

# Chapter.7 Clustering - DBSCAN



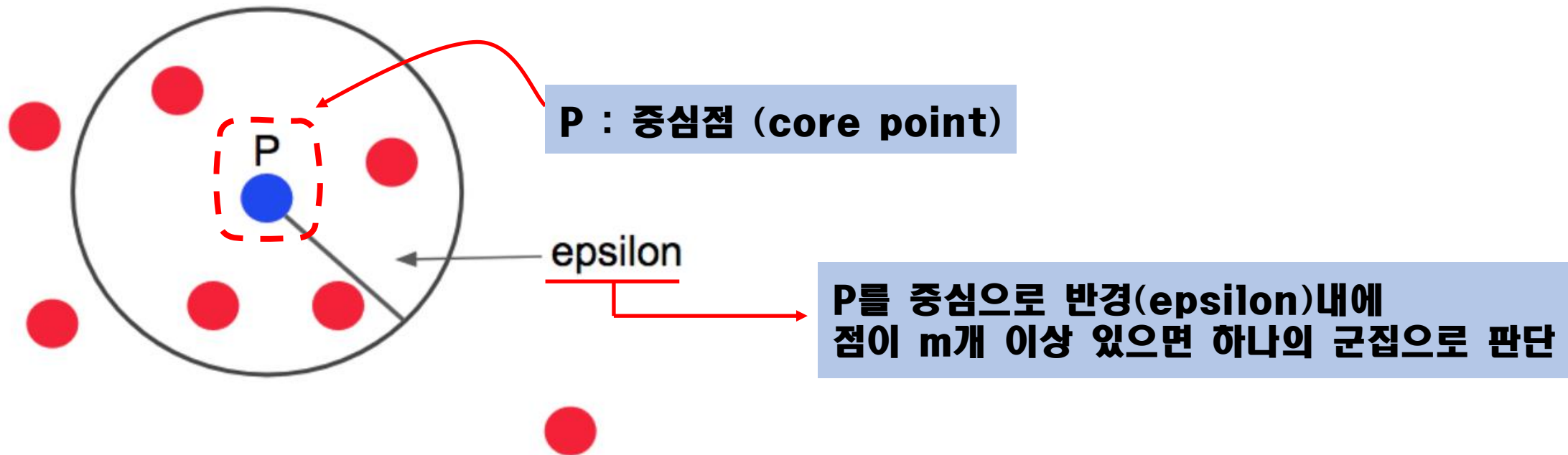
**Kyungnam DA.  
study group**

# 1. DBSCAN (밀도 기반 클러스터링)

DBSCAN ( Density-Based Spatial Clustering of Application with Noise)

군집간 거리를 이용한 k-means, Hierarchical clustering과 달리 밀도 기반의 클러스터링 방법  
관찰값이 세밀하게 몰려 있어서 밀도가 높은 부분을 클러스터링 하는 방식

