

17기 정규세션

TOBIG's 16기 박한나

# Clustering

군집화

# CONTENTS

---

Unit 01 | Clustering

---

Unit 02 | Hierarchical Clustering

---

Unit 03 | K-Means Clustering

---

Unit 04 | DBSCAN

---

Unit 05 | 최적의 군집 수 및 모델 평가

---

# **UNIT 01 | CLUSTERING**

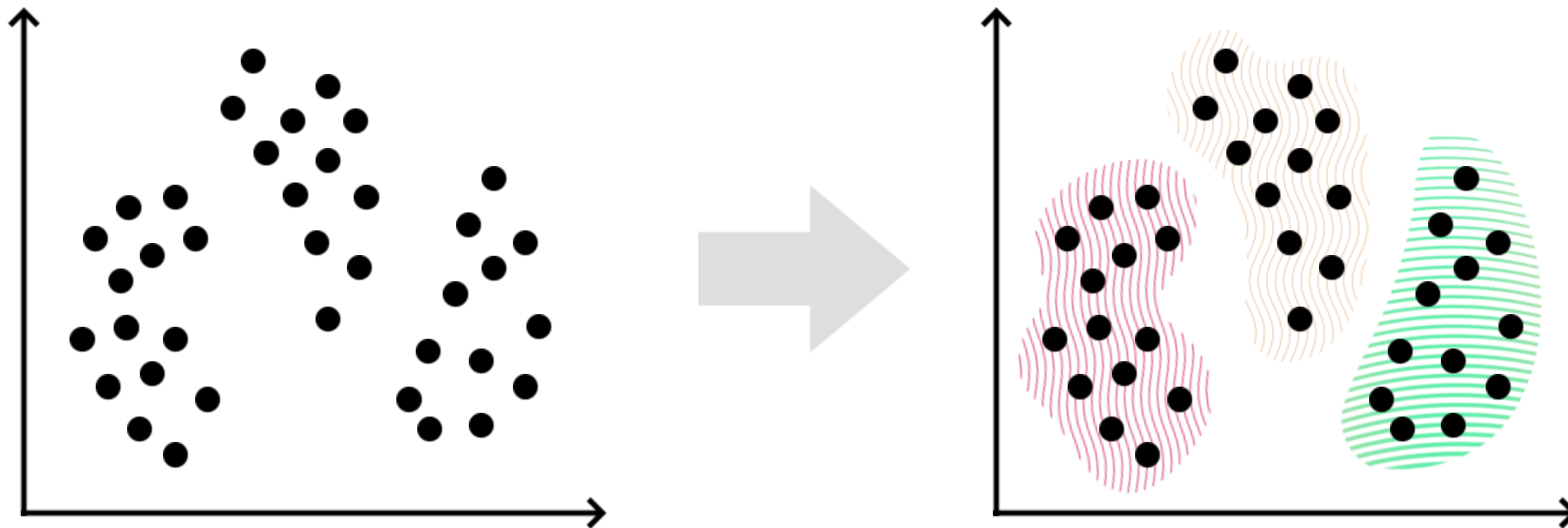
# UNIT 01 | CLUSTERING

## CLUSTERING

군집화

### CLUSTERING 개념

- ✓ 유사한 속성들을 갖는 관측치들을 묶어 데이터를 몇 개의 군집(그룹)으로 나누는 것



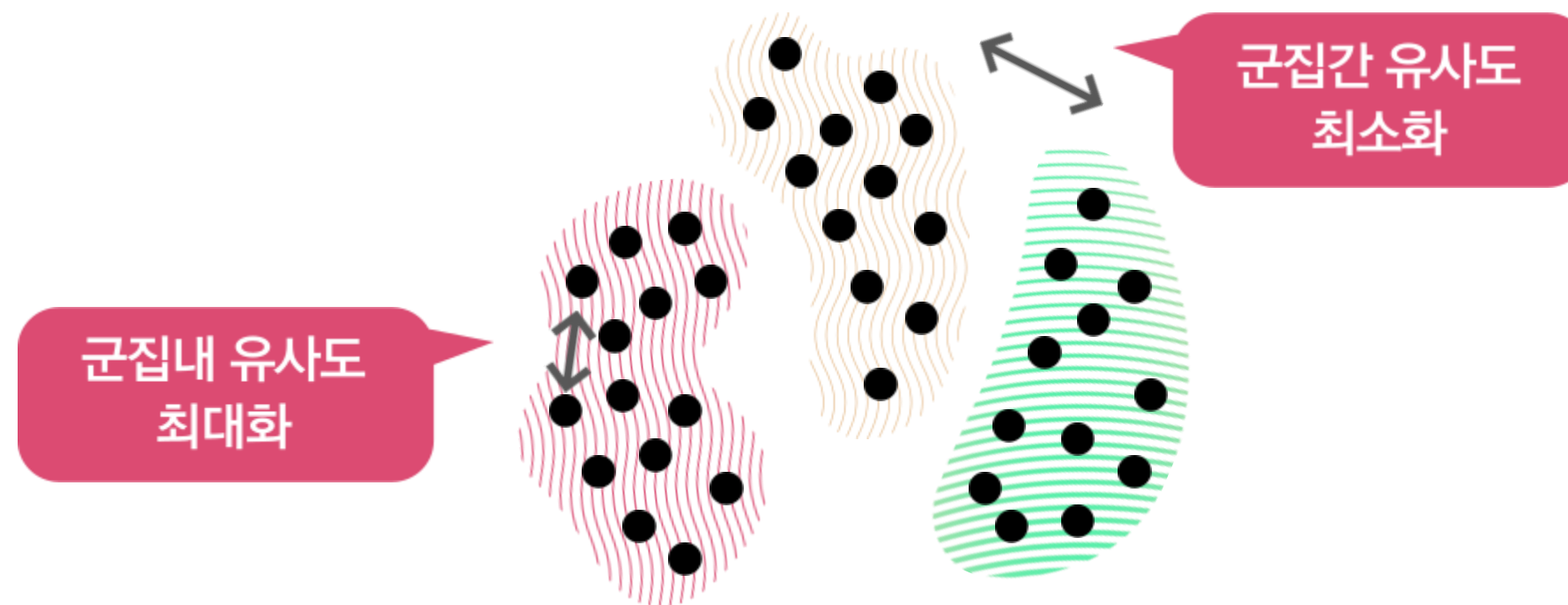
# UNIT 01 | CLUSTERING

## CLUSTERING

군집화

### CLUSTERING 기준

- ✓ 동일한 군집에 소속된 관측치들은 서로 유사할 수록 좋다.
- ✓ 상이한 군집에 소속된 관측치들은 서로 다를수록 좋다.



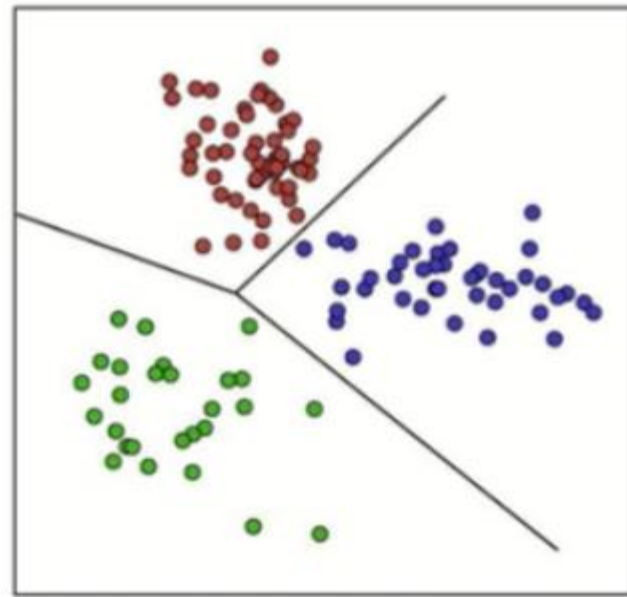
# UNIT 01 | CLUSTERING

## CLUSTERING

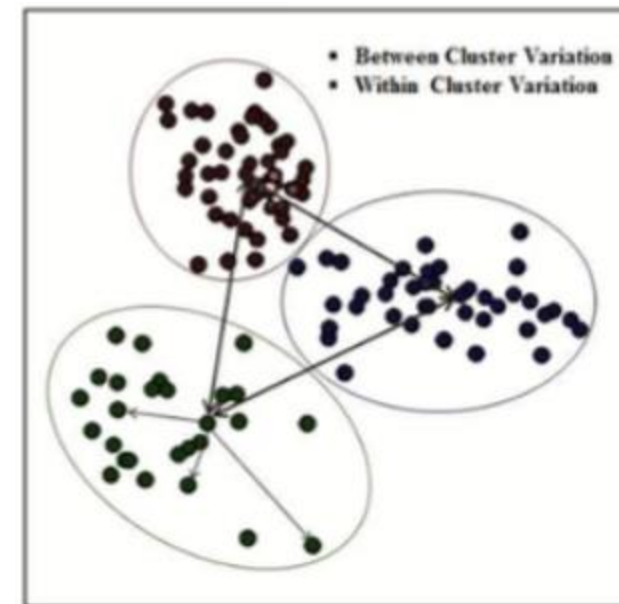
군집화

### 분류 (CLASSIFICATION) VS 군집화 (CLUSTERING)

- ✓ **분류:** 사전 정의된 범주가 있는 (labeled) 데이터로부터 예측 모델을 학습하는 문제 (지도학습: Supervised learning)
- ✓ **군집화:** 사전 정의된 범주가 없는 (unlabeled) 데이터에서 최적의 그룹을 찾아나가는 문제 (비지도학습: Unsupervised learning)



분류



군집화

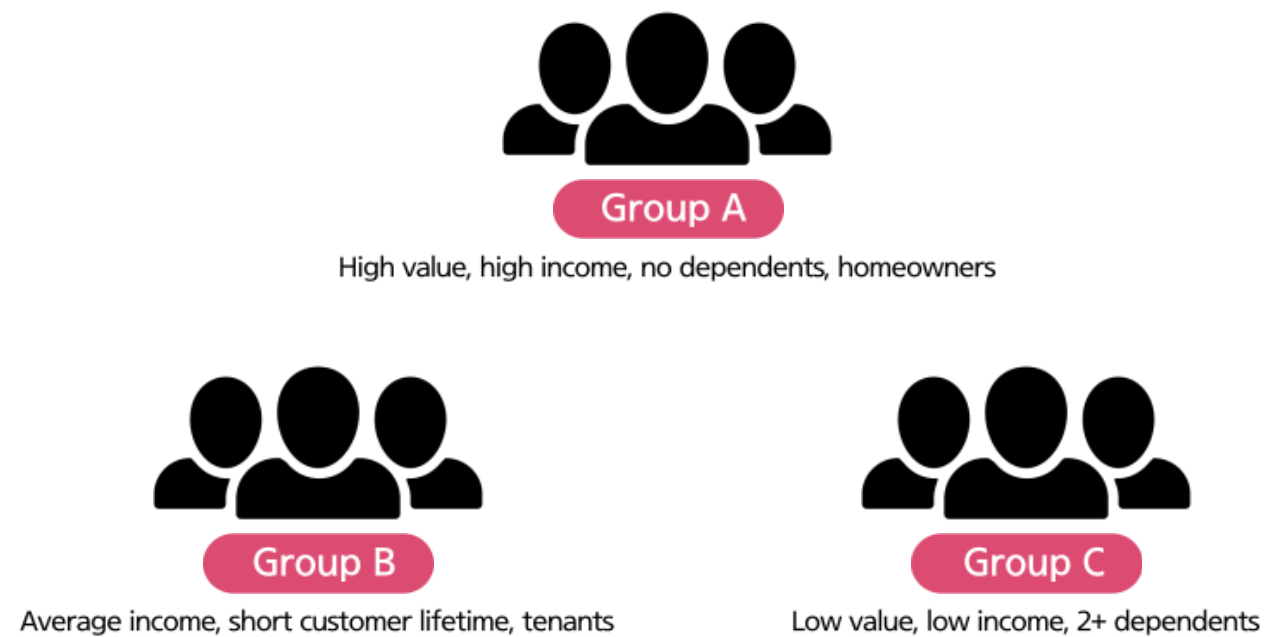
# UNIT 01 | CLUSTERING

## CLUSTERING

군집화

### CLUSTERING 적용 사례

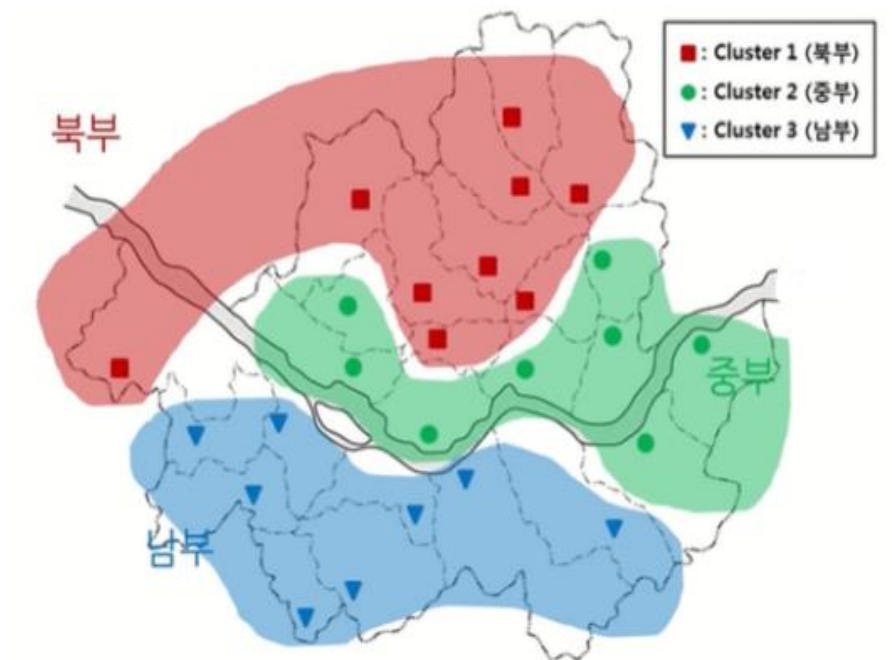
#### 1 고객 세분화 (Customer Segmentation)



#### 2 유사문서 군집화



#### 3 서울시 오존농도 패턴 군집화 (25개 구)



# UNIT 01 | CLUSTERING

## CLUSTERING

군집화

### CLUSTERING 수행 시 주요 고려사항

- ✓ 어떤 거리 척도를 사용하여 유사도를 측정할 것인가?
- ✓ 어떤 클러스터링 알고리즘을 사용할 것인가?
- ✓ 어떻게 최적의 군집 수를 결정할 것인가?
- ✓ 어떻게 군집화 결과를 측정/평가할 것인가?



