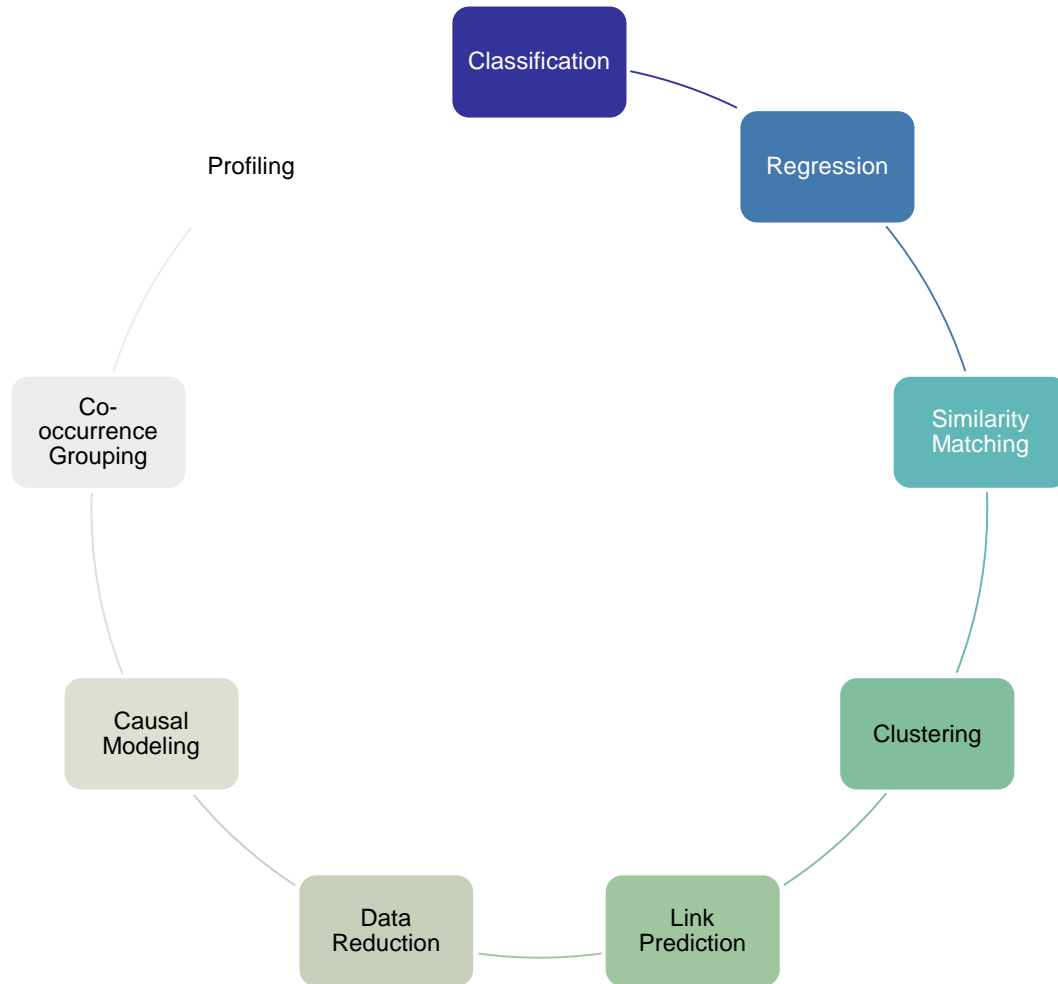




Lecture 2

k평균 군집분석

Analytics Tasks



類類相從

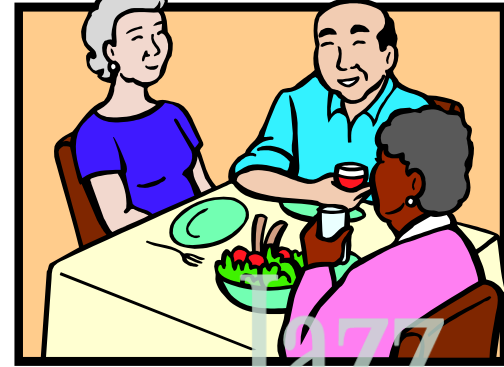
Pop Music

Hip Hop



Electronic

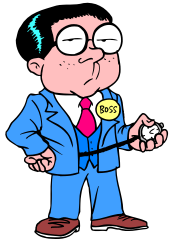
Classical



Jazz



類似渡



철수:
나이=35세
연소득=\$95000



영희:
나이=39세
연소득=\$51000

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

