

Advanced Discovery: DBSCAN Algorithm

Density-Based Spatial Clustering of Applications with Noise

Core Concept: Density = Connectivity

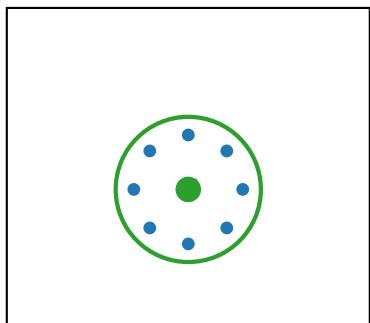
ε (eps) = radius
Neighborhood size



MinPts = minimum
Points in neighborhood

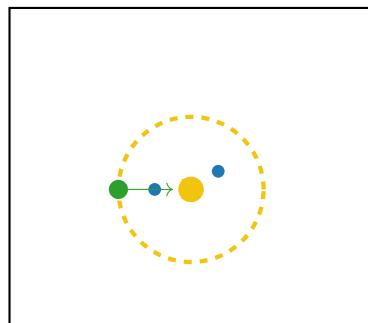
Three Types of Points

Core Point



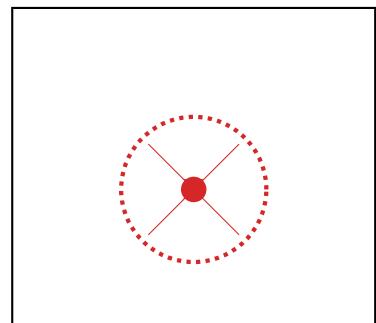
\geq MinPts in ε

Border Point



$<$ MinPts but
reachable from core

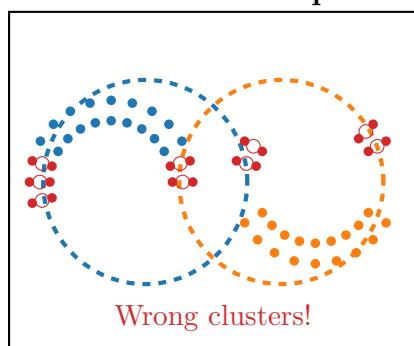
Noise Point



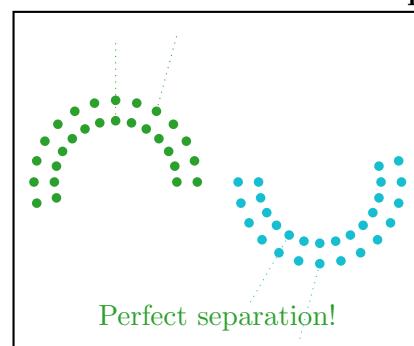
