

YBIGTA

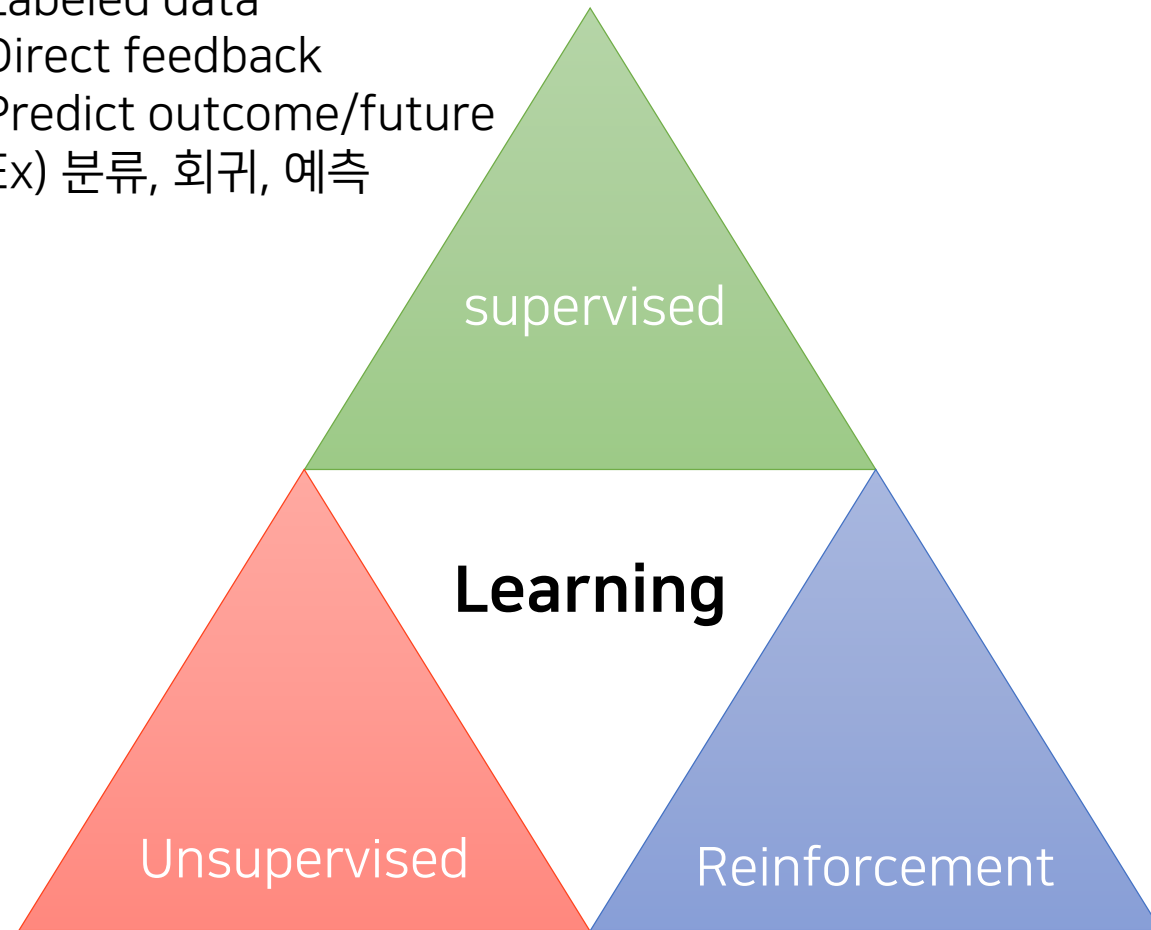
Machine Learning

사이언스팀 백진우

목차

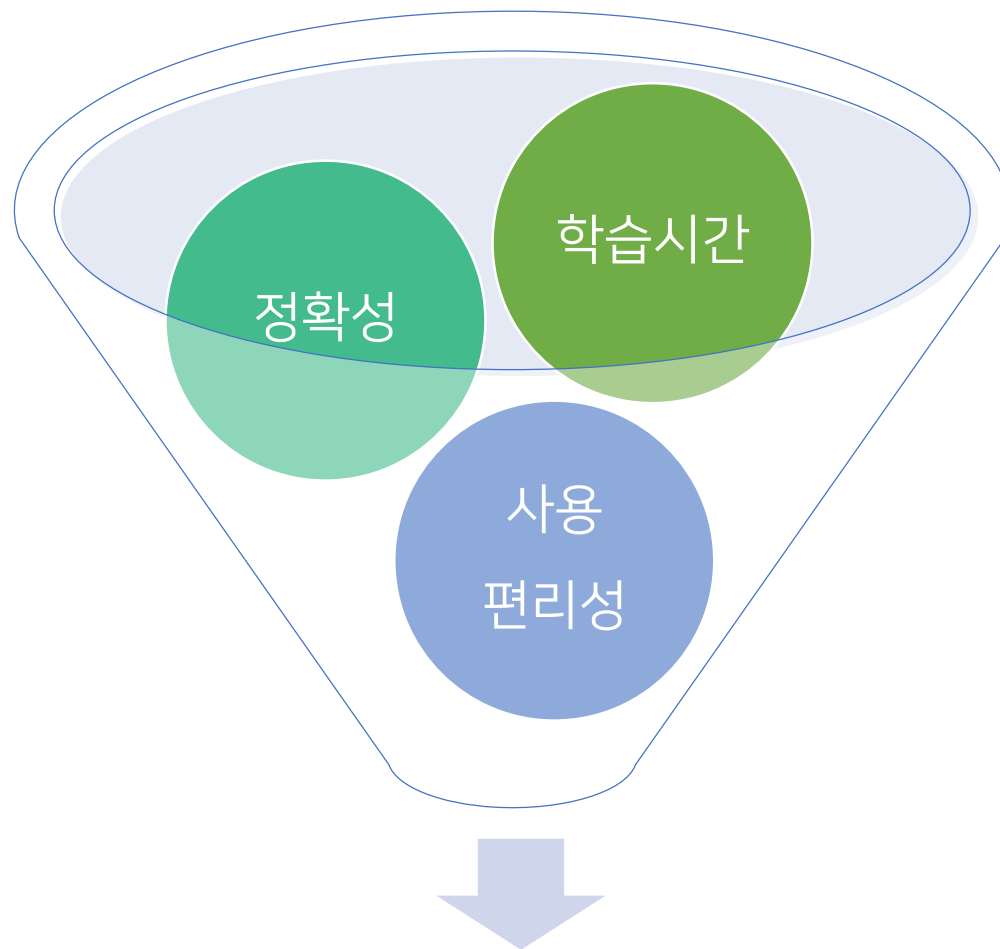
1. 머신러닝 알고리즘 유형
2. 알고리즘 선택 시 고려 사항
3. 머신러닝 알고리즘 치트 시트
4. 머신러닝 알고리즘 설명
5. 머신러닝 알고리즘 문제

