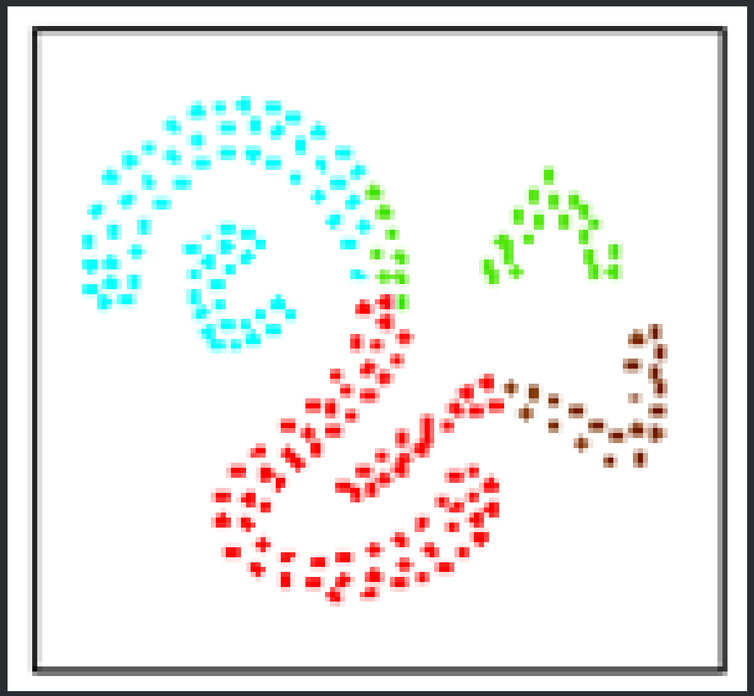
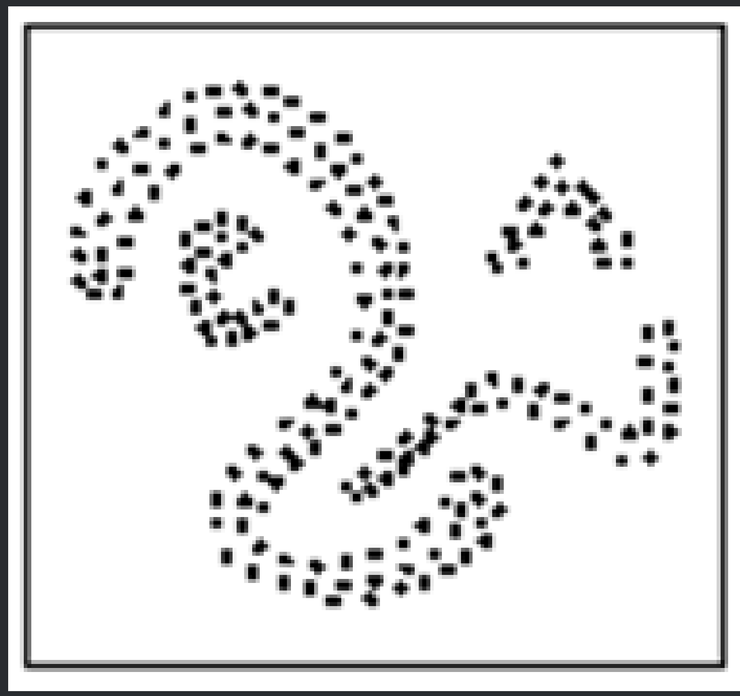
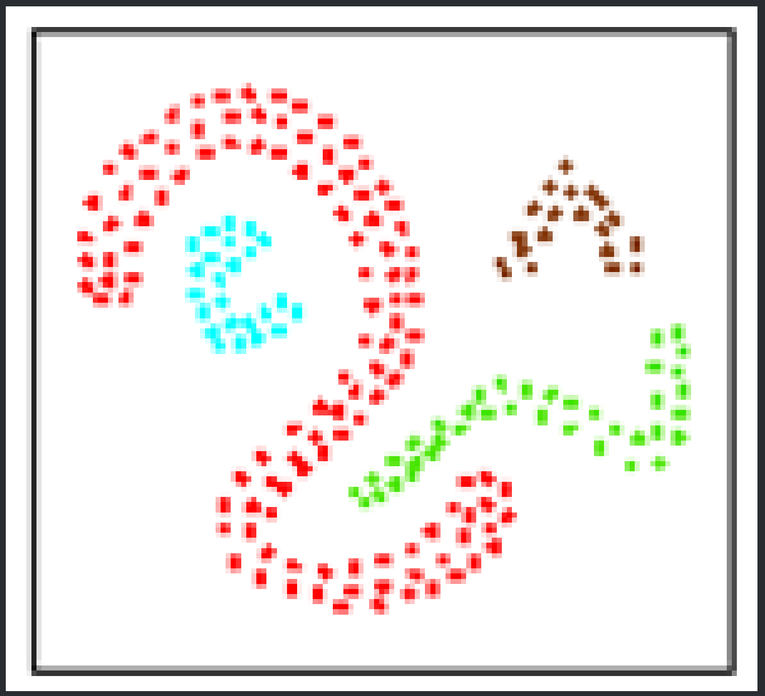
DBSCAN





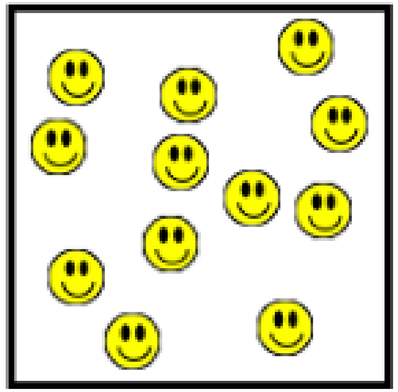


Density-Based Spatial Clustering of Applications with Noise

그룹화의 방식

Density of Matter

How tightly packed matter is. The amount of mass in a given space.

