

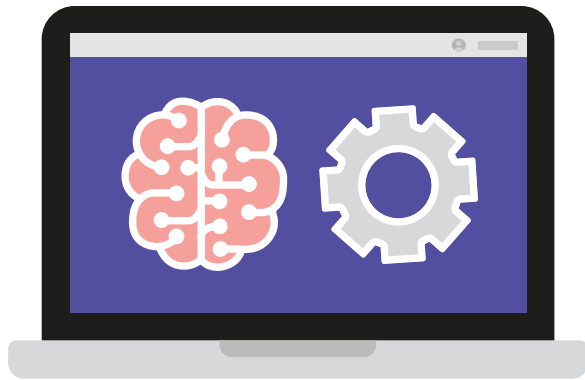
```
/* elice */
```

# 인공지능/머신러닝 기초

# Module 1: 인공지능 Overview



# 인공지능/머신러닝 기초



## 교육 기간

2월 7일 ~ 3월 11일 (4주)

---



## 선생님

김수인 선생님

---



## 라이브 수업

수요일 21:00 ~ 22:30

# 수강 대상



인공지능/머신러닝의 실체가 궁금한 개발자



인공지능/머신러닝을 본격적으로 공부 하기  
위한 **입문과정**을 찾는 분



파이썬 기본 문법을 아는 **누구나**

# 선생님 소개



## 김수인 선생님

<http://suin.kim/>

(현)

- 엘리스 Research Lead
- KAIST 전산학부 박사과정

(전)

- Research Assistant,  
Qatar Computing Research Institute
- Research Intern, Microsoft Research Asia
- SWE, Google Korea

/\* elice \*/

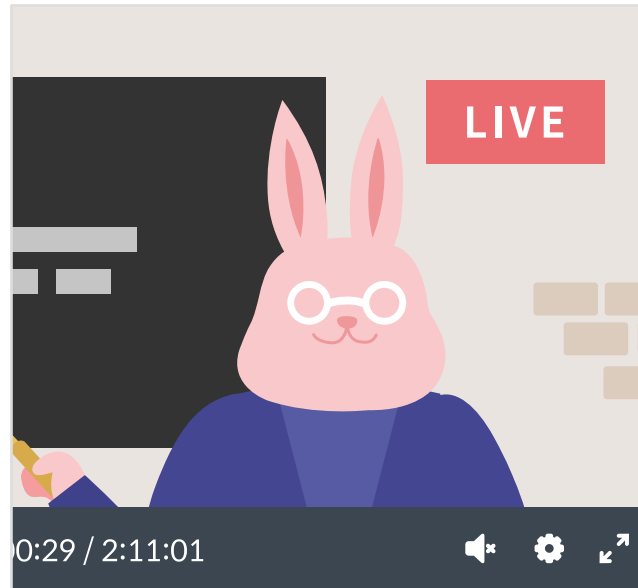
# 수업 방식

01	안녕 엘리스!	영상 보기
	1주차 강의자료	✓
	[실습1] 안녕 토끼!	✓
	[실습2] 크로케 경기장	✓
	[실습3] 도도한 도도새	✓

## 필수 학습

### 수업 1주 전

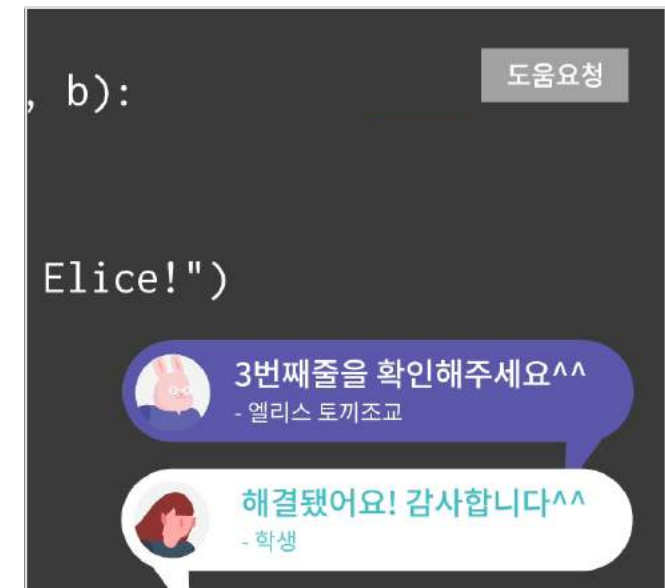
교육 자료·녹화 영상  
· 프로젝트 문제 공개



## 라이브 코딩

### 수업 당일

라이브 코딩으로  
프로젝트 문제 풀이,  
실시간 Q&A



## 다시 보기·헬프 센터

### 수업 후

다시 보기 영상,  
헬프 센터 항상 오픈

/\* elice \*/

# 주차별 커리큘럼

1 ○

## 기초 선형대수학 및 인공지능/머신러닝 개요

인공지능/머신러닝의 기본 개념과 앞으로 진행할 수업의 기본이 되는 기초 선형대수학, NumPy 라이브러리에 대해 다룹니다.

2 ○

## 선형회귀법

관찰한 데이터를 기반으로 새로운 데이터에 대한 예측을 진행할 수 있는 선형회귀법 (Linear Regression) 에 대해 배웁니다.

# 주차별 커리큘럼

3



## 확률론 및 나이브베이지스 분류기

동전 던지기 시뮬레이터, 인공지능 암 진단, Bag of Words를 구현하며 확률론과 나이브베이지스 분류법을 공부합니다.

4



## 군집화 알고리즘 및 챌린지! 데이터 사이언스 프로젝트

군집화 알고리즘의 대표적인 예인 K-means 알고리즘을 이용해 데이터를 최적화합니다. 마지막으로 지금까지 배운 인공지능 알고리즘을 이용해 엘리스 데이터 사이언스 챌린지 문제를 처음부터 끝까지 풀어봅니다.

# 4주 뒤

AI/ML 기본 개념을 이해하게 됩니다.

코드에 AI/ML 원리를 적용할 수 있습니다.

엘리스 데이터 챌린지에 도전할 수 있습니다.