- Labeled data
- Direct feedback
- Predict outcome/future
- Ex) 분류, 회귀, 예측



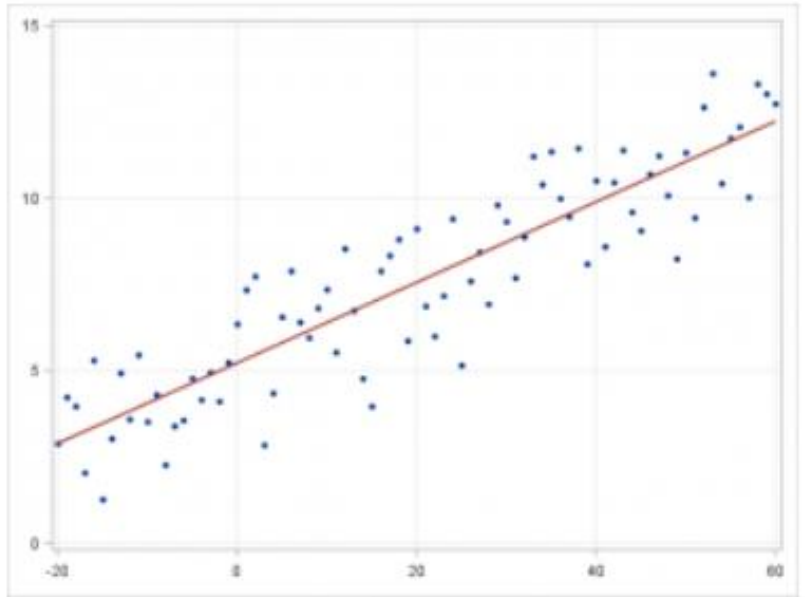
- No labels
- No feedback
- "Find hidden structure"
- Ex) 클러스터링, 차원 축소