Not reachable
from any core

DBSCAN vs K-Means: Shape Discovery

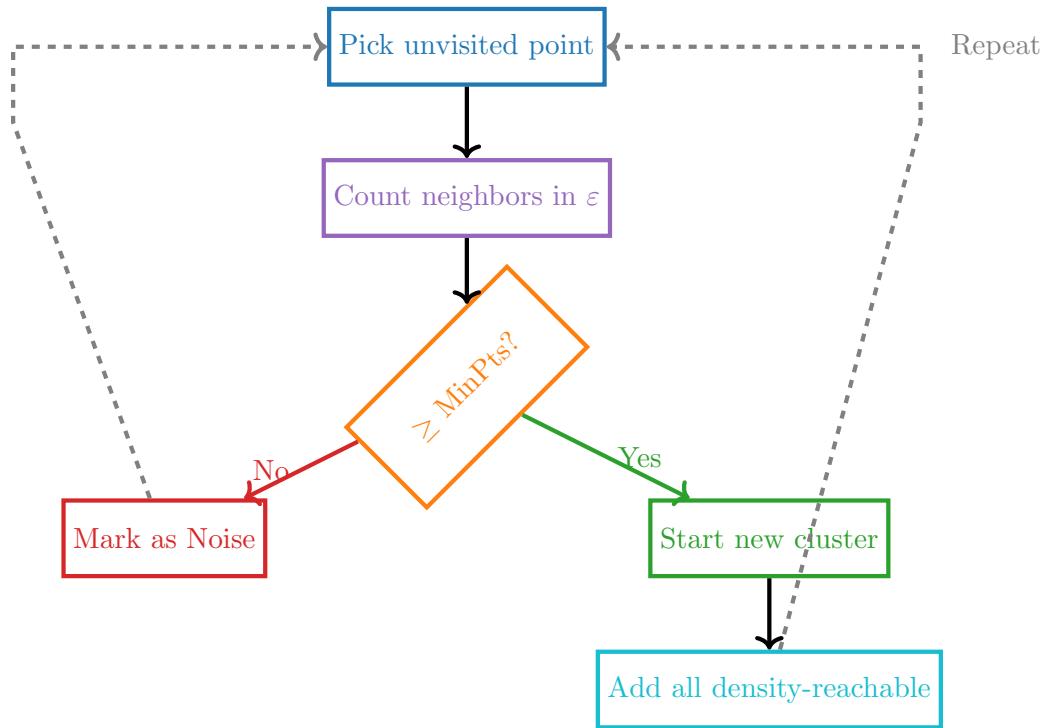
K-Means: Forces Spheres



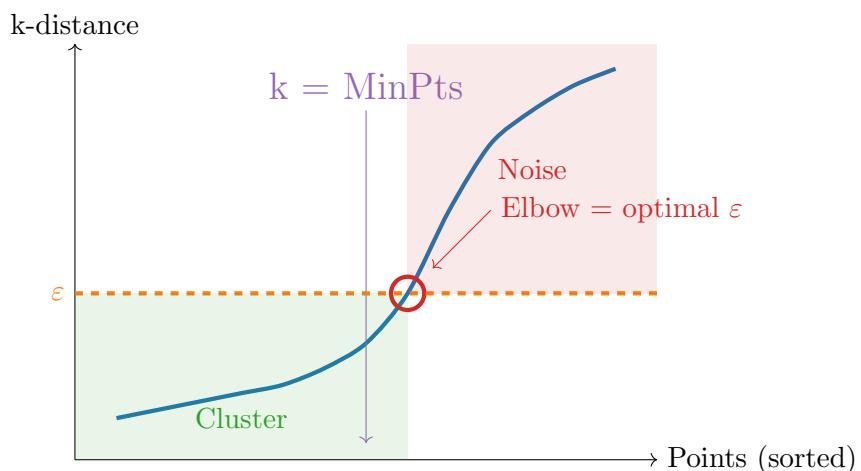
DBSCAN: Finds True Shape



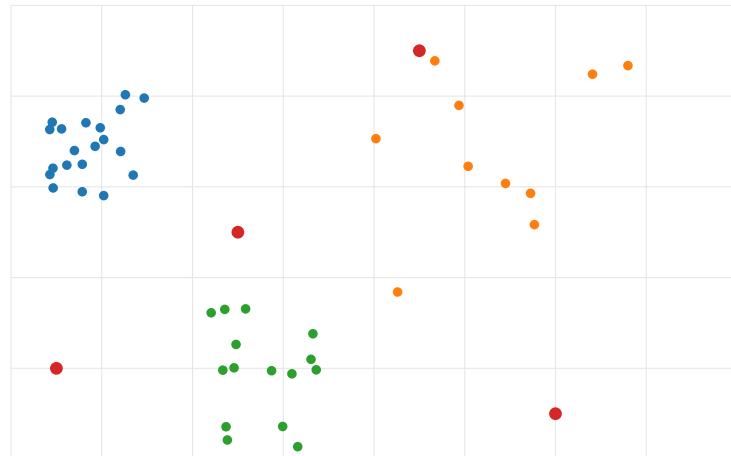
Algorithm Flow: Growing Clusters



Parameter Selection: The ϵ -MinPts Dance



Discovery Challenge: Optimal Parameters



One ε , One MinPts: Can you find all 3 clusters?

$\varepsilon = \underline{\hspace{2cm}}$ MinPts = $\underline{\hspace{2cm}}$

Next: Hierarchical - When you need clusters at every scale!