❖ Distance (철수, 영희) = $\text{sqrt} [(35-39)^2 + (95000-51000)^2]$



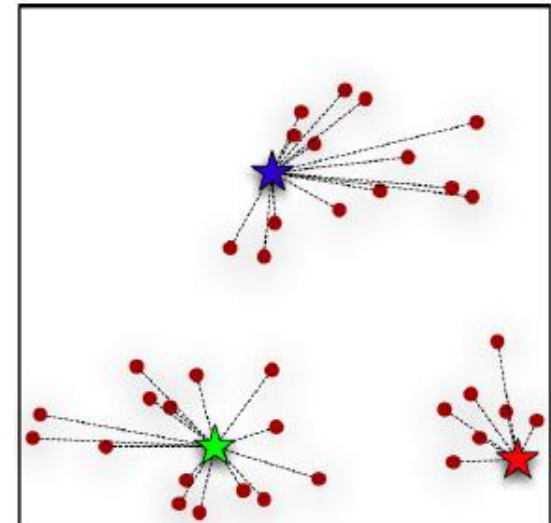
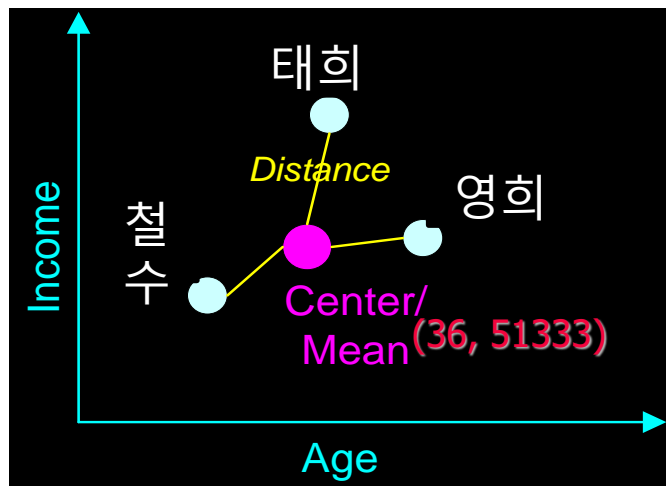
군집중심점

❖ 이름: 나이, 연소득

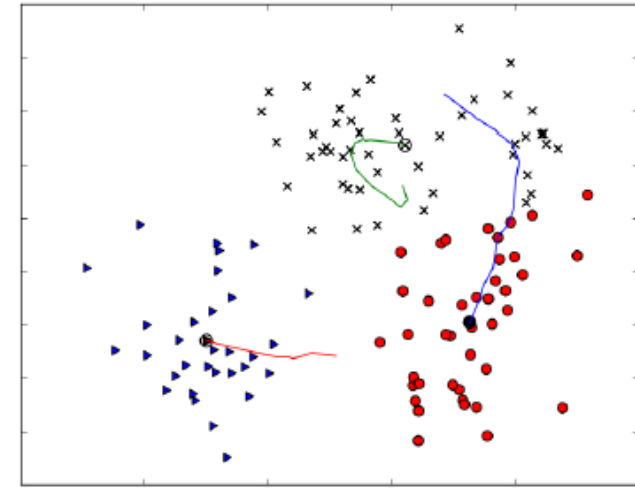
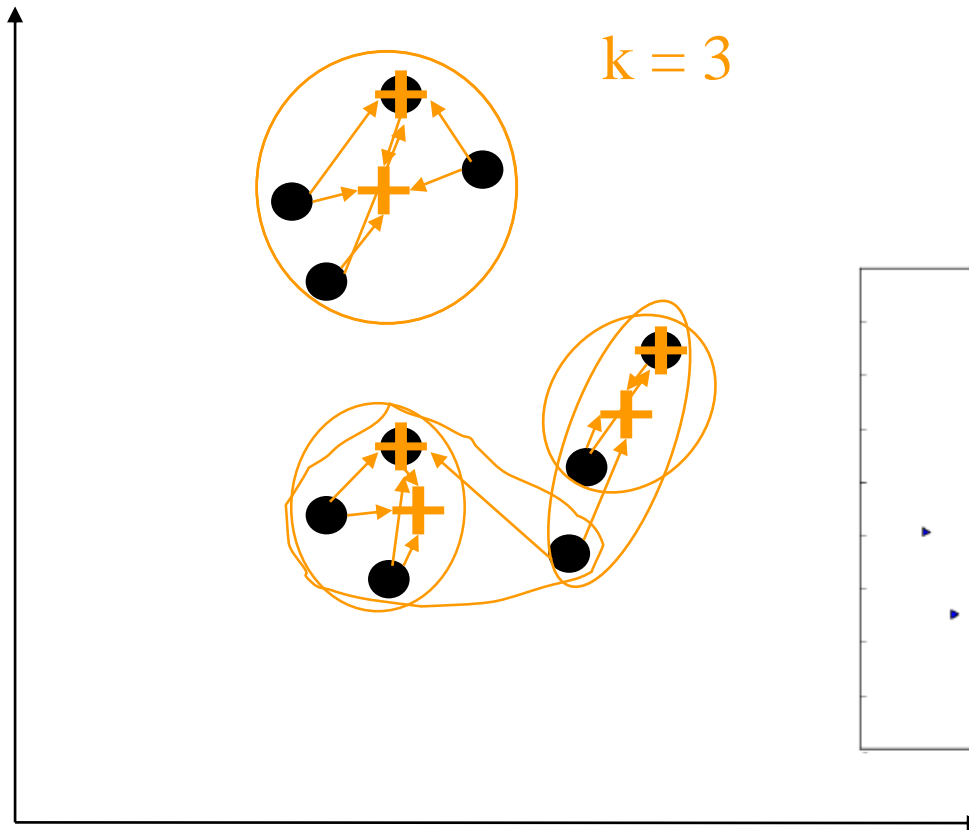
- 철수: 33, \$48000
- 태희: 36, \$55000
- 영희: 39, \$51000