즉, 어느점을 기준으로 반경  $\epsilon$  내에 점이  $n$ 개 이상 있으면 하나의 군집으로 인식하는 방식

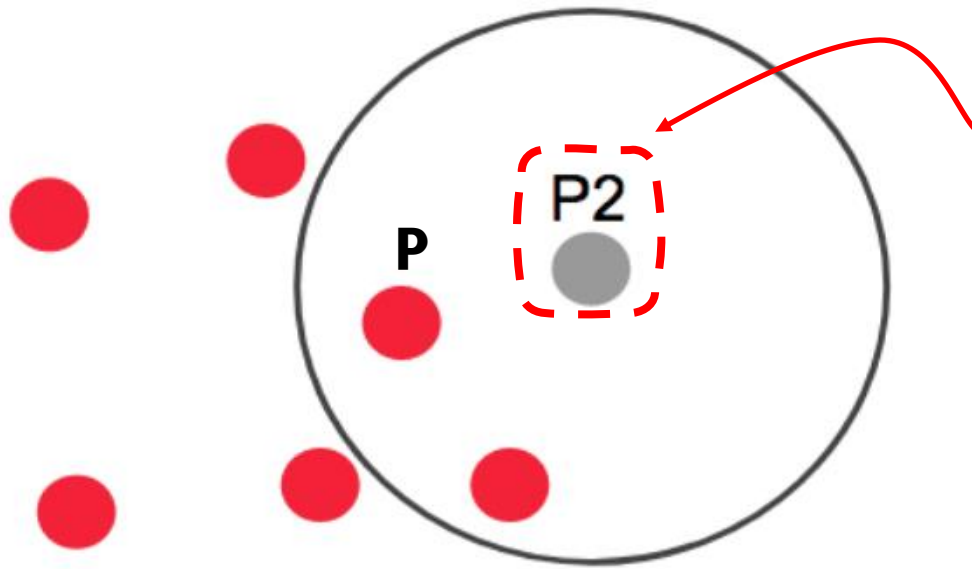


# 1. DBSCAN (밀도 기반 클러스터링)

DBSCAN ( Density-Based Spatial Clustering of Application with Noise)

군집간 거리를 이용한 k-means, Hierarchical clustering과 달리 밀도 기반의 클러스터링 방법  
관찰값이 세밀하게 몰려 있어서 밀도가 높은 부분을 클러스터링 하는 방식

즉, 어느점을 기준으로 반경  $x$  내에 점이  $n$  개 이상 있으면 하나의 군집으로 인식하는 방식



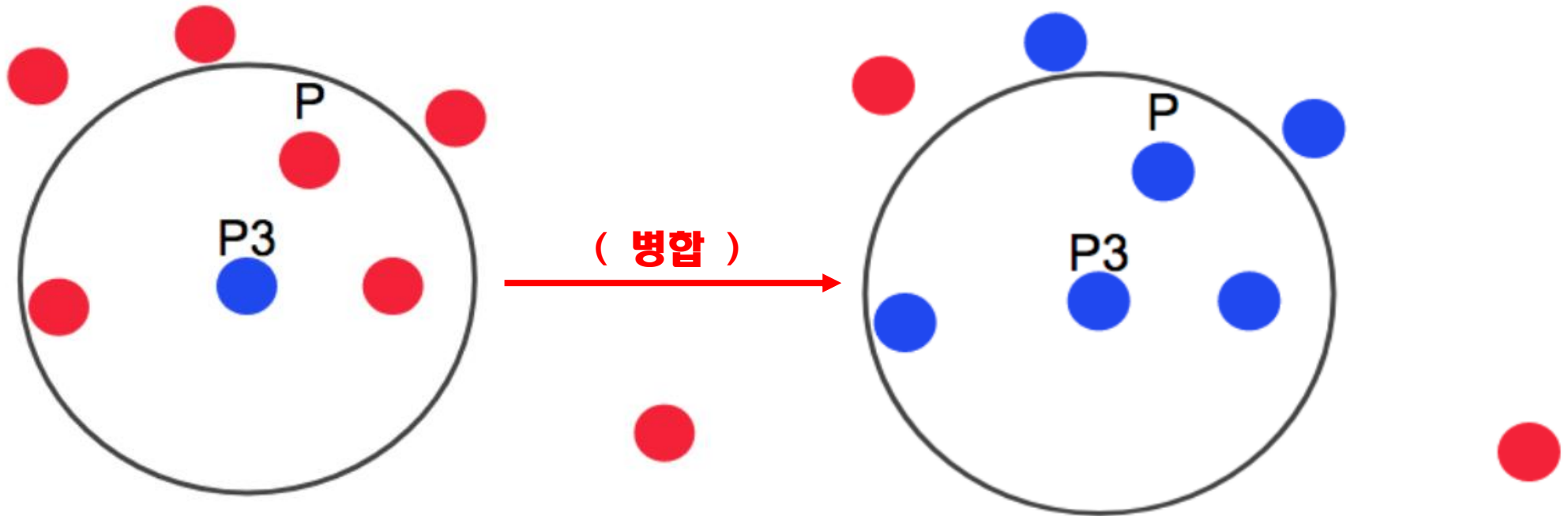
Center point는 되지 못하지만  $P$ 의 군집에 속하는 점 :  
border point(경계점)

# 1. DBSCAN (밀도 기반 클러스터링)

DBSCAN ( Density-Based Spatial Clustering of Application with Noise)

군집간 거리를 이용한 k-means, Hierarchical clustering과 달리 밀도 기반의 클러스터링 방법  
관찰값이 세밀하게 몰려 있어서 밀도가 높은 부분을 클러스터링 하는 방식