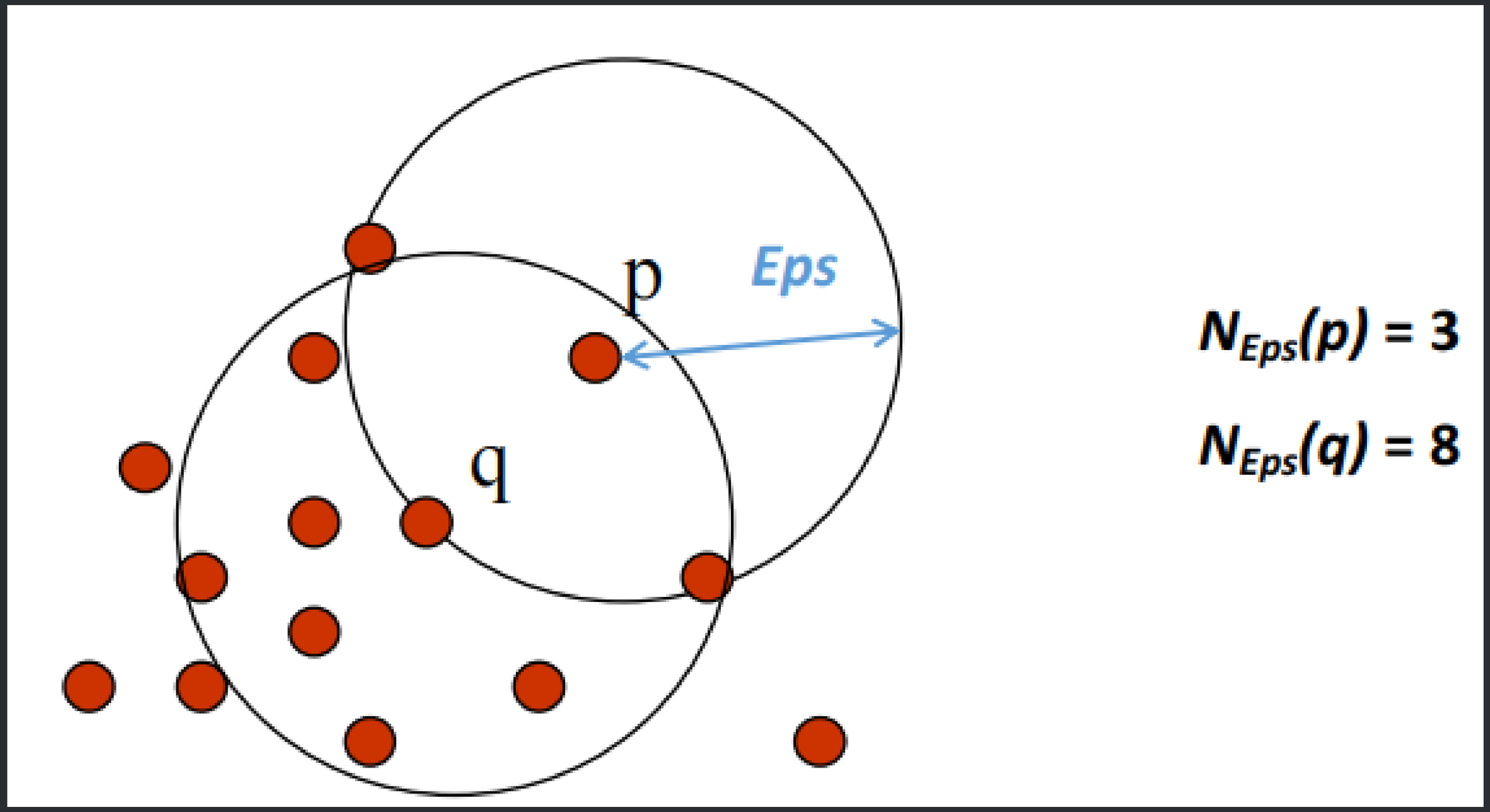


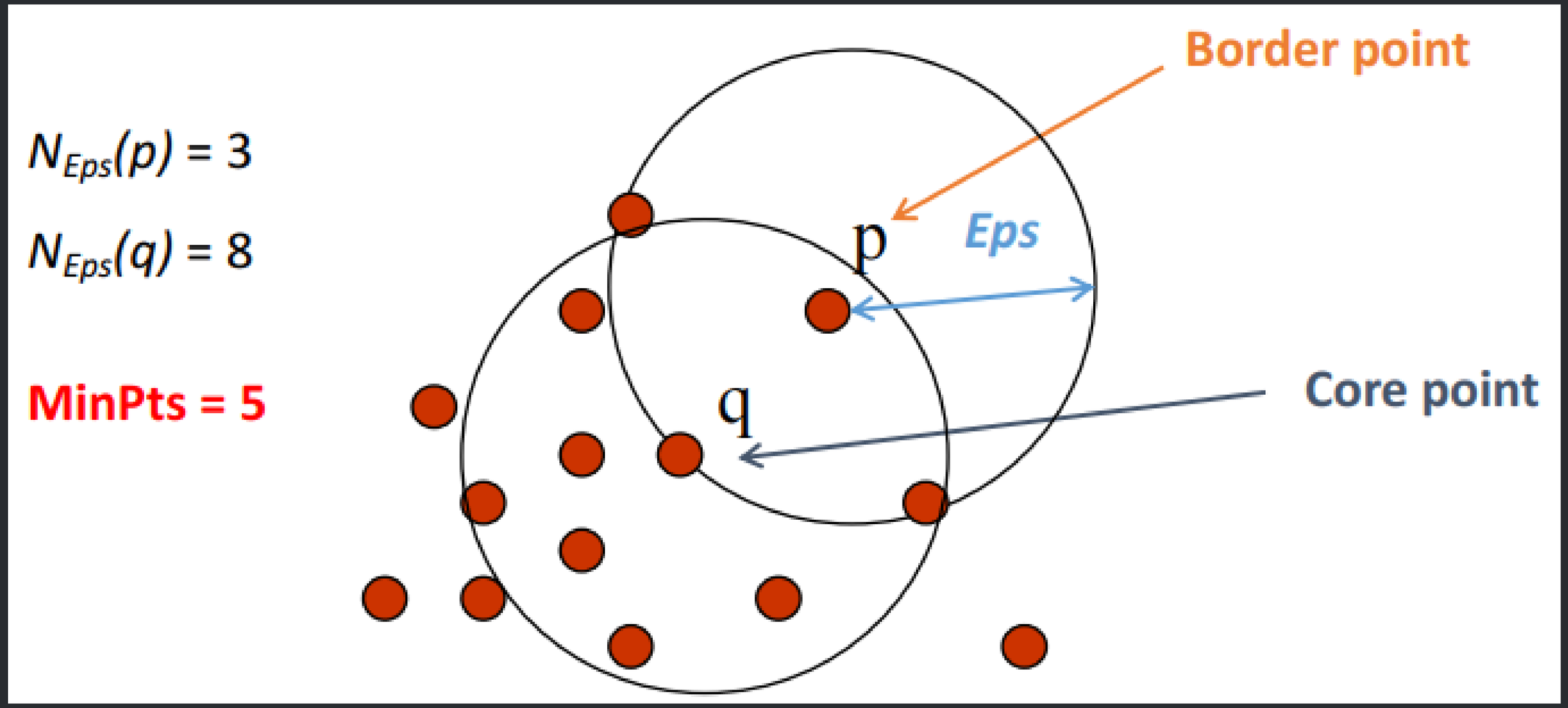
Gas

Liquid

Solid

Less dense → **More dense**


p
Eps
q