- Decision process
- Reward system
- Learn series of actions
- Ex) Q-Learning

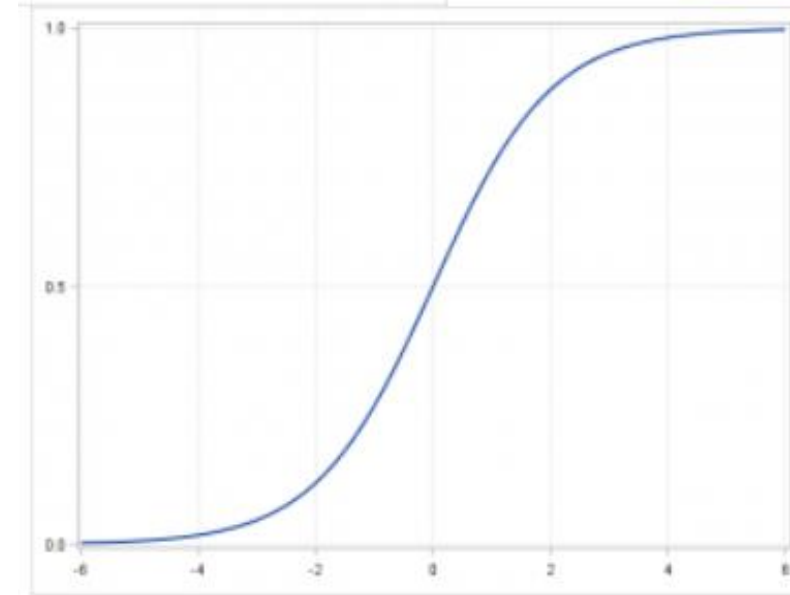




선형회귀 및 로지스틱 회귀

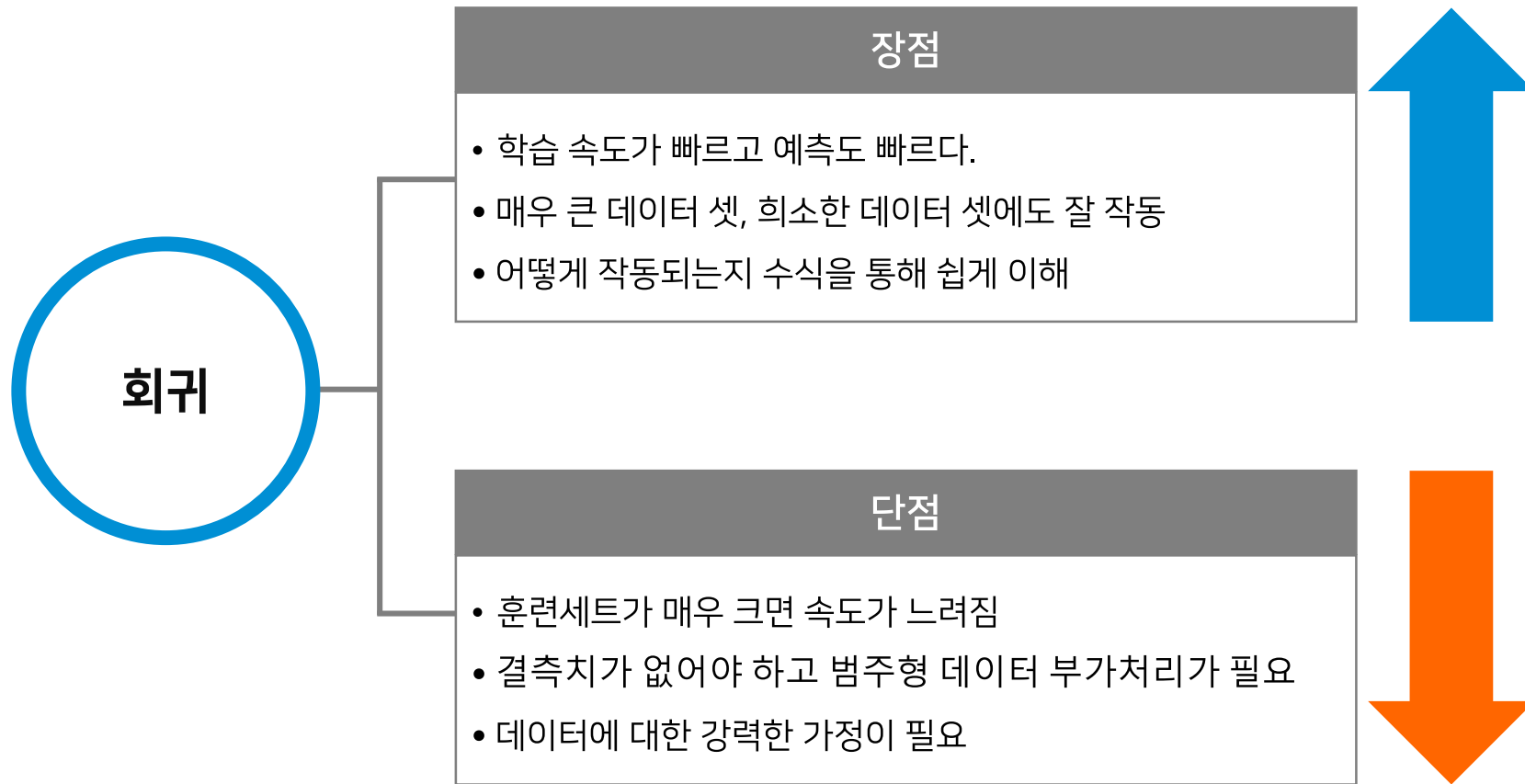


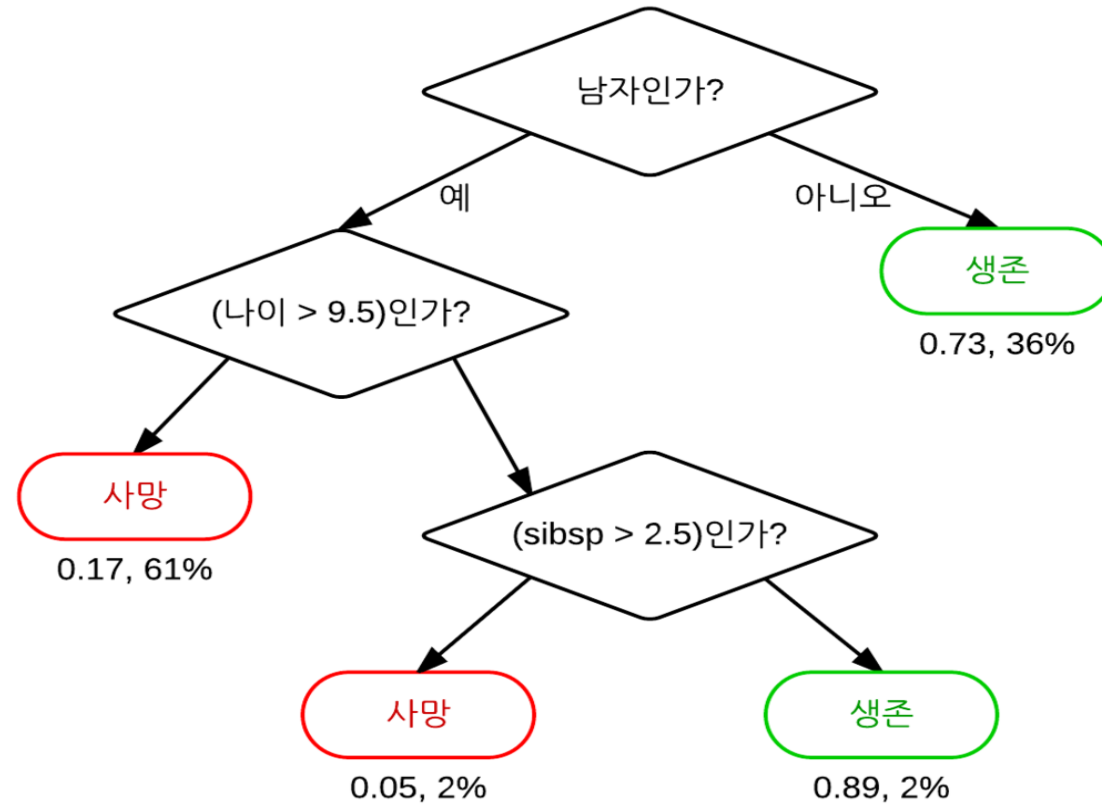
- 연속적인 종속 변수 Y와 한 개 이상의 예측 변수인 X 사이의 관계를 모델링