즉, 어느점을 기준으로 반경  $x$  내에 점이  $n$  개 이상 있으면 하나의 군집으로 인식하는 방식

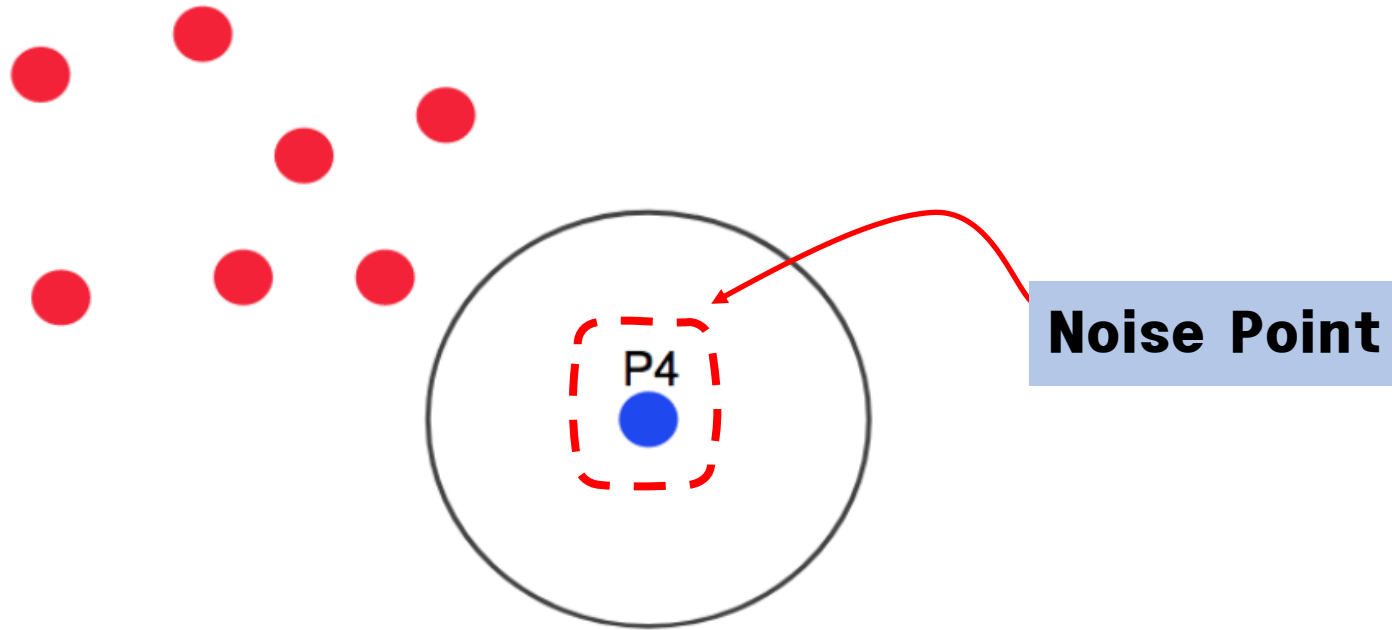


# 1. DBSCAN (밀도 기반 클러스터링)

DBSCAN ( Density-Based Spatial Clustering of Application with Noise)

군집간 거리를 이용한 k-means, Hierarchical clustering과 달리 밀도 기반의 클러스터링 방법  
관찰값이 세밀하게 몰려 있어서 밀도가 높은 부분을 클러스터링 하는 방식