$N_{Eps}(p) = 3$
$N_{Eps}(q) = 8$



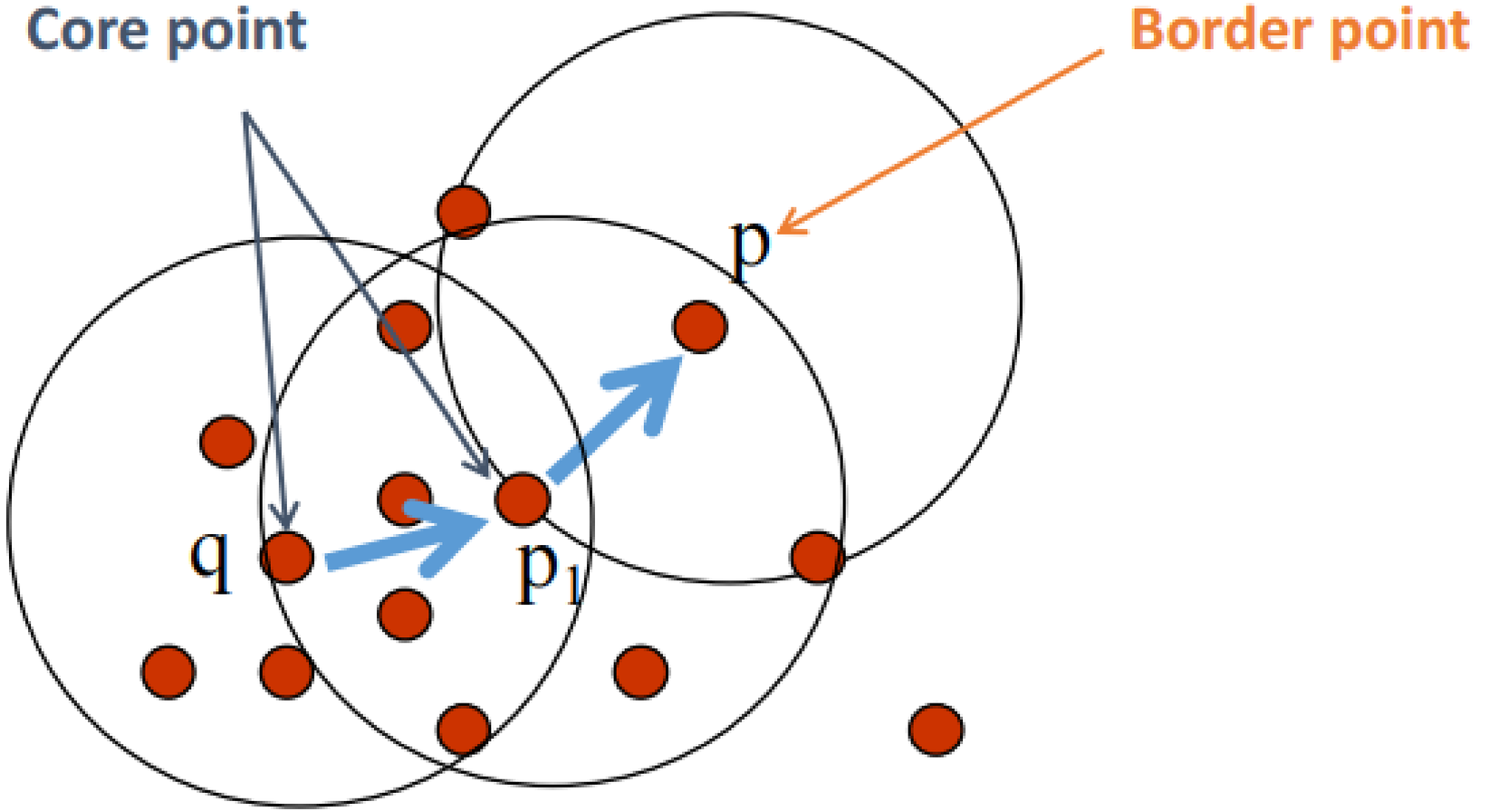
MinPts를 만족한 Core Point의 radius 안에
다른 CorePoint가 포함될 때 해당 포인트를 CorePoint로 하여

