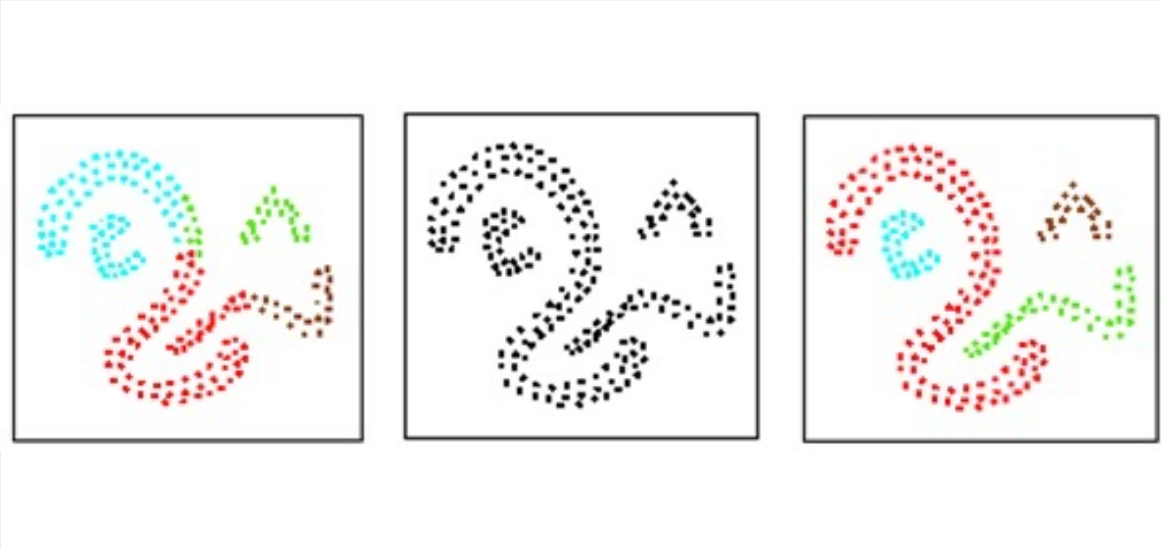
DBSCAN

1

DBSCAN



2

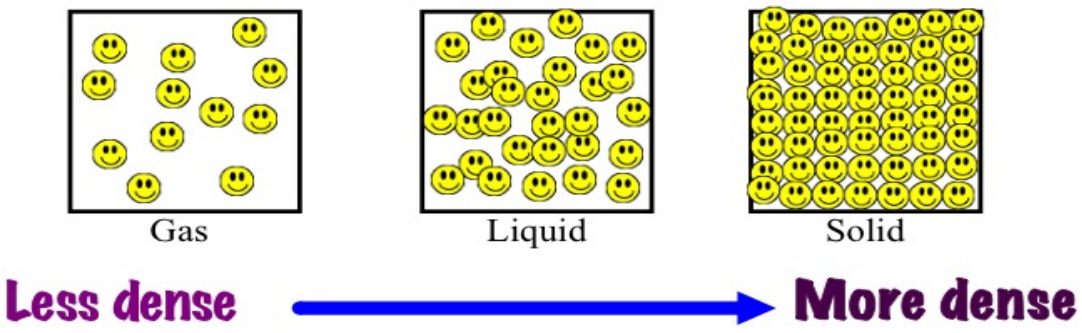
Density-Based Spatial Clustering of Applications with Noise

