

Part.05

Clustering

| Hierarchical clustering

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

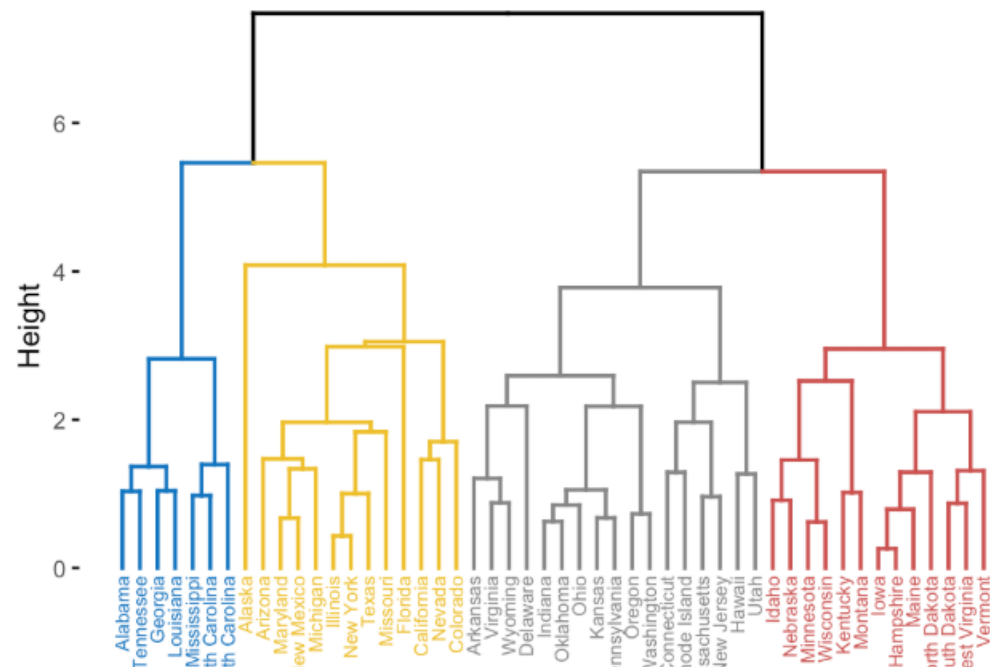
강사. 이경택

I Hierarchical clustering 소개

■ Hierarchical clustering

- 개체들을 가까운 집단부터 순차적/계층적으로 차근차근 묶어 나가는 방식
- 유사한 개체들이 결합되는 dendrogram 을 통해 시각화 가능
- 사전에 군집의 개수를 정하지 않아도 수행가능

Cluster Dendrogram

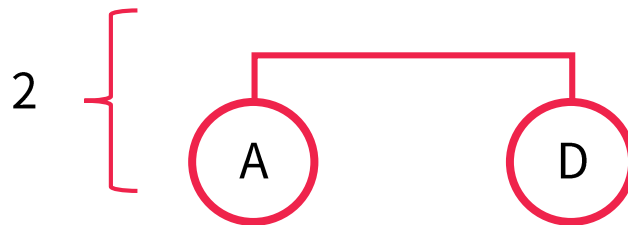


I Hierarchical clustering 소개

■ Hierarchical clustering

- 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 거리가 인접한 관측치끼리 cluster 형성
- 유사도 행렬 update

| | A | B | C | D |
|---|---|----|----|----|
| A | | 20 | 7 | 2 |
| B | | | 10 | 25 |
| C | | | | 3 |
| D | | | | |

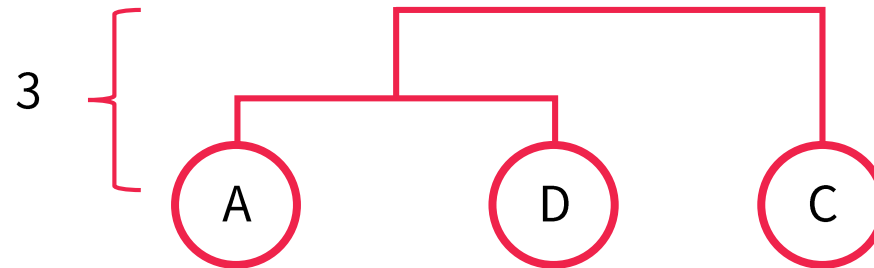


I Hierarchical clustering 소개

■ Hierarchical clustering

- 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 거리가 인접한 관측치끼리 cluster 형성
- 유사도 행렬 update