# UNIT 01 | CLUSTERING

## CLUSTERING

군집화

### CLUSTERING 유사도 척도

- 1 유클리디안 거리 (Euclidean Distance)
- 2 맨해튼 거리 (Manhattan Distance)
- 3 코사인 거리 (Cosine Distance)

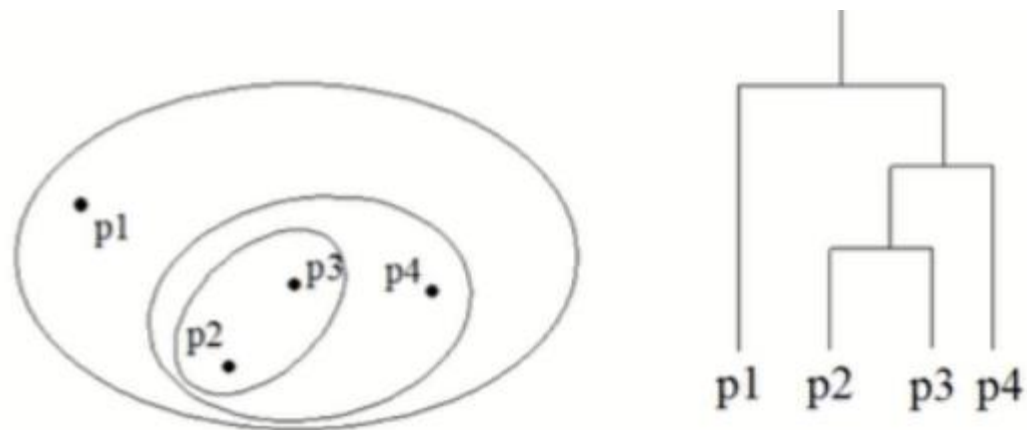
## CLUSTERING

군집화

### CLUSTERING 알고리즘 종류

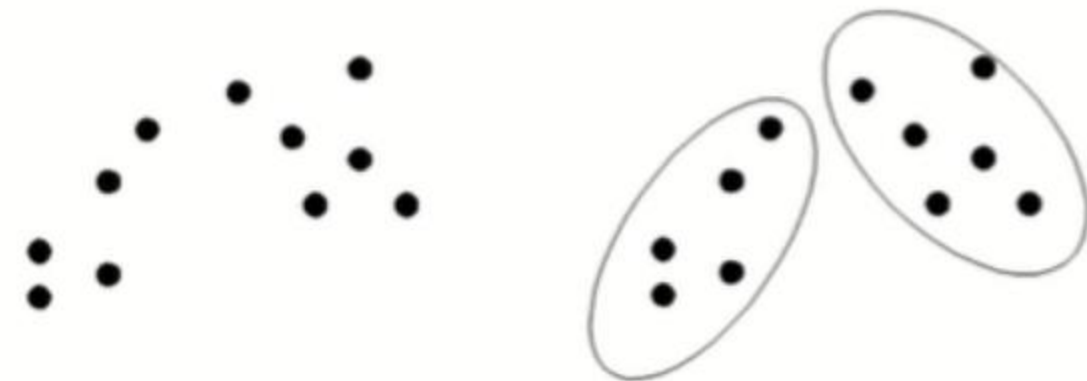
#### 1 계층적 클러스터링

- ✓ 개체들을 가까운 집단부터 차근차근 묶어나가는 방식
- ✓ 군집화 결과 뿐만 아니라 유사한 개체들이 결합되는 dendrogram 생성



#### 2 분리형 클러스터링

- ✓ 전체 데이터의 영역을 특정 기준에 의해 동시에 구분
- ✓ 각 개체들은 사전에 정의된 개수의 군집 중 하나에 속하게 됨



# UNIT 01 | CLUSTERING

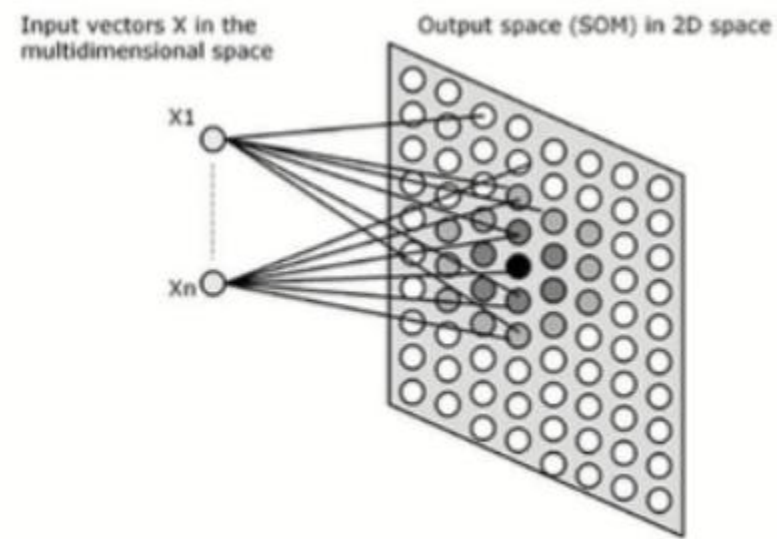
## CLUSTERING

군집화

### CLUSTERING 알고리즘 종류

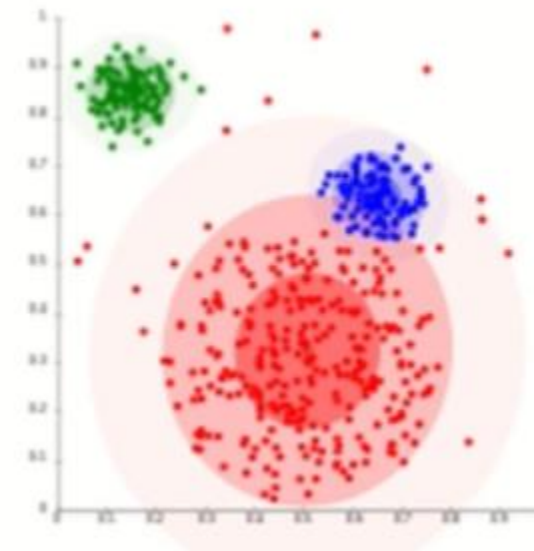
#### 3 자기조직화 지도

- ✓ 2차원의 격자에 각 개체들이 대응하도록 인공신경망과 유사한 학습을 통해 군집 도출



#### 4 분포 기반 클러스터링

- ✓ 데이터의 분포를 기반으로 높은 밀도를 갖는 세부 영역들로 전체 영역을 구분



# **UNIT 02 |**

# **HIERARCHICAL CLUSTERING**

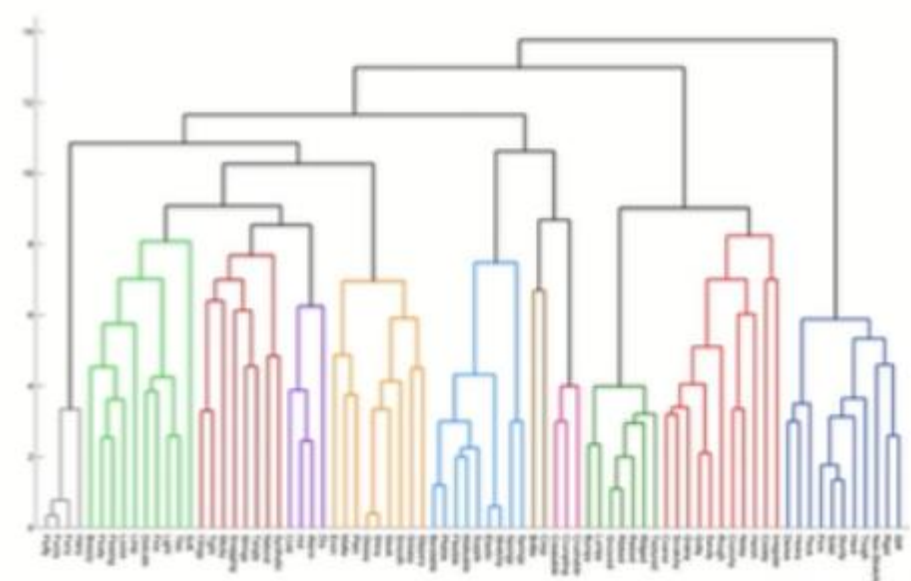
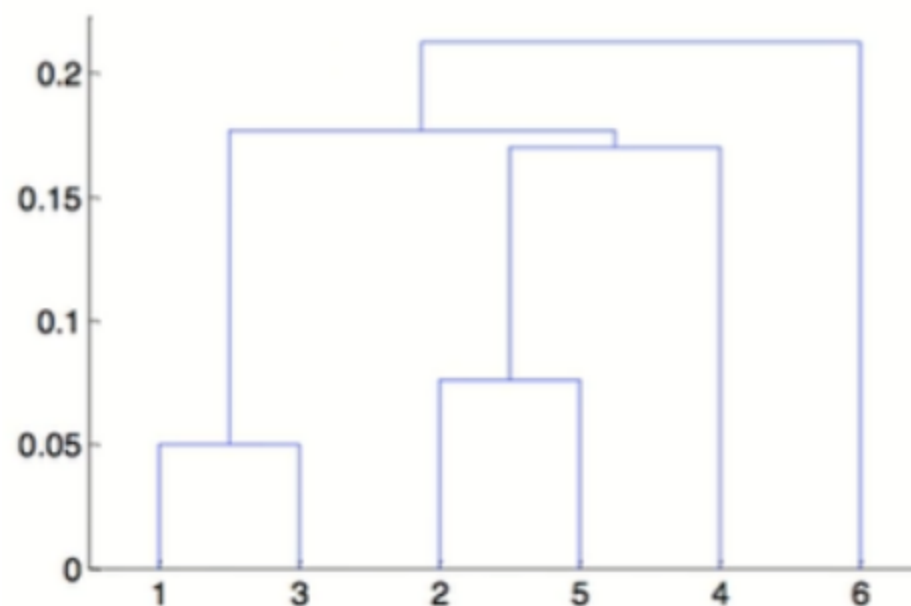
# UNIT 02 | HIERARCHICAL CLUSTERING

## HIERARCHICAL CLUSTERING

계층적 군집화

### HIERARCHICAL CLUSTERING 계층적 군집화

- ✓ 계층적 트리모형을 이용하여 개별 개체들을 순차적/계층적으로 유사한 개체/군집과 통합한다.
- ✓ 덴드로그램(Dendrogram)을 통해 시각화가 가능하다.
  - 덴드로그램: 개체들이 결합되는 순서를 나타내는 트리 형태의 구조
- ✓ 사전에 군집의 수를 정하지 않아도 수행 가능하다.
  - 덴드로그램 생성 후 적절한 수준에서 자르면 그에 해당하는 군집화 결과 생성



## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING 수행 예시

계층적 군집화

#### 수행 예시

- 1 모든 개체들 사이의 거리에 대한 유사도 행렬 계산

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

A

D

B

C

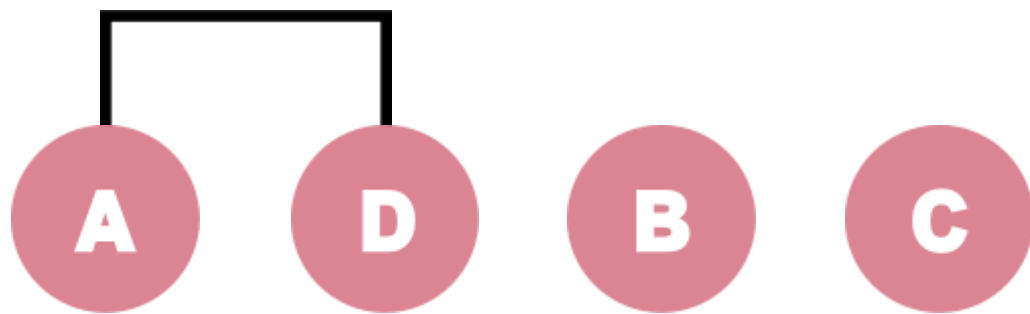
## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING 수행 예시

계층적 군집화

#### 수행 예시

- 1 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 2 거리가 인접한 관측치끼리 군집 형성



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

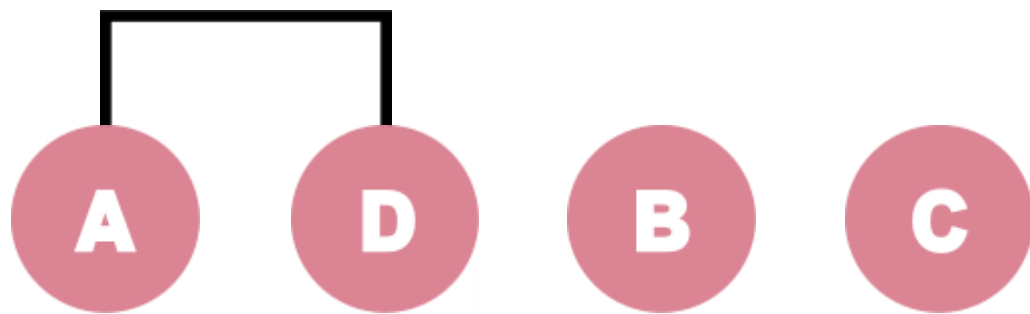
## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING 수행 예시

계층적 군집화

#### 수행 예시

- 1 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 2 거리가 인접한 관측치끼리 군집 형성
- 3 유사도 행렬 업데이트



	AD	B	C	
AD		20	3	
B			10	
C				



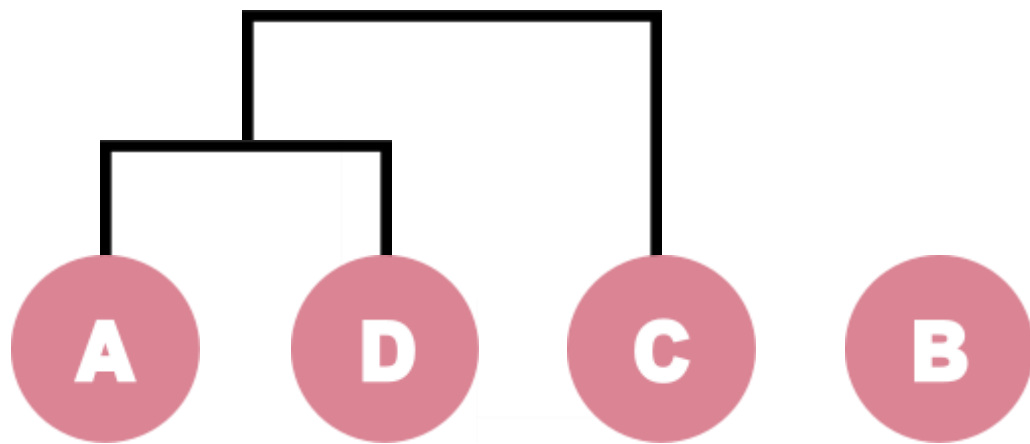
## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING 수행 예시

계층적 군집화

#### 수행 예시

- 1 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 2 거리가 인접한 관측치끼리 군집 형성
- 3 유사도 행렬 업데이트
- 4 위 과정 반복



	AD	B	C	
AD		20	3	
B			10	
C				

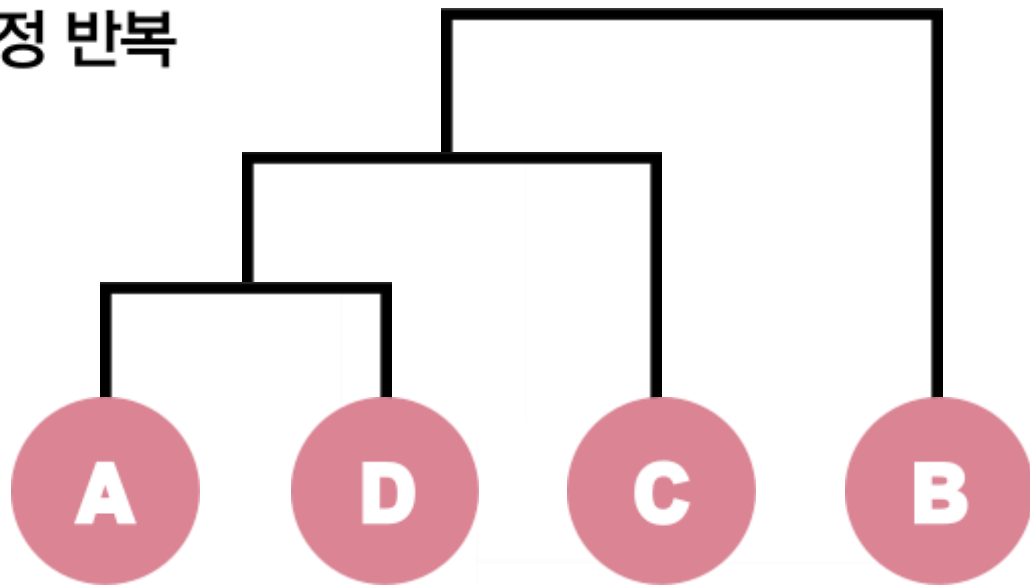
## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING 수행 예시

계층적 군집화

#### 수행 예시

- 1 모든 개체들 사이의 거리에 대한 유사도 행렬 계산
- 2 거리가 인접한 관측치끼리 군집 형성
- 3 유사도 행렬 업데이트
- 4 위 과정 반복