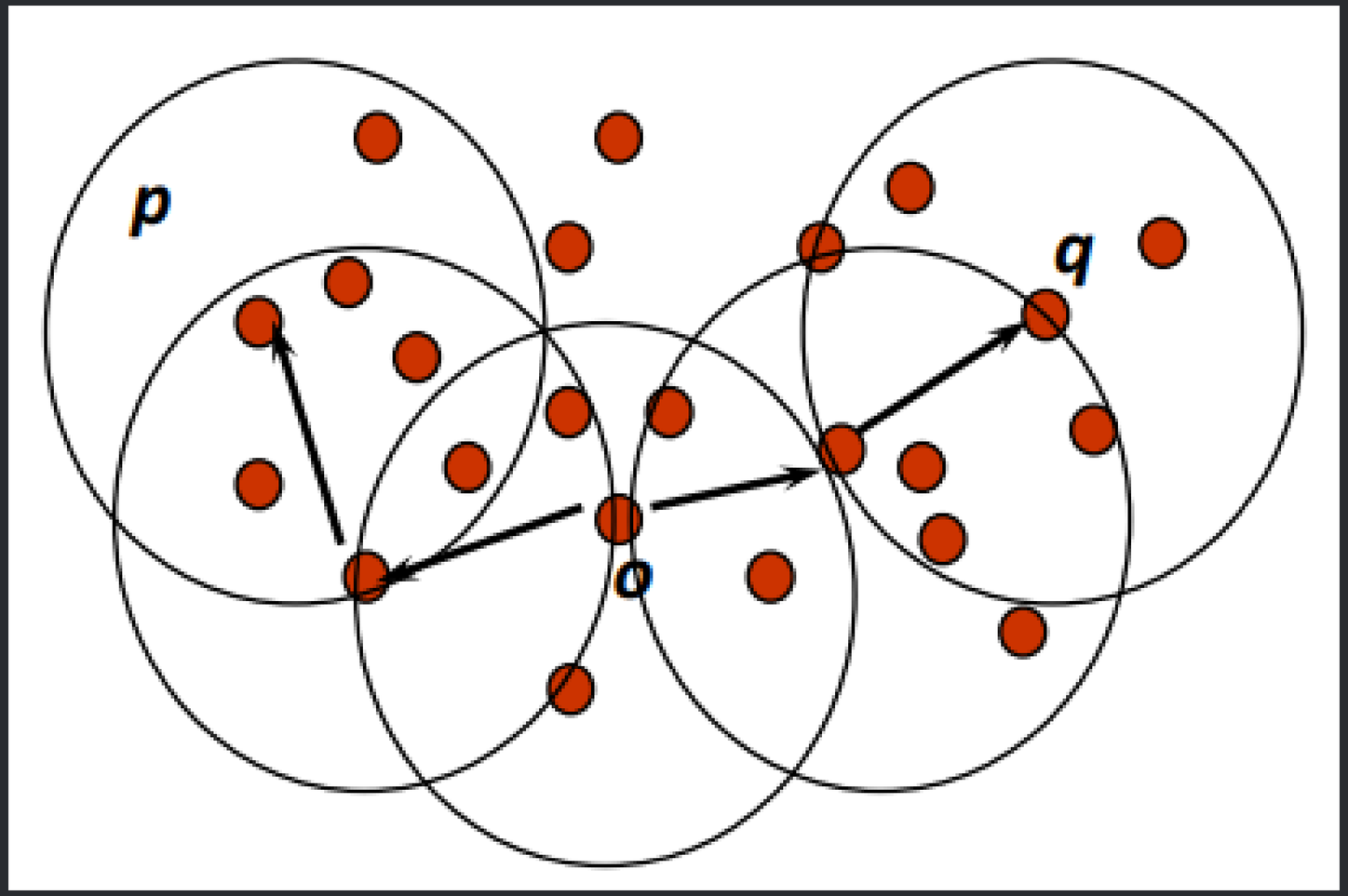
$N_{Eps}(p) = 3$

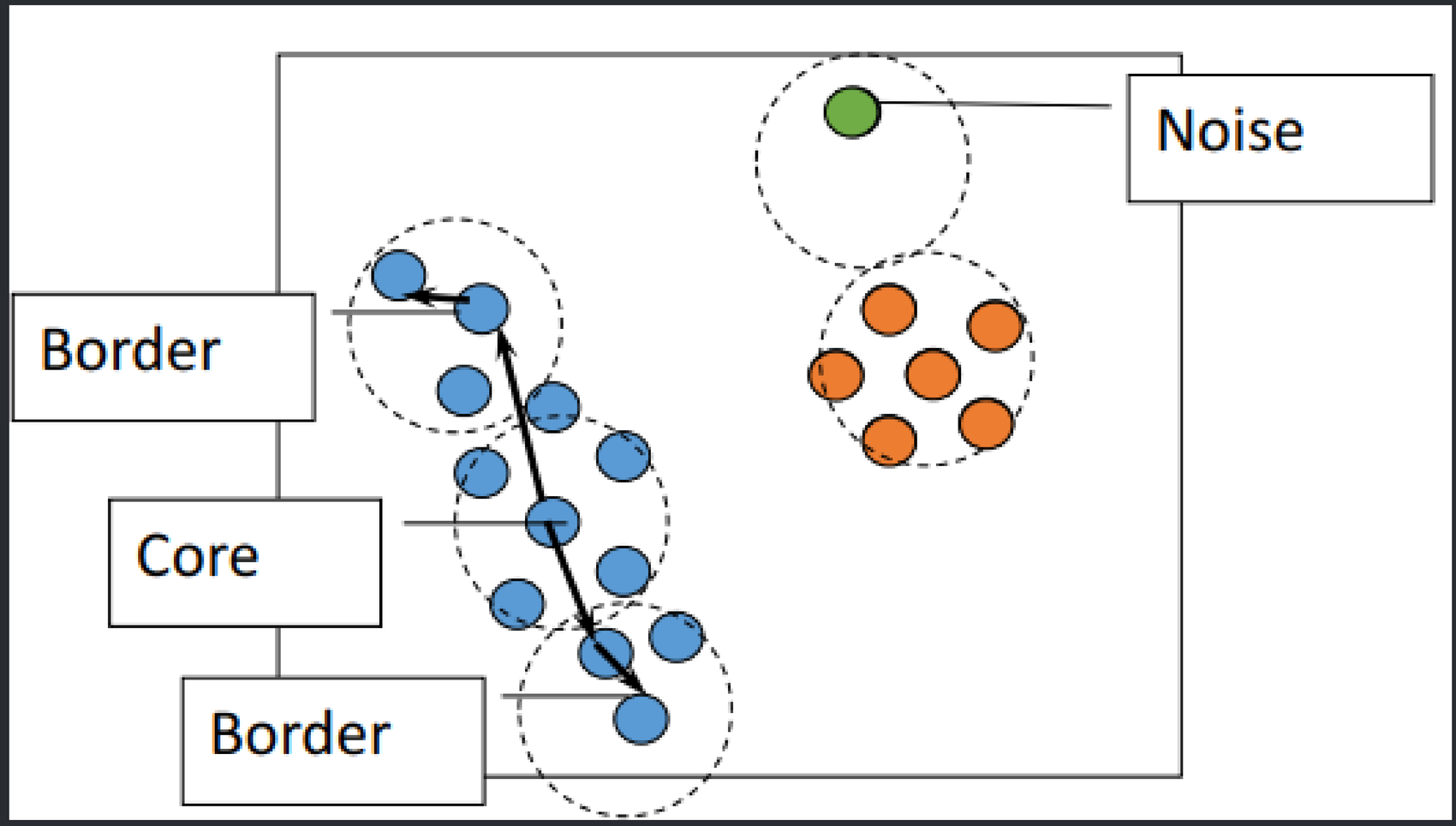
$N_{Eps}(p_1) = 8$

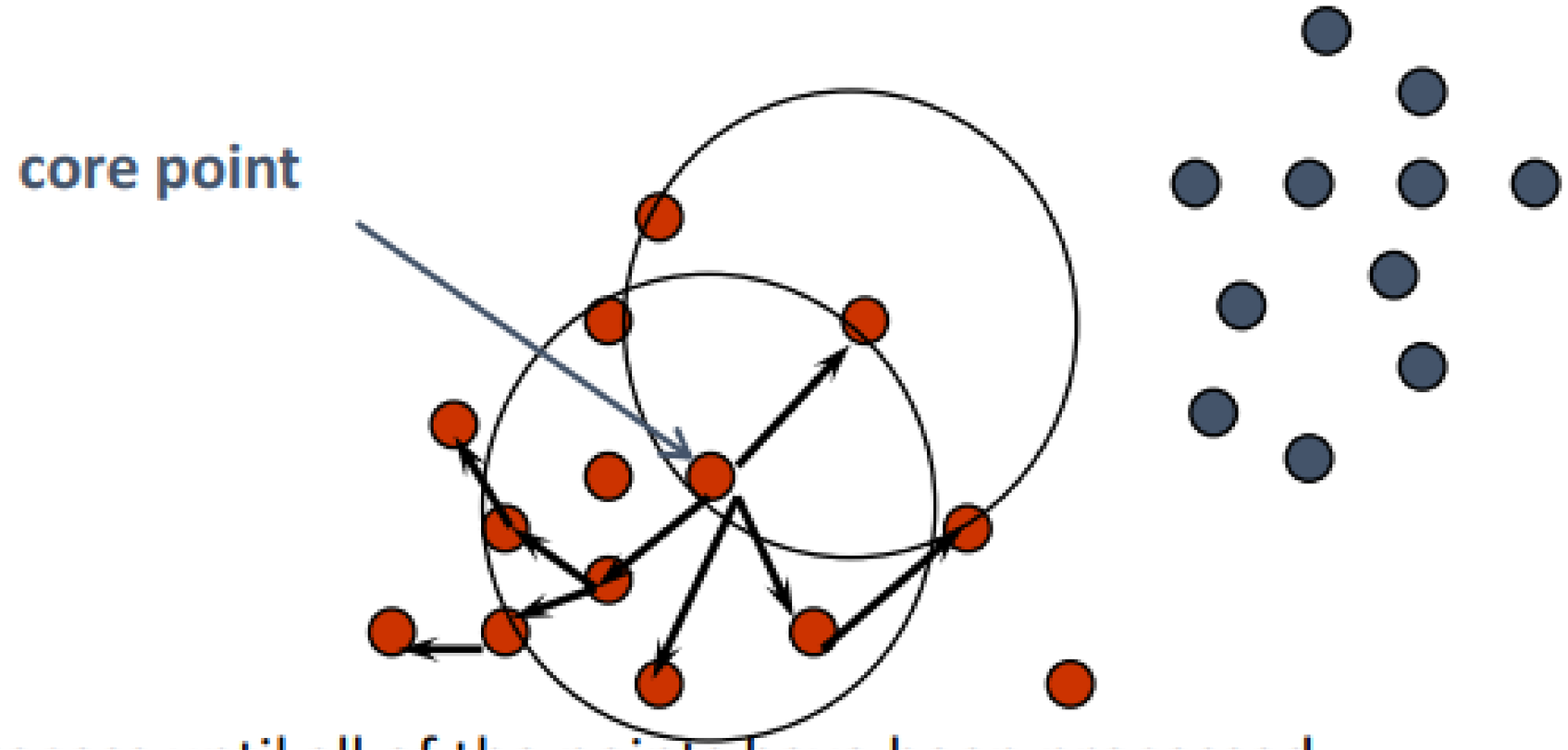
$N_{Eps}(q) = 8$

MinPts = 5