3

그룹화

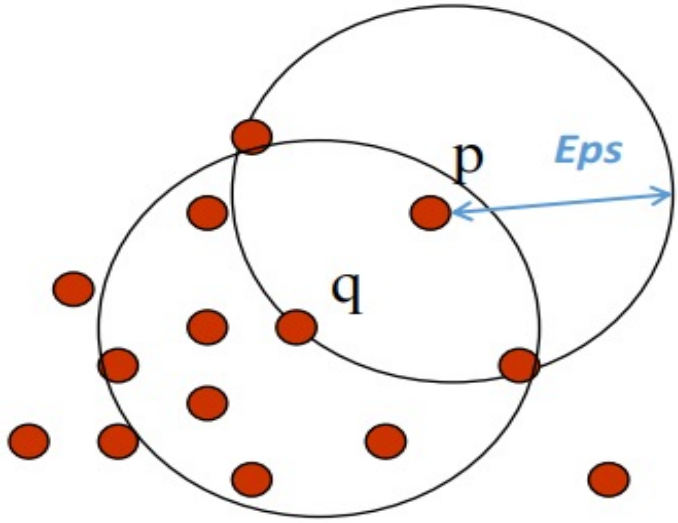
Density of Matter

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



4

그룹화



$N_{Eps}(p) = 3$

$N_{Eps}(q) = 8$

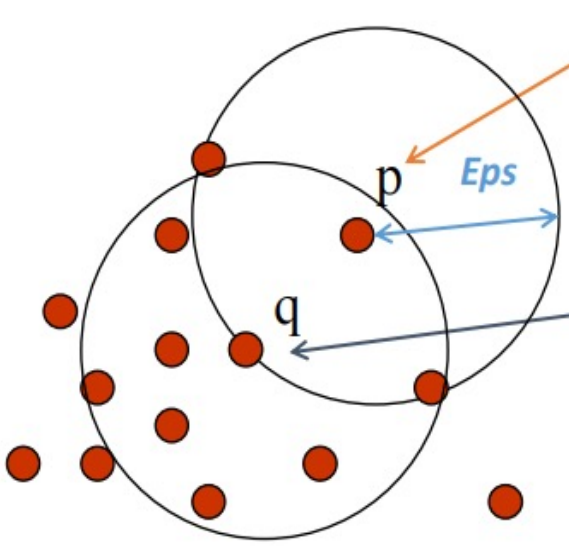
5

그룹화