	ADC	B		
ADC		10		
B				

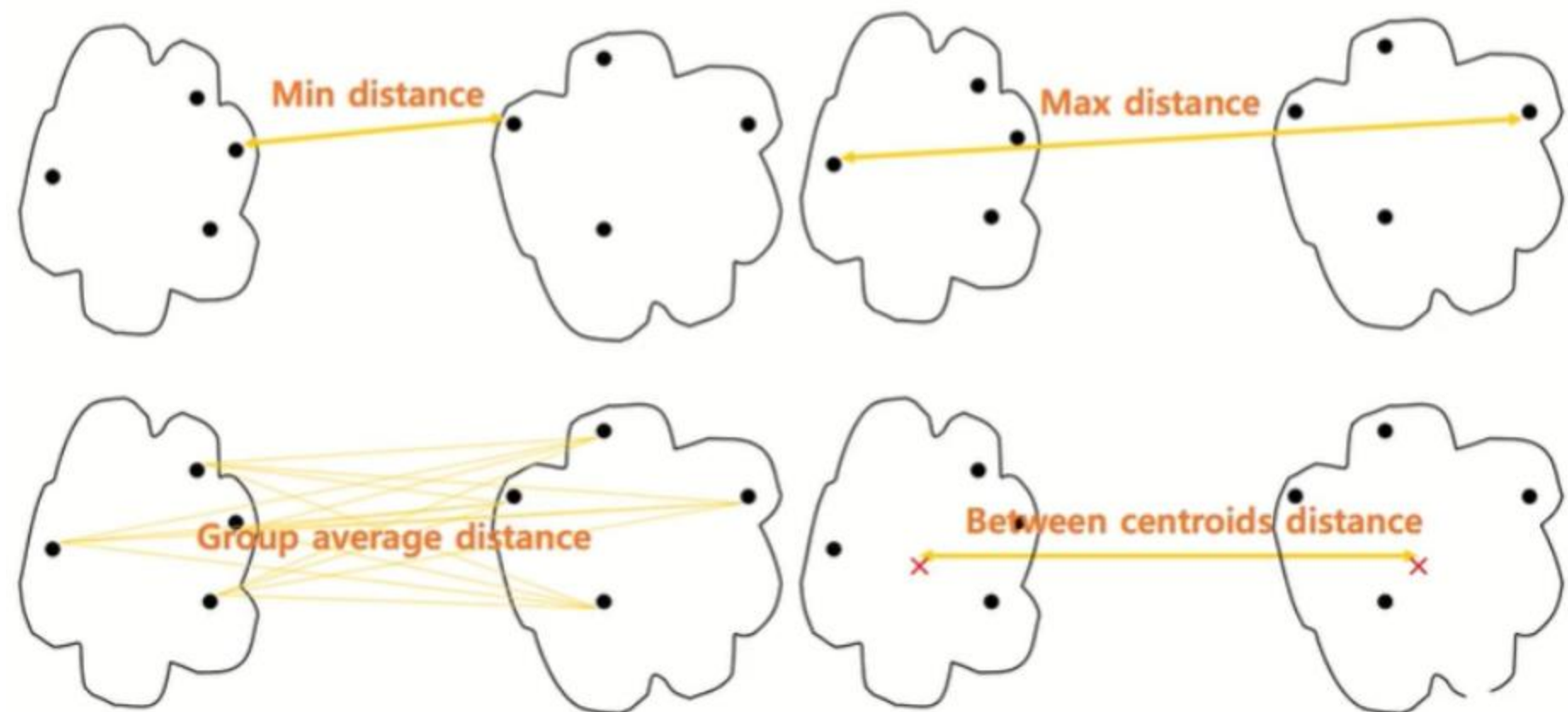
# UNIT 02 | HIERARCHICAL CLUSTERING

## HIERARCHICAL CLUSTERING

계층적 군집화

### 군집 간 거리 측정 두 군집 사이의 유사성/거리 측정

- 1 min(단일 연결)
- 2 max(완전 연결)
- 3 group average(평균 연결)
- 4 between centroid
- 5 Ward's



## UNIT 02 | HIERARCHICAL CLUSTERING

### HIERARCHICAL CLUSTERING

계층적 군집화

#### WARD'S METHOD

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

$m_A$  is the center of cluster A.

# **UNIT 02 |**

# **K-MEANS CLUSTERING**

## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### K-MEANS CLUSTERING K-평균 군집화

- ✓ 대표적인 분리형 군집화 알고리즘
- ✓ 각 군집은 하나의 중심을 가진다.
- ✓ 각 개체는 가장 가까운 중심에 할당되며, 같은 중심에 할당된 개체들이 모여 하나의 군집을 형성한다.
- ✓ 사전에 군집의 수 K가 정해져야 알고리즘을 실행할 수 있다.

$$X = C_1 \cup C_2 \cdots C_k, \quad C_i \cap C_j = \phi, \quad i \neq j$$
$$\arg \min_c \sum_{i=1}^K \sum_{x_j \in C_i} \|x_j - c_i\|^2$$

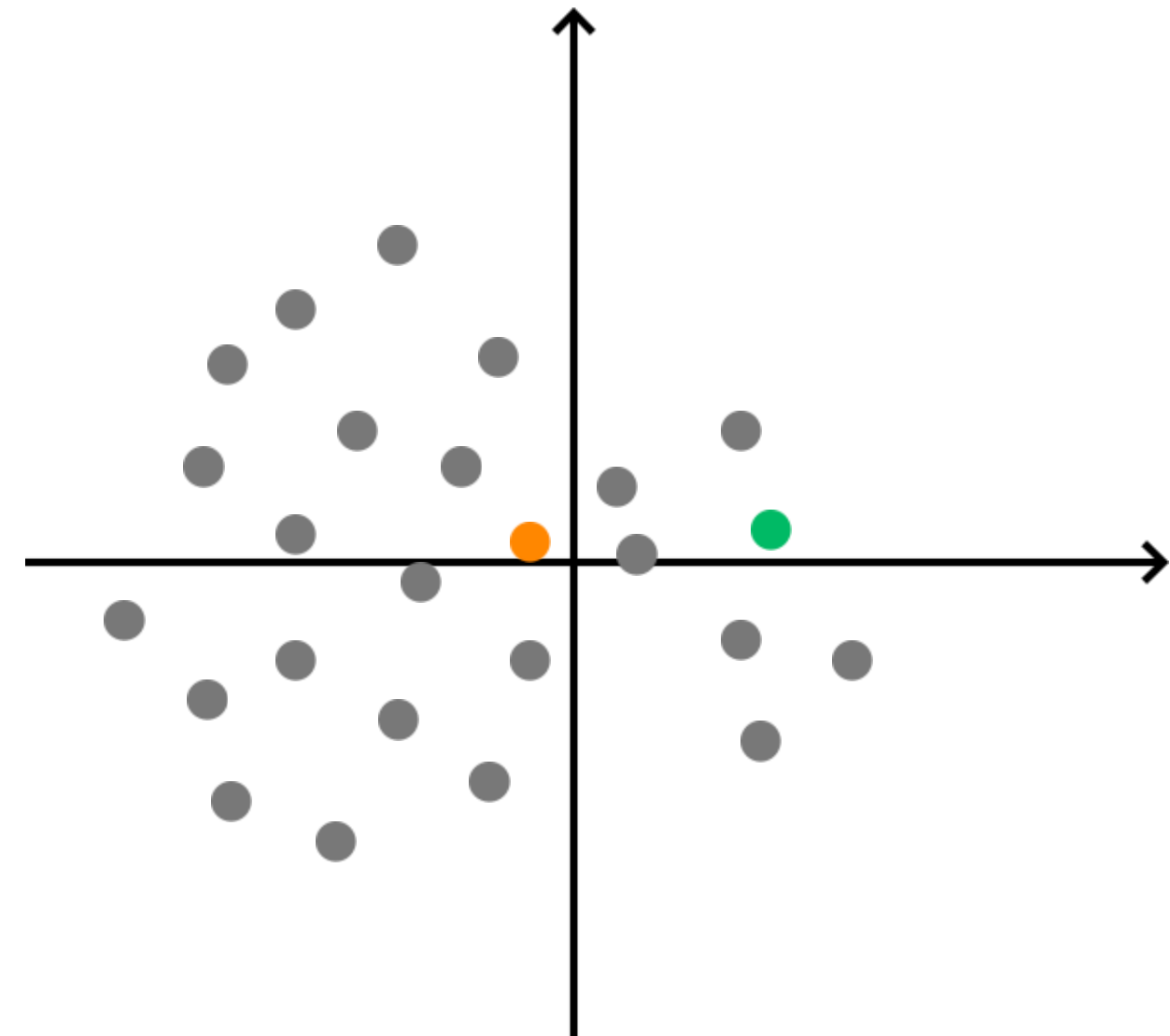
## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### 수행 예시 (K=2)

- 1 2개의 중심을 임의로 생성



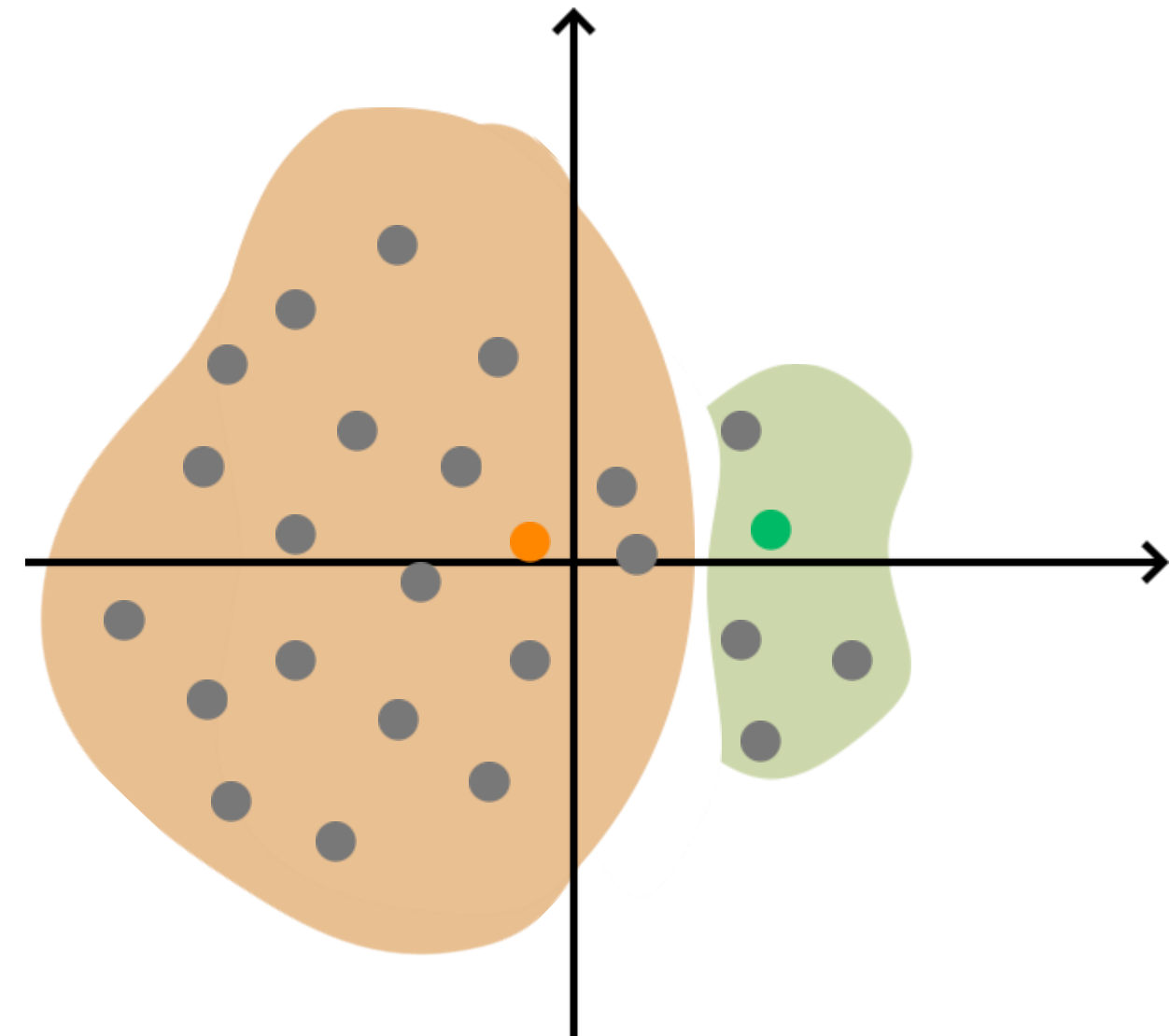
## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### 수행 예시 (K=2)

- 1 2개의 중심을 임의로 생성
- 2 생성된 중심을 기준으로 모든 관측치에 군집 할당





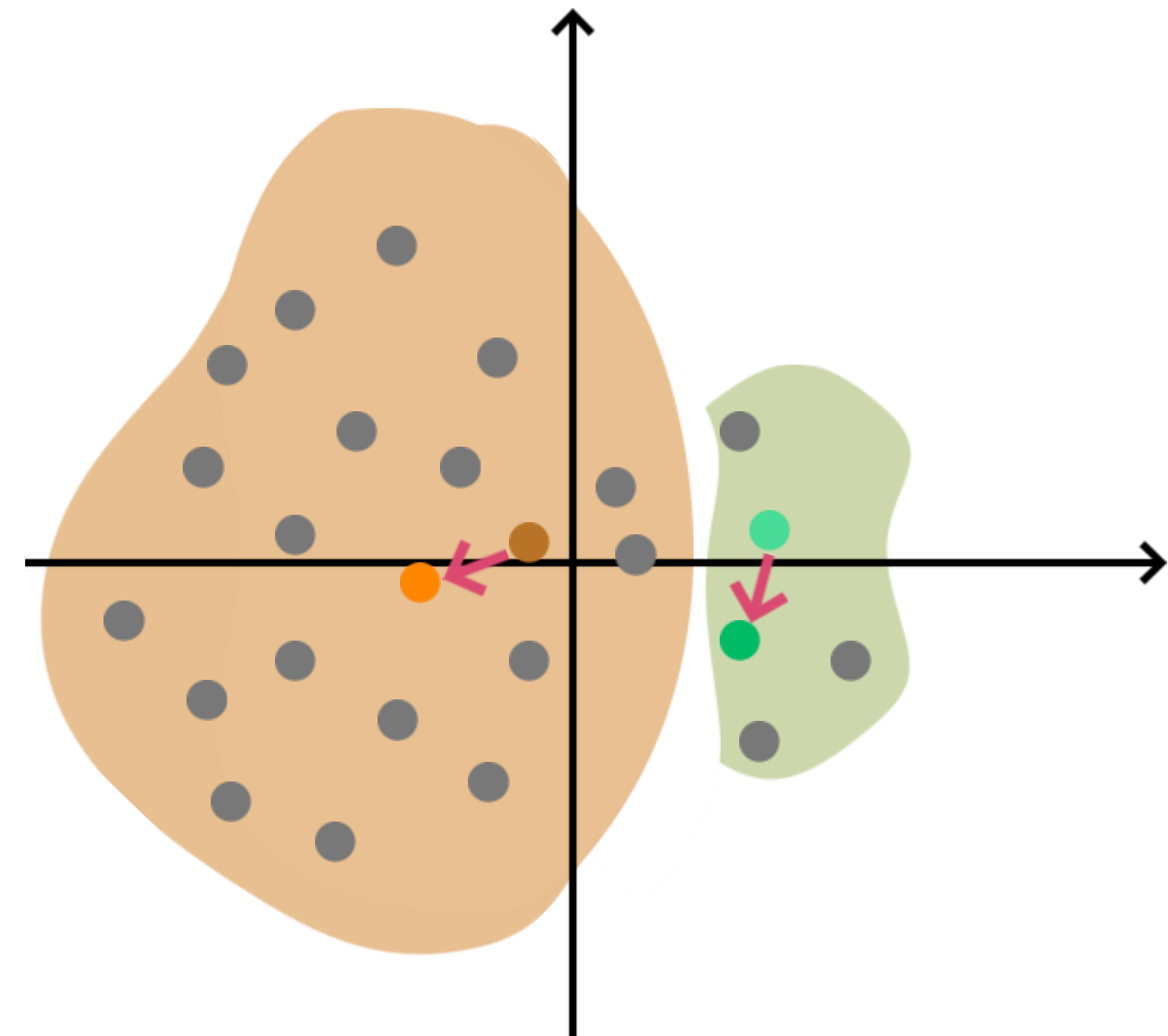
## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### 수행 예시 (K=2)