`/* elice */`

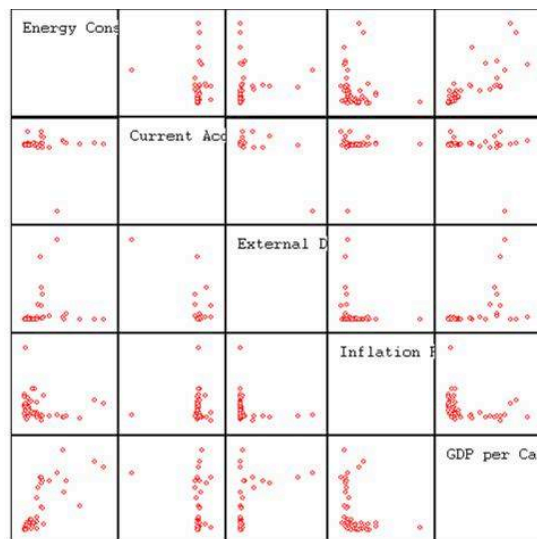


# 인공지능/머신러닝 개론

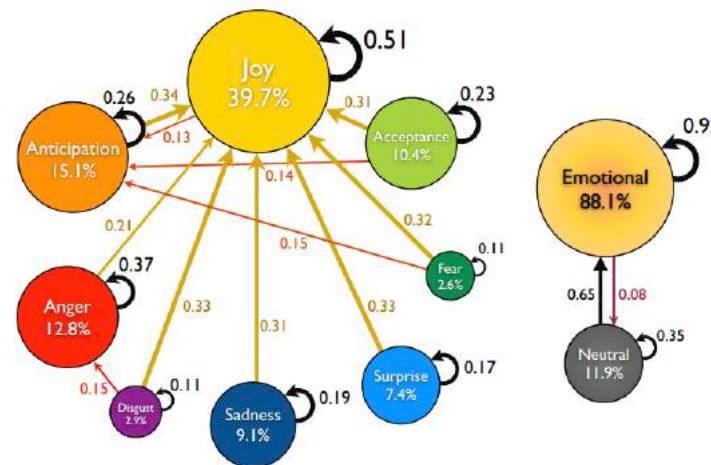
# Data Science

## 데이터 중심의 과학

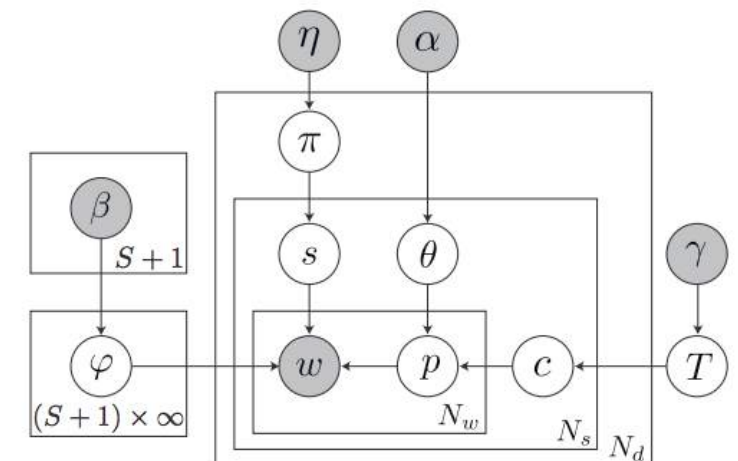
데이터에서 지식과 통찰력을 발견하는 과학적인 방법론



통계학

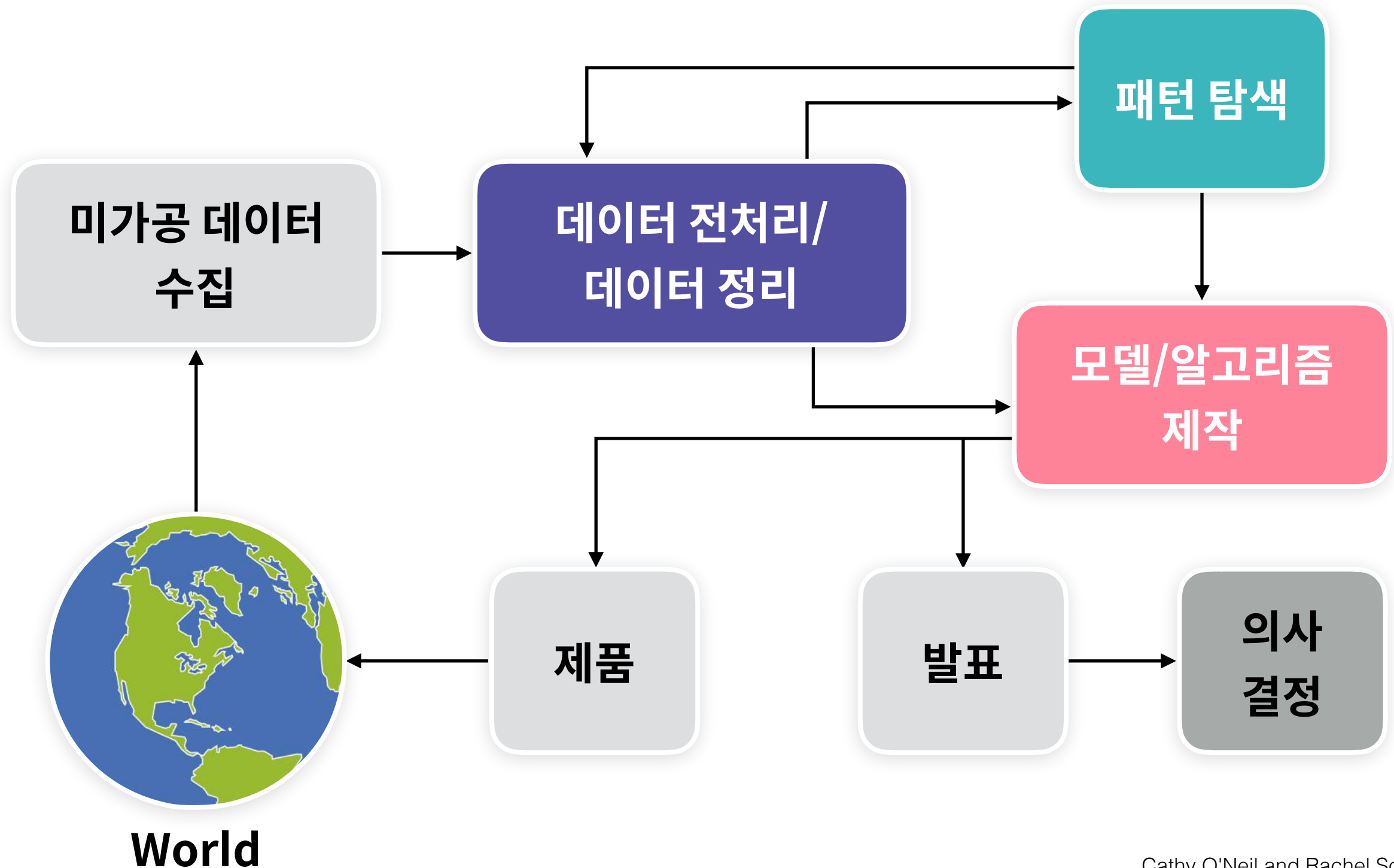


패턴 인식



머신 러닝

# Data Science Process



Cathy O'Neil and Rachel Schutt, 2013

# Early Data Science

19세기 중반 영국 런던

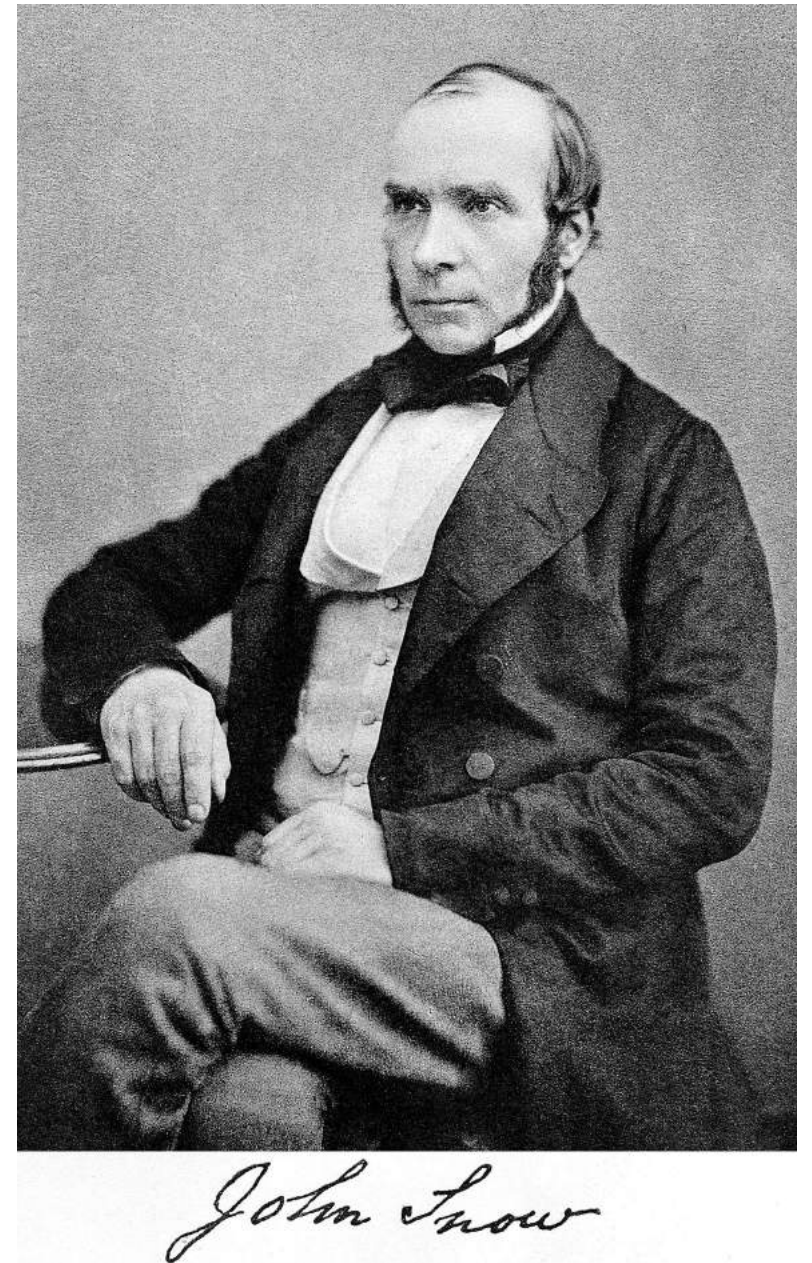
콜레라 발병

기존의 가설:

- Miasma라는 독이 공기를 통해 전파

John Snow:

- “콜레라는 왜 그리고 어떻게 전파되는가?”