$N_{Eps}(p) = 3$

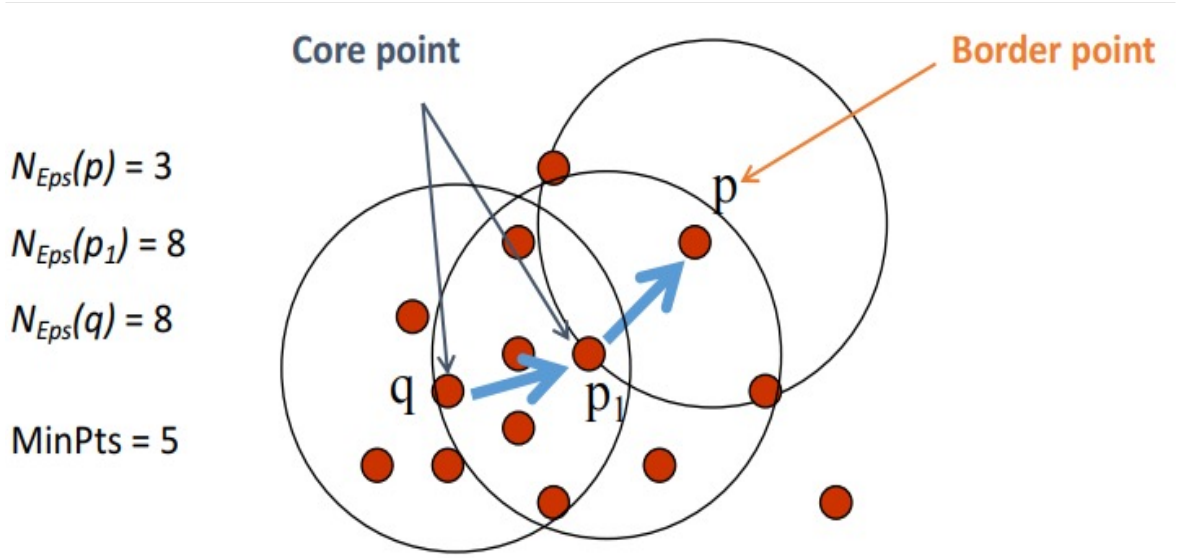
$N_{Eps}(q) = 8$

MinPts = 5



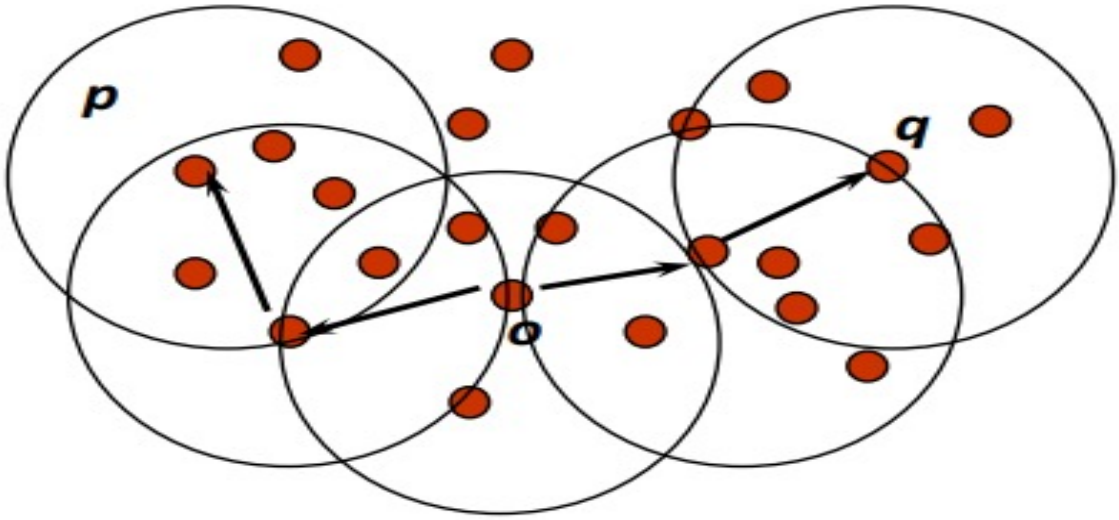
6

그룹화


Core point
Border point
$N_{Eps}(p) = 3$
$N_{Eps}(p_1) = 8$
$N_{Eps}(q) = 8$
MinPts = 5
p
q
p_1