| | AD | B | C | |
|----|----|----|----|--|
| AD | | 20 | 3 | |
| B | | | 10 | |
| C | | | | |
| D | | | | |

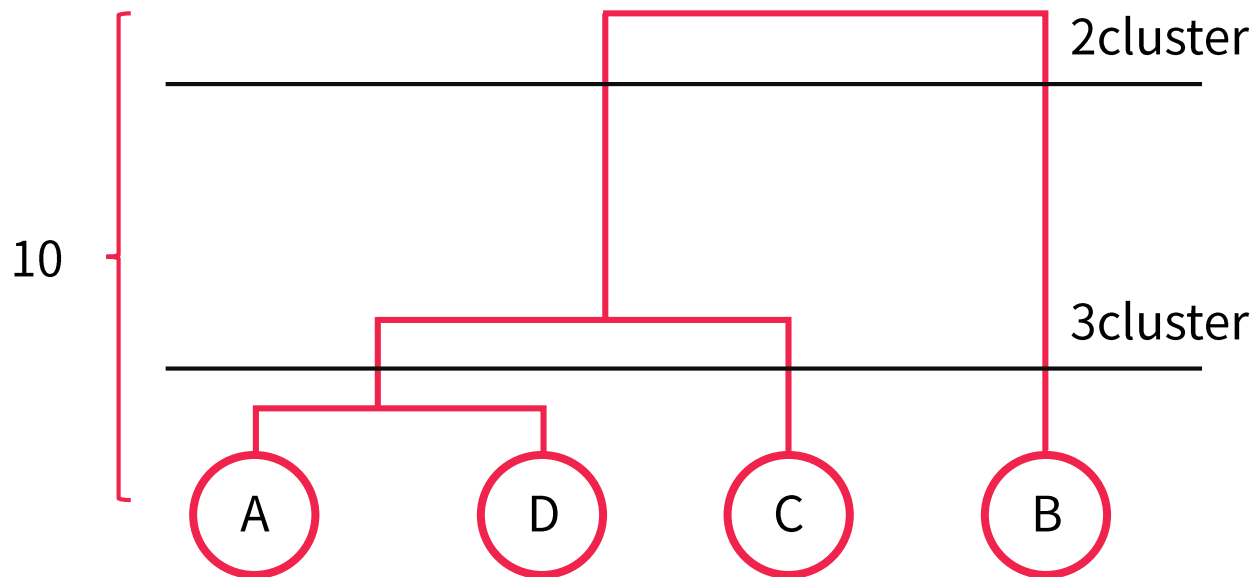


I Hierarchical clustering 소개

■ Hierarchical clustering

- 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 거리가 인접한 관측치끼리 cluster 형성
- 유사도 행렬 update

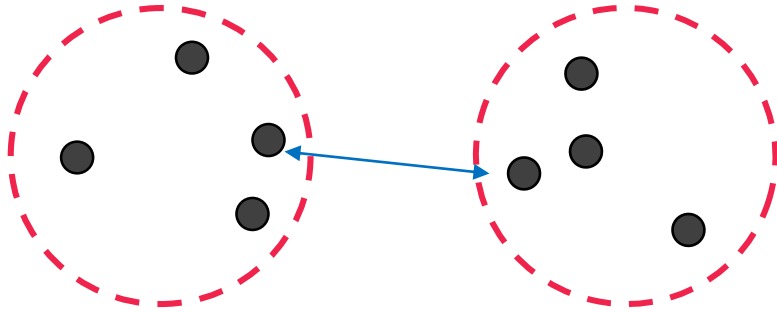
| | ADC | B | | |
|-----|-----|----|--|--|
| ADC | | 10 | | |
| B | | | | |
| | | | | |
| | | | | |