- 1 2개의 중심을 임의로 생성
- 2 생성된 중심을 기준으로 모든 관측치에 군집 할당
- 3 각 군집의 중심을 다시 계산



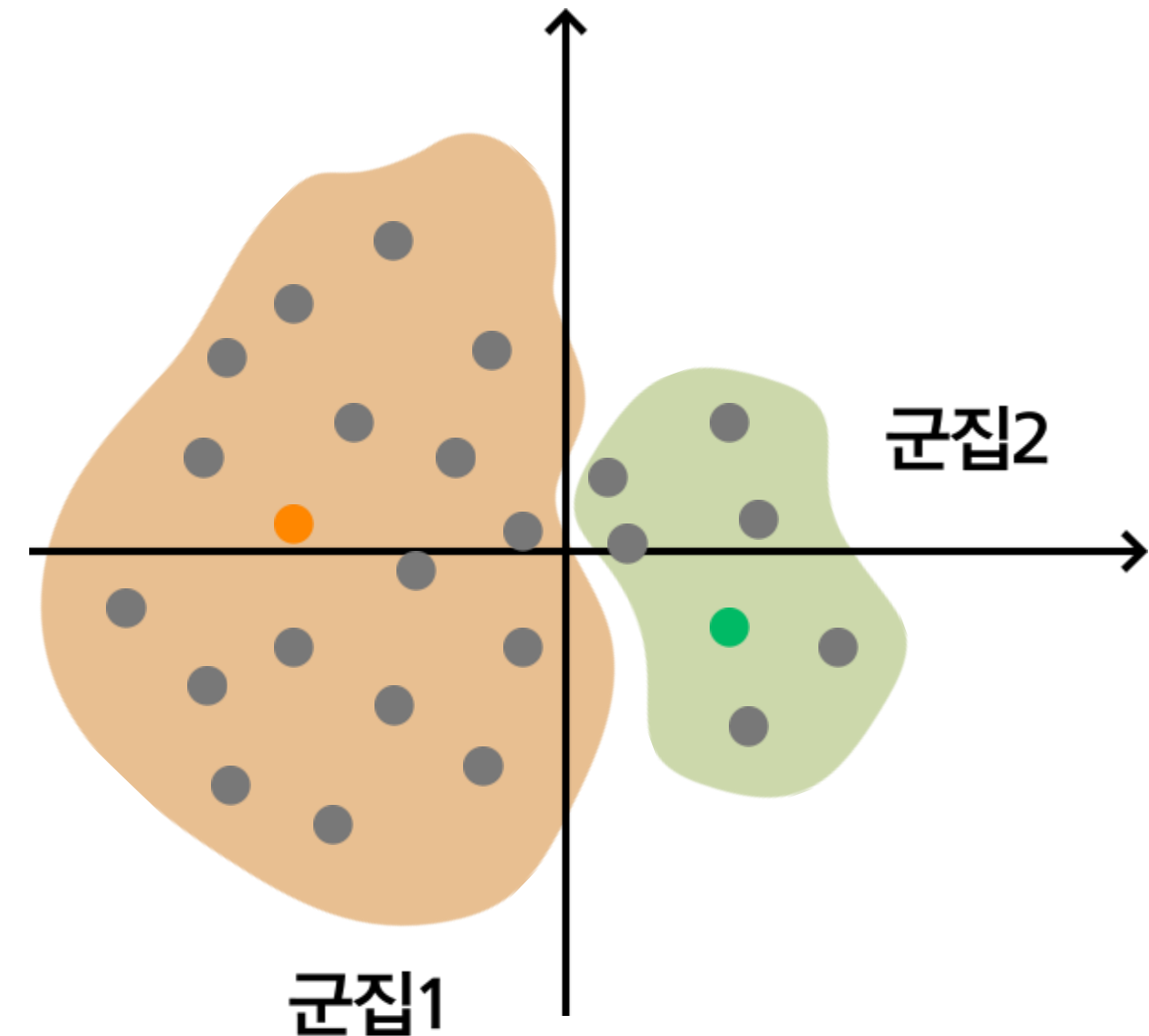
## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### 수행 예시 (K=2)

- 1 2개의 중심을 임의로 생성
- 2 생성된 중심을 기준으로 모든 관측치에 군집 할당
- 3 각 군집의 중심을 다시 계산
- 4 중심이 변하지 않을 때까지 위의 과정을 반복



## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

#### 수행절차

- 1 초기 중심 K개를 임의로 생성한다.
  - 2 개별 관측치로부터 각 중심까지의 거리를 계산한 후, 가장 가까운 중심이 이루는 군집에 관측치를 할당한다.
  - 3 각 군집이 중심을 다시 계산한다.
  - 4 중심이 변하지 않을 때까지 2, 3의 과정을 반복한다.
- ✓ 초기 중심은 종종 무작위로 설정된다. 군집화 결과가 초기 중심 설정에 따라 다르게 나타나는 경우가 발생할 수도 있다.

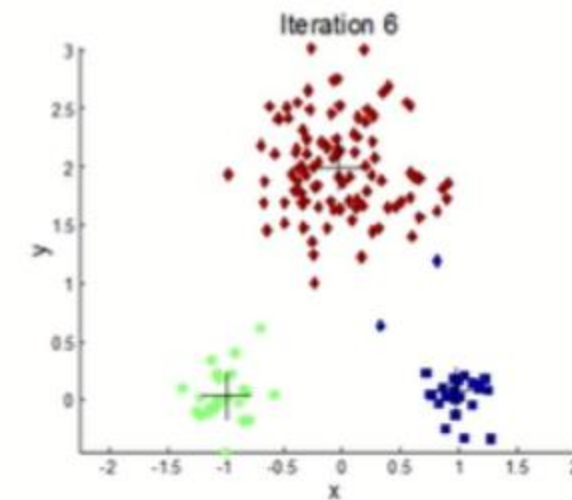
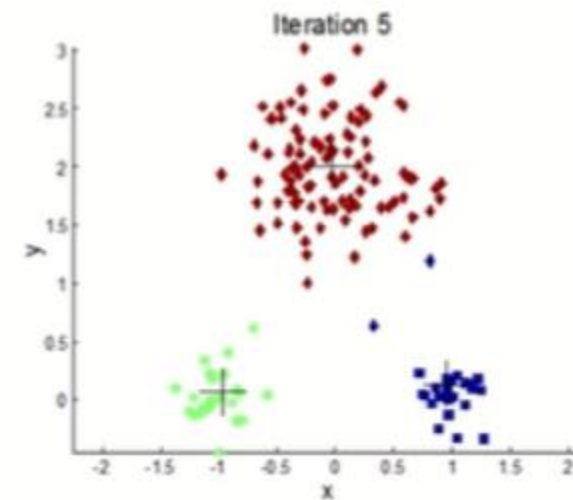
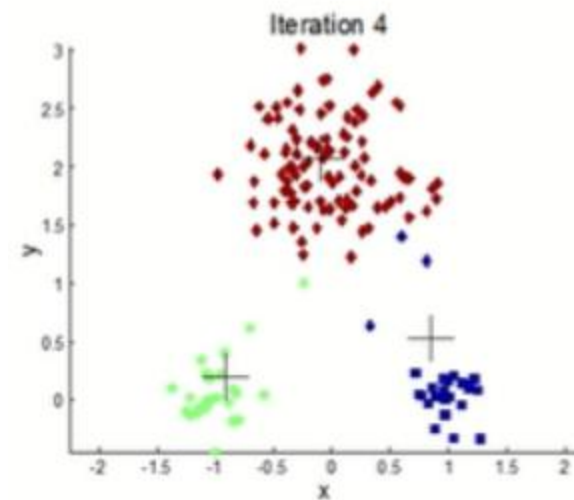
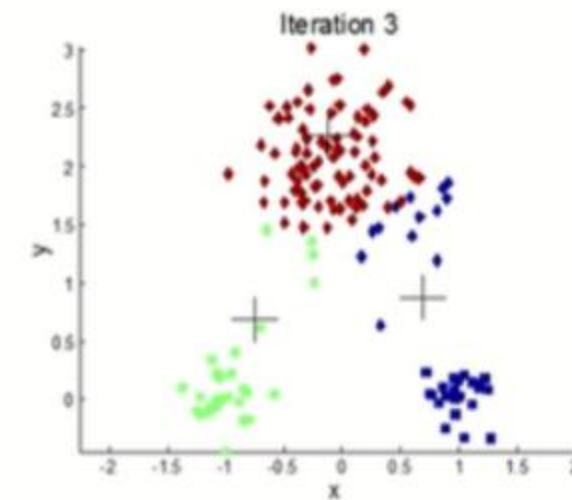
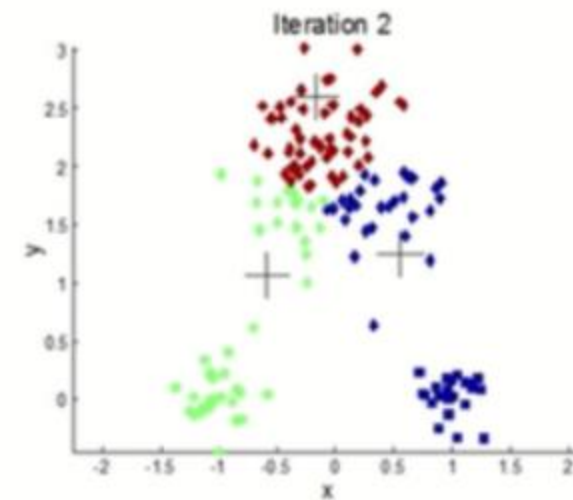
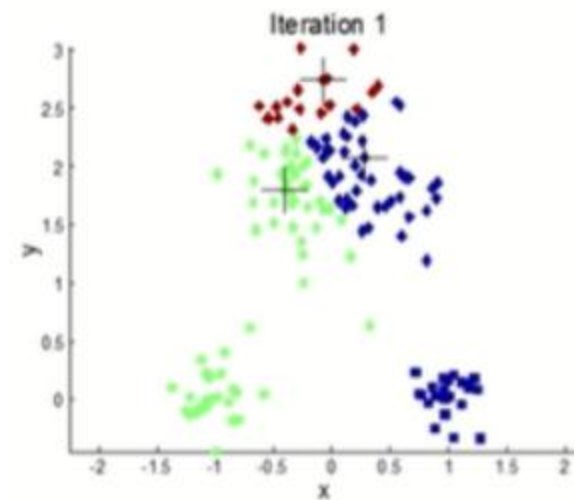
# UNIT 03 | K-MEANS CLUSTERING

## K-MEANS CLUSTERING

K 평균 군집화

### 초기 중심 설정이 최종 결과에 미치는 영향

✓ 바람직한 결과



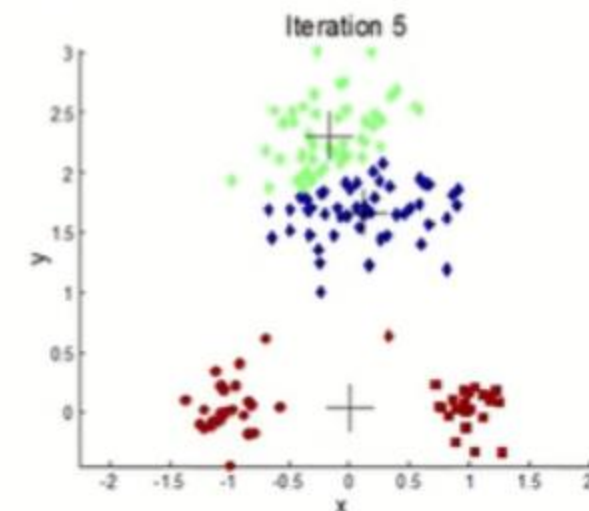
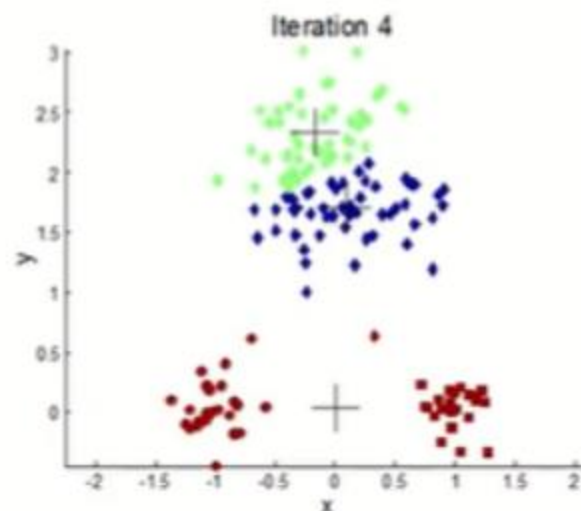
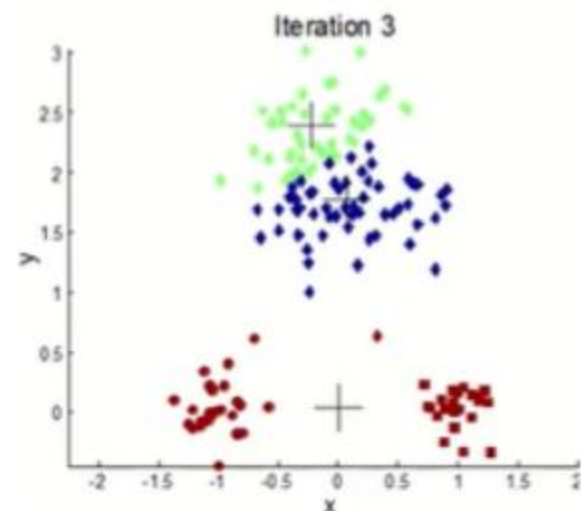
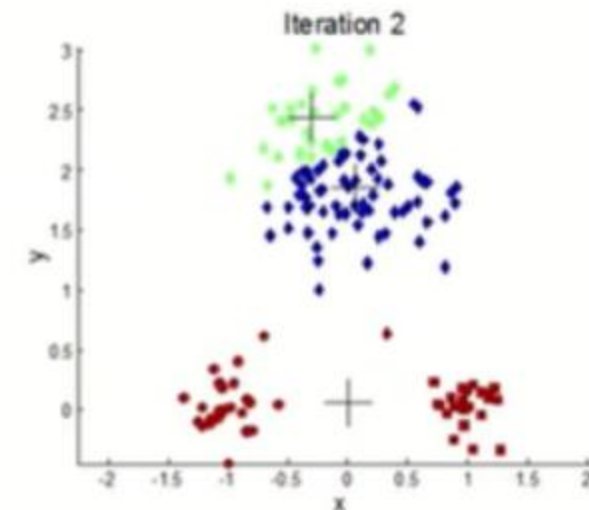
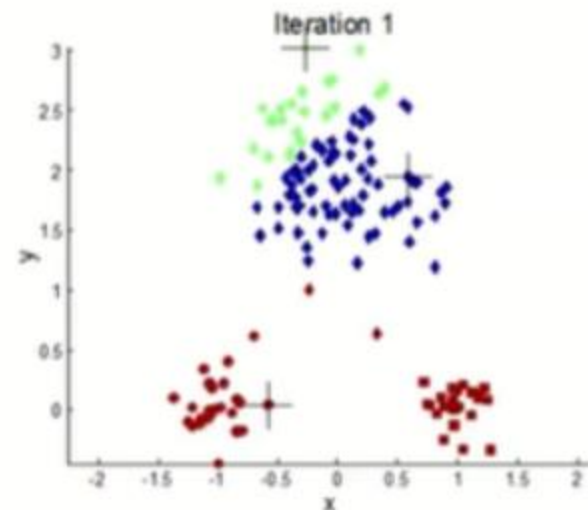
# UNIT 03 | K-MEANS CLUSTERING

## K-MEANS CLUSTERING

K 평균 군집화

### 초기 중심 설정이 최종 결과에 미치는 영향

❌ 바람직하지 않은 결과



## UNIT 03 | K-MEANS CLUSTERING

### K-MEANS CLUSTERING

K 평균 군집화

**"무작위 초기 중심 설정의 위험을 피하고자 다양한 연구가 존재한다."**

- 1 반복적으로 수행하여 가장 여러 번 나타나는 군집을 사용
  - 2 전체 데이터 중 일부만 샘플링하여 계층적 군집화를 수행한 뒤 초기 군집 중심 설정
  - 3 데이터 분포의 정보를 사용하여 초기 중심 설정
- ✓ 하지만 많은 경우 초기 중심 설정이 최종 결과에 큰 영향을 미치지 않는다.