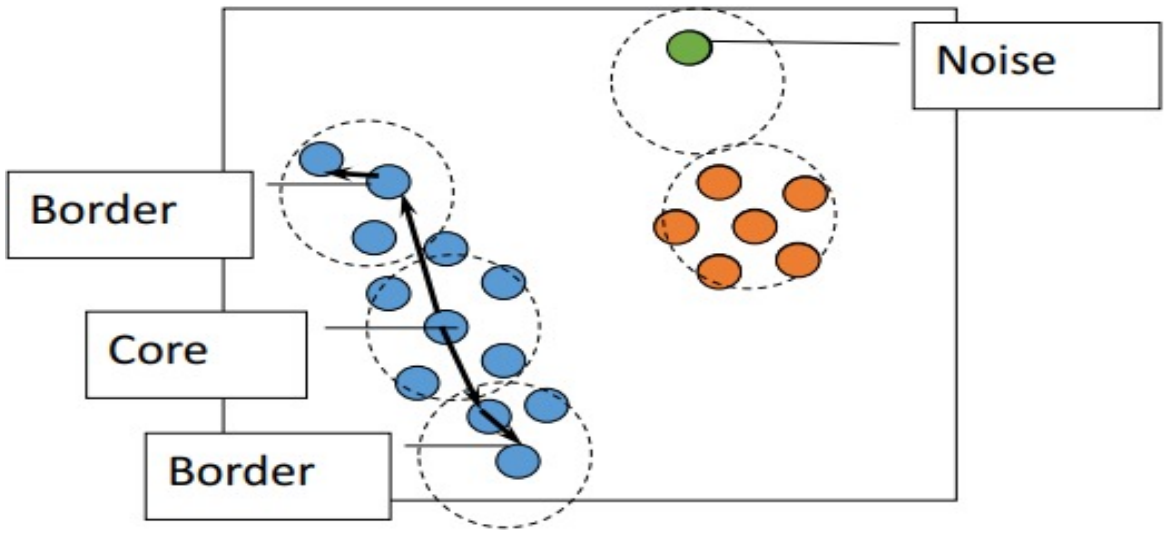
7

그룹화


p
q
o

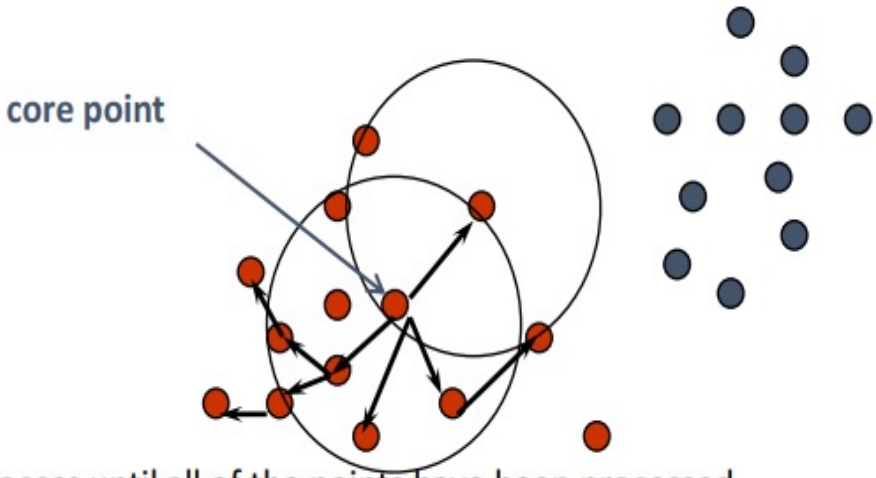
8

그룹화



9

그룹화

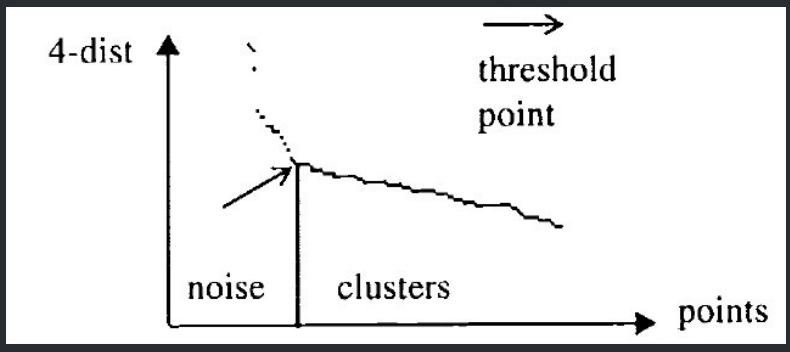


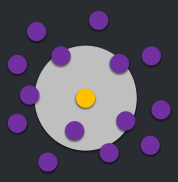
◦ Continue the process until all of the points have been processed.