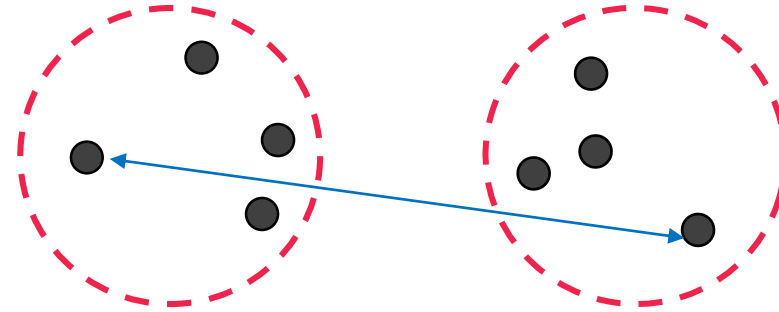
I Hierarchical clustering 소개

- Clustering간의 거리

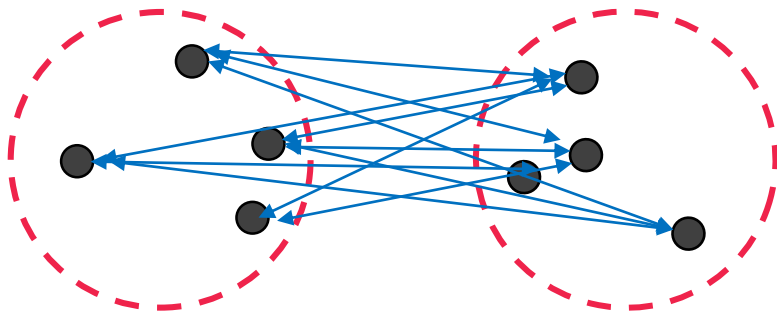
Min (Single Link)



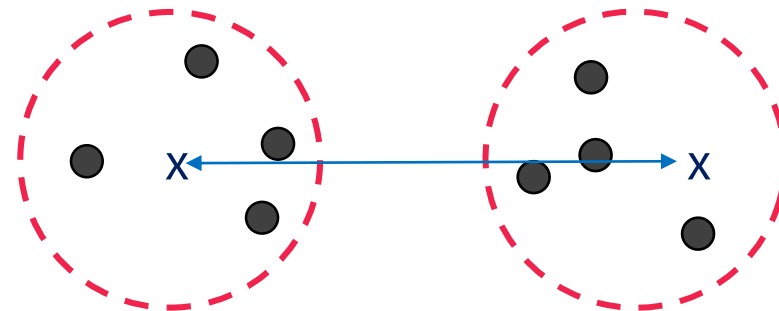
Max(Complete Link)



Average Link



Centroids



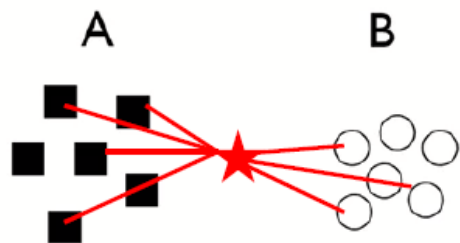
I Hierarchical clustering 소개

■ Clustering간의 거리

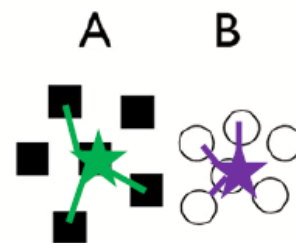
Ward's method : 두개의 클러스터가 병합 되었을 때 증가되는 변동성의 양

$$\text{Ward Distance} = \sum_{i \in A \cup B} \|x_i - m_{A \cup B}\|^2 - \left\{ \sum_{i \in A} \|x_i - m_A\|^2 + \sum_{i \in B} \|x_i - m_B\|^2 \right\}$$

m_A is the center of cluster A.



$$\text{Ward's distance} = 10 - (3 + 2) = 5$$



$$\text{Ward's distance} = 7 - (3 + 2) = 2$$

Part.05

Clustering

| DBSCAN clustering

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택