## UNIT 03 | K-MEANS CLUSTERING

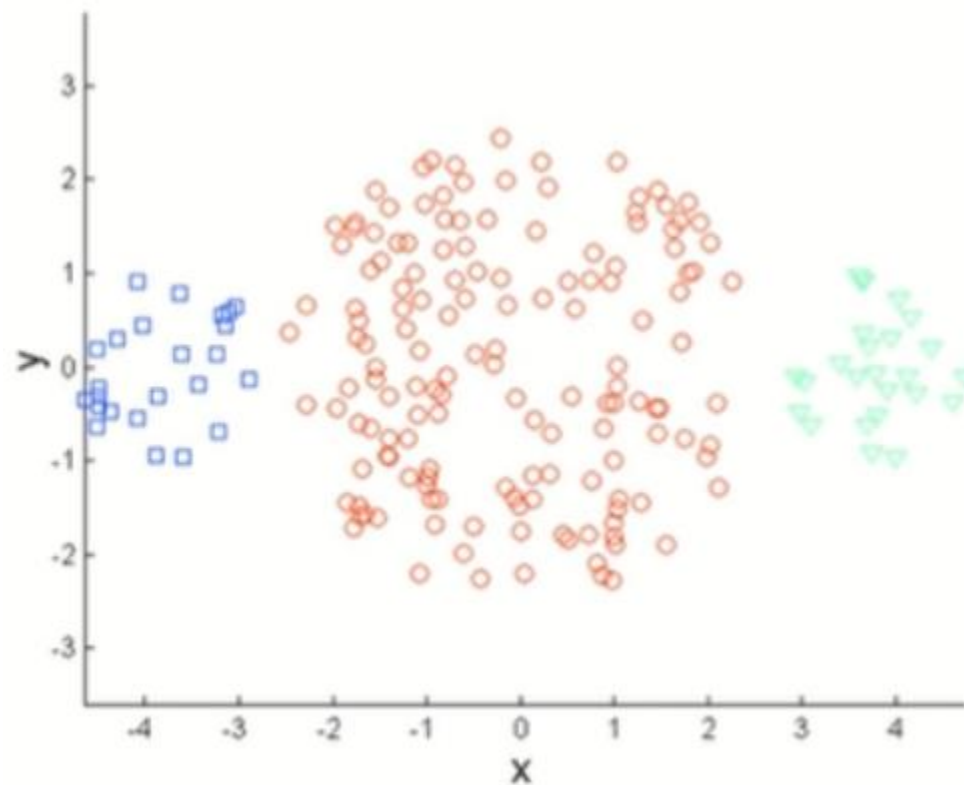
### K-MEANS CLUSTERING

K 평균 군집화

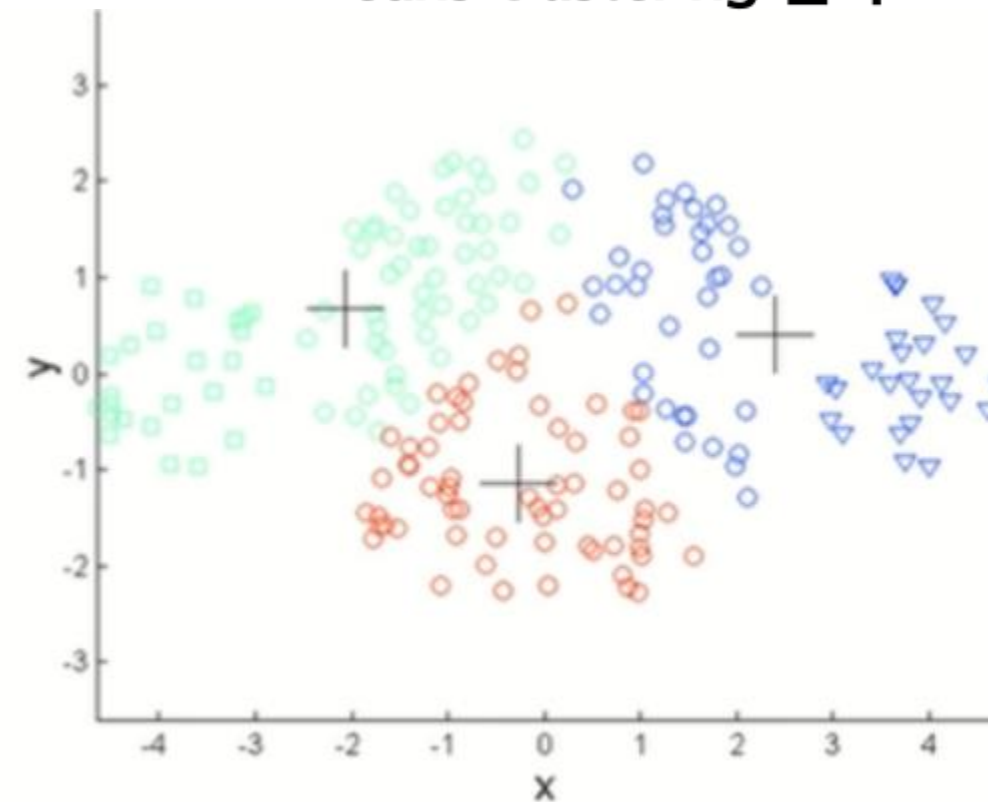
### K-MEANS CLUSTERING의 문제점

- 1 서로 다른 크기의 군집을 잘 찾아내지 못한다.

정답



K-Means Clustering 결과



# UNIT 03 | K-MEANS CLUSTERING

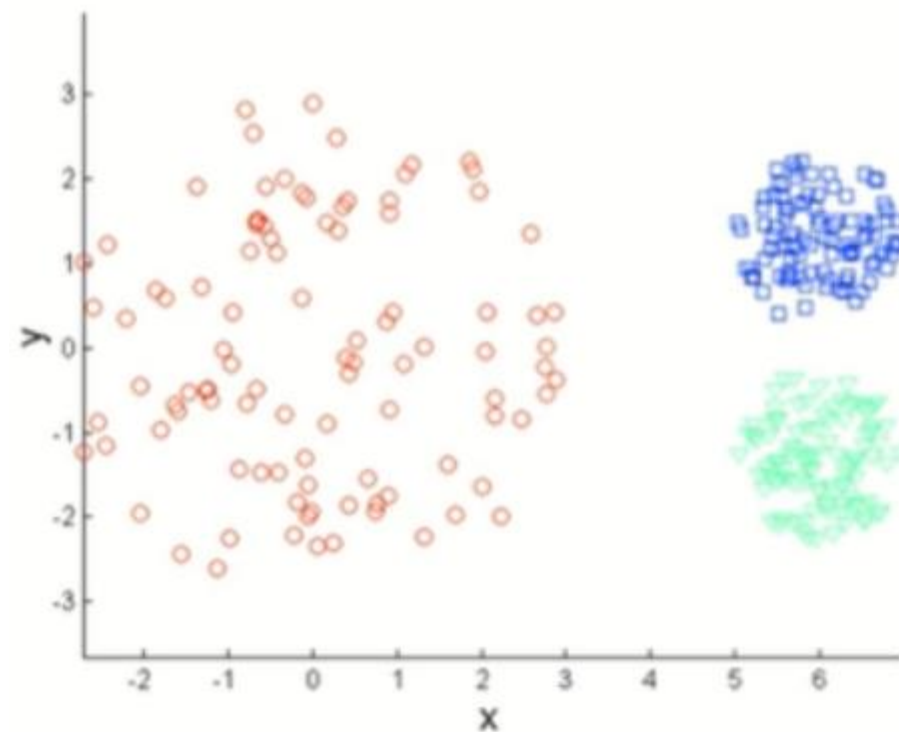
## K-MEANS CLUSTERING

K 평균 군집화

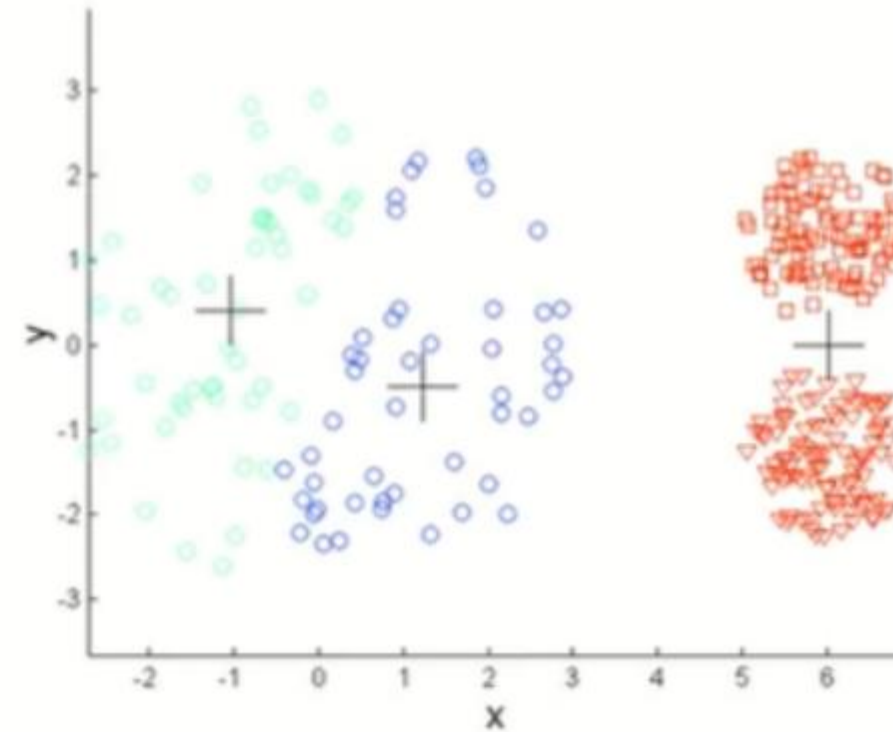
### K-MEANS CLUSTERING의 문제점

- 2 서로 다른 밀도의 군집을 잘 찾아내지 못한다.

정답



K-Means Clustering 결과





# UNIT 03 | K-MEANS CLUSTERING

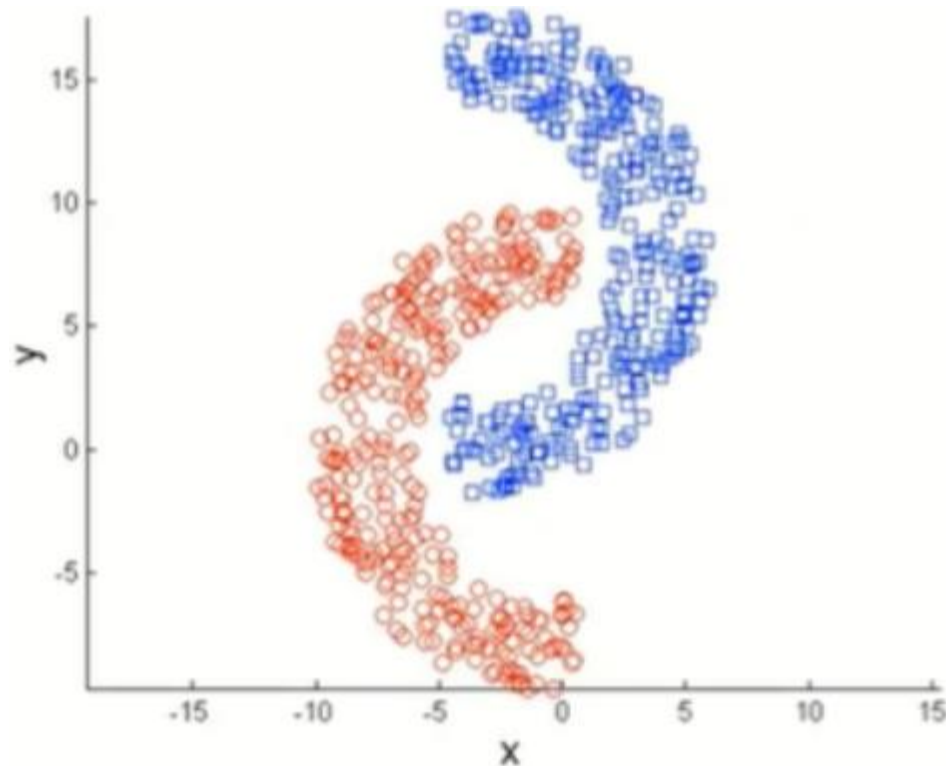
## K-MEANS CLUSTERING

K 평균 군집화

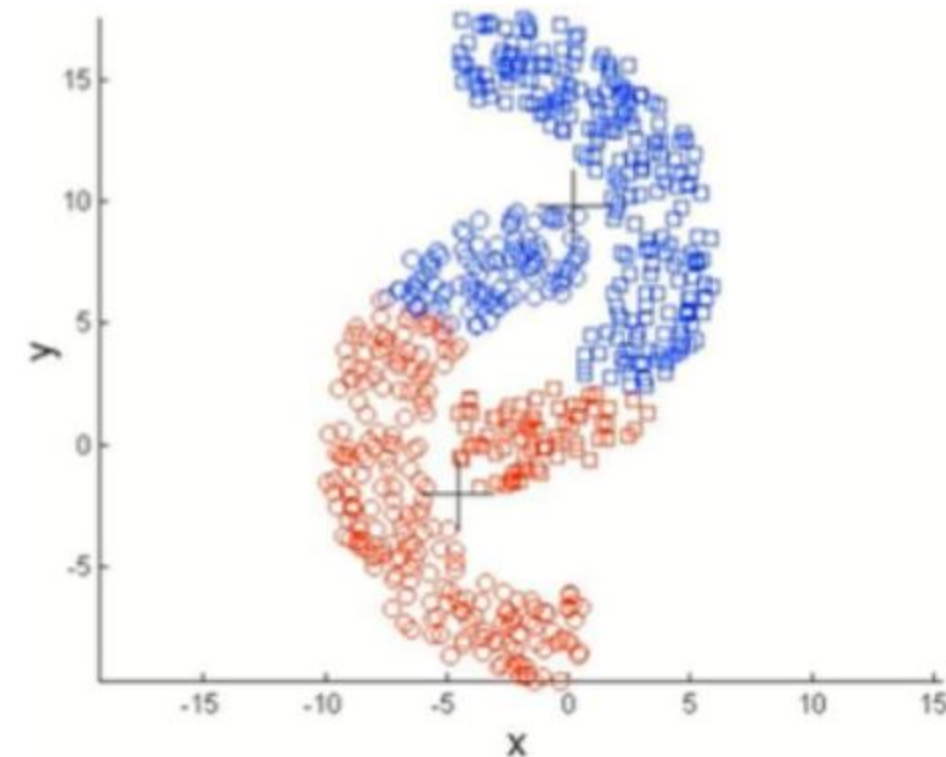
### K-MEANS CLUSTERING의 문제점

- 3 지역적 패턴이 존재하는 군집을 판별하기 어렵다.

정답



K-Means Clustering 결과



# **UNIT 04 | DBSCAN**

## UNIT 05 | 최적의 군집 수 및 모델 평가

### DBSCAN

Density-Based Spatial Clustering of Application with Noise

### DBSCAN

공간상에 높은 밀도를 가지고 모여 있는 관측치들을 하나의 그룹으로 간주하고 낮은 밀도를 가지고 홀로 있는 관측치는 이상치 또는 잡음으로 분류하는 밀도 기반 군집화 알고리즘

#### 특징

- ✔ 클러스터의 개수를 미리 지정할 필요가 없다.
- ✔ 복잡한 형상도 찾을 수 있으며, 어떤 클래스에도 속하지 않는 포인트를 구분할 수 있다.
- ✔ 병합 군집이나 k-평균보다는 다소 느리지만 비교적 큰 데이터셋에도 적용할 수 있다

#### 아이디어

특성 공간에서 가까이 있는 데이터가 많아 붐비는 지역(밀집지역)의 포인트를 찾는다. 이러한 밀집 지역이 한 클러스터를 구성하며 비교적 비어있는 지역을 경계로 다른 클러스터와 구분한다.