# Early Data Science

“지도에 발병자를 표시해 보자”

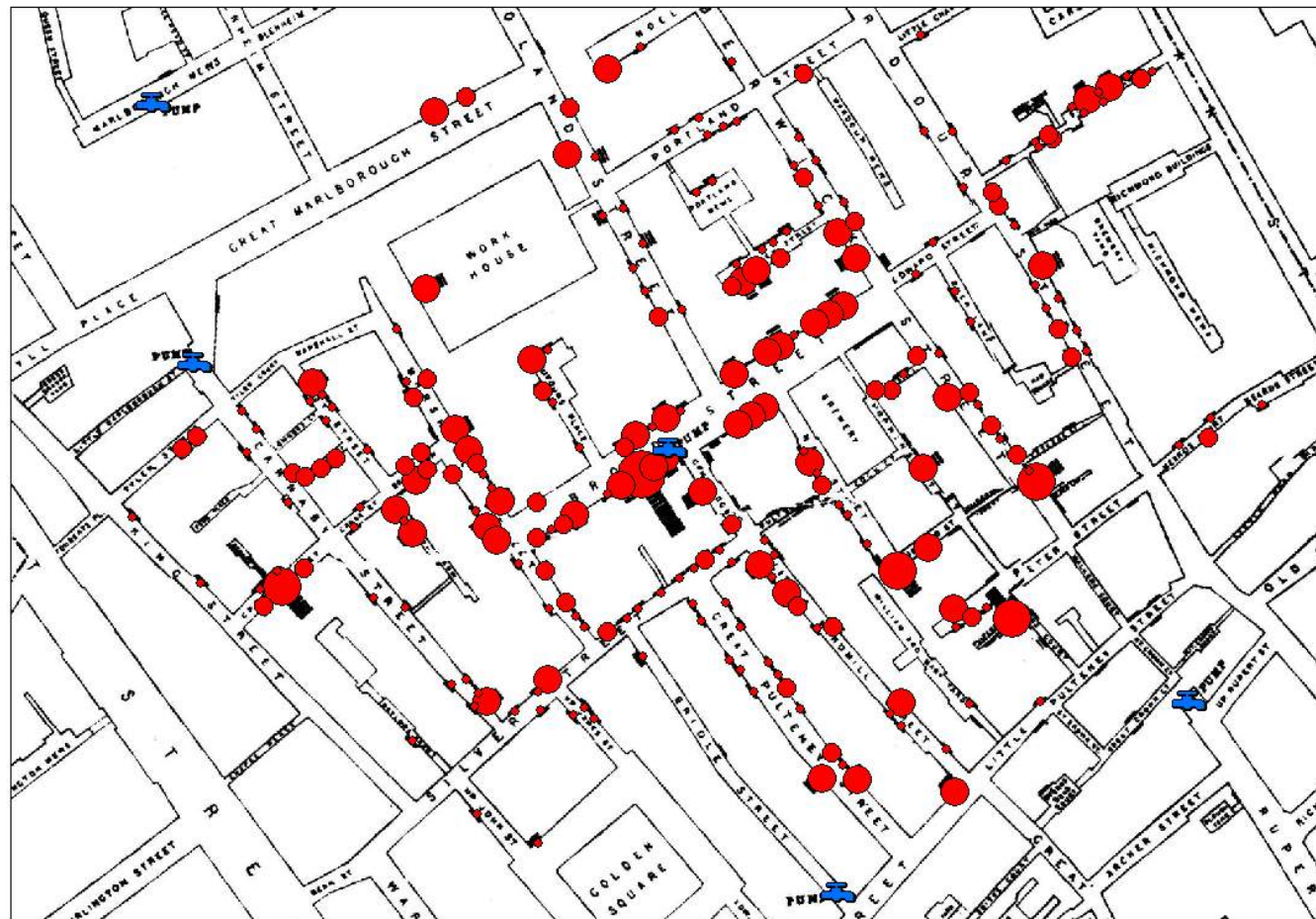


# Early Data Science

콜레라로 인한 사망자는 **한 펌프를 기준으로** 퍼짐

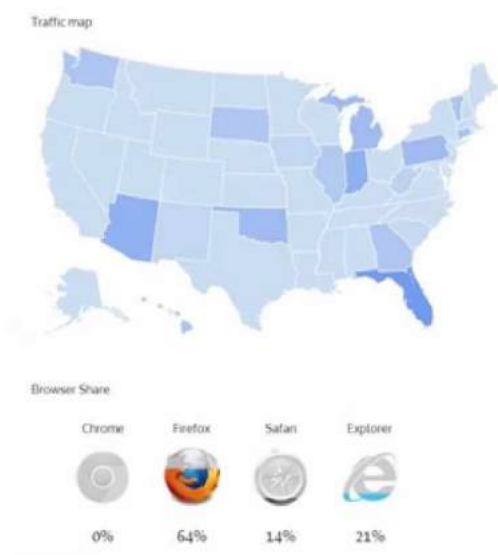
다른 펌프를 이용한 사용자는 사망하지 않음

“콜레라를 옮기는 매개체는 공기가 아니라 물이다”



# 머신 러닝

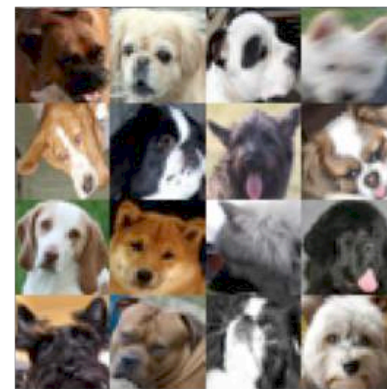
명시적으로 프로그래밍을 하지 않고도  
컴퓨터가 학습할 수 있는 능력을 갖게 하는 것



데이터 마이닝



VS



Feature 학습



추천

# 왜 머신 러닝이 필요한가?

명시적으로 프로그래밍을 하지 않고도  
컴퓨터가 **학습할 수 있는 능력**을 갖게 하는 것



여러분은 페이스북의 데이터 사이언티스트로서  
매일 발생하는 상태와 댓글을 분석하고자 한다.



# 왜 머신 러닝이 필요한가?

명시적으로 프로그래밍을 하지 않고도  
컴퓨터가 **학습할 수 있는 능력**을 갖게 하는 것



여러분은 페이스북의 데이터 사이언티스트로서  
매일 발생하는 상태와 댓글을 분석하고자 한다.

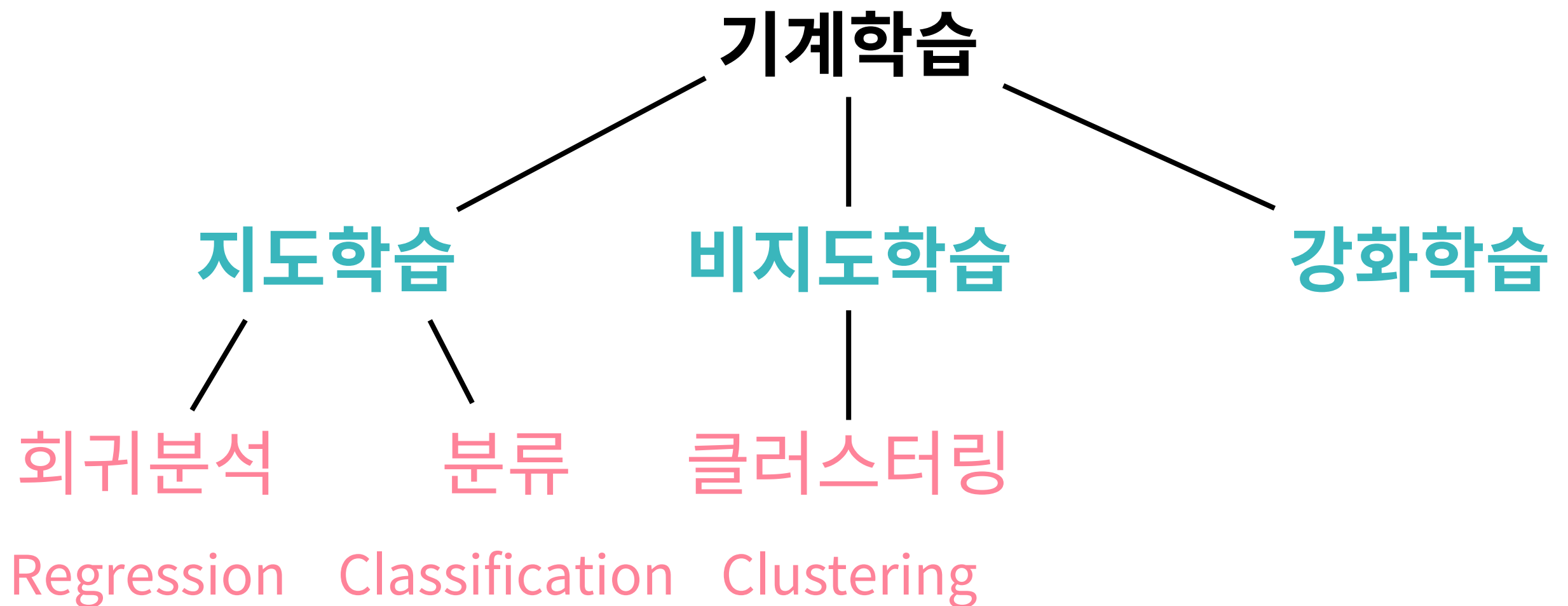
...

