

## 비지도 학습

:지도 학습과 달리 정답이 없는 데이터를 비슷한 부류끼리 군집화 하여 새로운 데이터에 대한 결과를 예측하는 방법

### ·군집

- 비슷한 샘플들을 클러스터로 모음
- 데이터 분석, 고객 분류, 추천 시스템 차원축소등에 사용하는 도구

### ·군집을 사용하는 다양한 어플리케이션

- 고객 분류
- 데이터 분석
- 차원 축소 기법
- 이상치 탐지
- 준지도 학습
- 검색 엔진
- 이미지 분할

### ·k- 평균

:주어진 데이터를 k개의 클러스터로 묶는 알고리즘, 각 클러스터와 거리 차이의 분산을 최소화 하는 방식으로 동작

### ·k- 평균단점

:군집의 크기가 서로 많이 다르면 잘 작동하지 않는다

하드 군집 - 각 샘플에 대해 가장 가까운 클러스터를 선택하는 방법  
소프트 군집 - 클러스터마다 샘플에 점수를 부여하고 샘플별로 각 군집 센트로이드와의 거리를 측정

## · k- 평균의 한계

- 알고리즘이 단순한 만큼 여러 가지 한계를 지니고 있다
- 최적이 아닌 솔루션을 피하려면 알고리즘을 여러 번 실행 해야한다

## · 이미지 분할

- 이미지를 세그먼트 여러 개로 분할 하는 작업
- 다양한 클러스터 개수로 k-평균을 사용해 만든 이미지 분할

## · 시맨틱 분할

- 이미지의 모든 픽셀에 레이블 또는 카테고리를 연관 짓는 작업
- 자율주행: 보행자들을 모두 하나의 영역 또는 각각의 영역으로 할당 가능

## · 색상 분할

- k-평균을 이용하여 색상분할 실행
- k-평균이 비슷한 크기의 클러스터를 만드는 경향 = 무당벌레를 하나의 클러스터로 만들지 못함

## · 레이블 전파

:그래프 기반의 데이터에서 사용되는 머신러닝 기법

- 레이블이 있는 몇몇 데이터 포인트로부터 시작하여 그래프의 다른 노드에 레이블을 전파하는 방식으로 작동

레이블 전파는 주로 커뮤니티 탐지, 분류, 클러스터링 등의 작업에서 활용

- 모든 상황에 적합한 것은 아니며 데이터의 특성과 문제에 따라 성능이 다를수 있다

## · DBSCAN

:특정 데이터를 중심으로 밀도가 높은 곳에 포함된 데이터에는 클러스터를 할당하고 밀도가 낮으면 그 데이터를 노이즈로 취급하는 알고리즘

- 장점:클러스터의 수를 정하지 않아도 되며 클러스터의 밀도에 따라 클러스터를 서로 연결하기 때문에 가하학적인 모양을 갖는 군집도 잘 찾을수 있다

-단점: 다른 밀도 분포를 가진 데이터의 군집 분석을 잘 하지 못한다

## · 가우시안 혼합

:샘플이 파라미터가 알려지지 않은 여러개의 혼합된 가우시안 분포에서 생성되었다고 가정하는 확률 모델

-가우시안 분포 = 정규분포

## · 클러스터

-하나의 가우시안 분포에서 생성된 모든 샘플들의 그룹

-일반적으로 타원형 모양이며 일반적으로 모양, 크기, 밀집도, 방향이 다르다\$