Part. 05

Clustering

# | K-means clustering

FASTCAMPUS

ONLINE

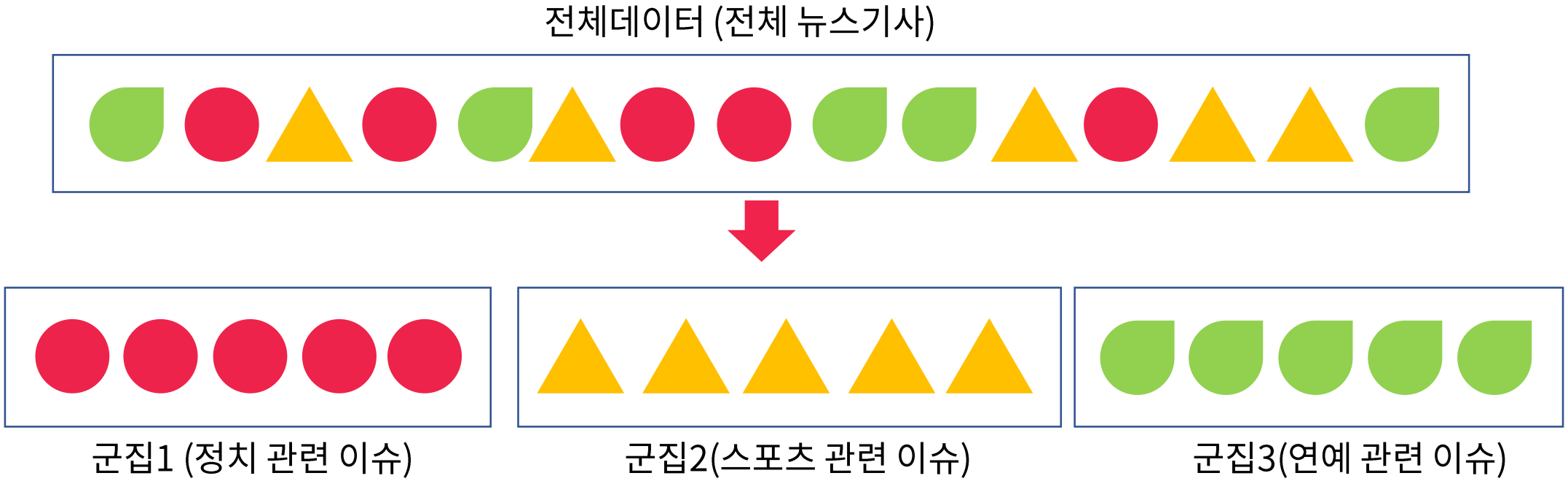
머신러닝과 데이터분석 A-Z|

강사. 이경택

# I K-means clustering



- Clustering(군집분석)이란
  - 각 데이터의 유사성을 측정하여 높은 대상 집단을 분류하고, 군집 간에 상이성을 규명하는 방법




  - 고객 segmentation을 통한 마케팅 활용 방안 / 군집 별 추가 분석수행

# I K-means clustering



- Clustering의 종류
  - K-means clustering : 데이터를 사용자가 지정한 k개의 군집으로 나눔
  - Hierarchical clustering (계층적 군집분석) : 나무 모양의 계층 구조를 형성해나가는 방법
  - DBSCAN : k개를 설정할 필요없이 군집화 할 수 있는 방법

# I K-means clustering



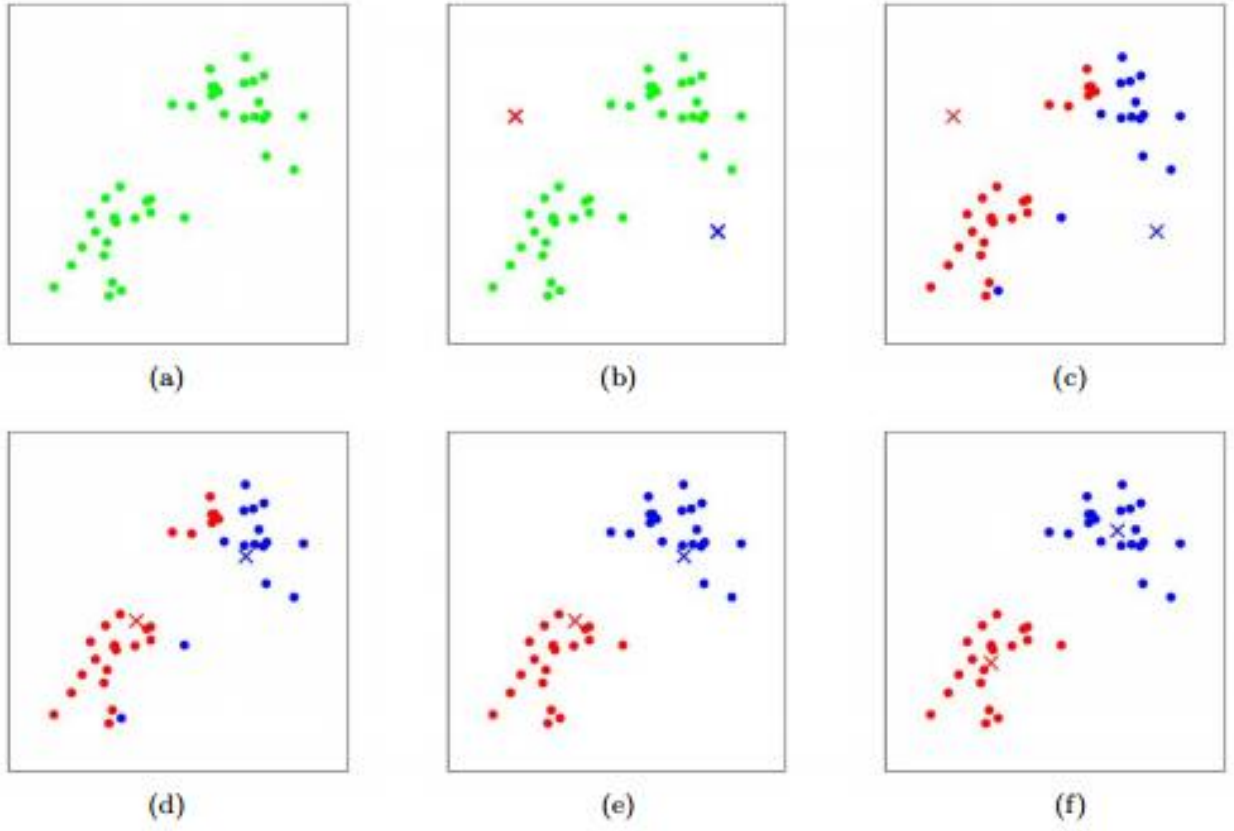
- K-means clustering
  - 각 군집에 할당된 포인트들의 평균 좌표를 이용해 중심점을 반복적으로 업데이트