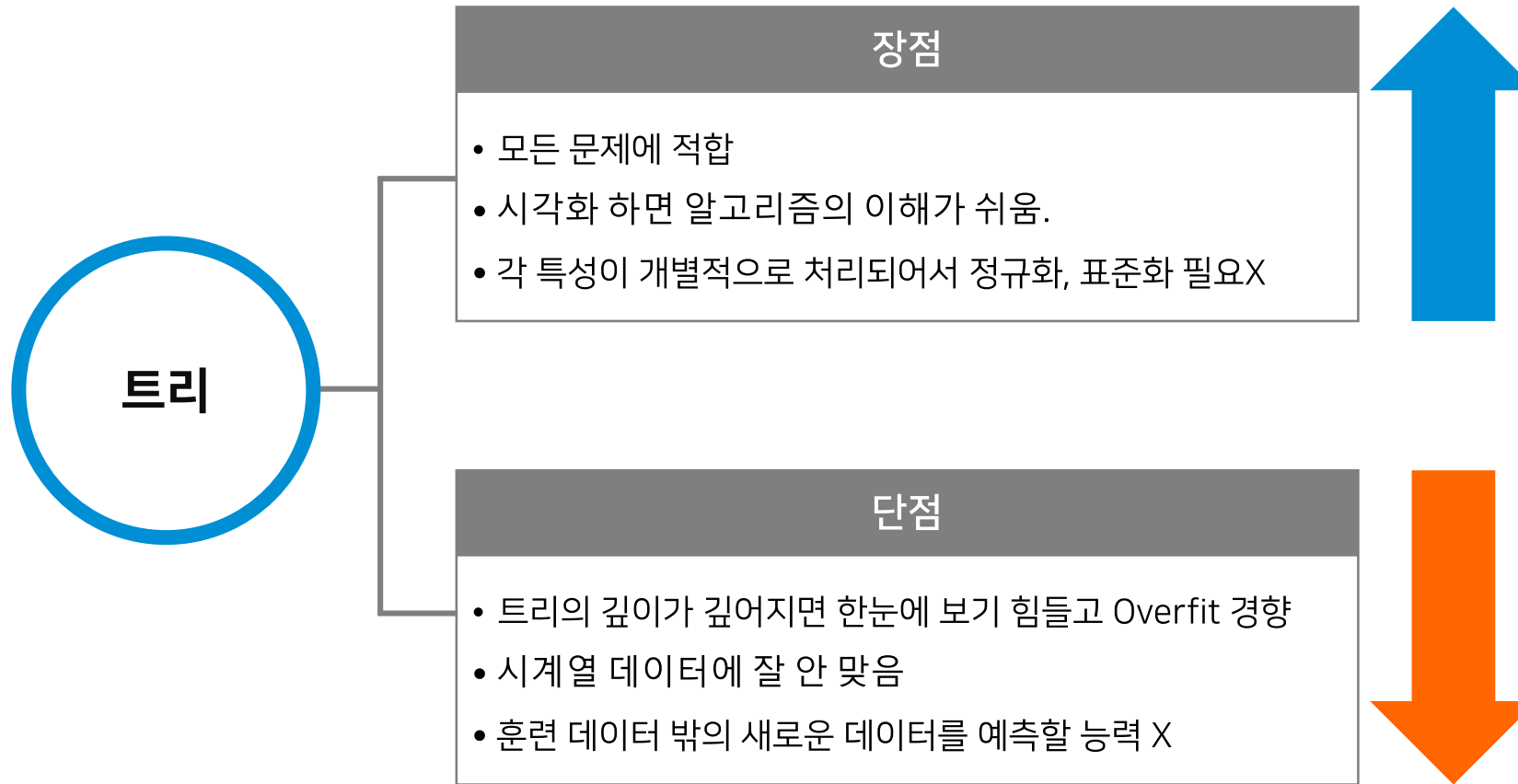
- Sigmoid 함수를 이용하여 두가지 클래스를 구분하는데 사용, Y값이 0과 1사이에 있음.