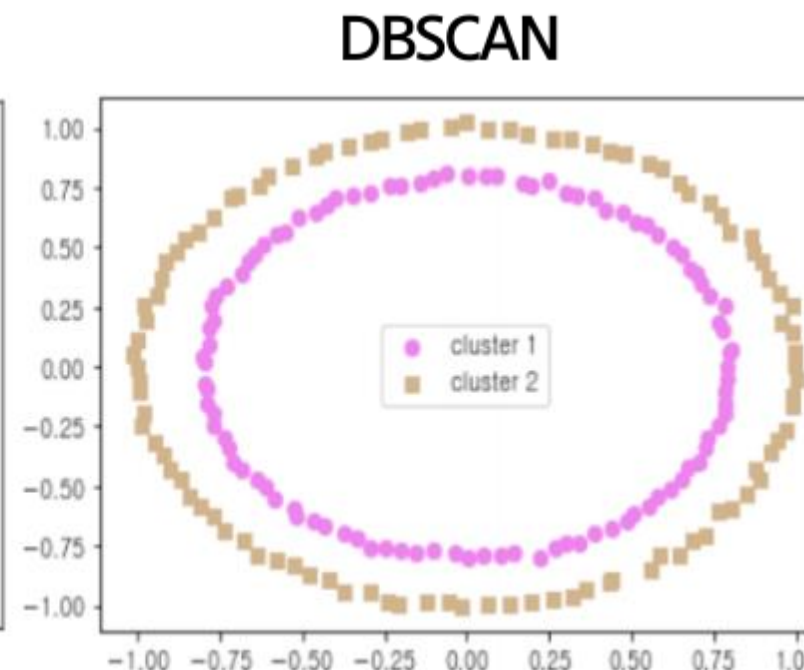
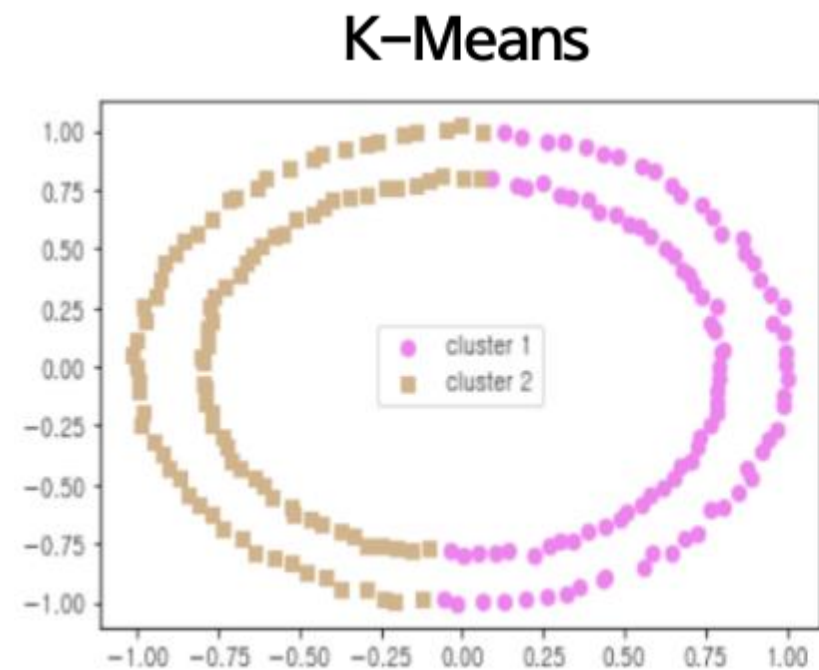
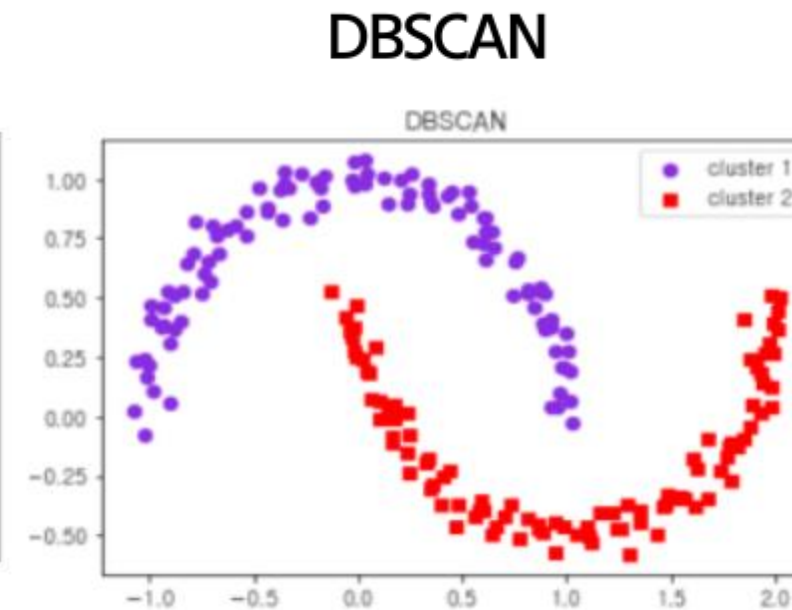
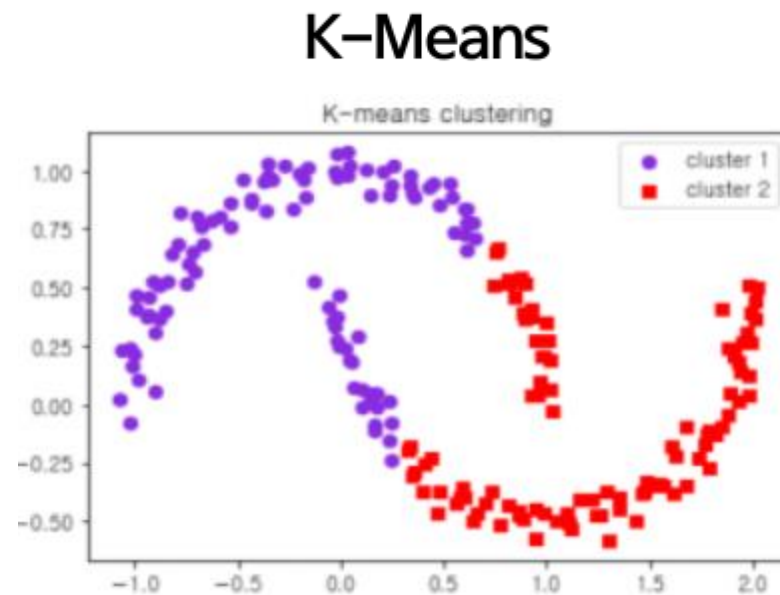
- 밀집 지역에 있는 포인트를 핵심 샘플(또는 핵심 포인트)라고 한다.

# UNIT 05 | 최적의 군집 수 및 모델 평가

## DBSCAN

Density-Based Spatial Clustering of Application with Noise

## DBSCAN



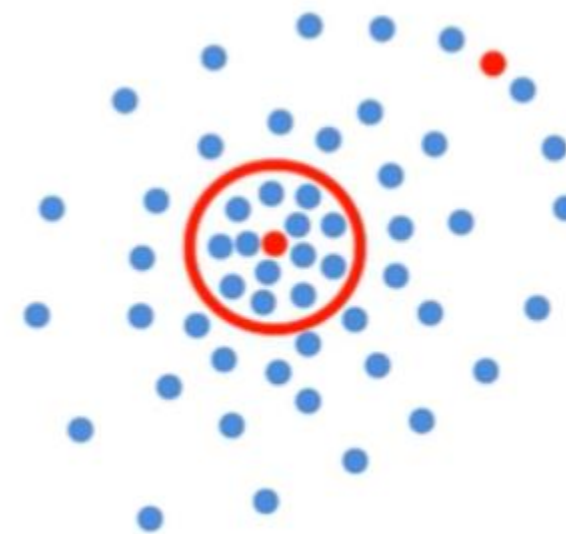
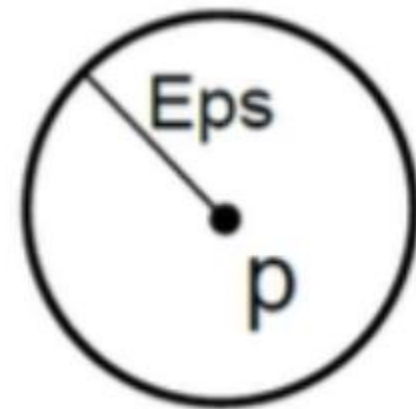
## UNIT 05 | 최적의 군집 수 및 모델 평가

### DBSCAN

Density-Based Spatial Clustering of Application with Noise

**매개변수:** eps와 min\_samples

한 데이터 포인트에서 eps 거리 안에 데이터가 min\_samples 개수만큼 들어 있으면 이 데이터 포인트를 핵심 샘플로 분류한다. Eps보다 가까운 핵심 샘플은 DBSCAN에 의해 동일한 클러스터로 합쳐진다.



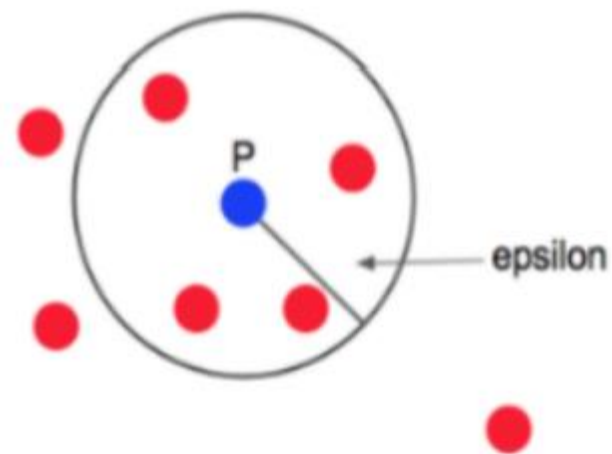
# UNIT 05 | 최적의 군집 수 및 모델 평가

## DBSCAN

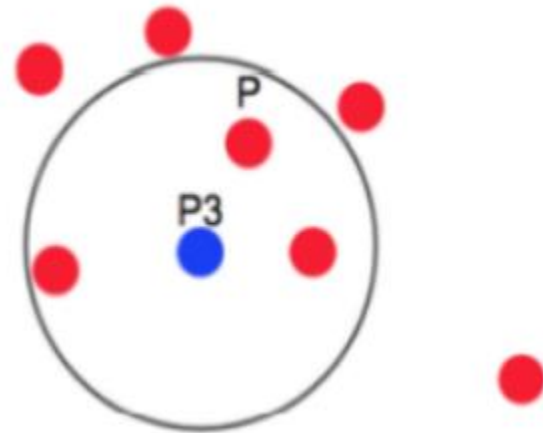
Density-Based Spatial Clustering of Application with Noise

### 포인트 종류

min\_samples = 4

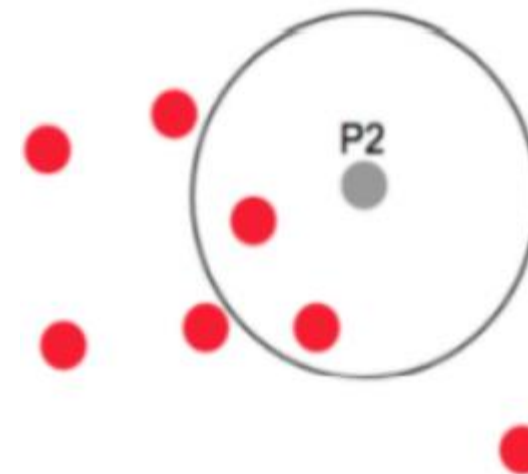


P3: 핵심 포인트



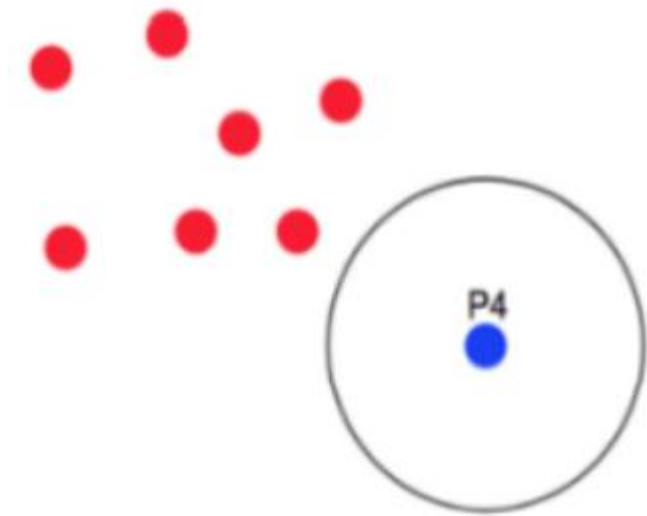
"eps 거리 내에 데이터가 (자신 포함)  
min\_samples개 이상 있는 포인트"

P2: 경계 포인트



"핵심 포인트를 이웃으로 갖고 있지만  
eps 거리 내에 데이터가 min\_samples개  
보다 적은 포인트"

P4: 잡음 포인트



"핵심 포인트를 이웃으로 갖고 있지 않고  
eps 거리 내에 데이터가 min\_samples개  
보다 적은 포인트"



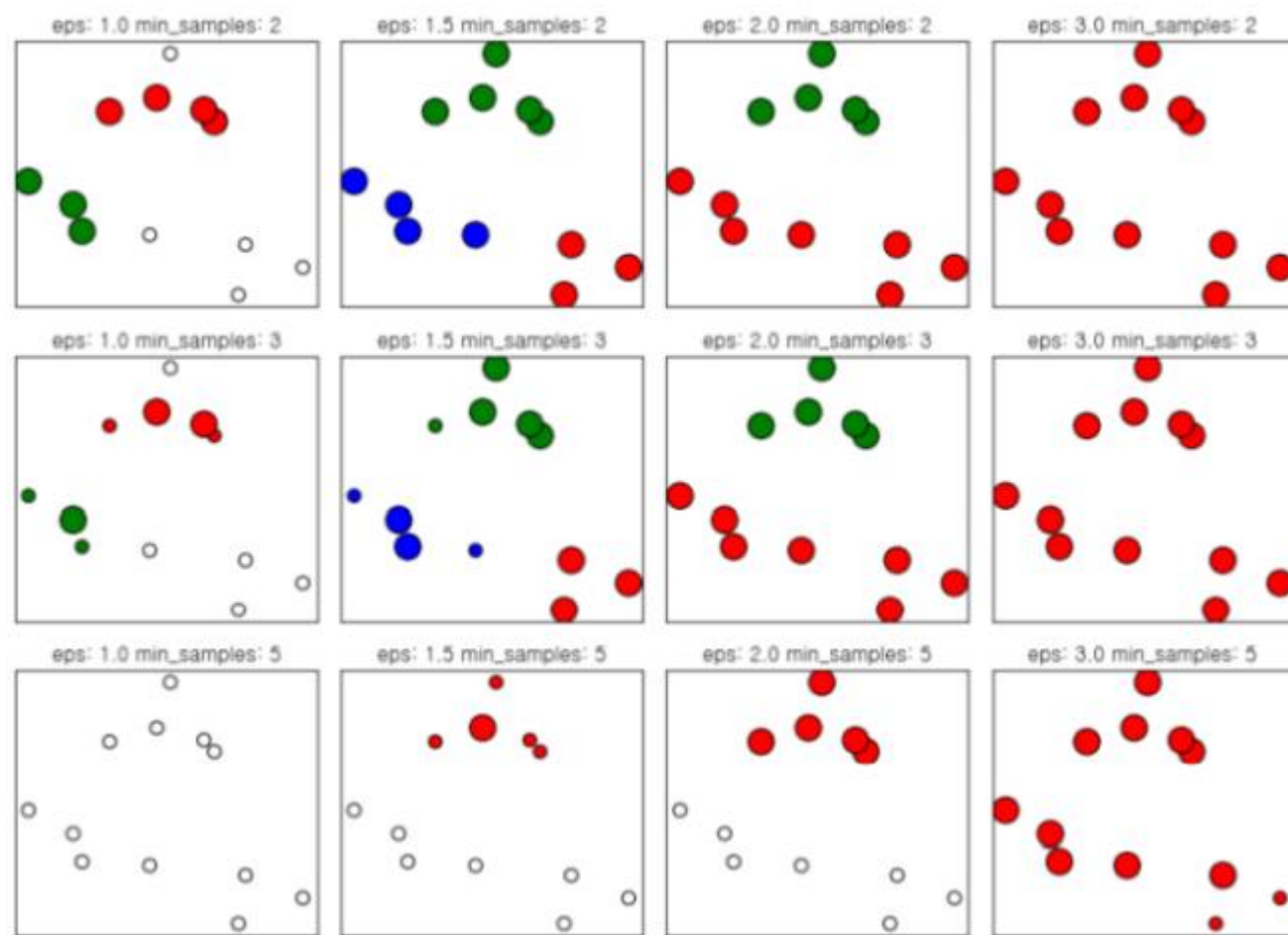
# UNIT 05 | 최적의 군집 수 및 모델 평가

## DBSCAN

Density-Based Spatial Clustering of Application with Noise

eps 증가: 하나의 군집에 더 많은 데이터가 포함 (=군집의 크기 증가)

min\_samples 증가: 핵심 포인트 줄고, 잡음 포인트 증가



- ✓ 2차원 데이터의 경우 최소 자료개수 min\_samples는 통상적으로 4를 사용한다.
- ✓ eps는 너무 작으면 많은 관측치가 잡음자료로 분류되고, 너무 크면 군집의 개수가 너무 작아지게 되어 eps는 DBSCAN의 성능에 결정적인 역할을 한다.

흰색이 아닌 큰 원: 핵심 포인트  
흰색이 아닌 작은 원: 경계 포인트  
흰색 원: 잡음 포인트

## UNIT 05 | 최적의 군집 수 및 모델 평가

### DBSCAN

Density-Based Spatial Clustering of Application with Noise

### 수행절차

- 1 무작위로 포인트를 선택한다.
- 2 그 포인트에서 eps 거리 안의 모든 포인트를 찾는다.
  - 2-1 만약 eps 거리 안에 있는 포인트 수가 min\_samples보다 적다면 그 포인트는 어떤 클래스에도 속하지 않는 잡음으로 레이블한다.
  - 2-2 eps 거리 안에 min\_samples보다 많은 포인트가 있다면 그 포인트는 핵심 샘플로 레이블하고 새로운 클러스터 레이블을 할당한다.
- 3 2-2의 핵심 포인트의 eps 거리 안의 모든 이웃을 살핀다.
  - 3-1 만약 어떤 클러스터에도 아직 할당되지 않았다면 바로 전에 만든 클러스터 레이블을 할당한다.
  - 3-2 만약 핵심 샘플이면 그 포인트의 이웃을 차례로 방문한다.
- 4 위 과정을 반복한다. 클러스터는 eps 거리 안에 더 이상 핵심 샘플이 없을 때까지 자라난다.
- 5 한 개의 클러스터가 형성 되었고, 그 다음 포인트를 찾아 위와 같은 과정을 반복한다.

