

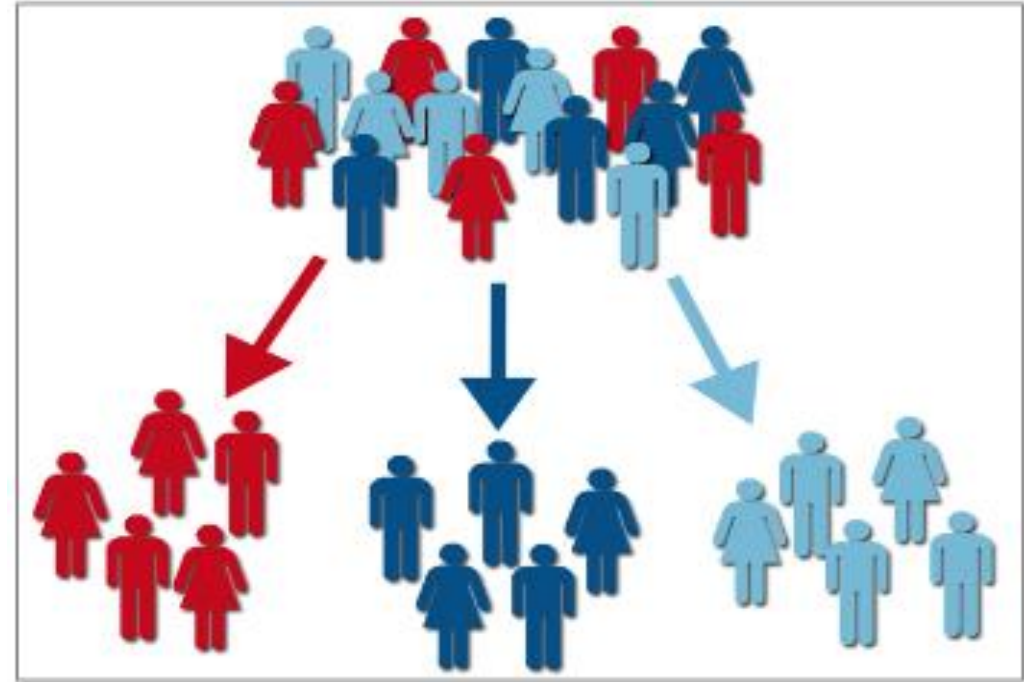
군집분석:

군집분석 소개와 두 가지 유형

숙명여자대학교 경영학부 오중산

군집분석 소개

- 군집분석(cluster analysis)이란?
 - ◆ IV를 토대로 사례를 몇 개 군집(cluster/class)으로 분류하는 통계 분석방법
 - ◆ Grouping n cases into k classes ($n > k$)



(출처) <https://brunch.co.kr/@gimmesilver/32>

군집분석 소개

- 군집분석 특징

- ◆ IV의 척도는 계량/비계량척도 모두 가능
- ◆ 모집단이 없고, DV도 없는 탐색적 분석방법
 - ML에서 비지도(unsupervised) 학습으로 분류됨

- 군집의 특성

- ◆ 동일 군집 사례는 서로 유사성이 크고, 다른 군집 간에는 이질성이 큼(類類相從)
- ◆ 예시: customer segmentation / product segmentation

군집분석 소개

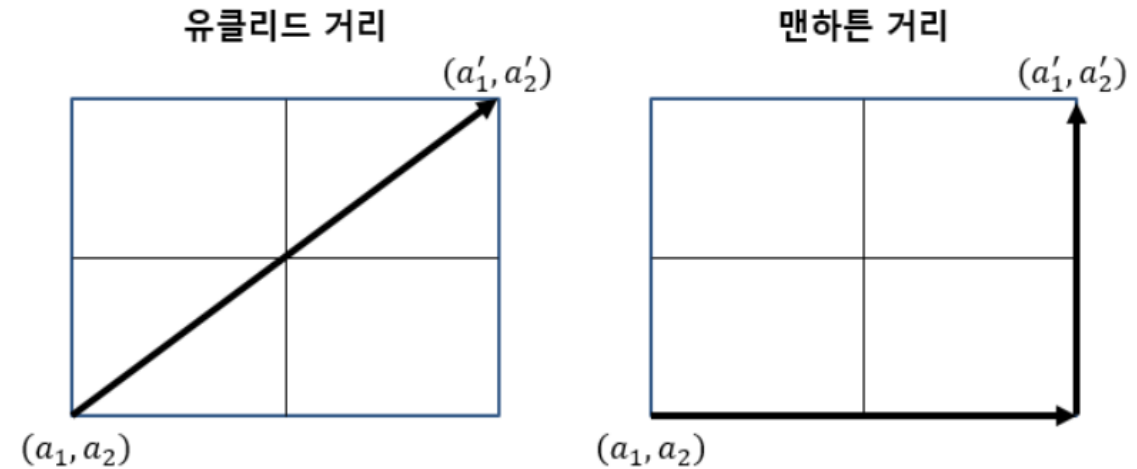
- 군집분석에서의 거리(distance)