한달 평균 **18억 명**의 active user

하루 평균 **3억 건**의 사진

1분 평균 **51만 건**의 댓글과 **30만 건**의 상태 업로드

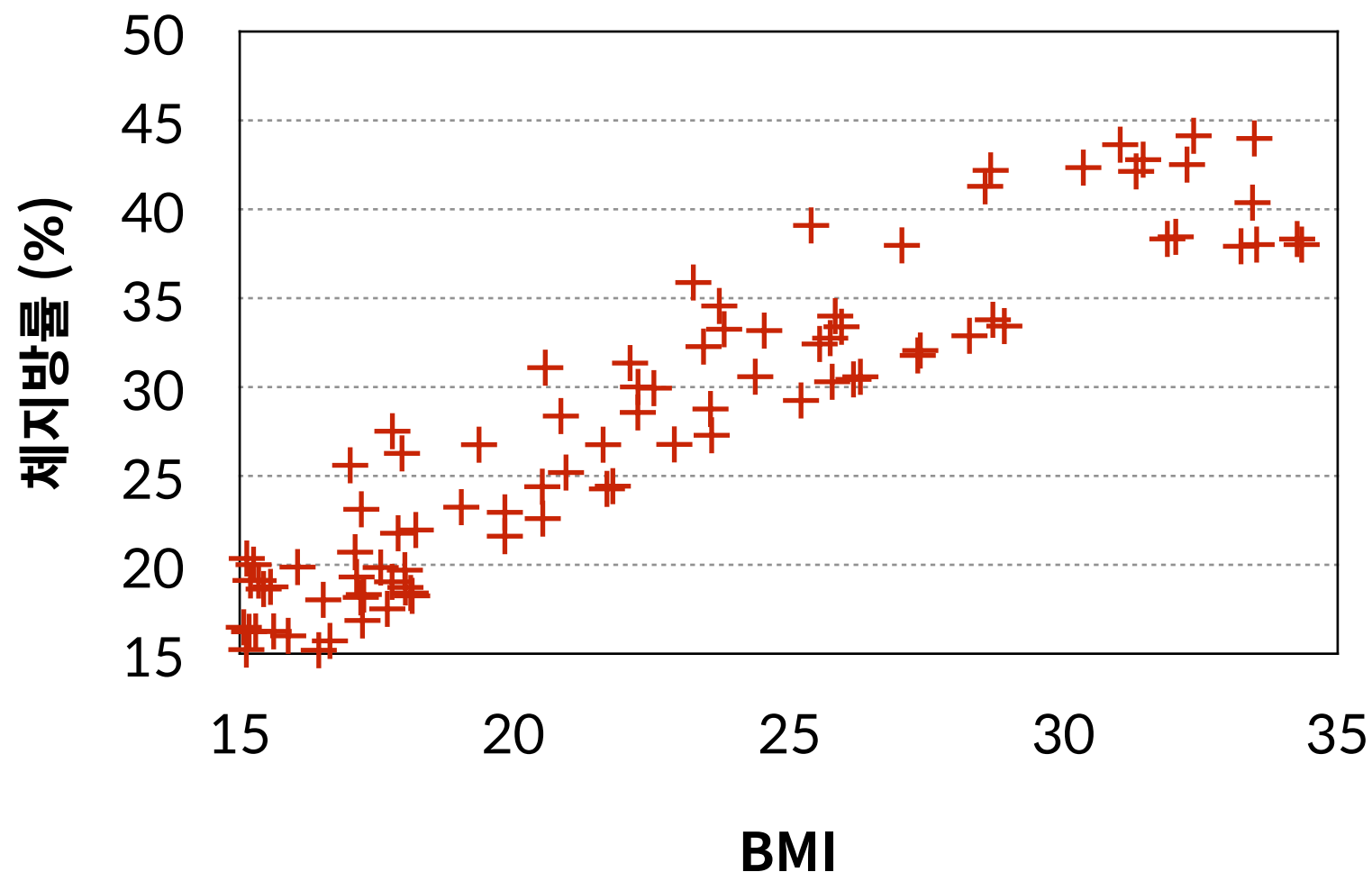
# 그럼 어떤 걸 할수 있죠?



# 지도 학습 - 회귀 문제

100명의 일반인을 대상으로 BMI와 체지방률을 측정했다.

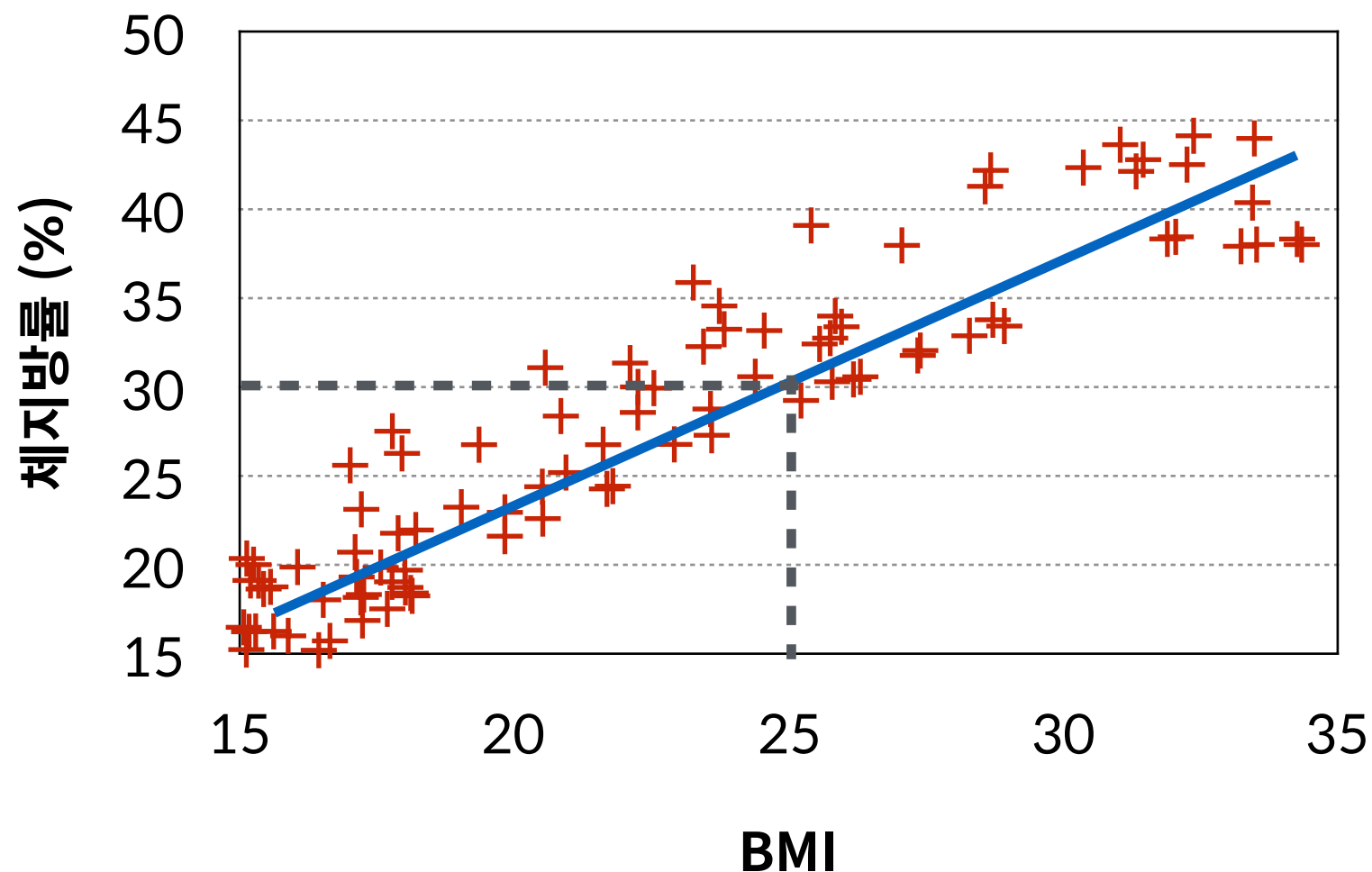
BMI 25인 사람의 체지방률을 예상할 수 있는가?



# 지도 학습 - 회귀 문제

직선을 하나 그려 보자. 이 직선은 어떻게 그릴까?

약 30%로 예측하면 적절할까?