선형회귀 및 로지스틱 회귀의 장단점

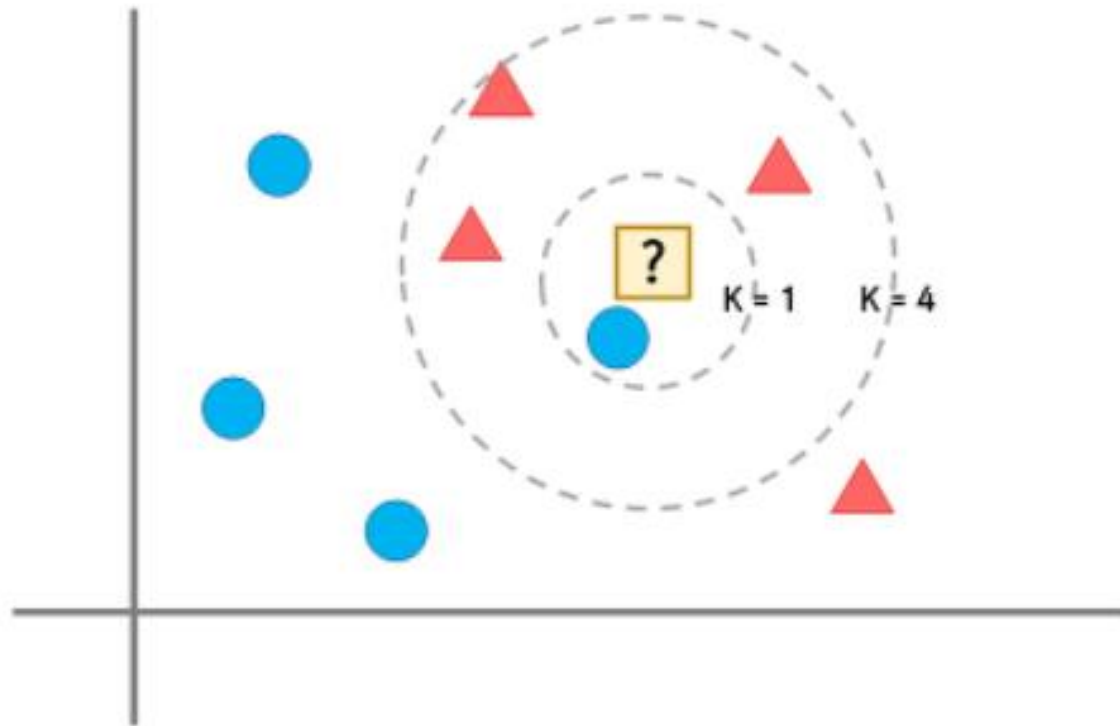




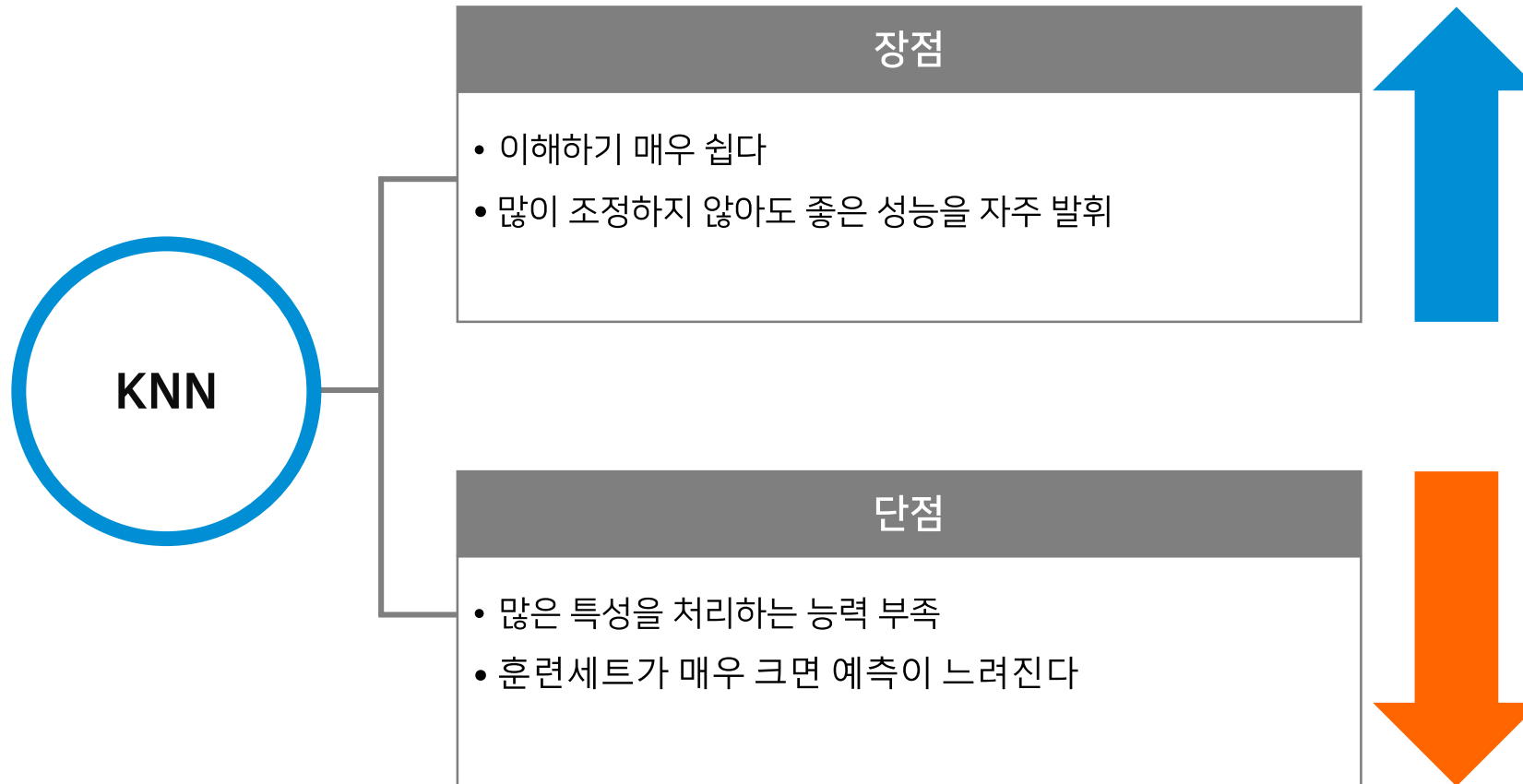
- 몇몇 입력 변수를 바탕으로 목표 변수의 값을 예측하는 모델을 생성하는 것을 목표로 한다.
- ex) decision tree