- ◆ 서로 다른 사례간의 유사성과 이질성은
IV를 기준으로 사례 간의 거리로 판단

- 두 사례 간에 거리가 가까울수록 유사함

- ◆ cluster 패키지의 daisy함수

- 계량척도 IV: 사례 간에 Manhattan distance 측정
- 비계량척도 IV: 사례 간에 dice distance 측정



$$d = \sqrt{(a'_1 - a_1)^2 + (a'_2 - a_2)^2}$$

$$d = (a'_1 - a_1) + (a'_2 - a_2)$$

(출처) <https://m.blog.naver.com/samsjang/220980080522>

군집분석 유형

- 군집분석 유형1: 계층적 군집분석(hierarchical clustering)

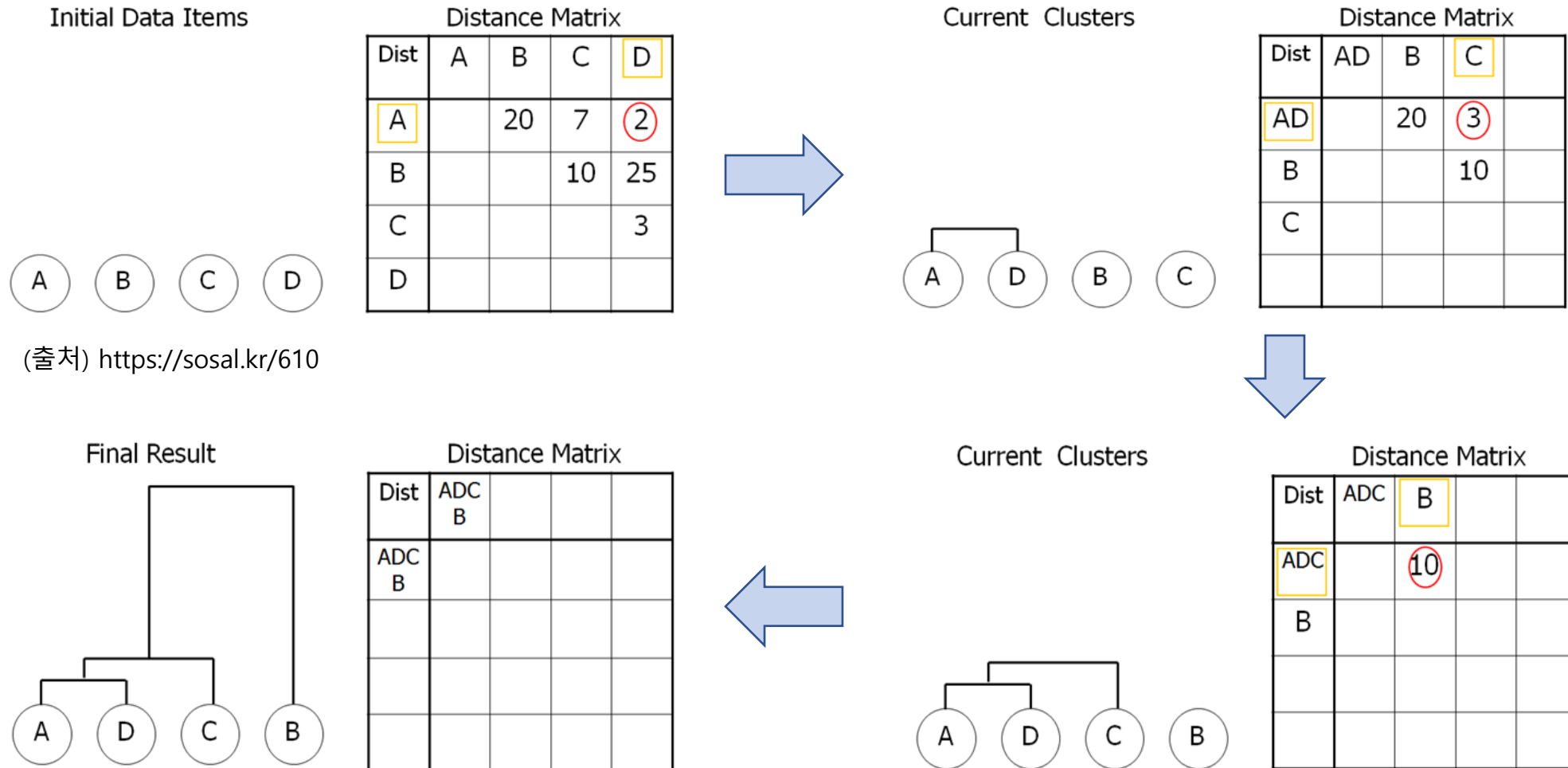
- ◆ 사전에 군집 개수(k)를 정하지 않은 군집분석

- 처음에는 n (사례 개수) = k (군집 개수)
- 가장 가까운 사례끼리 순차적으로 군집으로 묶으면서 ‘꼬리잡기’ 놀이처럼 점차 k 가 줄어듦
- 여러 단계를 거치며 최종적으로 모든 사례가 하나의 군집으로 묶이게 됨