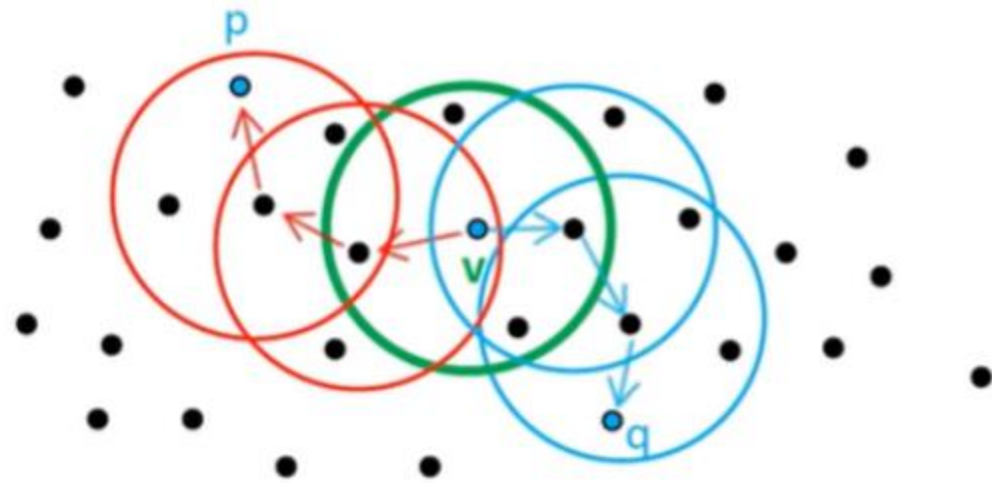


## UNIT 05 | 최적의 군집 수 및 모델 평가

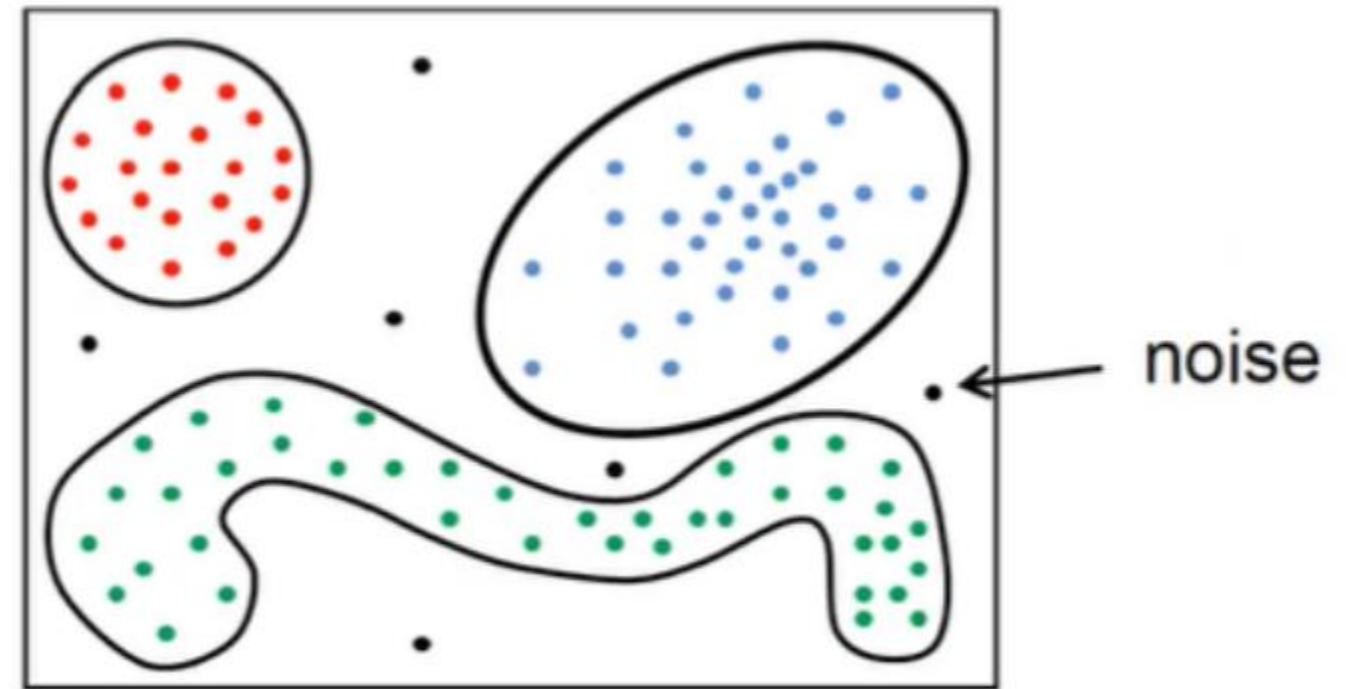
### DBSCAN

Density-Based Spatial Clustering of Application with Noise

### 수행절차



MinPts = 5



## UNIT 05 | 최적의 군집 수 및 모델 평가

### DBSCAN

Density-Based Spatial Clustering of Application with Noise

### DBSCAN의 한계점

- 1 사전에 데이터에 대한 충분한 이해도를 갖고 있지 않다면 `eps`와 `min_samples`의 값을 정하기 어렵다.
- 2 연산량이 많아 K-Means에 비해 속도가 느리다.
- 3 차원의 저주 문제
  - 차원 수가 낮은 데이터 문제가 되지 않지만 고차원 데이터로 갈수록 학습 데이터 양이 급증해 많은 연산이 필요하다.
  - 유클리디안 거리를 사용하는 모든 모델의 공통적인 한계점이다.

# **UNIT 05 |**

## **최적의 군집 수 및 모델 평가**

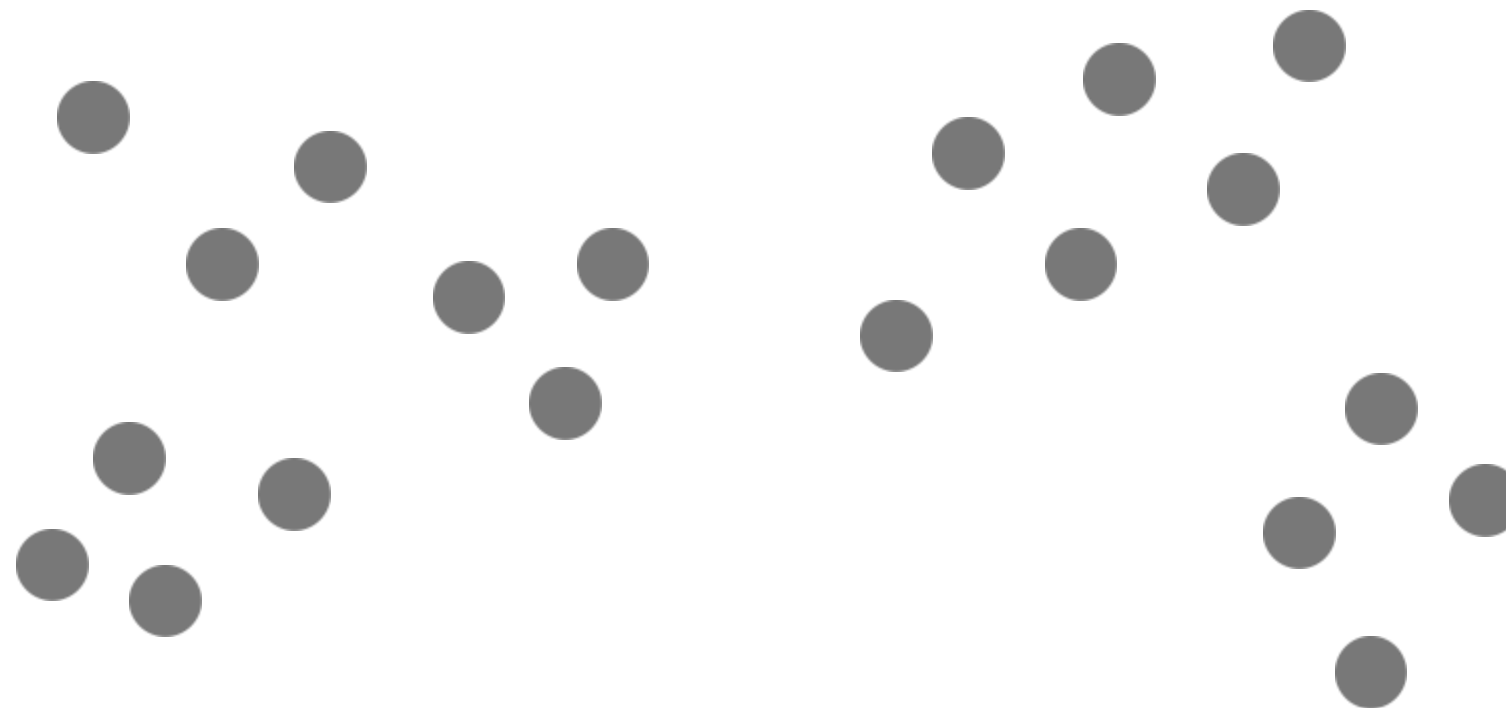
## UNIT 05 | 최적의 군집 수 및 모델 평가

### 최적의 군집 수 결정

Clustering

### 어떻게 최적의 군집 수를 결정할 것인가?

- ✓ 예) 20개의 관측치가 존재할 때, 최적의 군집 수는?



## UNIT 05 | 최적의 군집 수 및 모델 평가

### 최적의 군집 수 결정

Clustering

### 어떻게 최적의 군집 수를 결정할 것인가?

- ✓ 예) 20개의 관측치가 존재할 때, 최적의 군집 수는?

군집 수 = 2



## UNIT 05 | 최적의 군집 수 및 모델 평가

### 최적의 군집 수 결정

Clustering

### 어떻게 최적의 군집 수를 결정할 것인가?

- ✓ 예) 20개의 관측치가 존재할 때, 최적의 군집 수는?

군집 수 = 3



## UNIT 05 | 최적의 군집 수 및 모델 평가

### 최적의 군집 수 결정

Clustering

### 어떻게 최적의 군집 수를 결정할 것인가?

- ✓ 예) 20개의 관측치가 존재할 때, 최적의 군집 수는?

군집 수 = 5



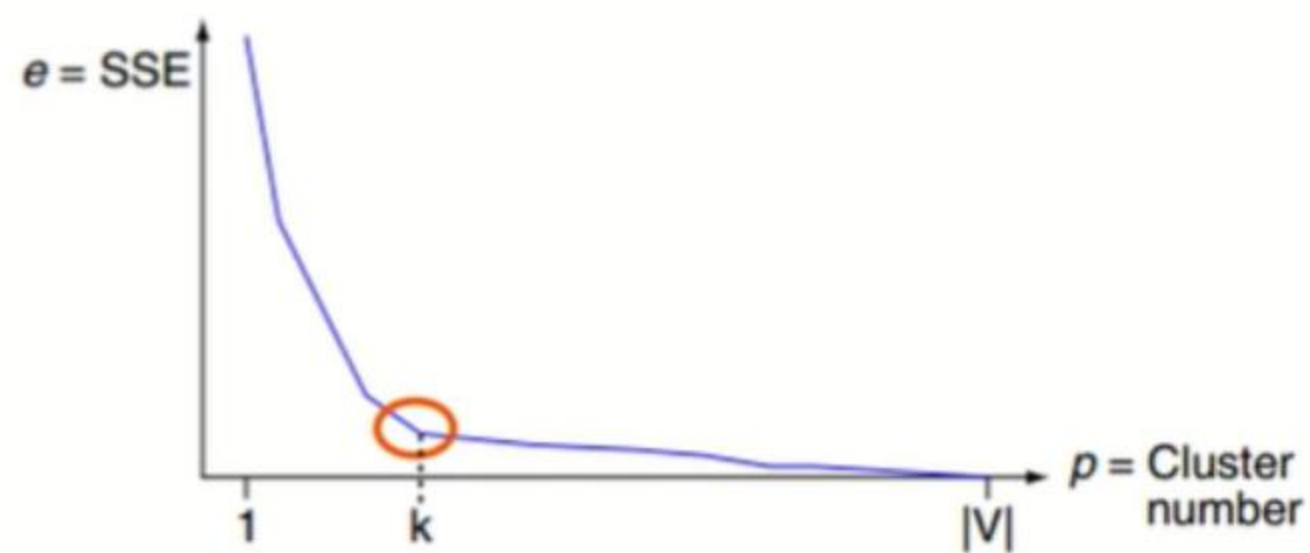
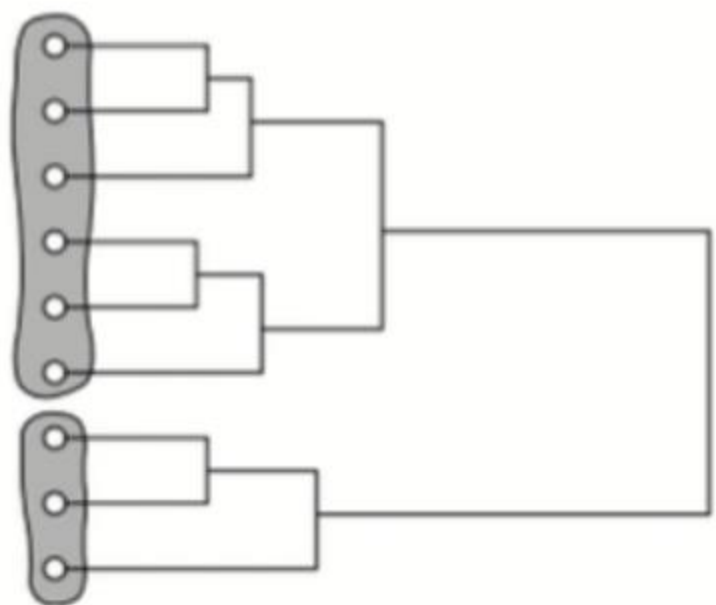
## UNIT 05 | 최적의 군집 수 및 모델 평가

### 최적의 군집 수 결정

Clustering

### 어떻게 최적의 군집 수를 결정할 것인가?

- ✓ 다양한 군집 수에 대해 성능 평가 지표를 도출하여 최적의 군집 수를 선택한다.
- ✓ Elbow point에서 최적 군집 수가 결정되는 경우가 일반적이다.





## UNIT 05 | 최적의 군집 수 및 모델 평가

### 결과 측정 및 평가

Clustering

**"분류 알고리즘처럼 모든 상황에 적용가능한 평가 지표가 부재한다."**



#### 내부 평가 지표

- Dunn Index, Silhouette, Sum of Squared Error 등



#### 외부 평가 지표

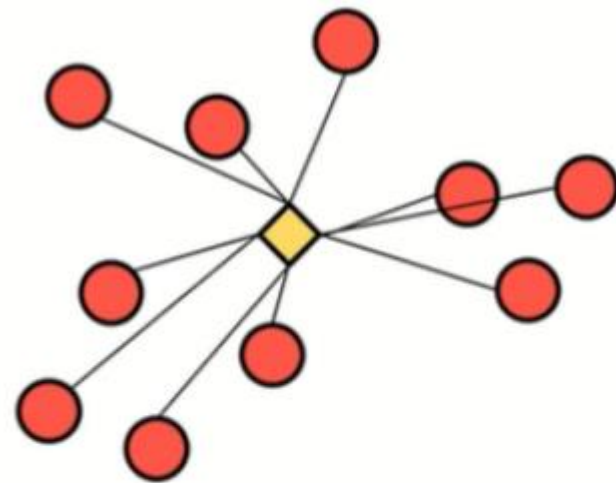
- Rank Index, Jaccard Coefficient, Folks and Mallows Index 등

## UNIT 05 | 최적의 군집 수 및 모델 평가

### 결과 측정 및 평가

Clustering

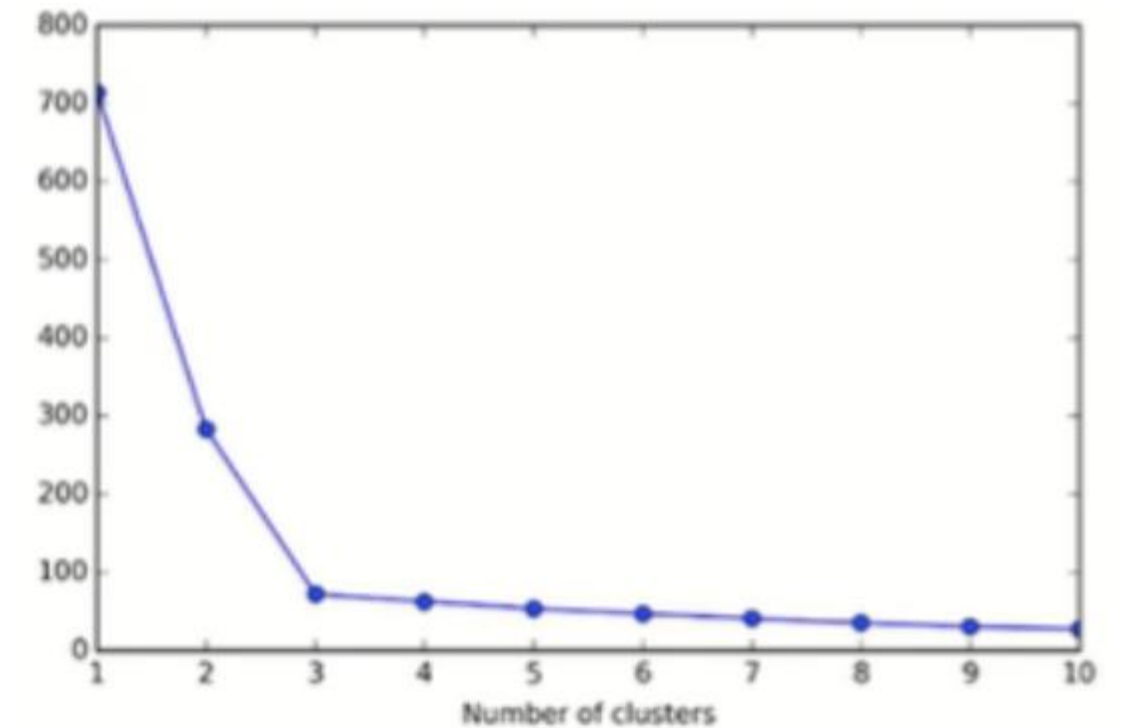
### 평가 지표1: SUM OF SQUARED ERROR (SSE)