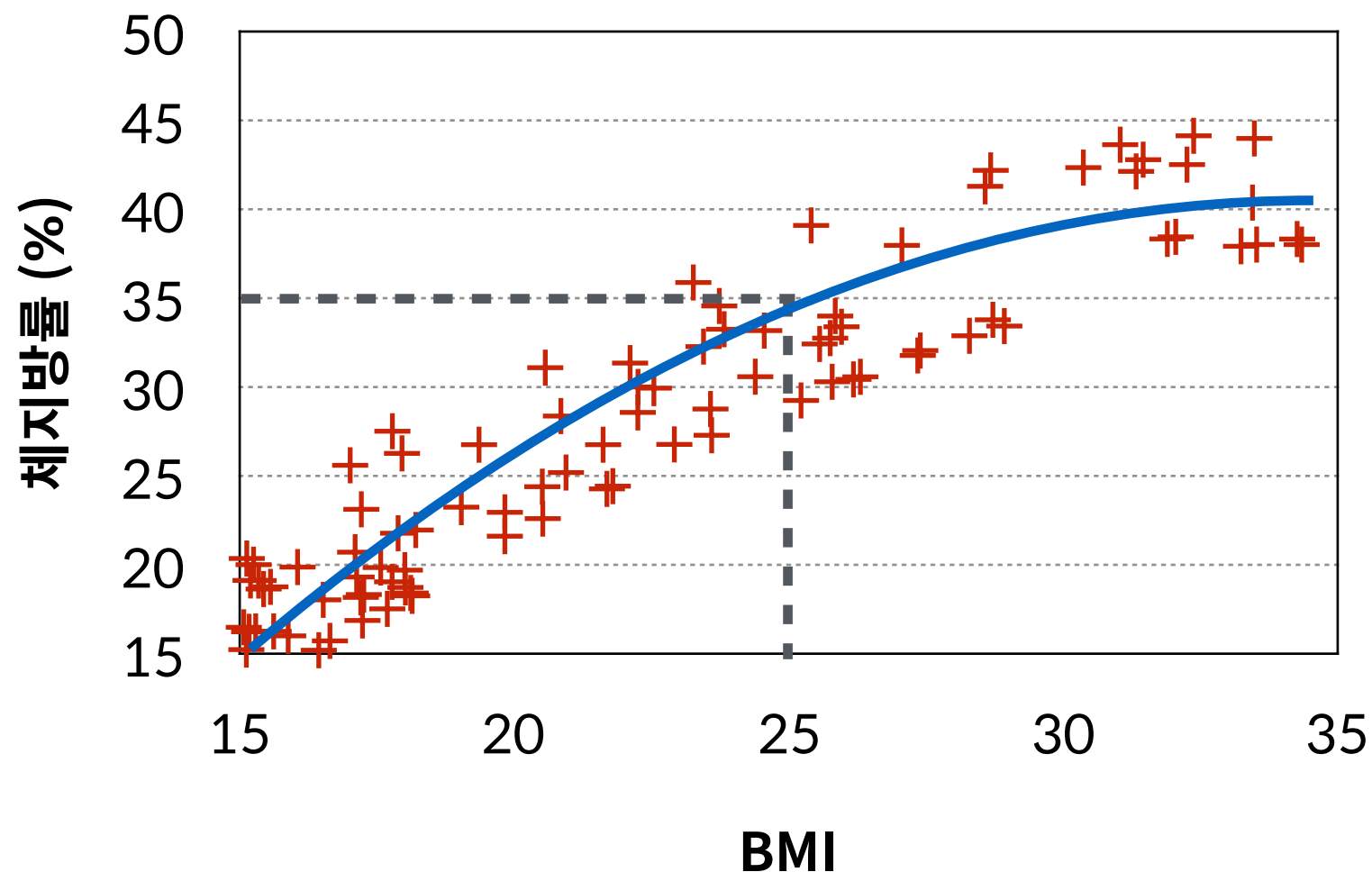


/\* elice \*/

# 지도 학습 - 회귀 문제

직선 말고 조금 더 제대로 할 수는 없을까?

2차곡선: 35%?

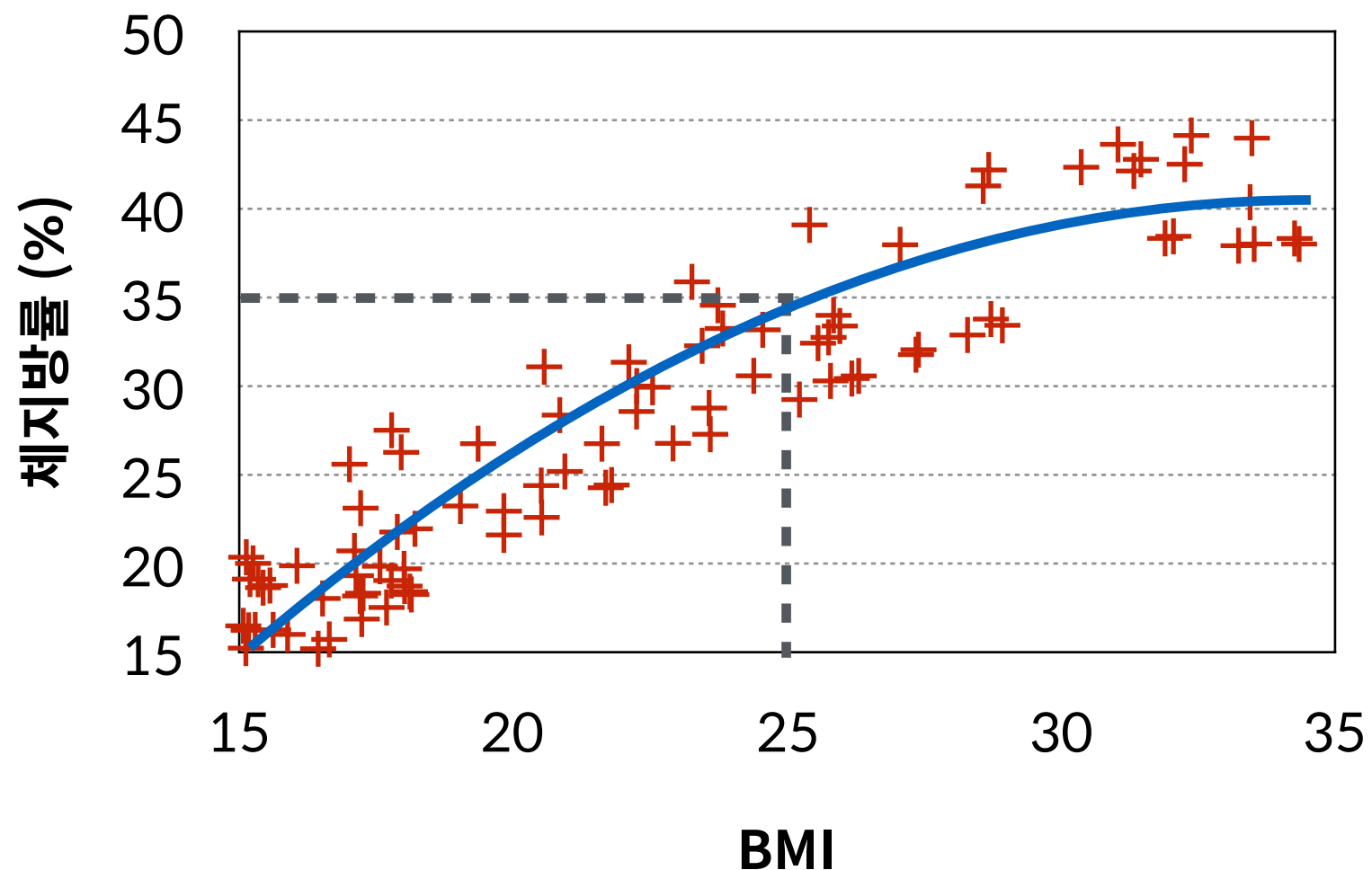


/\* elice \*/

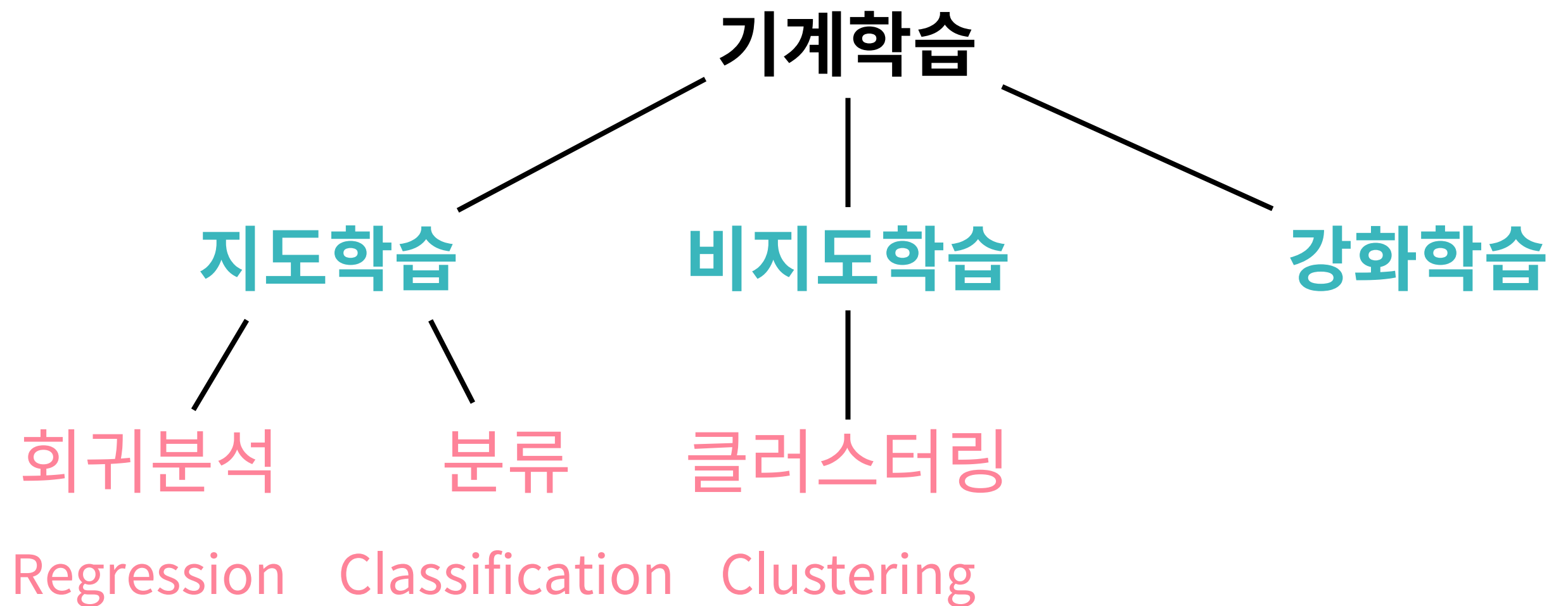
# 지도 학습 - 회귀 문제

회귀분석:

결과값을 가장 잘 예측하는 방법을 찾는다.