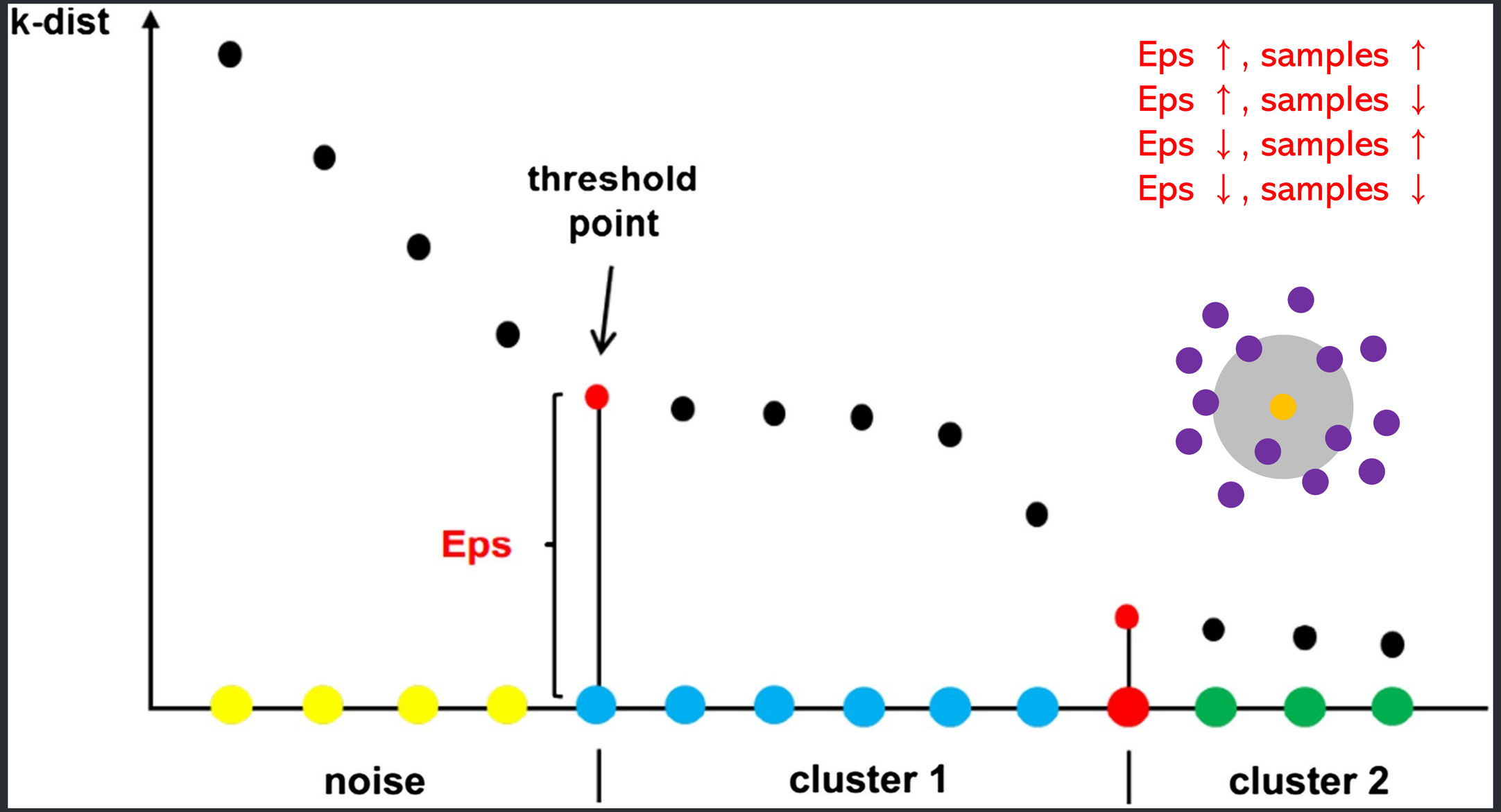

p
q
o


Noise
Border
Core
Border

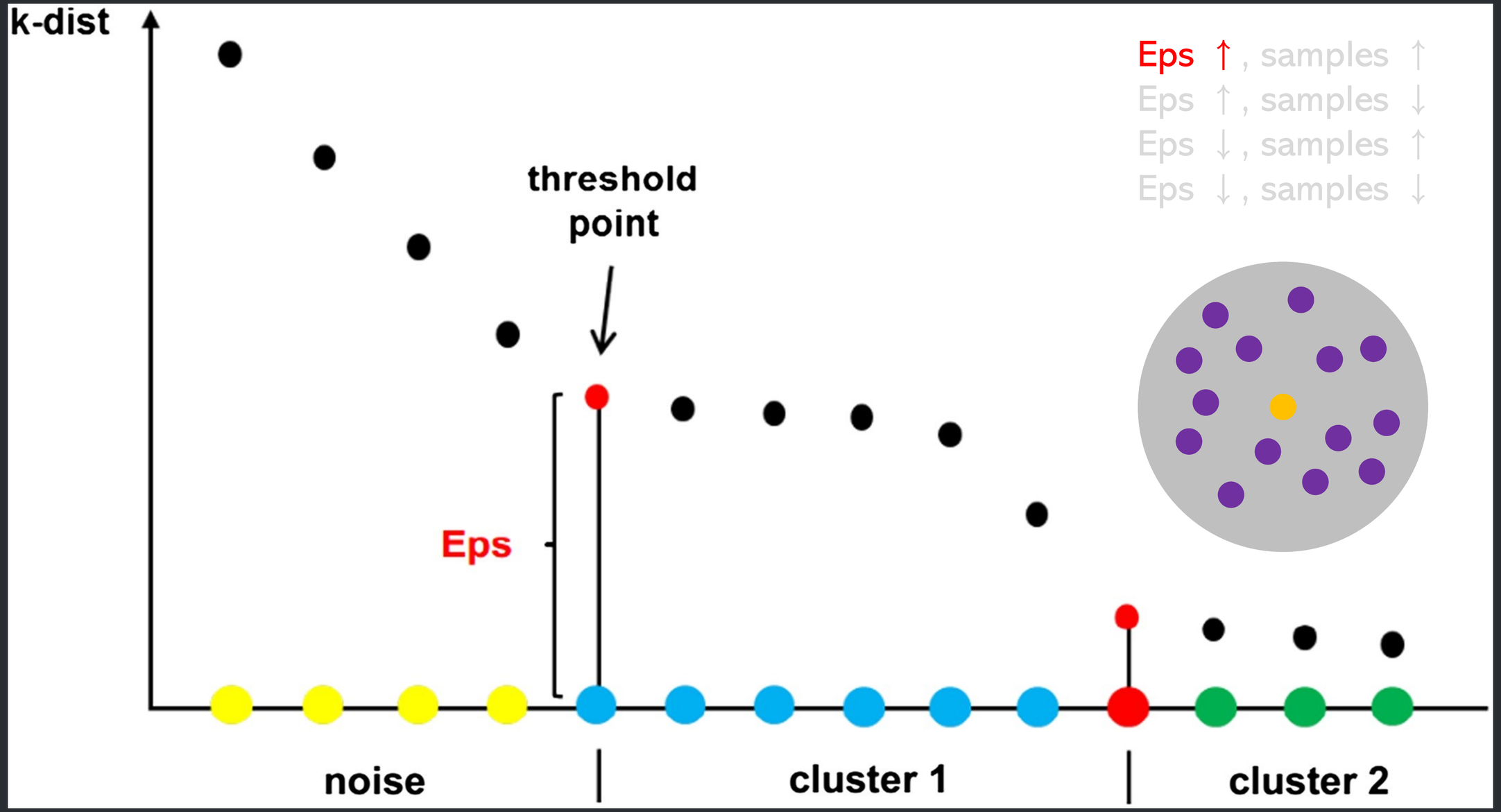


- Continue the process until all of the points have been processed.