● : 관측치 ( $x$ )

◆ : 중심 ( $c_i$ )

$$SEE = \sum_{i=1}^K \sum_{x \in C_i} dist(x, c_i)^2$$



## UNIT 05 | 최적의 군집 수 및 모델 평가

### 결과 측정 및 평가

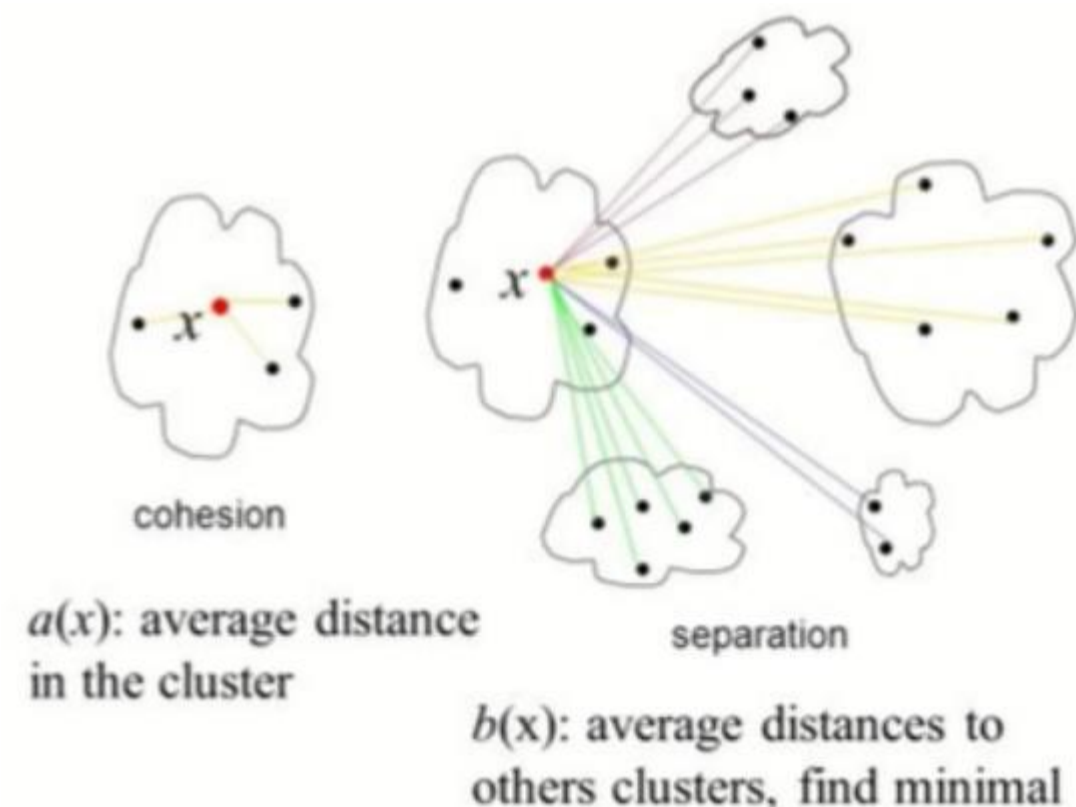
Clustering

### 평가 지표2: SILHOUETTE 통계량

- ✓ a(i): 관측치 i로부터 같은 군집 내에 있는 모든 다른 개체들 사이의 평균 거리 - 군집 내 응집도(cohesion)
- ✓ b(i): 관측치 i로부터 다른 군집 내에 있는 개체들 사이의 평균 거리 중 최솟값 - 군집 간 분리도(separation)
- ✓ 일반적으로  $\bar{S}$ 의 값 0.5보다 크면 군집 결과가 타당하다고 볼 수 있다.
- ✓ -1에 가까우면 클러스터링이 전혀 되지 않았다고 볼 수 있다.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad -1 \leq s(i) \leq 1 \quad \bar{S} = \frac{1}{n} \sum_{i=1}^n S(i)$$

군집 안의 데이터가 자신이 속한 군집 안의 다른 데이터와 얼마나 유사하며, 다른 군집에 속한 데이터와 얼마나 차이가 나는지 측정



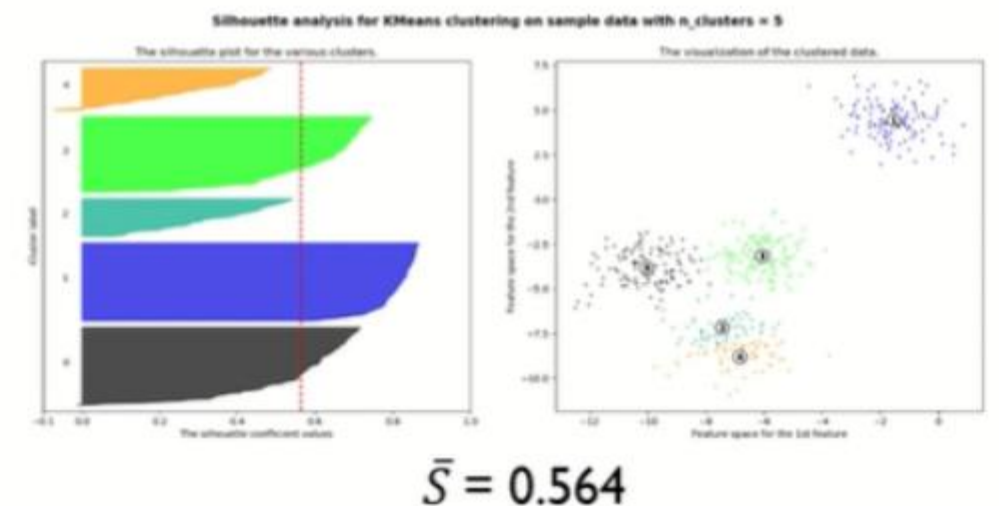
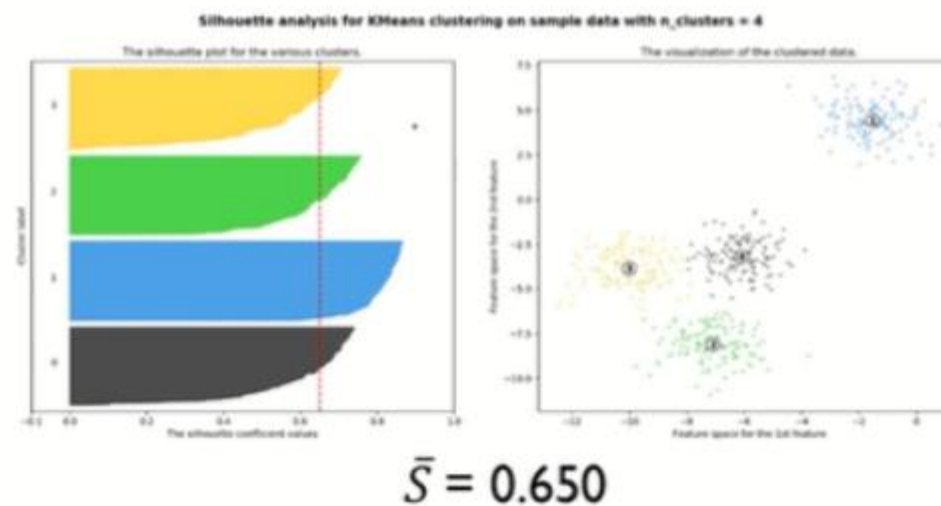
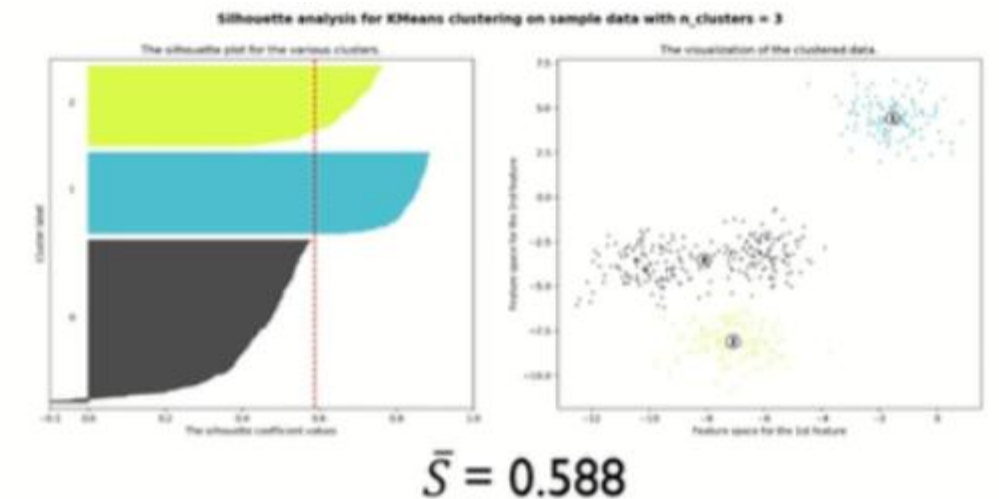
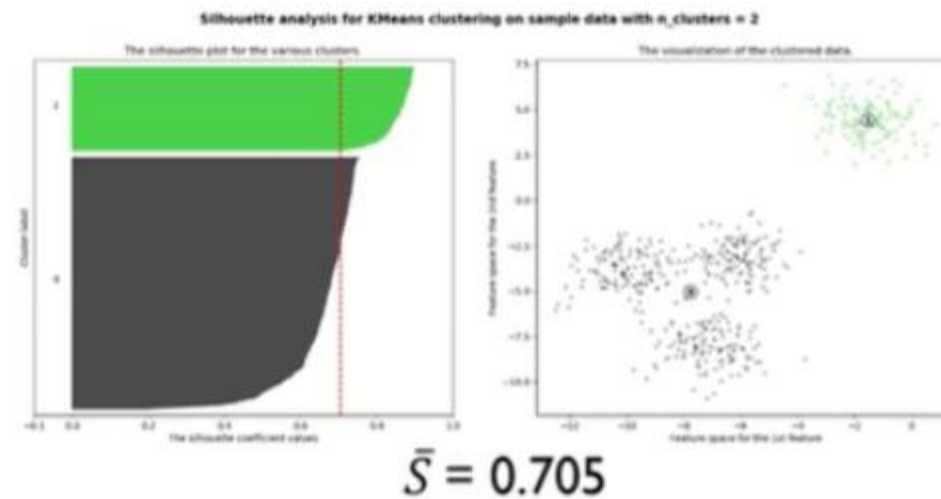
# UNIT 05 | 최적의 군집 수 및 모델 평가

## 결과 측정 및 평가

Clustering

### 평가 지표2: SILHOUETTE 통계량

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad -1 \leq s(i) \leq 1$$



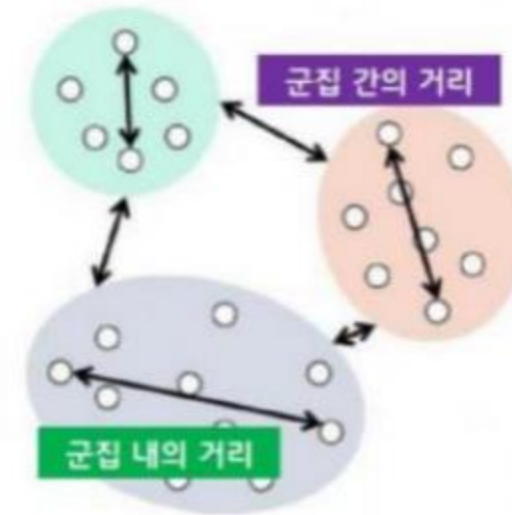
## UNIT 05 | 최적의 군집 수 및 모델 평가

### 결과 측정 및 평가

Clustering

### 평가 지표3: DUNN'S INDEX

$$DI = \frac{\text{군집과 군집 사이의 거리 중 최솟값}}{\text{군집 내 객체 간 거리 중 최댓값}}$$



- ✔ 군집과 군집 사이의 거리가 클수록, 군집 내 객체가 거리가 작을 수록 좋은 모델이므로 DI가 큰 모델이 좋은 모델이다.

### "각자 해보기!"

#### ❶ Clustering 실습 1

- 계층적 군집화, K-평균 군집화, DBSCAN 간단하게 구현해보기

#### ❷ Clustering 실습 2

- 계층적 군집화, K-평균 군집화, DBSCAN 파라미터를 바꿔가며 비교해보기
- 고려대학교 김성범 교수님 [파이썬 실습] 군집 분석 자료  
[https://www.youtube.com/watch?v=\\_U7AUJqzIJs&list=PLpIPLT0Pf7lo8pMhxJ6vhM1chReYa8KIn&index=1](https://www.youtube.com/watch?v=_U7AUJqzIJs&list=PLpIPLT0Pf7lo8pMhxJ6vhM1chReYa8KIn&index=1)
- 파이썬 및 모델 구현에 능숙하신 분들은 해당 파일을 꼭 따라서 실습을 진행해주시고, 구현이 어려우신 분들은 위 링크의 강의를 참고해주세요.

# 과제

## 과제

Homework



### Clustering 해보기

1. Preprocessing / EDA
2. Clustering (수업시간에 배운 세 가지 방법 + @)
3. Evaluation

데이터: <https://www.kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python>

## REFERENCE

### 참고자료

Reference

- ✔ 투빅스 15기 김현지님 강의자료
- ✔ 이 자료는 고려대학교 김성범 교수님의 핵심 머신러닝 군집분석 (Clustering Analysis) 강의를 참고하여 제작했습니다.  
[https://www.youtube.com/watch?v=8zB-\\_LrAraw](https://www.youtube.com/watch?v=8zB-_LrAraw)
- ✔ 투빅스 14기 김민경님 강의자료
- ✔ 고려대학교 강필성 교수님 비즈니스 애널리틱스 Clustering - DBSCAN 강의  
<https://www.youtube.com/watch?v=PuvH38UpgNU>
- ✔ 파이썬 라이브러리를 활용한 머신러닝: 사이킷 런 핵심 개발자가 쓴 머신러닝과 데이터 과학 실무서



## REFERENCE

### 참고자료

Reference

#### 공부자료 추천

- ✓ 이 강의자료와 같은 강의를 글로 정리해놓은 블로그  
<https://ratsgo.github.io/machine%20learning/2017/04/16/clustering/>  
<https://ratsgo.github.io/machine%20learning/2017/04/18/HC/>  
<https://ratsgo.github.io/machine%20learning/2017/04/19/KC/>  
<https://ratsgo.github.io/machine%20learning/2017/05/01/SOM/>
- ✓ 고려대학교 강필성 교수님 다변량데이터분석 강의:  
<https://www.youtube.com/watch?v=k885zMo0jQs&list=PLetSlH8YjlfWKLpMp-r6enJvnk6L93wz2&index=32>  
<https://www.youtube.com/watch?v=QJB6eClNQVI&list=PLetSlH8YjlfWKLpMp-r6enJvnk6L93wz2&index=33>  
<https://www.youtube.com/watch?v=sMMbAgKVwAk&list=PLetSlH8YjlfWKLpMp-r6enJvnk6L93wz2&index=34>  
[https://www.youtube.com/watch?v=O\\_EigN9iF6E&list=PLetSlH8YjlfWKLpMp-r6enJvnk6L93wz2&index=35](https://www.youtube.com/watch?v=O_EigN9iF6E&list=PLetSlH8YjlfWKLpMp-r6enJvnk6L93wz2&index=35)

**Q&A**

**들어주셔서 감사합니다!**