10

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓

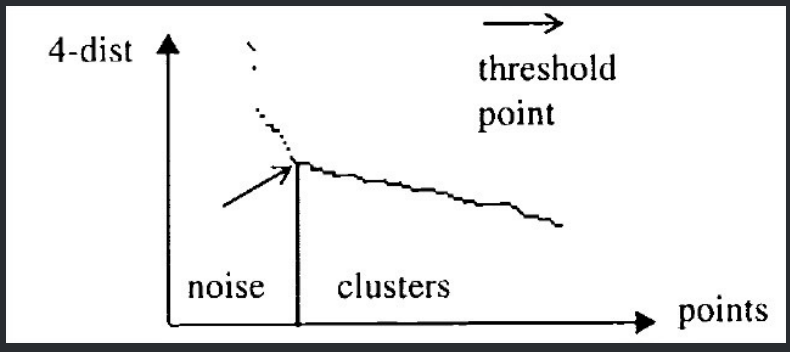


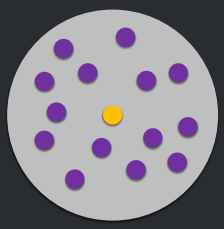


11

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓

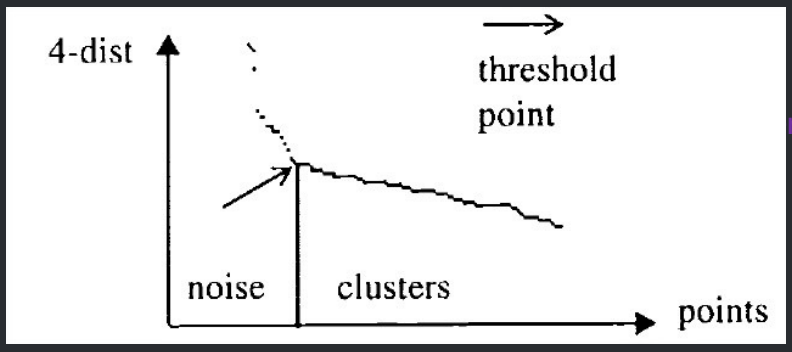




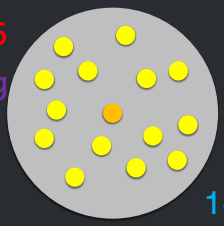
12

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓



15
ring

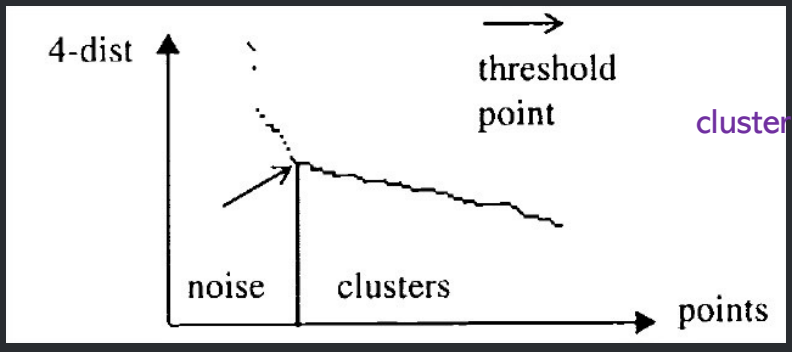


14

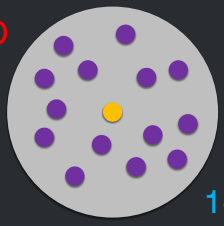
13

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓



10
clustering

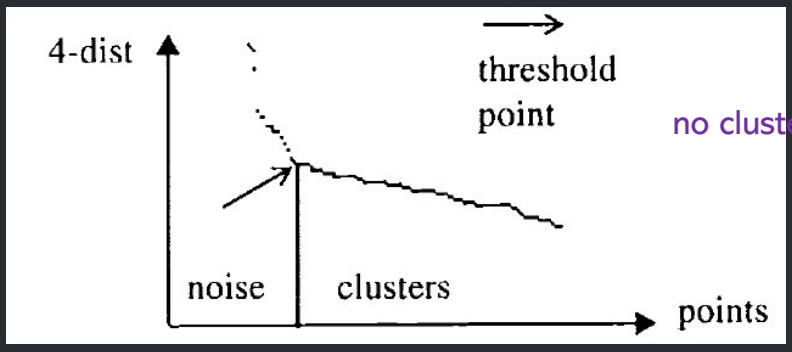


14

14

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓



15

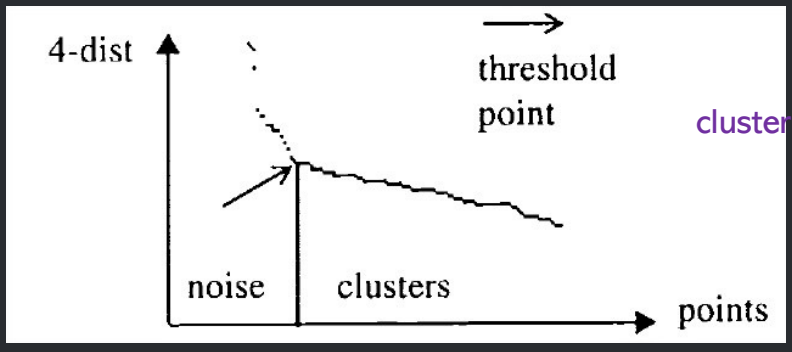
no clustering

6

15

Heuristics

Eps ↑, samples ↑
Eps ↑, samples ↓
Eps ↓, samples ↑
Eps ↓, samples ↓



5

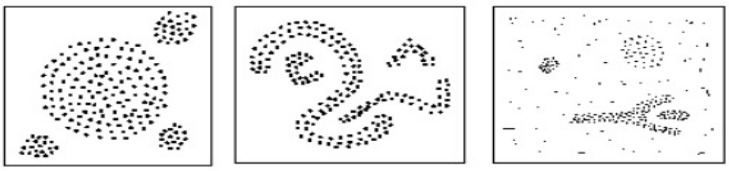
clustering

6

16

DBSCAN VS. OTHER CLUSTERING

Data sets

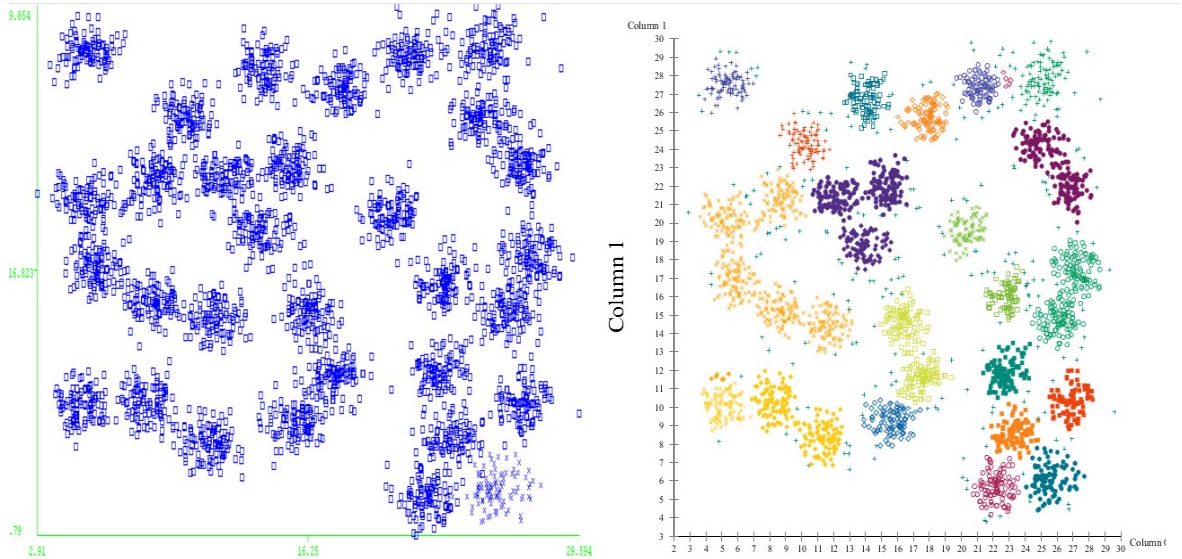


Algorithms
(Non-convex shapes)

DBSCAN
(Arbitrary shapes)

17

DBSCAN



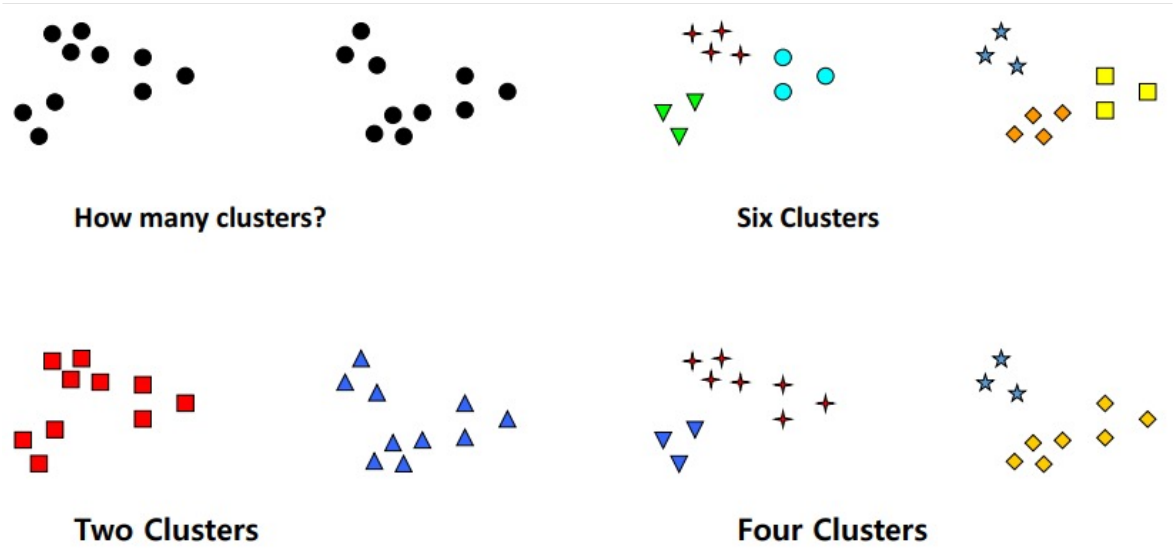
18

DBSCAN

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

20

QUESTION: WHAT IS THE OPTIMAL CLUSTERING NUMBER?



21

클러스터링한 결과를 어떻게 판단할 것인가? (NO BEST RULES)

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

22