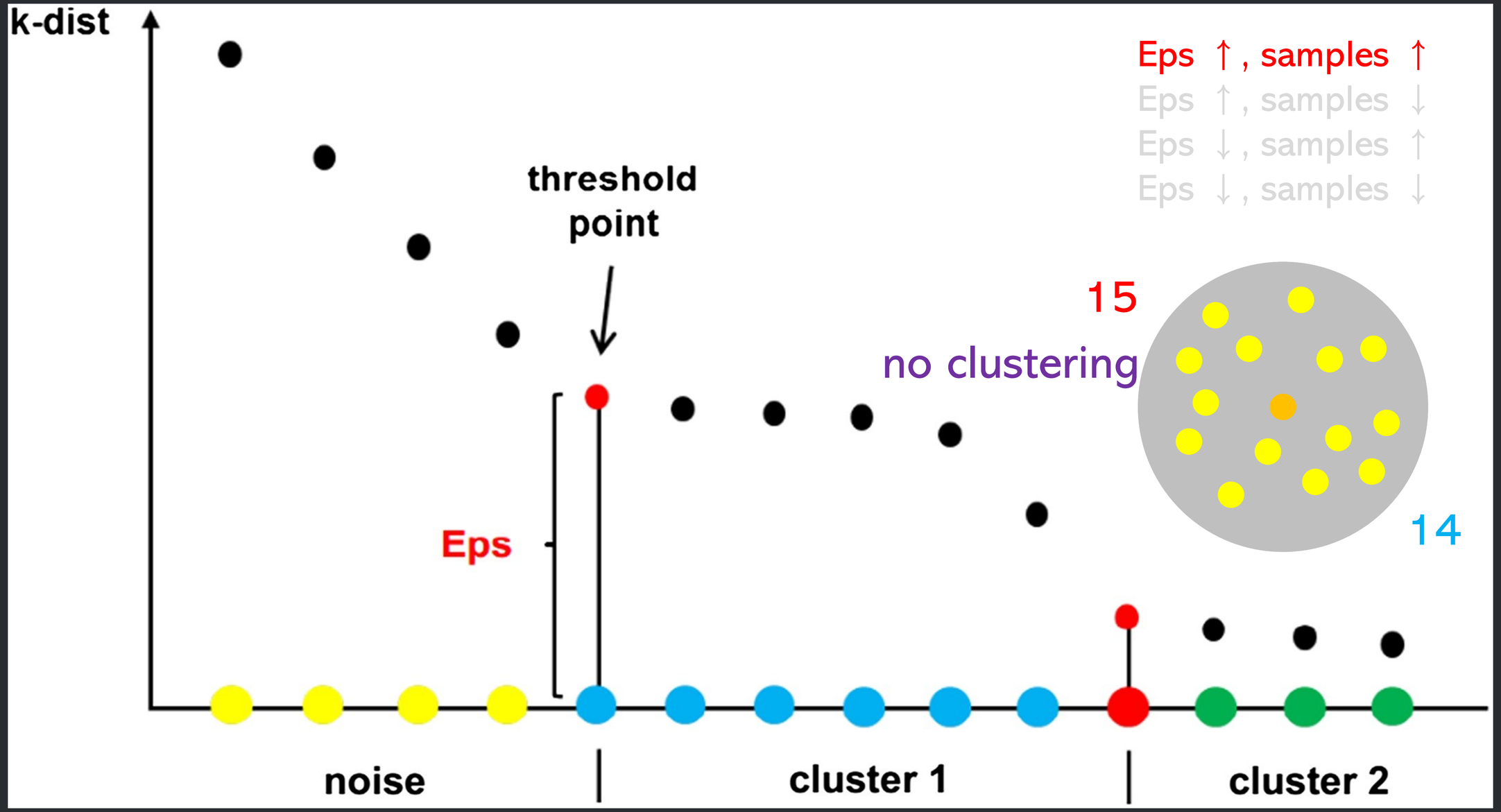
Heuristics


k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓

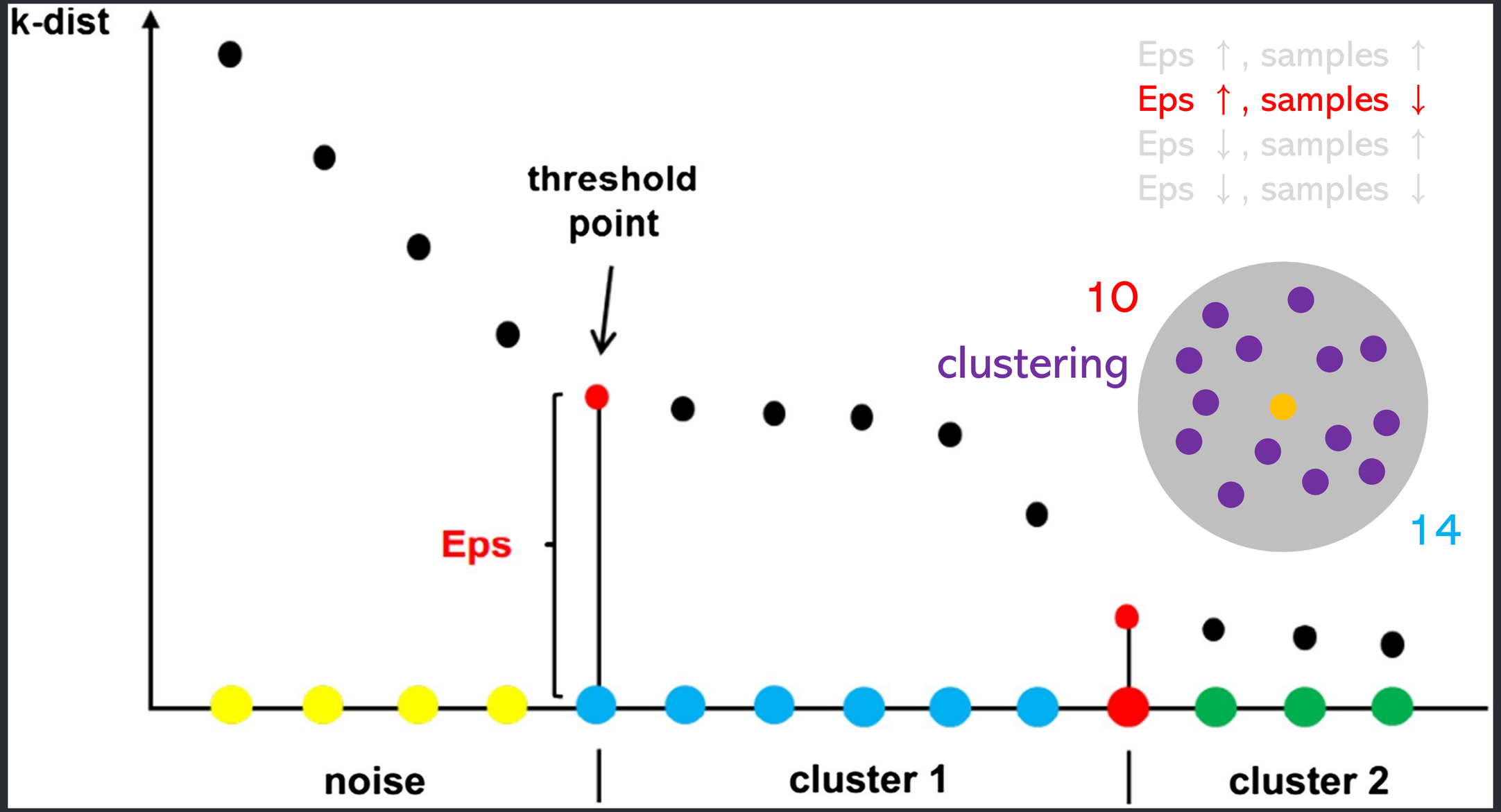
Heuristics


k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓

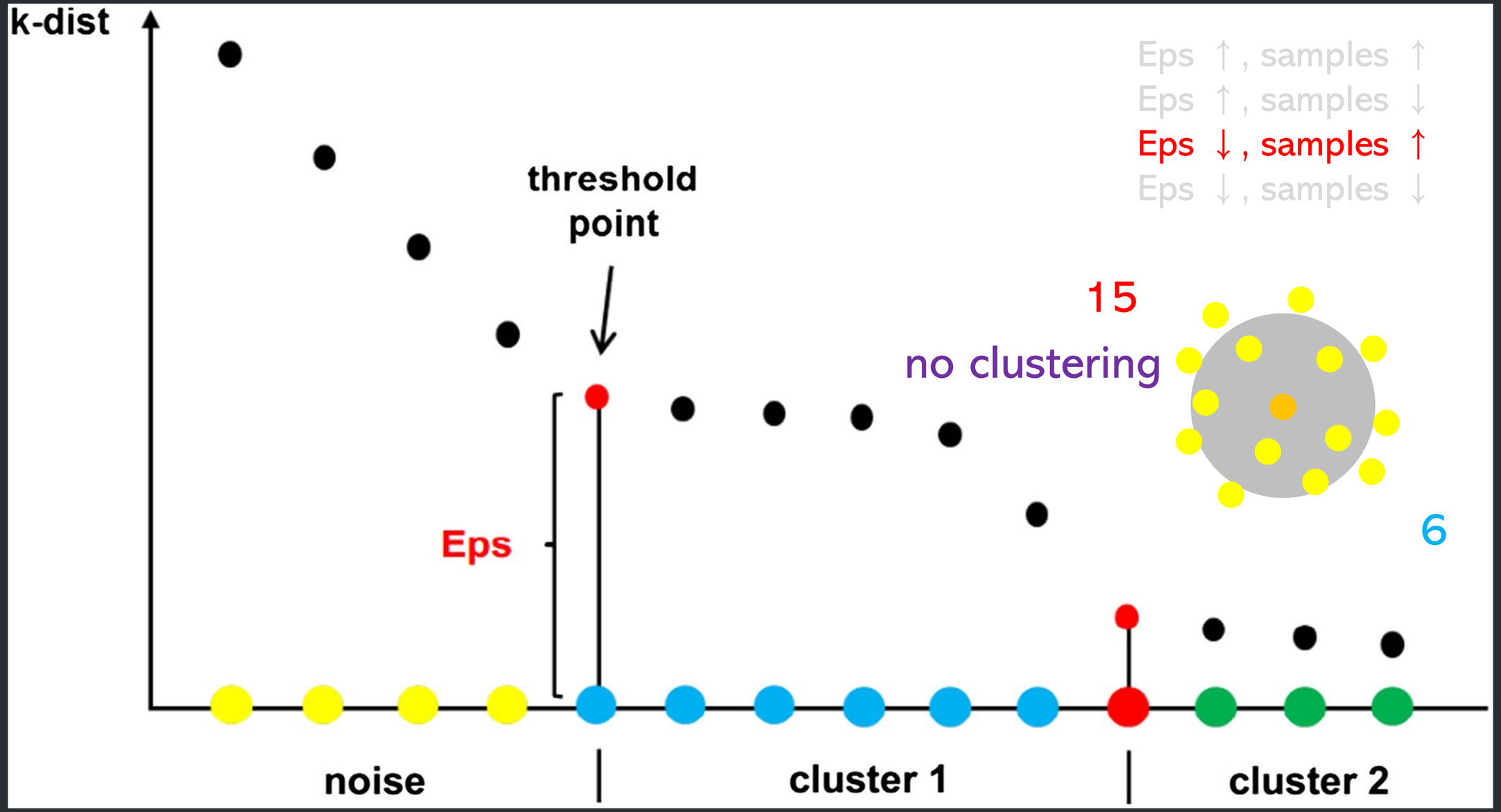
Heuristics


k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓
15
no clustering
14

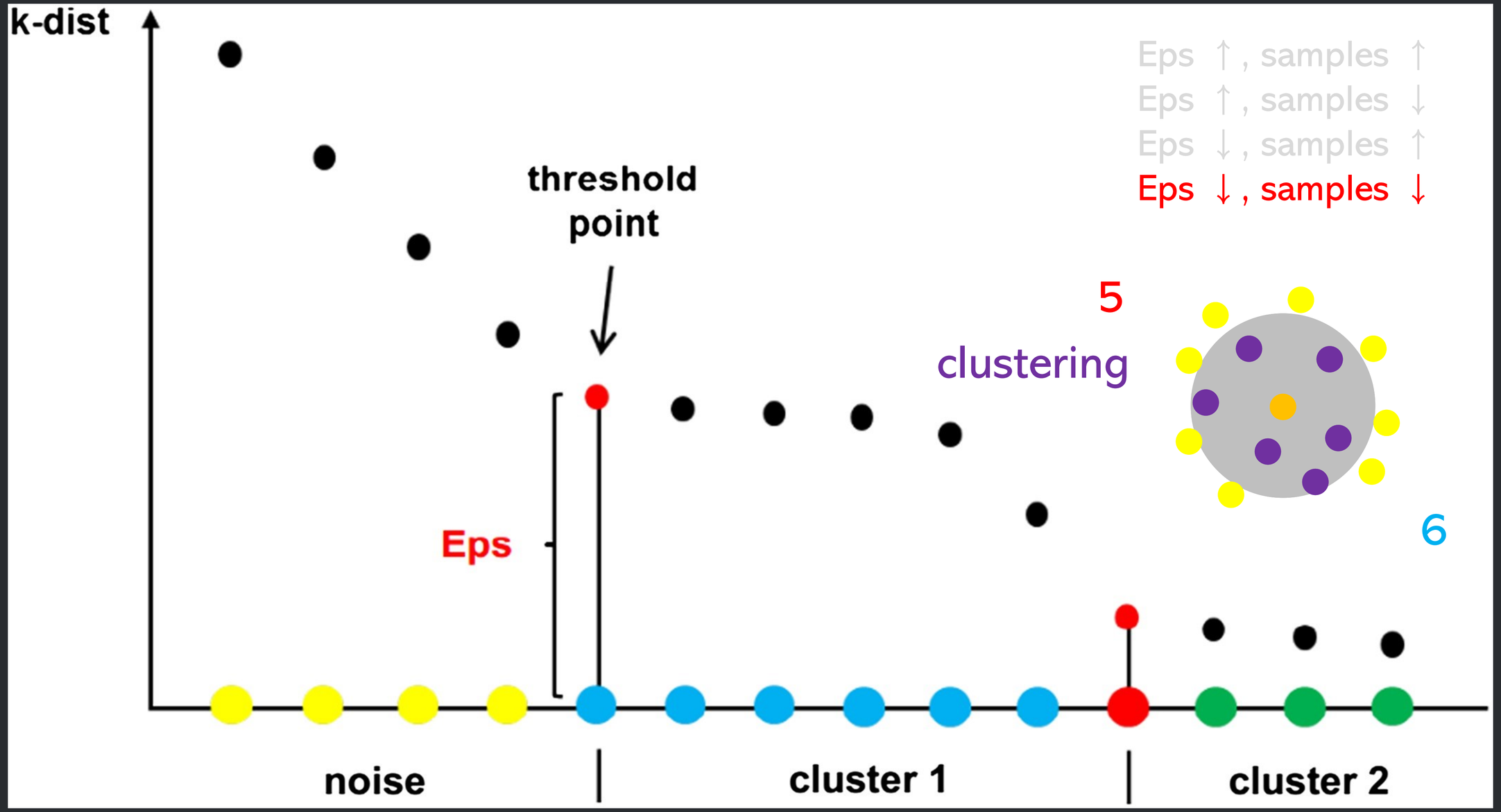
Heuristics


k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓
10
clustering
14

Heuristics

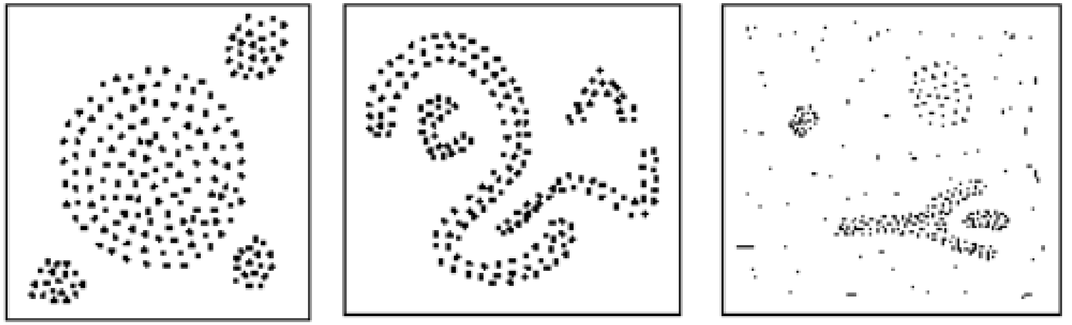

k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑ , samples ↑
Eps ↑ , samples ↓
Eps ↓ , samples ↑
Eps ↓ , samples ↓
15
no clustering
6

Heuristics


k-dist
threshold point
Eps
noise
cluster 1
cluster 2
Eps ↑ , samples ↑
Eps ↑ , samples ↓
Eps ↓ , samples ↑
Eps ↓ , samples ↓
5
clustering
6

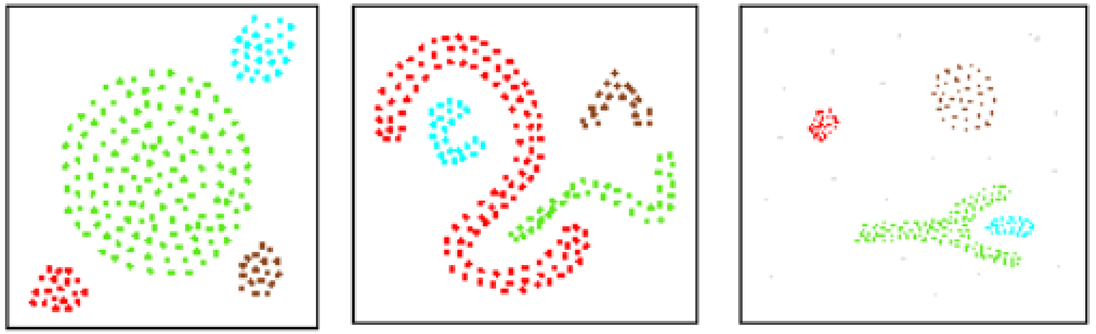
- **DBSCAN vs. Other clustering techniques**