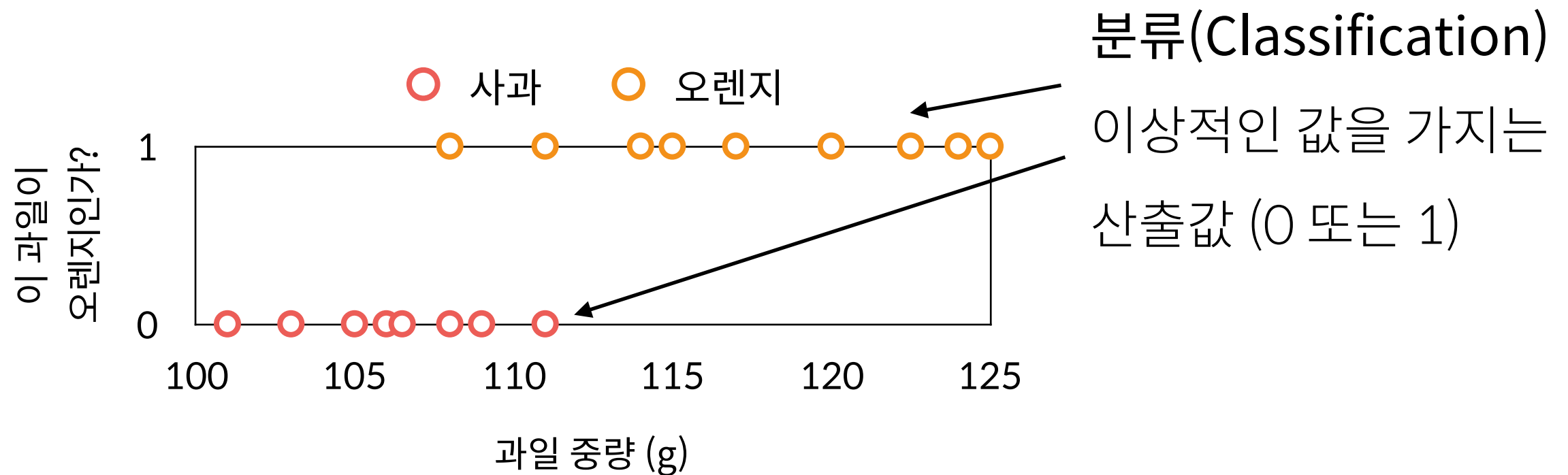


/\* elice \*/



# 지도 학습 - 분류 문제

분류 문제는 새로운 데이터가 어떤 클래스(class)에  
해당할지 판단하는 문제

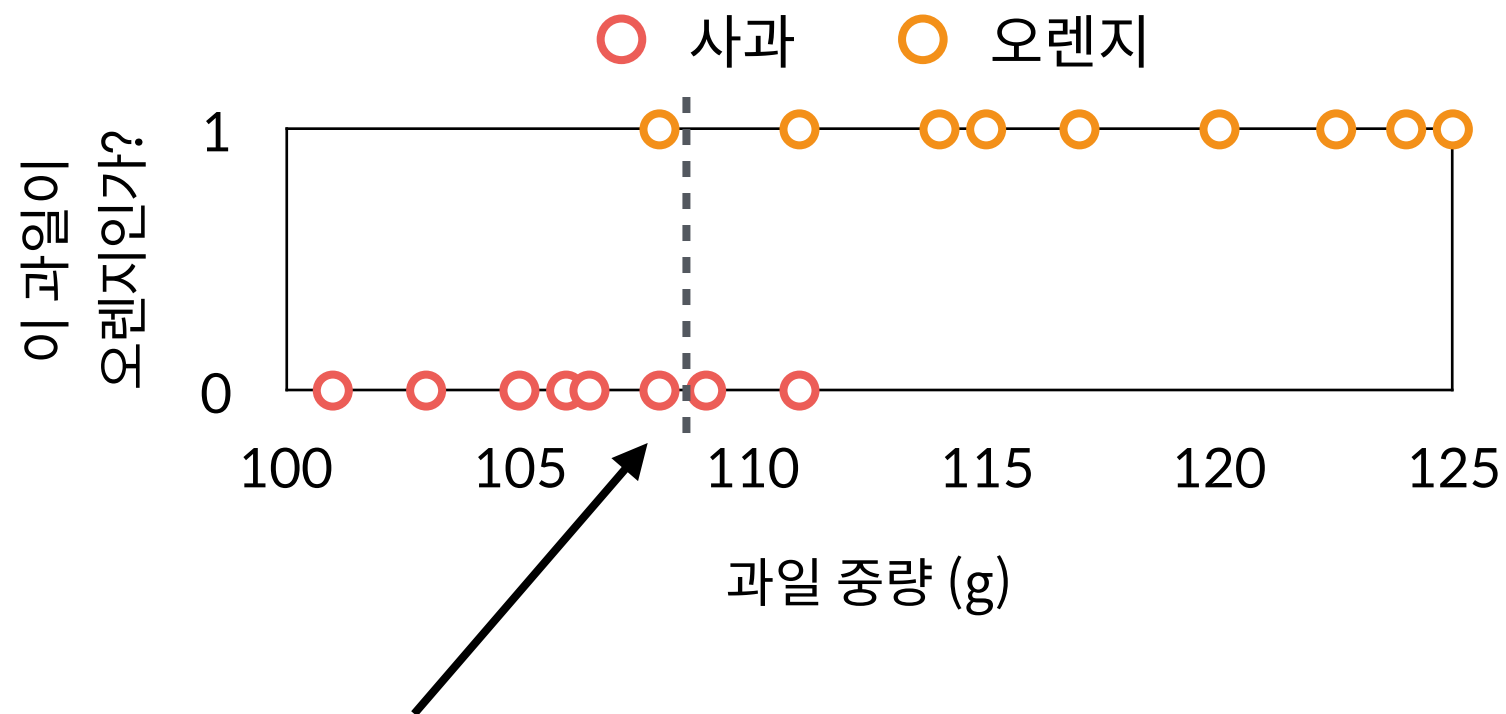


과일의 종류가 여러가지라면, 산출값이 여러가지일 수 있다.

예) 0: 사과, 1: 오렌지, 2: 멜론, 3: 수박



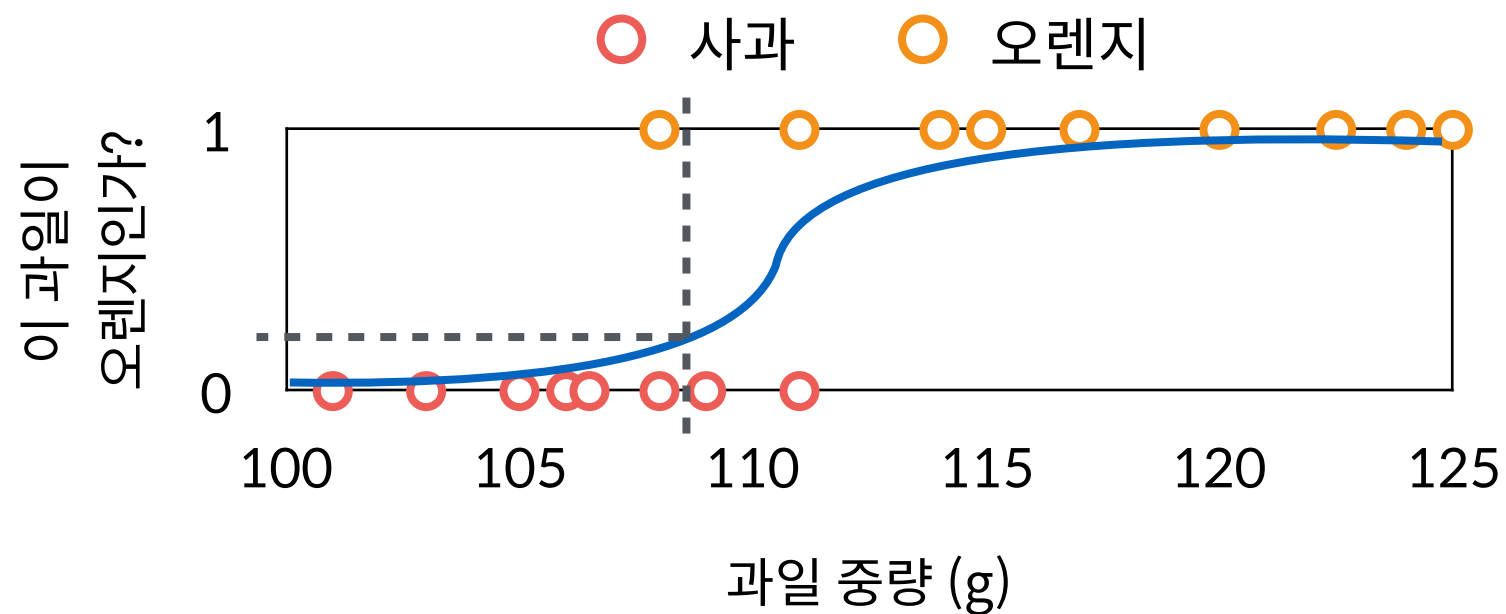
# 지도 학습 - 분류 문제



이 중량을 가진 과일은 사과일까, 오렌지일까?

