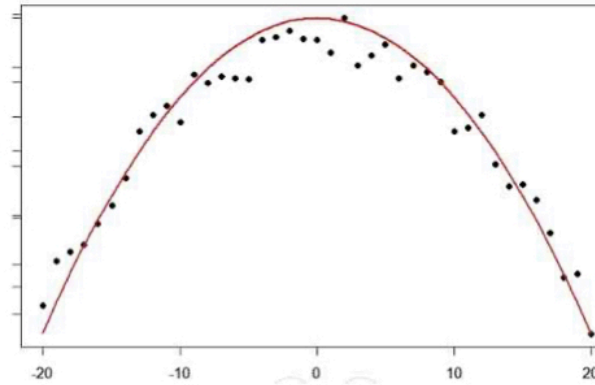
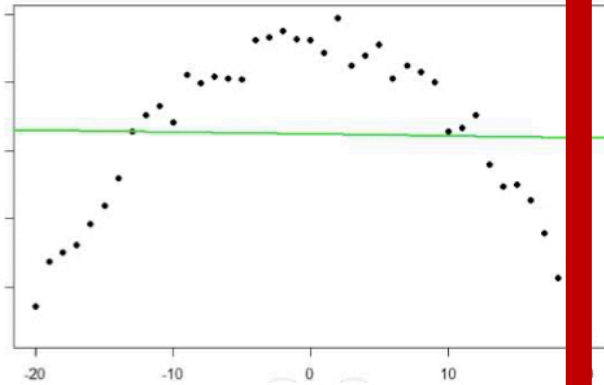
New data에 대한 예측을 잘 한다.



Overfitting

New data에 대한 예측을 잘 못한다.