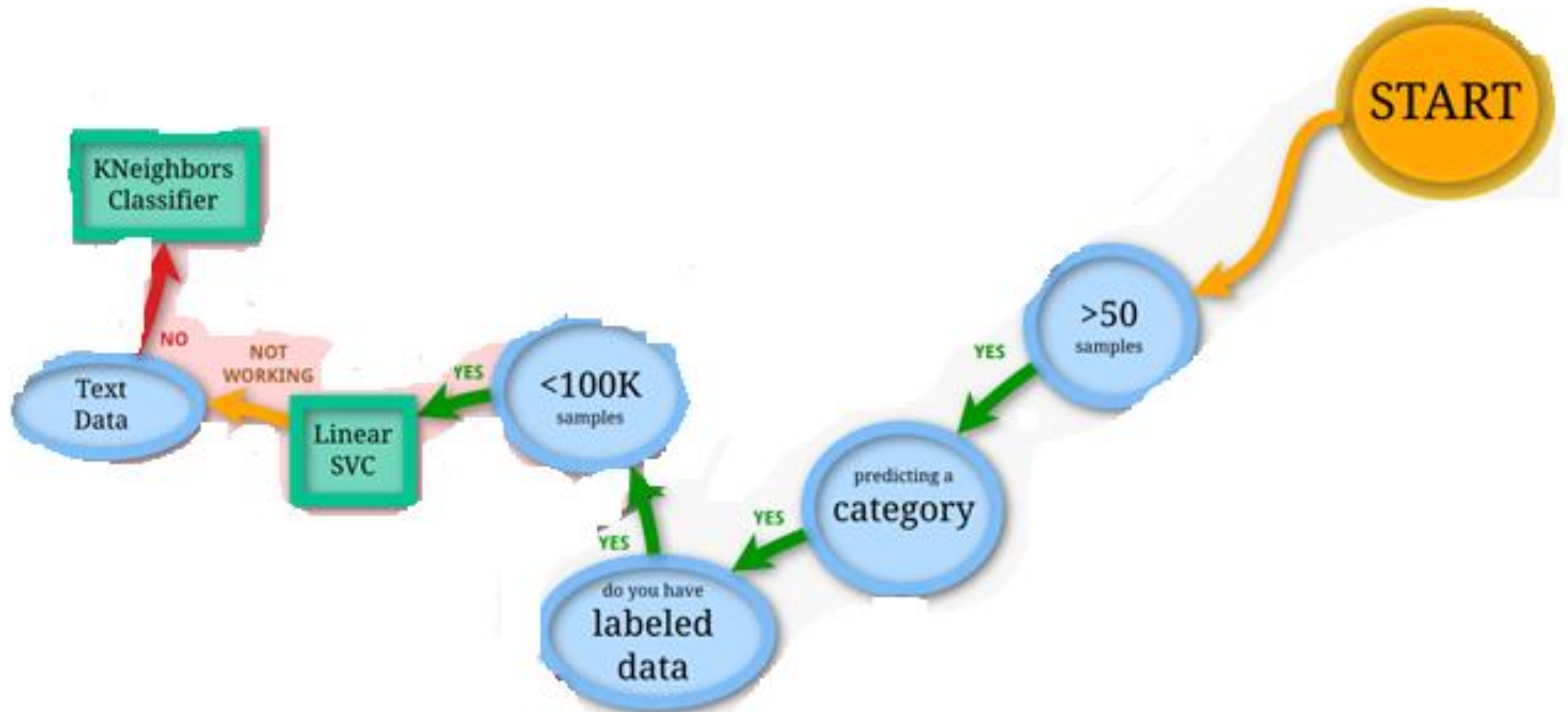
- Overfitting을 방지하기 위해서 랜덤포레스트와 그래디언트 부스팅 기법을 사용



