

## 0. 클러스터링? (Clustering)

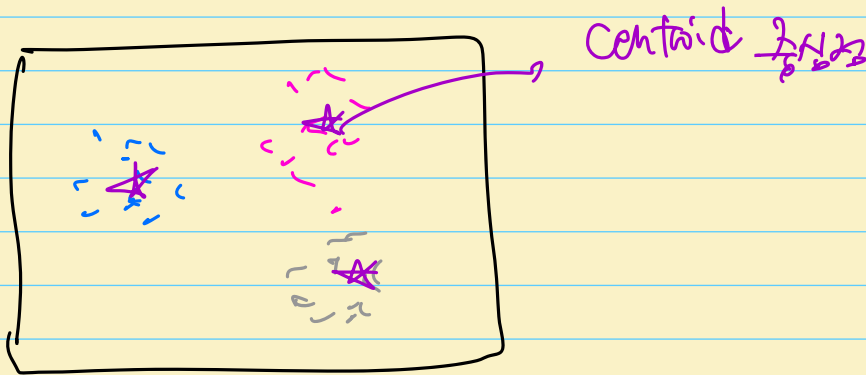
- 비머트 클러스터링: 대용량인 기술타입에 대한 레이블이 지정되어있지 않은 레이블을 그룹화하는 분석 알고리즘
- 레이터나 특성은 고차원 데이터 집합(클러스터)은 광범위, 데이터 집합에 대한 주어진 중심점을 갖는다

## 1. K-평균 (K-Means)

- 중심점 (Centroid)라는 특정한 일의 지점을 선택해 해당 중심에 가장 가까운 데이터 선택
  - 선택된 포인트의 중심점과 이동한 중심점에서 다시 가장 가까운 선택
- 반복

반복) 가장 많이 사용

장점) 속도가 빠르다 ↑ 정확도 ↓ (PCA와 연관성 있음)  
단점) 클러스터링이 빠르다 ↑ 속도가 느리다 ↓  
특히 클러스터링이 빠르다 ↑ 속도가 느리다 ↓



## 2. 평가지표 (평가기준)

- 조정된 랜드 인덱스 (Adjusted Rand Index)
- 조정된 상호정보량 (Adjusted Mutual Information)
- 실루엣 계수 (Silhouette Coefficient)

### 3. Mean Shift (Mean Shift)

- K-평균과 유사하나, 개체가 아닌 '평균'을 중심으로 중심을 이동
- bandwidth  $< 0$  이자, 평상에서 특정 개체 주변, 중심 이동
- 컴퓨터 비전 분야 사용

상황) 특정 개체 개수  $X$ , 특정 분포로 개체 분포를 가정  $X \rightarrow$  유한한 개체 수 가능

이상의 평상  $\downarrow$   
 분포를 따르  $X$

상황)  $\uparrow$   
 bandwidth 크기 다른 평상으로 맞추기

### 4. GMM (Gaussian Mixture Model)

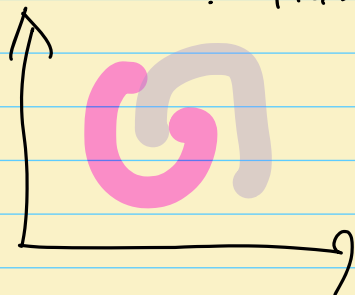
- 데이터가 여러 개의 가우시안 분포를 가지는 데이터에서 생성된 데이터를 가정할 수 있음
- ↓  
 가우시안 분포

→ 여러 개의 데이터 분포에서 가우시안 분포를 추출

→ 학습 과정을 통해 EM (Expectation and Maximization) 방법 사용

### 5. DBSCAN

- Density Based Spatial Clustering of Applications with Noise
- 밀도 기반 클러스터링
- 노이즈를 포함하여, 개체 분포를 특정 개체로 분류 가능



→ 이진 평상으로 가능

ex)

K-means

DBSCAN

Hierarchical Clustering

Affinity Propagation

Spectral Clustering

가을크기  
하늘

가을크기  
하늘

가을크기  
하늘

노스랑

노스랑

노스랑