❖ 군집중심점

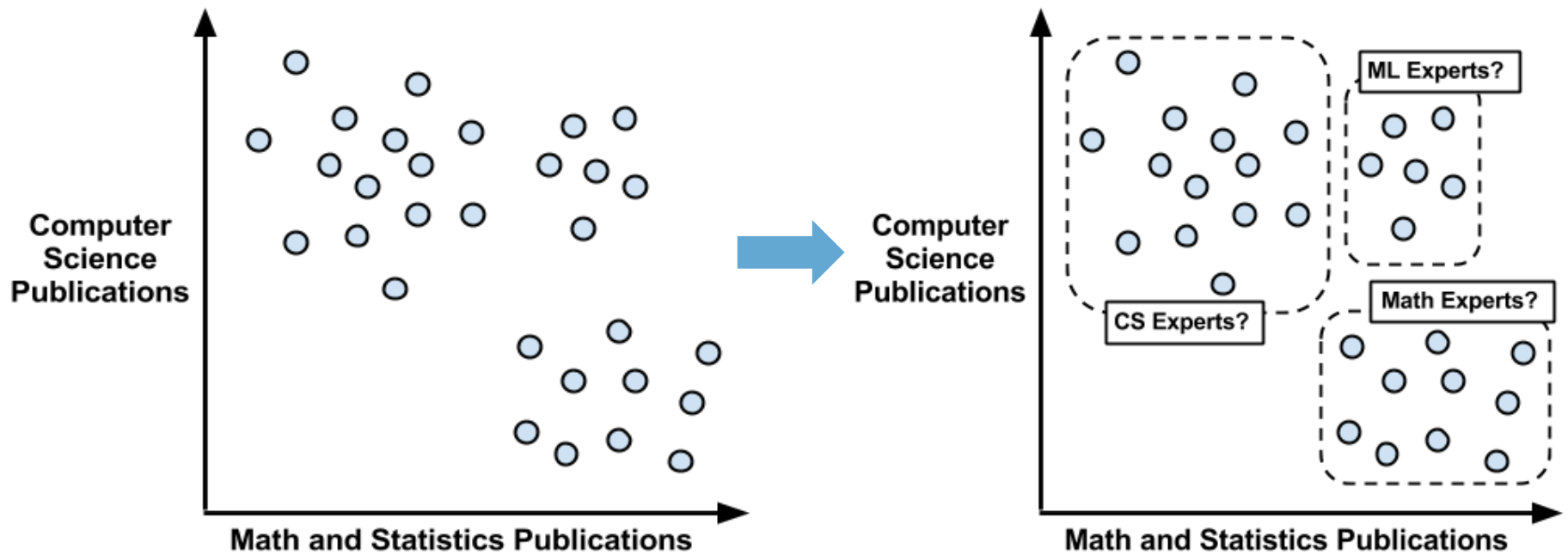
- 나이 = $(33 + 36 + 39) / 3 = 36$
- 연소득 = $(48000 + 55000 + 51000) / 3 = \51333