Regression

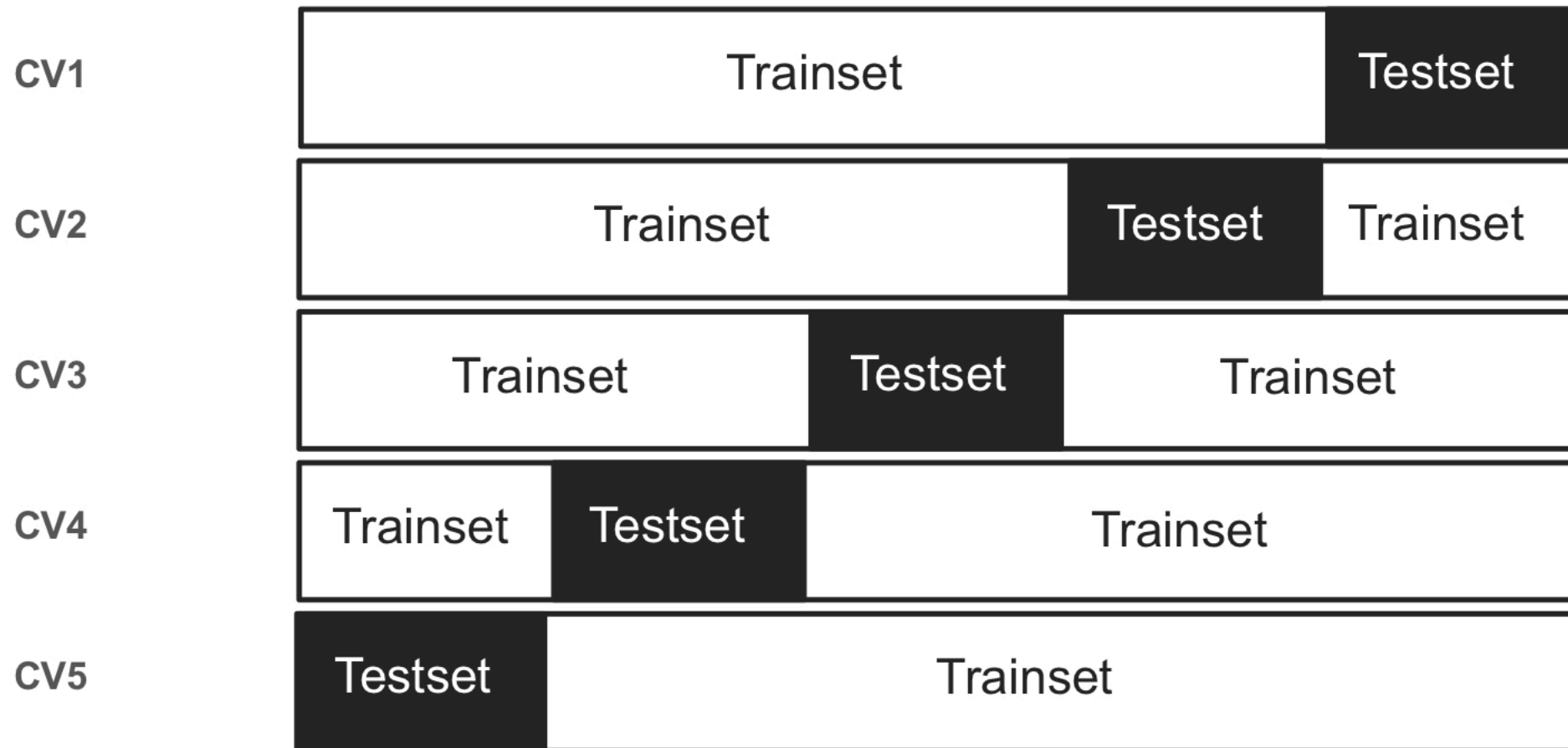


[그림 1] 과소적합, 적합, 과적합(과대적합) 예시

그렇다면, 우리의 예측이 맞았는지 어떻게 알 수 있을까?

교차 검증(Cross Validation)

- K-fold CV(K=5)



기계 학습의 장단점

장점

1. 가설을 만들고 검증할 필요가 없이 바로 예측이 가능하다
2. 특정한 의도 없이 수집된 다양한 형태의 빅데이터를 분석할 수 있다. (변수 수 > 사례 수)

기계 학습의 장단점

단점 및 주의점

1. 각 변수의 설명력을 알 수 없다(블랙박스)
2. 분석이 자료 주도적으로 진행된다. 하지만, 내용학적 지식이 부족한 경우 잘못된 모형이 도출될 수 있다.

