

[0209]_Clustering_summary_9th_SunjaeYoo

📅 날짜	@2023년 2월 9일
➤ 과목	
📁 구분	DSL
☑ 복습 1	<input type="checkbox"/>
☑ 복습 2	<input type="checkbox"/>
☑ 복습 3	<input type="checkbox"/>
📝 첨부	
📌 학기	etc

1 Clustering

레이블이 없는 데이터 집합을 유사한 데이터들의 그룹으로 나누는 것.

2 종류

2.1 K-means Clustering

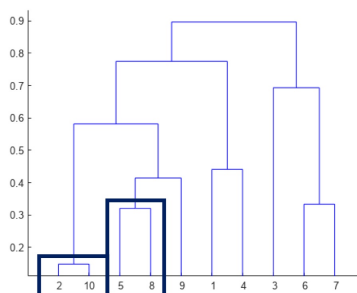
k개의 클러스터를 중심으로 묶는 알고리즘

- 1) k개의 중심점 설정
- 2) 중심점 업데이트

→ 각 객체와 그룹 / 그룹의 중심 간 유클리디안 거리 합이 최소가 되는 방향으로 군집화 !

2.2 Hierarchical Clustering

- 계층적 트리 모양을 이용해 개별 객체들을 유사한 객체와 통합하는 알고리즘
- 덴드로그램을 통해 시각화 가능



- 덴드로그램의 높이는 객체 간의 거리를 의미한다.

덴드로그램 : 개체들이 결합되는 순서를 나타내는 트리 형태의 구조

덴드로그램을 적절한 수준에서 자르면 군집화 결과 생성

1) 거리에 대한 유사도 행렬 계산

거리 계산 방법은 여러가지가 있음

- Euclidean (최단연결법, 최장연결법, 평균연결법, 중심연결법)
- Ward's Linkage

2) 거리가 인접한 관측치끼리 군집화

3) 유사도 행렬 업데이트

2.3 DBSCAN

점이 세밀하게 몰려 있어 밀도가 높은 부분을 클러스터링.

한 점을 기준으로 반경 x 내에 점이 n 개 이상 있으면 하나의 군집으로 인식.

- core point끼리도 같은 군집이 될 수 있음