이것이 사과일 확률과 오렌지일 확률은 각각 얼마인가?

# 지도 학습 - 분류 문제

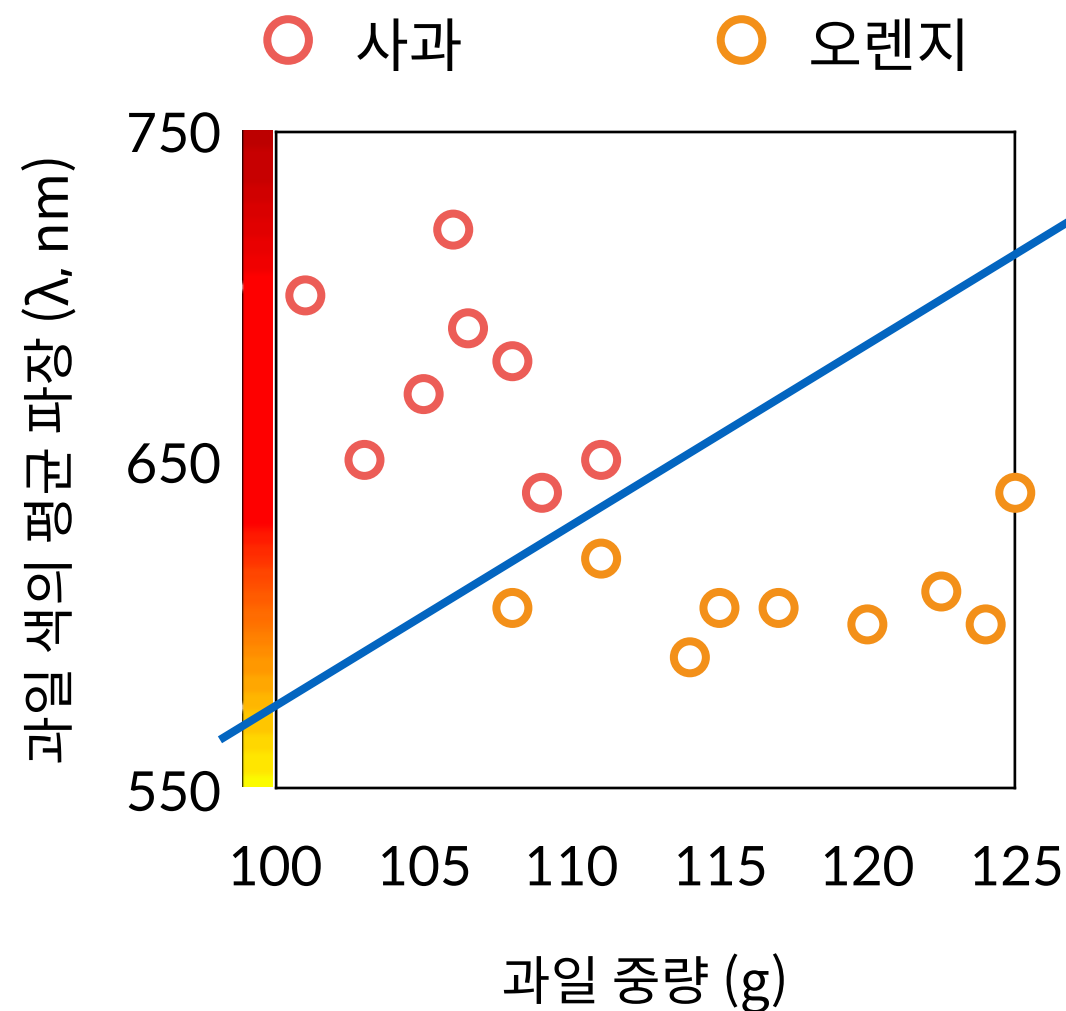


이 중량을 가진 과일은 사과일까, 오렌지일까?

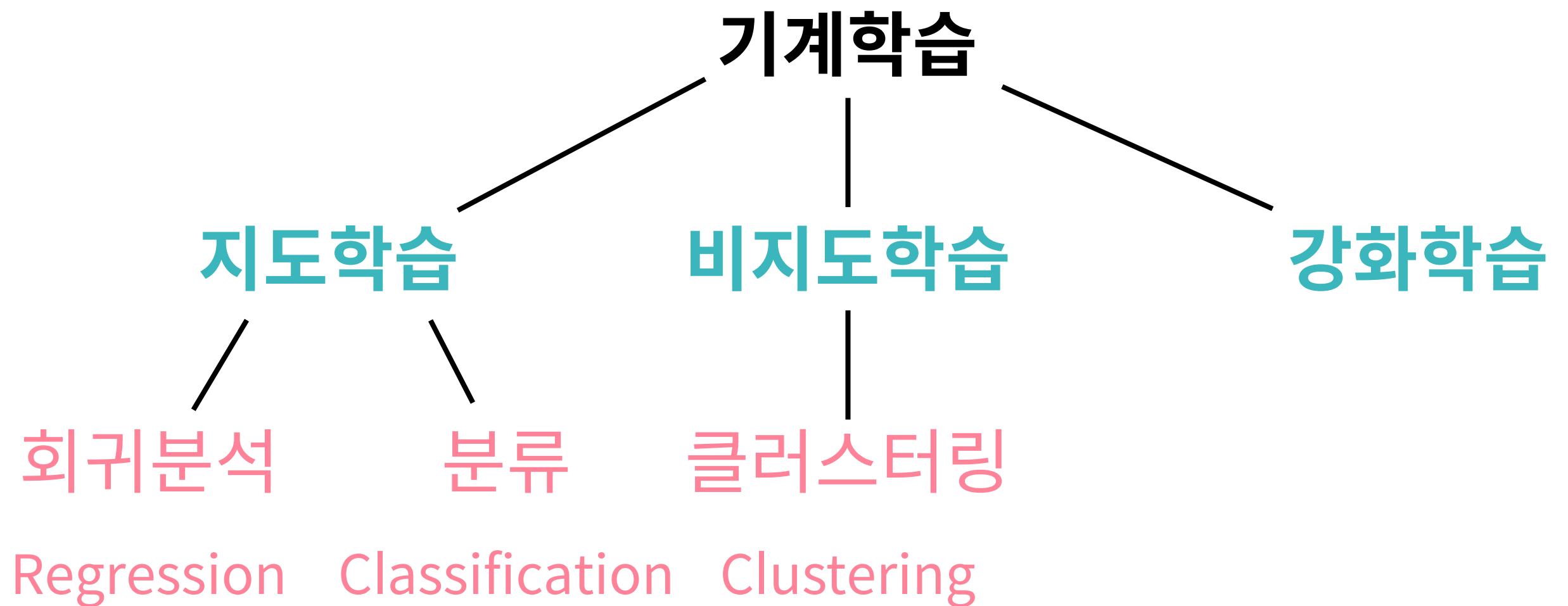
$$P(\text{오렌지}) = 0.2, \quad P(\text{사과}) = 1 - 0.2 = 0.8$$

# 지도 학습 - 분류 문제

데이터는 여러 개의 차원을 가질 수 있다.