  - Step1 – 각 데이터 포인트 i에 대해 가장 가까운 중심점을 찾고, 그 중심점에 해당하는 군집 할당
  - Step2 – 할당된 군집을 기반으로 새로운 중심 계산, 중심점은 군집 내부 점들 좌표의 평균(mean) 으로 함
  - Step3 – 각 클러스터의 할당이 바뀌지 않을 때까지 반복

# I K-means clustering



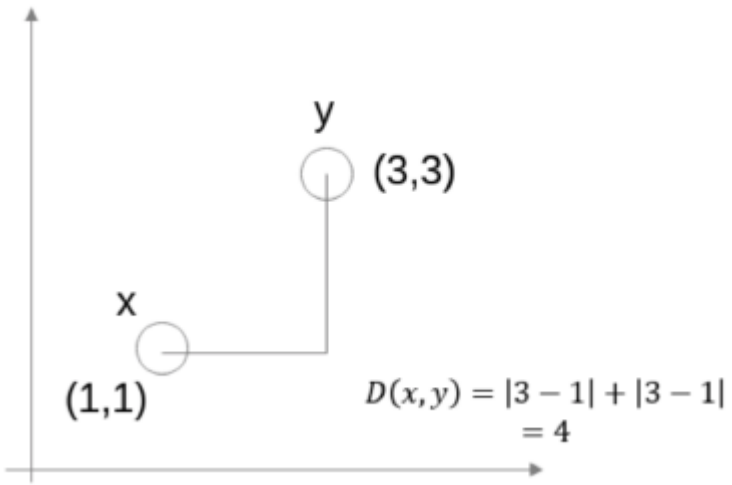
- K-means clustering
  - Step1 – 각 데이터 포인트 i에 대해 가장 가까운 중심점을 찾고, 그 중심점에 해당하는 군집 할당
  - Step2 – 할당된 군집을 기반으로 새로운 중심 계산, 중심점은 군집 내부 점들 좌표의 평균(mean) 으로 함
  - Step3 – 각 클러스터의 할당이 바뀌지 않을 때 까지 반복



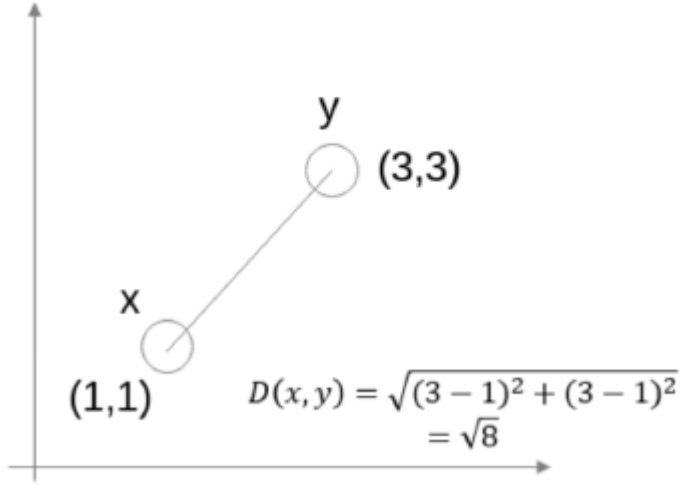
# I K-means clustering



- 점과 점사이의 거리 측정
  - Manhattan distance – 각 축에 대해 수직으로만 이동하여 계산하는 거리 측정 방식
  - Euclidean distance – 점과 점 사이의 가장 짧은 거리를 계산하는 거리 측정 방식




Manhattan distance




Euclidean distance

Part. 05

Clustering

# | 최적의 k를 찾는 방법

FASTCAMPUS
ONLINE
머신러닝과 데이터분석 A-Z|

강사. 이경택