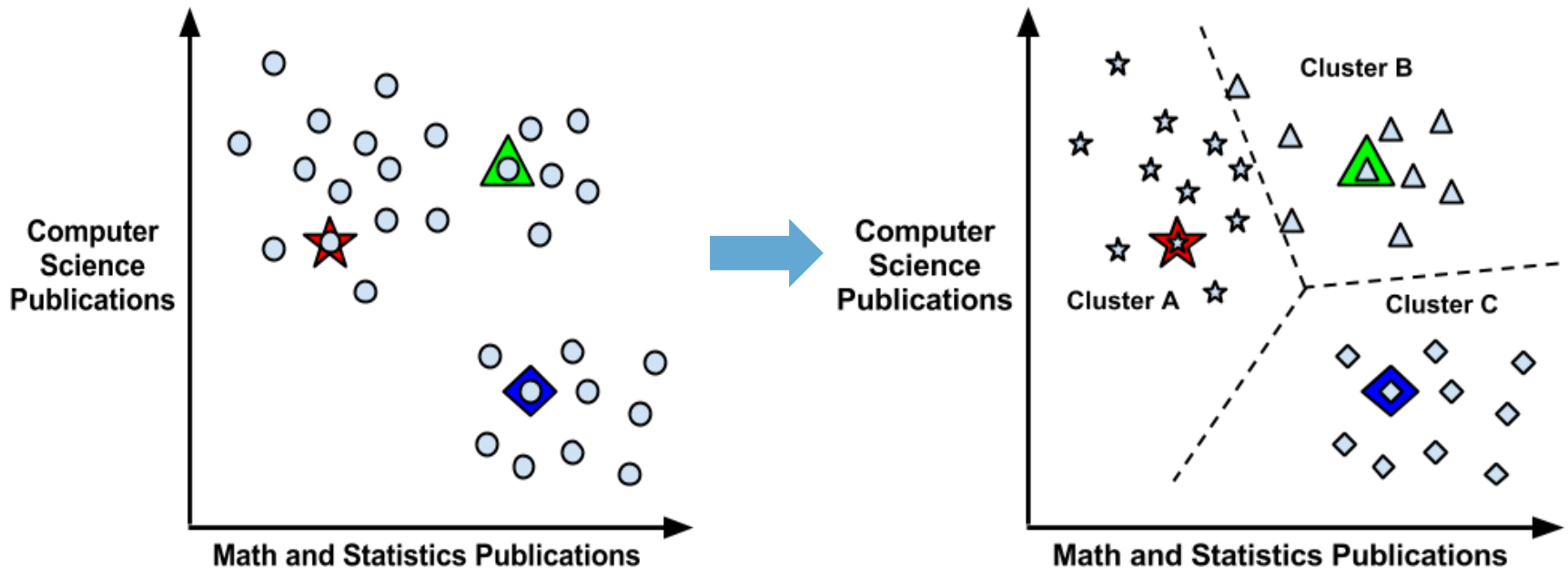
K평균군집분석의 원리



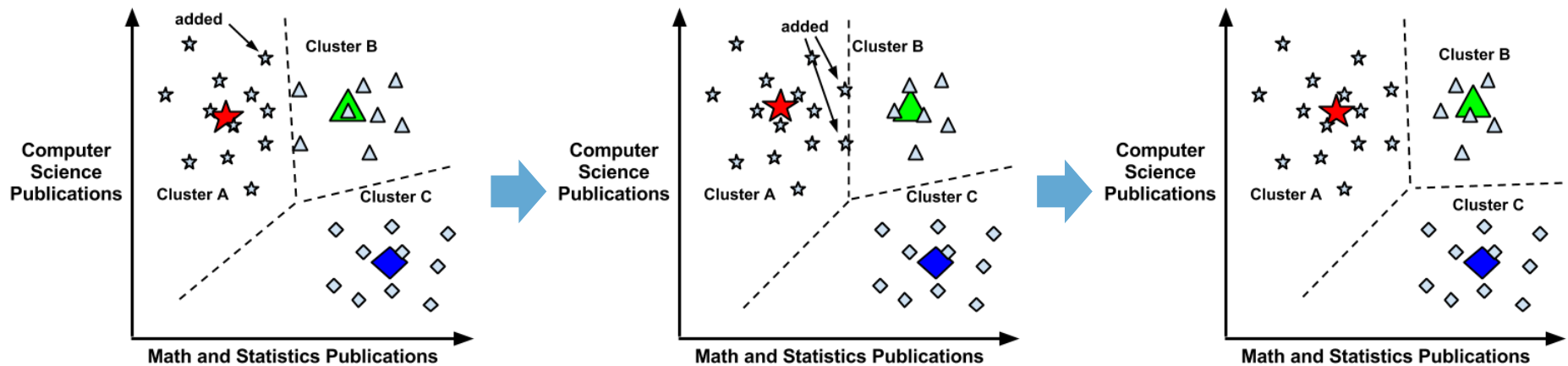
K평균군집분석의 사례



K평균군집분석의 사례

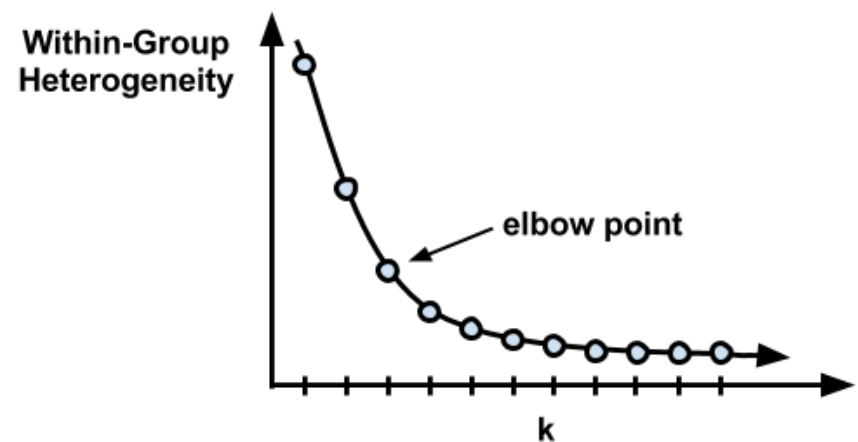
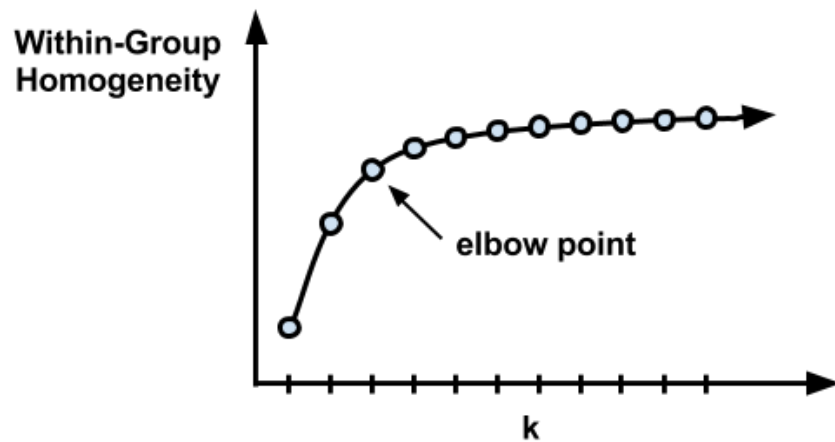


K평균군집분석의 사례



K평균군집분석의 과정

❖ Elbow method



군집분석의 활용

❖ 장단점

● 장점

- 사전에 그룹분류에 대한 정보가 없는 데이터를 가지고 사용자가 추구하는 바에 맞게 그룹을 나눌 수 있음
- 군집화 과정이 비교적 이해하기 용이함

● 단점

- K평균군집분석의 경우 사용자가 사전 지식 없이 그룹의 수를 정해주어야 하기 때문에 결과가 잘 나오지 않거나, 분석 결과에 대한 해석이 어려워질 수 있음
- 초기 군집중심점에 따라 결과가 좌우될 수 있음