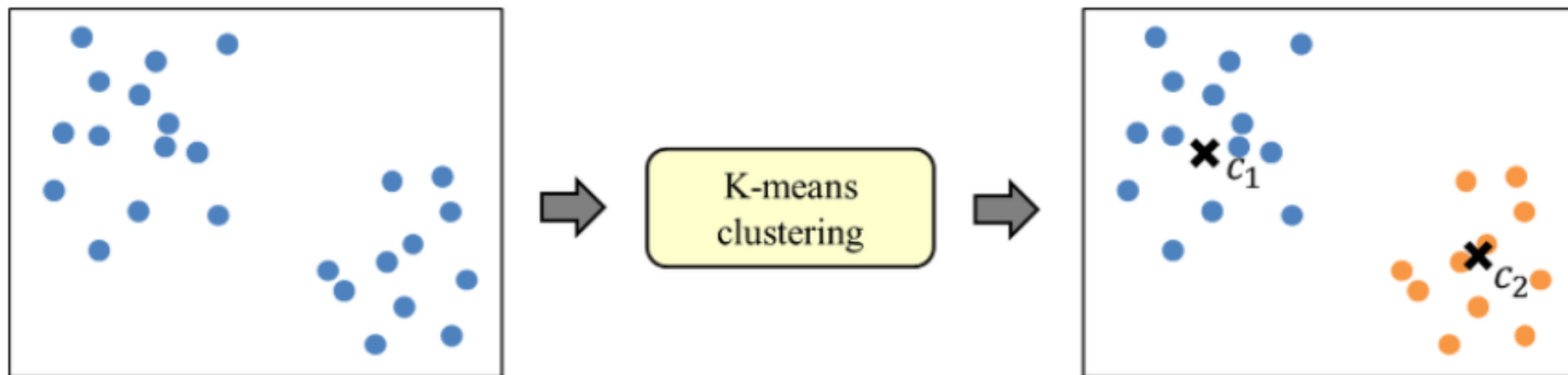
- 새로운 데이터를 입력 받았을 때 가장 가까이 있는 것이 무엇인냐를 중심으로 새로운 데이터의 종류를 정해주는 알고리즘(labeld data)



KNN 알고리즘 치트 시트(classification)