- ◆ 데이터 크기가 비교적 적을 경우에 적합함

군집분석 유형

- 군집분석 유형1: 계층적 군집분석(hierarchical clustering)



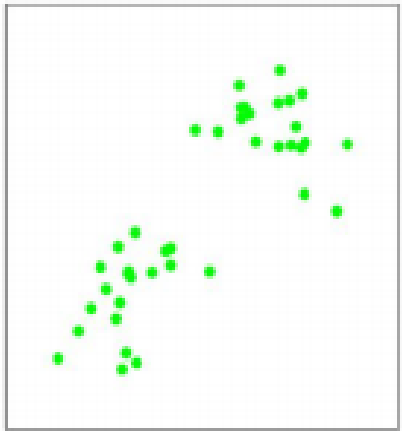
군집분석 유형

- 군집분석 유형2: K-평균 군집분석(K-means clustering)
 - ◆ 사전에 군집 개수(k)를 정해 놓은 군집분석
 - 실제 분석과정에서는 k 를 미리 정하지 않고, 최적의 k 를 찾기도 함
 - ◆ 데이터 크기가 비교적 큰 경우에 적합하지만, 이상치로 인해 왜곡될 수 있음
 - ◆ 비계량척도로 측정한 IV는 사용할 수 없는 한계가 있음

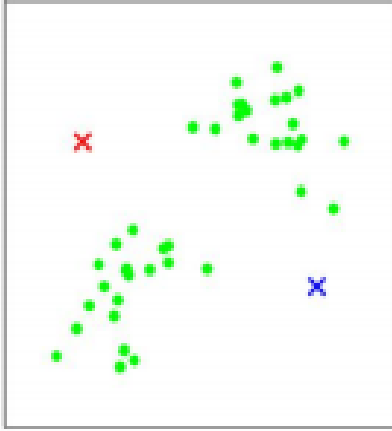
군집분석 유형

- 군집분석 유형2: K-평균 군집분석(K-means clustering)

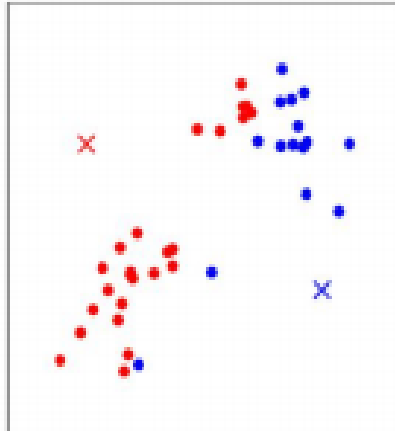
◆ $k = 2$ 예시: 무게중심(centroid)을 기준으로 군집에 변화가 없을 때까지 반복



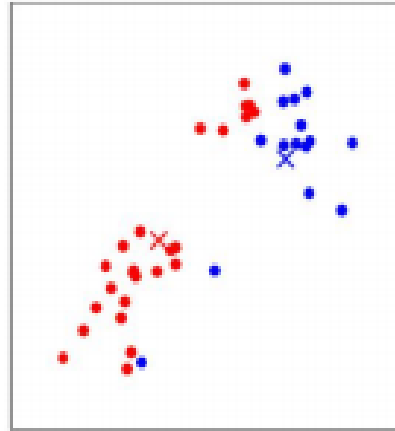
(a)



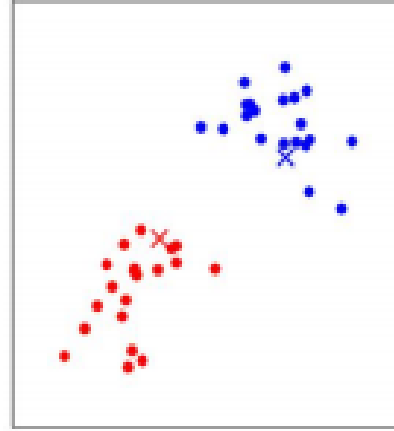
(b)



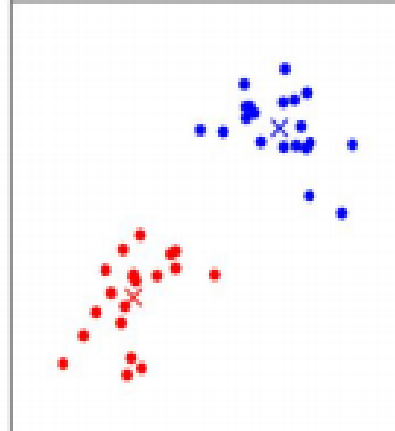
(c)



(d)



(e)



(f)