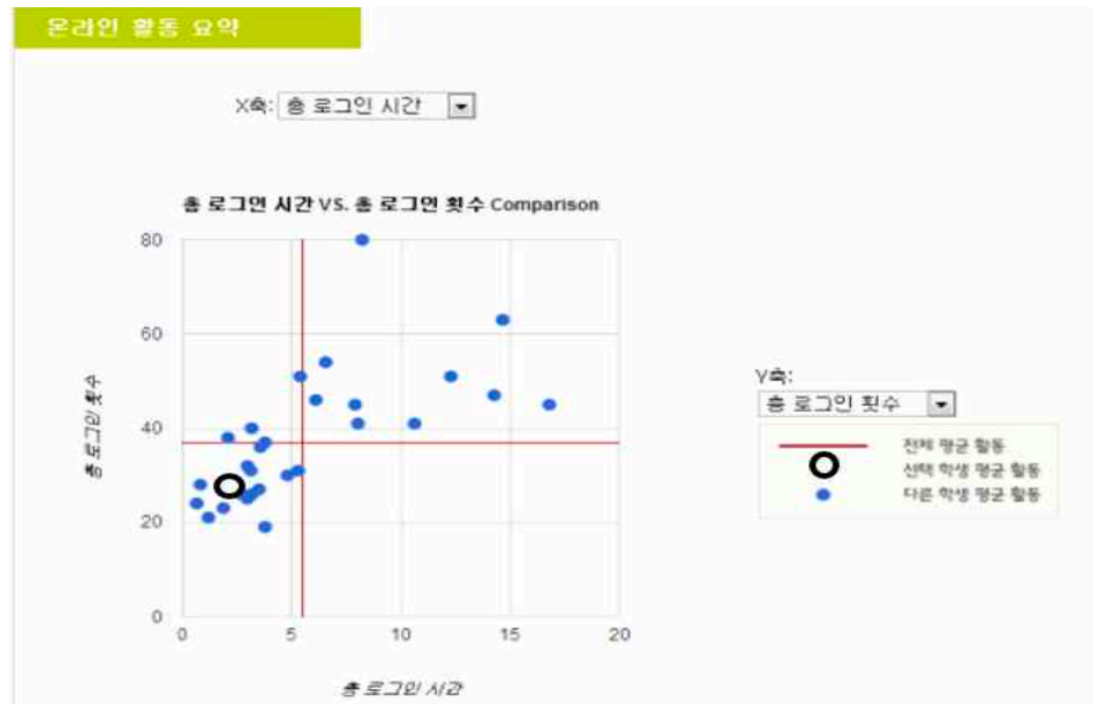
03

학습분석학 연구 실태

1. LMS 기반 대시보드 설계 관련 연구

대학의 LMS 대시보드 설계 연구

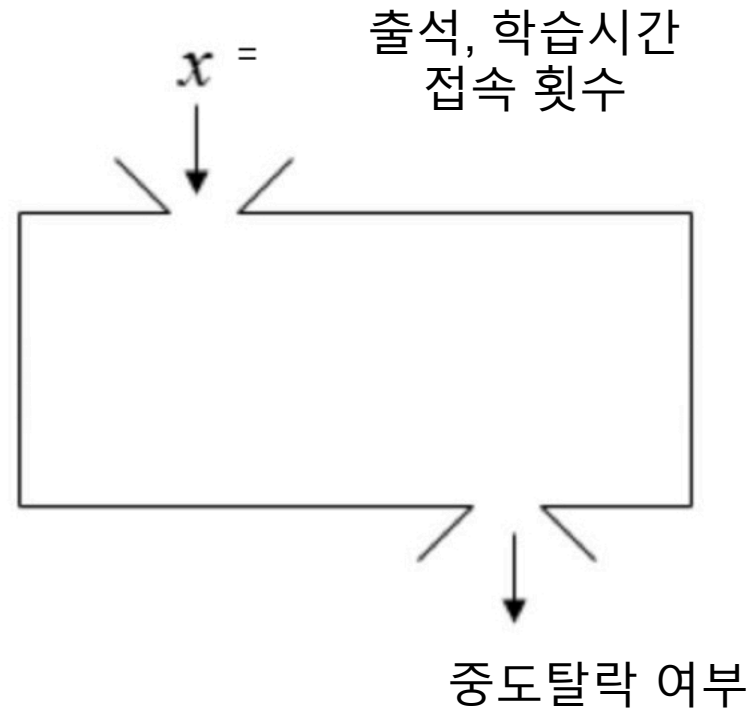
- 초점집단인터뷰, 심층 인터뷰 등을 통한 요구 분석-> 대시보드 설계 -> 사용성 평가(박연정, 조일현, 2014a; 임규연 외(2018); 임성태, 김은희(2017); 박연정, 조일현, 2014b 등)



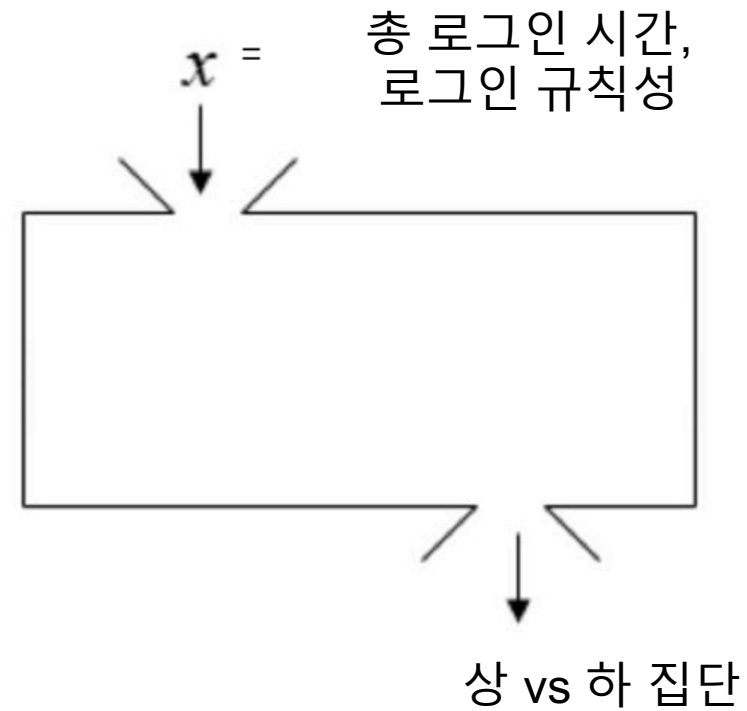
박연정, 조일현 (2014a)

2. LMS 로그파일 분석 연구

LMS 로그 데이터 분석을 통한 예측 모형



유지원(2014)



Kim, Park, Yoon & Jo
(2016)

04

년의

논의

- 여전히 다중회귀분석과 같은 전통적인 통계 기법 사용 – 가설을 세우고 검증해야하기 때문에 연구에 한계가 있음
- 기계 학습을 적용하면 기존 연구에서 파악하지 못한 새로운 패턴 파악할 수 있을 것으로 기대
- 기계학습을 통해 개별 학습자의 수행을 **실시간 측정, 평가**하고 다음 수행을 **예측**함으로써 **피드백 제공** 가능 기대

THANK YOU

2019-2학기 iLED 스터디
(191113, 송유경)