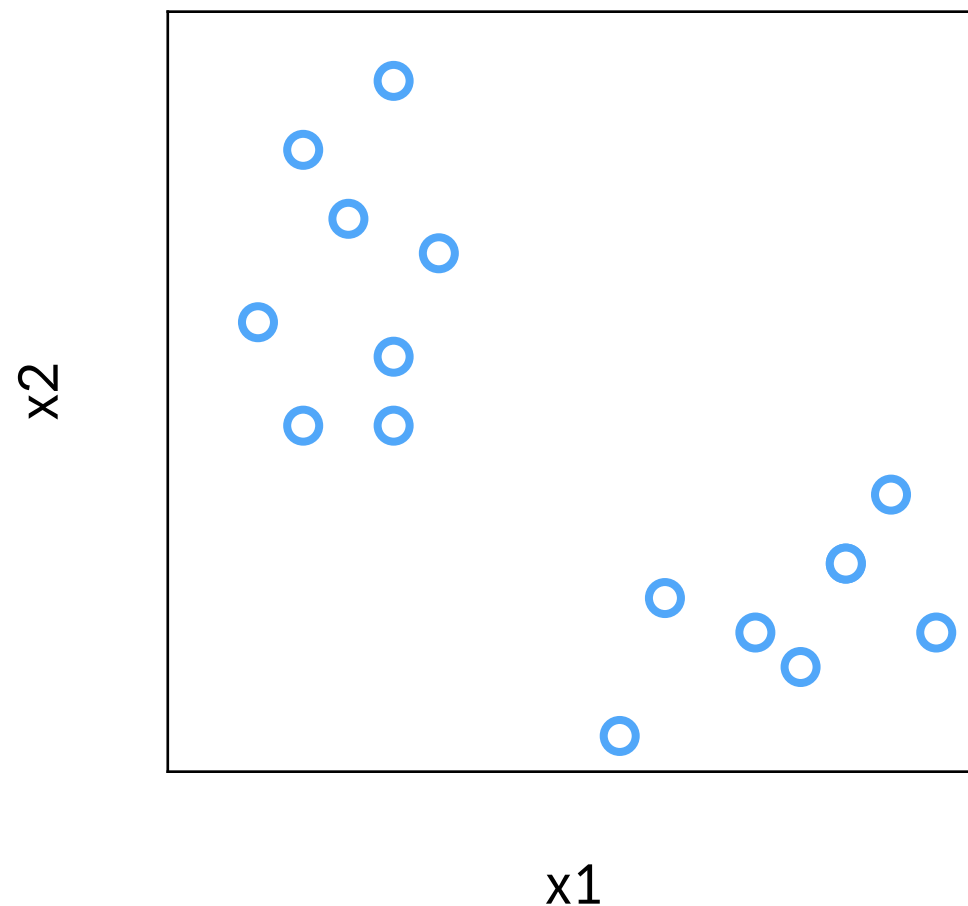
`/* elice */`



# 비지도 학습

이런 데이터가 있다.

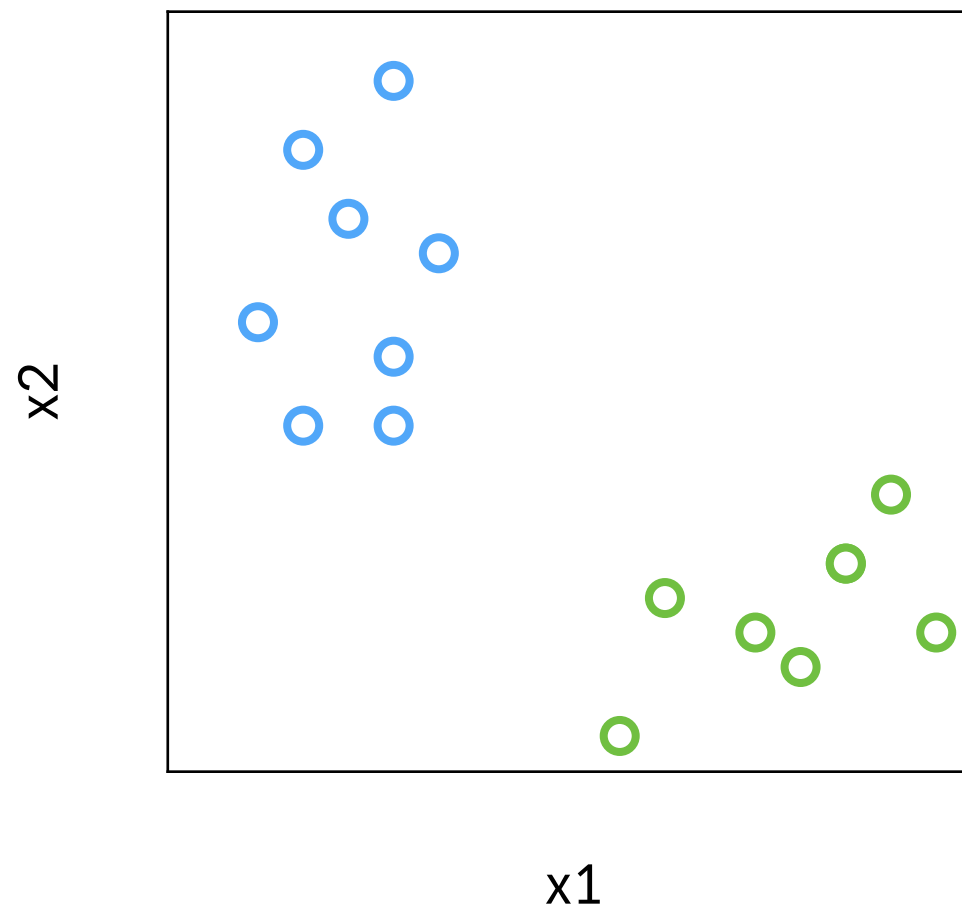
**무엇을 할 수 있나?**



# 비지도 학습

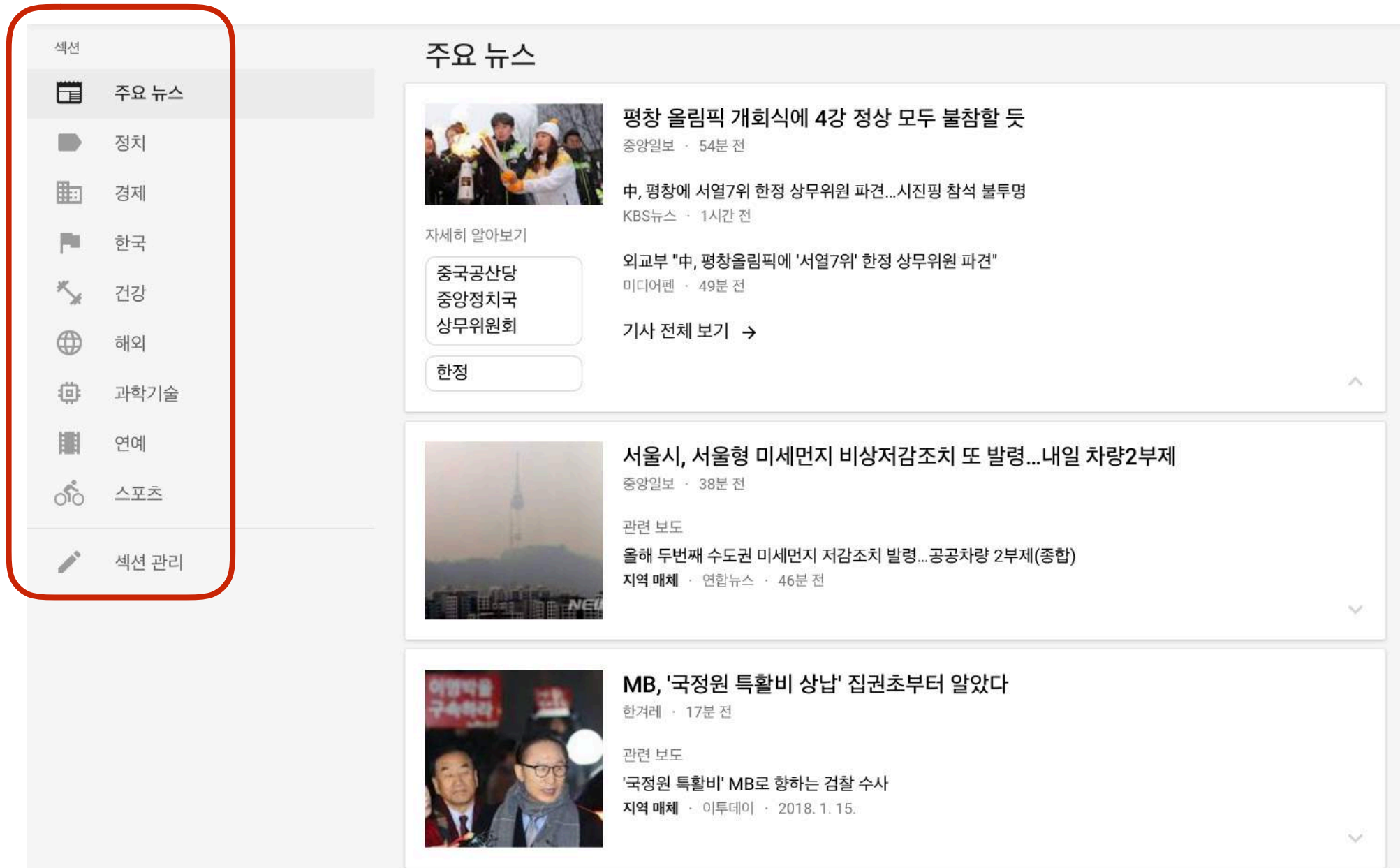
데이터에 대한 정보가 주어지지 않는다

→ 비슷한 것들끼리 묶어낸다



# 비지도 학습 - 군집화

뉴스 기사와 같은 텍스트에도 적용할 수 있다



# 강화학습

어떤 환경 안에서 정의된 **에이전트**가 **현재의 상태**에서,  
선택 가능한 행동들 중 **보상**을 최대화하는 **행동** 혹은  
행동 순서를 선택하는 방법