즉, 어느점을 기준으로 반경  $x$  내에 점이  $n$  개 이상 있으면 하나의 군집으로 인식하는 방식

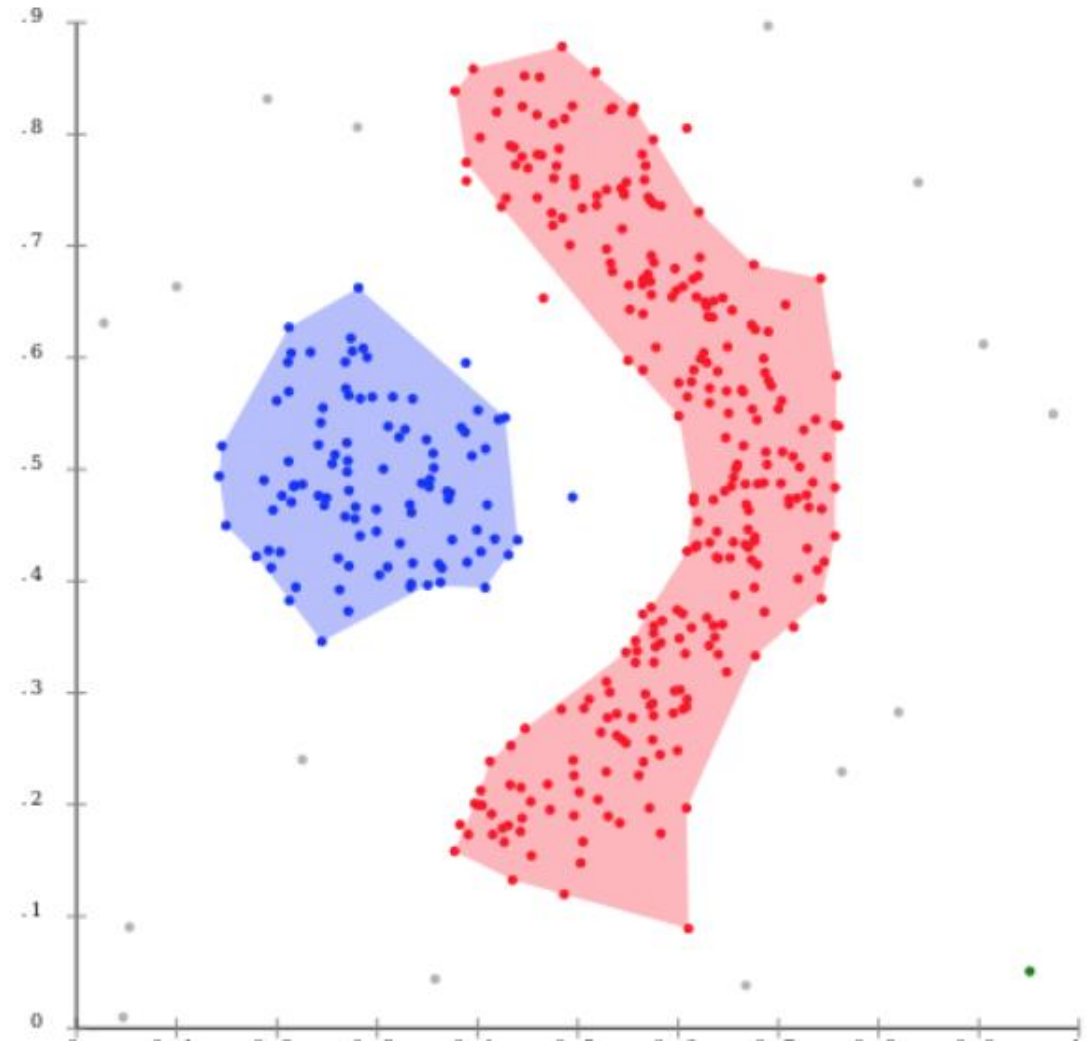


# 1. DBSCAN (밀도 기반 클러스터링)

DBSCAN ( Density-Based Spatial Clustering of Application with Noise)

점을 중심으로 epsilon 반경내에 minPts(최소 점)

Core point가 서로 다른 core point 군집의 일부  
군집에 속하지만 스스로 core point가 되지 못하는  
어느 군집에도 속하지 않는 점을 noise point



# Summary

각 군집화 기법은 장/단점을 가지고 있으므로 **군집화 하려는 데이터의 특성에 맞게 선택**

**K-means** : 거리 기반으로 군집 중심점을 이동시키며 군집화, 매우 직관적이거나 복잡한 데이터에 적용 힘들

**Mean Shift** : 데이터가 모여 있는 밀도가 가장 높은 쪽으로 군집 중심점을 이동하며 군집화

**GMM** : 전체 데이터 세트에서 서로 다른 정규 분포 형태를 추출해 각각 군집화, k-means보다 유연함

**DBSCAN** : 밀도기반 군집화, 직관적인 알고리즘임에도 복잡한 데이터세트에도 효과적인 군집화 가능