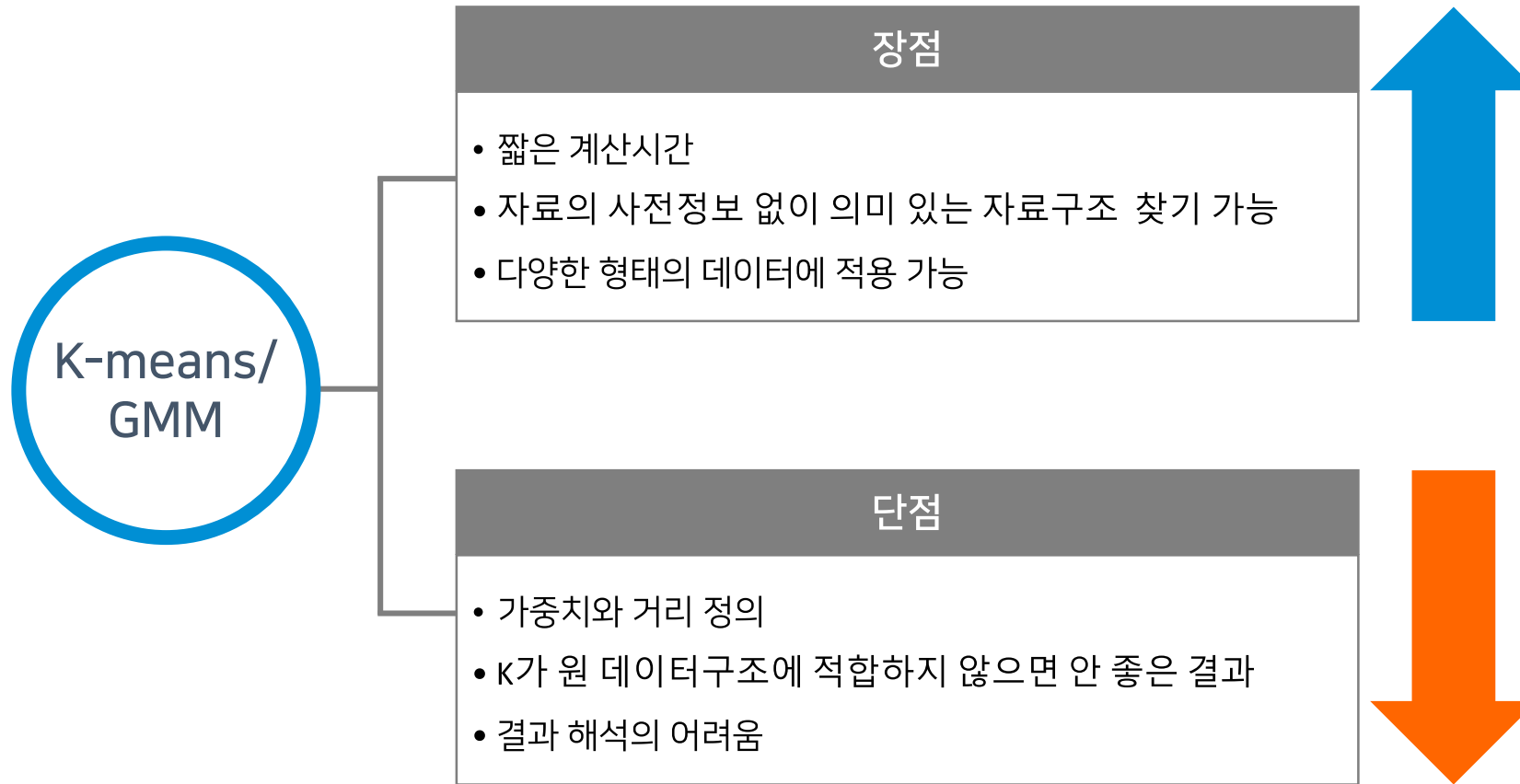


K-means/GMM 클러스터링

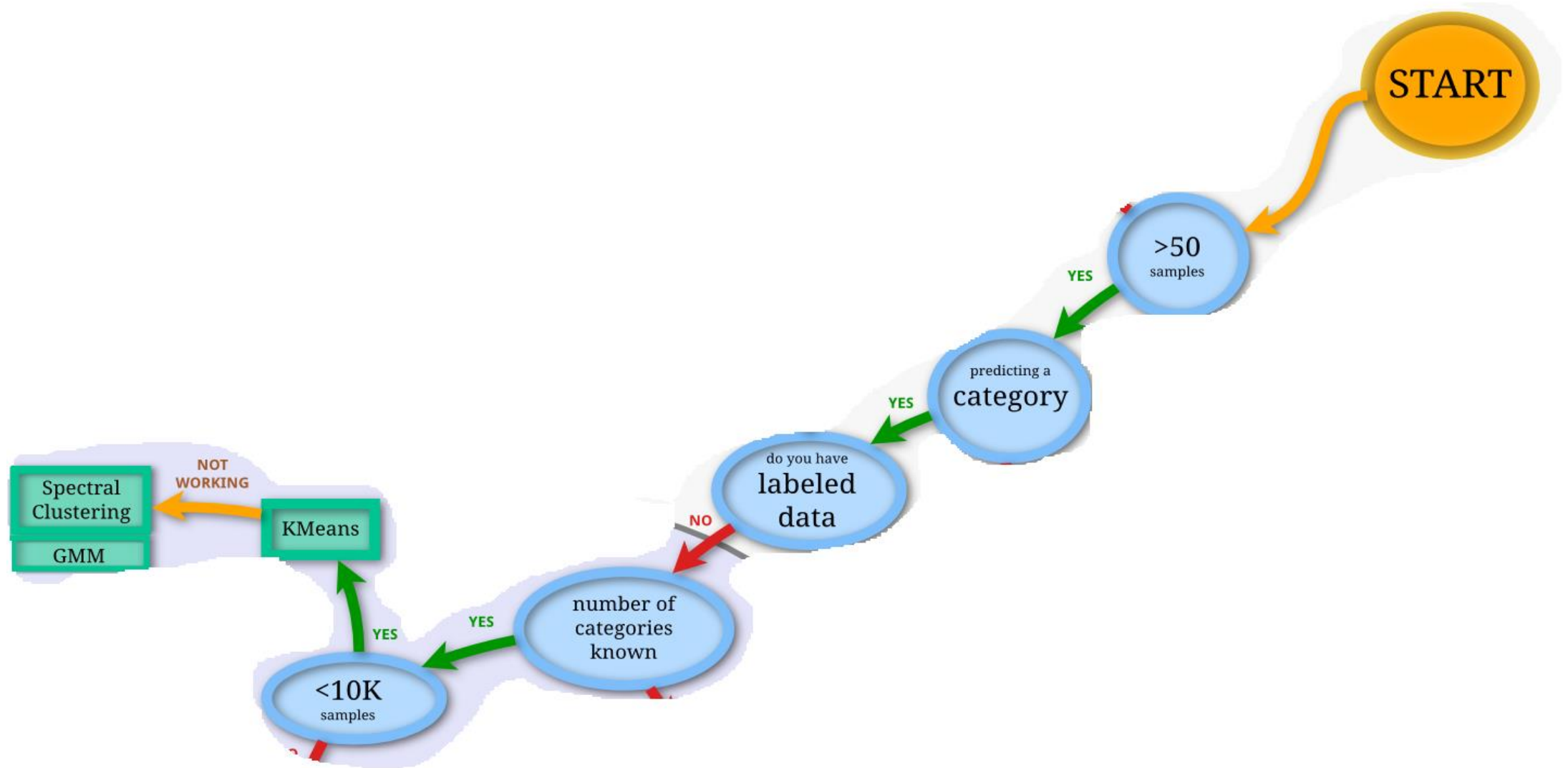


- K-평균/GMM 클러스터링의 목표는 n 개의 관측치를 k 개의 클러스터로 나누는 것입니다.
- 라벨이 안된 데이터를 분류하는데 사용
- K-평균은 표본을 하나의 클러스터에만 강하게 결속시키는 '하드 할당'
- GMM은 각 표본이 확률 값을 가짐으로써 어느 한 클러스터에만 결속되지 않는 '소프트 할당'

K-means/GMM 클러스터링 장단점



K-means/GMM 알고리즘 치트 시트(clustering)



SMS
spam

Iris

NOT WORKING 일 경우랑 아닐 경우
둘다 가정해서 풀기!

Working

Linear
SVC

KMeans

NOT WORKING

Naive
Bayes

Spectral
Clustering,
GMM

감사합니다.