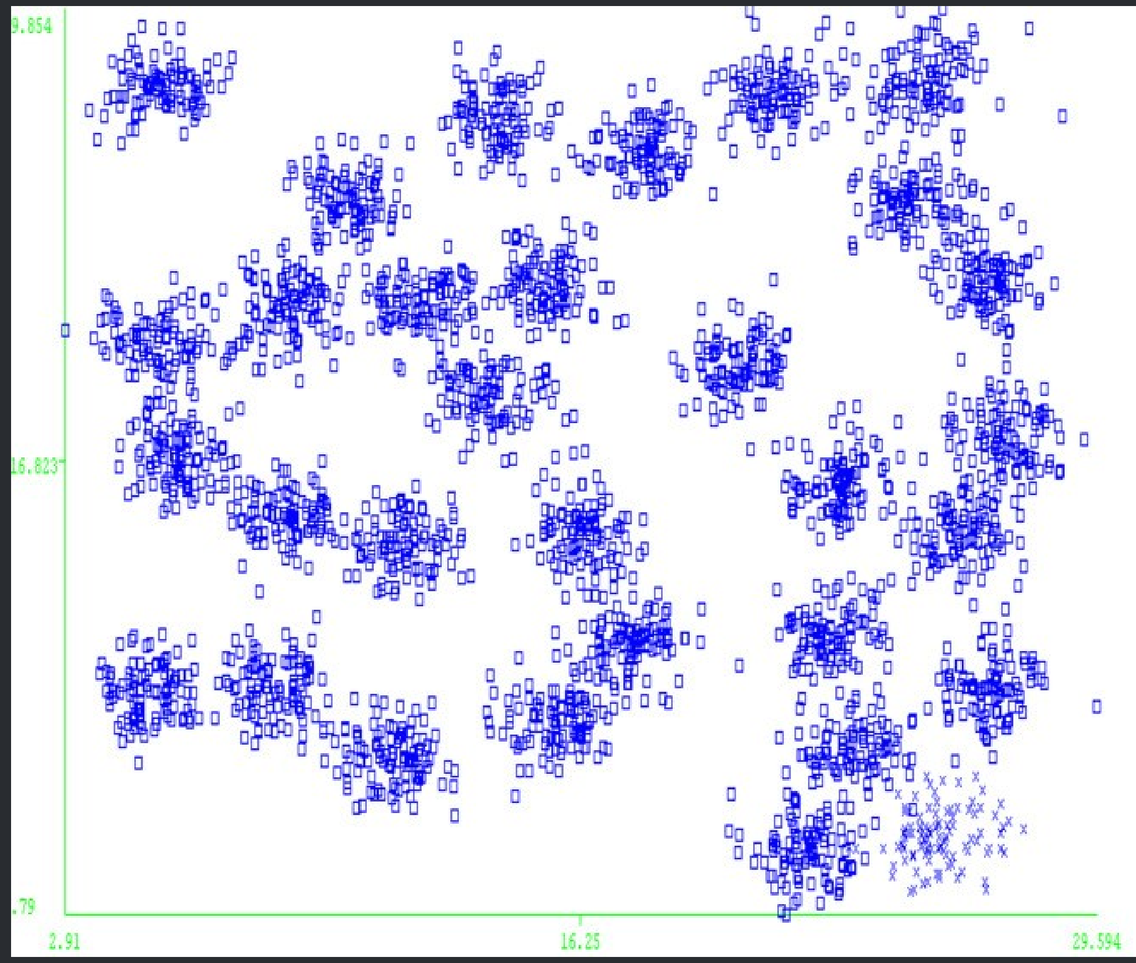
Data sets

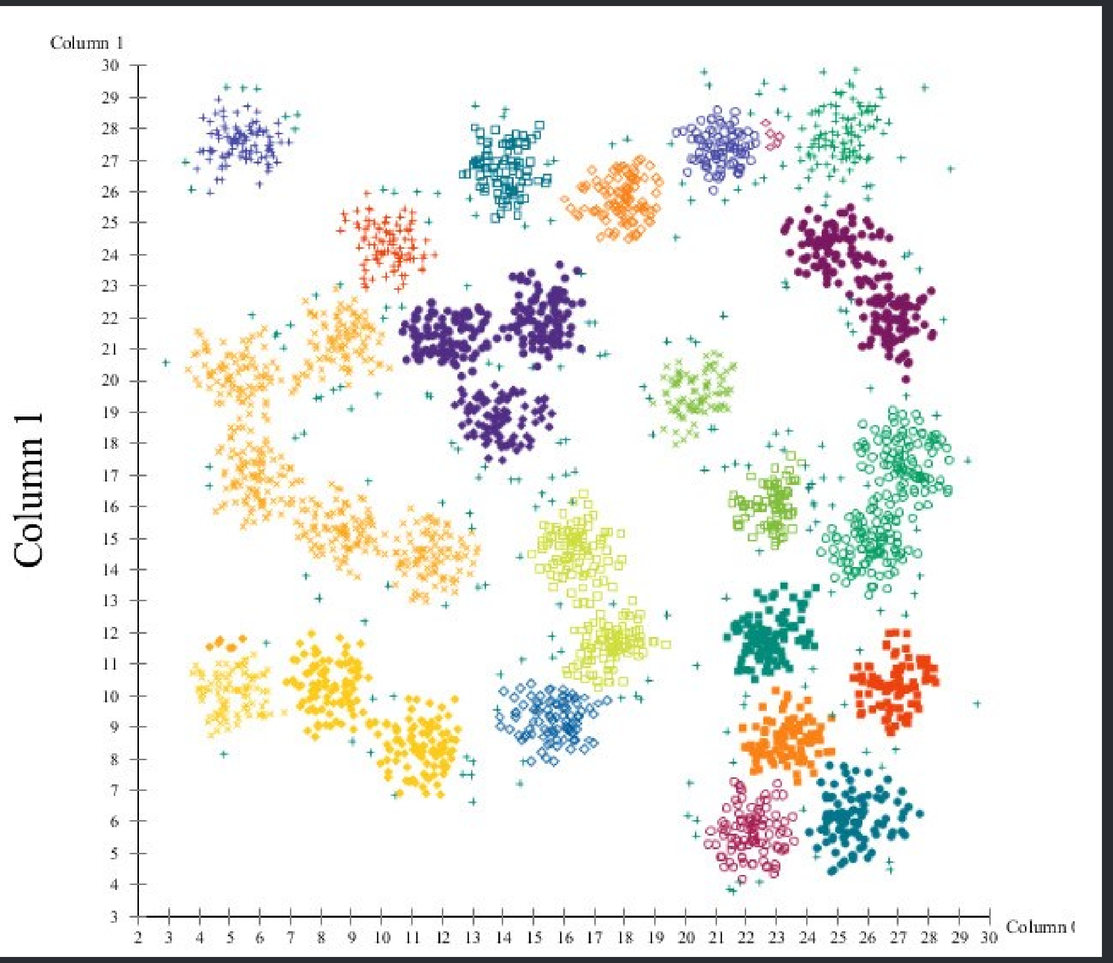


Algorithms
(Non-convex shapes)

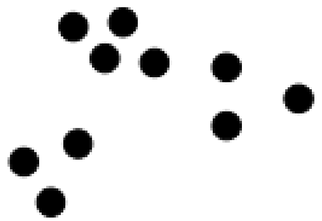
DBSCAN
(Arbitrary shapes)

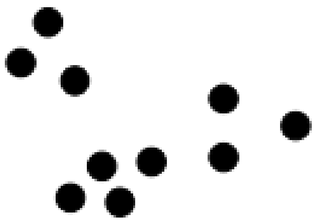



9.854
16.823
.79
2.91
16.25
29.594

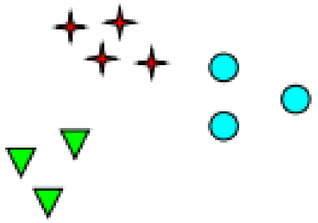

Column 1
Column 1
Column (

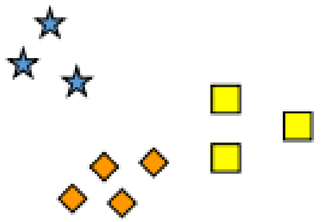
DBSCAN	
장점	노이즈에 매우 둔감한 군집화 가능
	임의의 모양을 갖는 군집을 생성할 수 있음
단점	Parameters에 따라 결과가 민감하게 작동
	높은 계산 비용



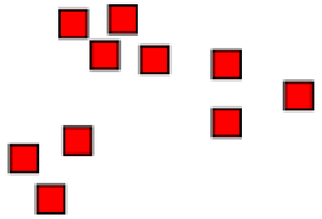


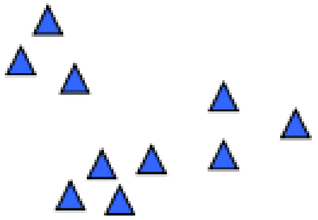
How many clusters?



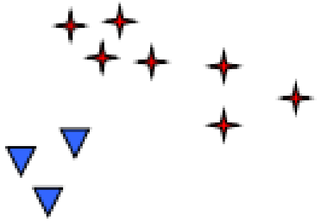


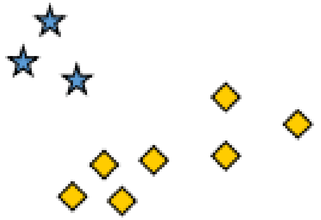
Six Clusters





Two Clusters





Four Clusters

External	• Rand Statistic • Jaccard Coefficient • Folks and Mallows index • (Normalized) Hurbert Γ statistic
Internal	• Cophenetic Correlation Coefficient • Sum of Squared error (SSE) • Cohesion and separation
Relative	• Dunn family of indices • Davies-Bouldin (DB) index • Semi-partial R-squared • SD validity index • Silhouette