

Density-based clustering

- Spherical-shape clusters

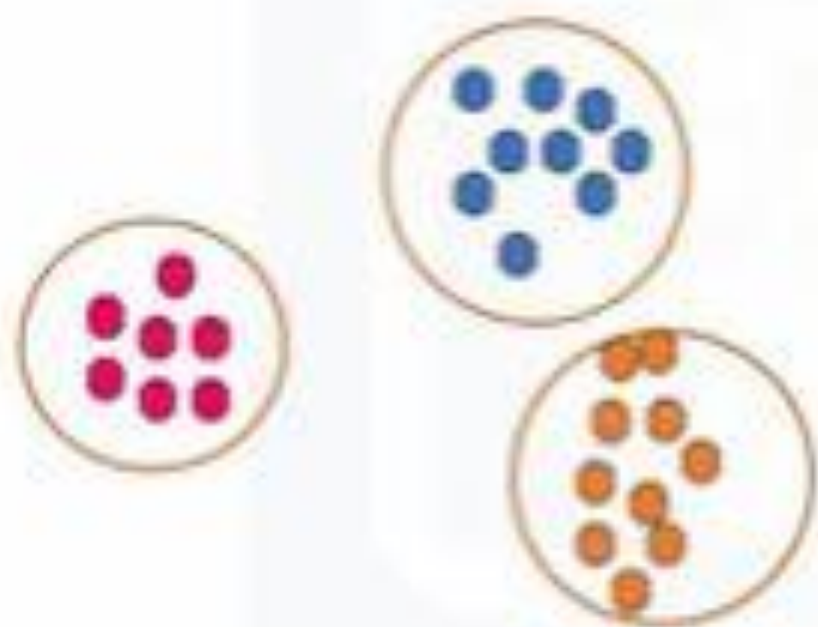


- Arbitrary-shape clusters

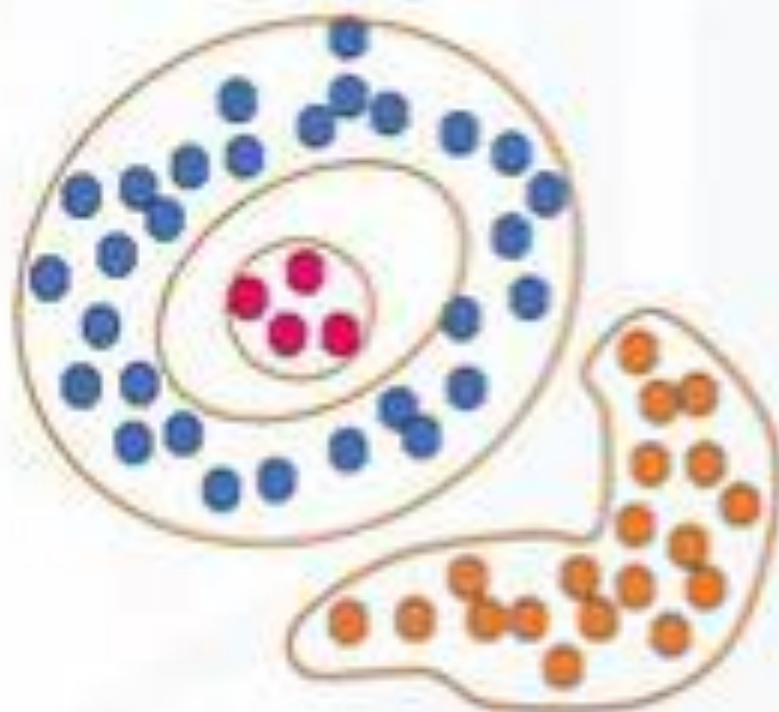


Density-based clustering

- Spherical-shape clusters

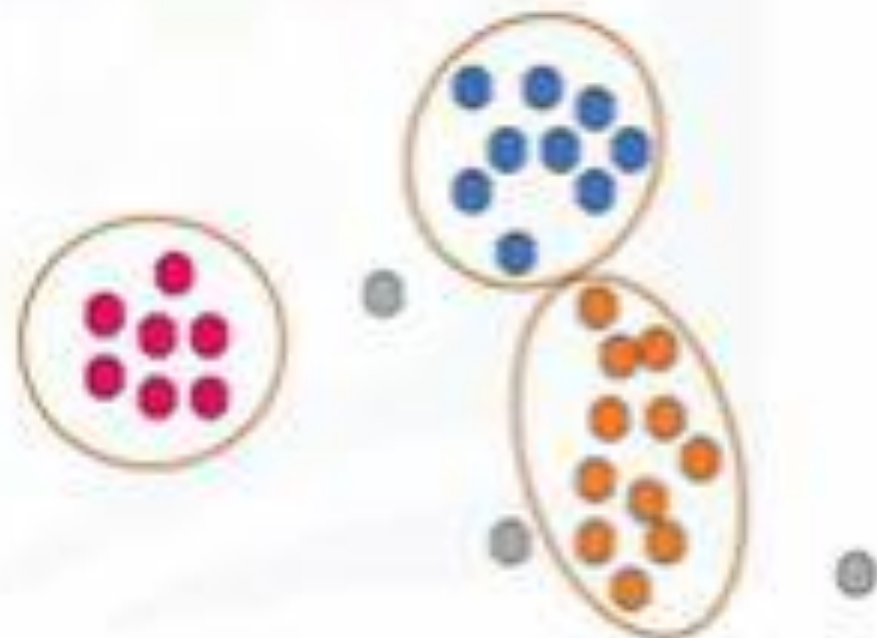
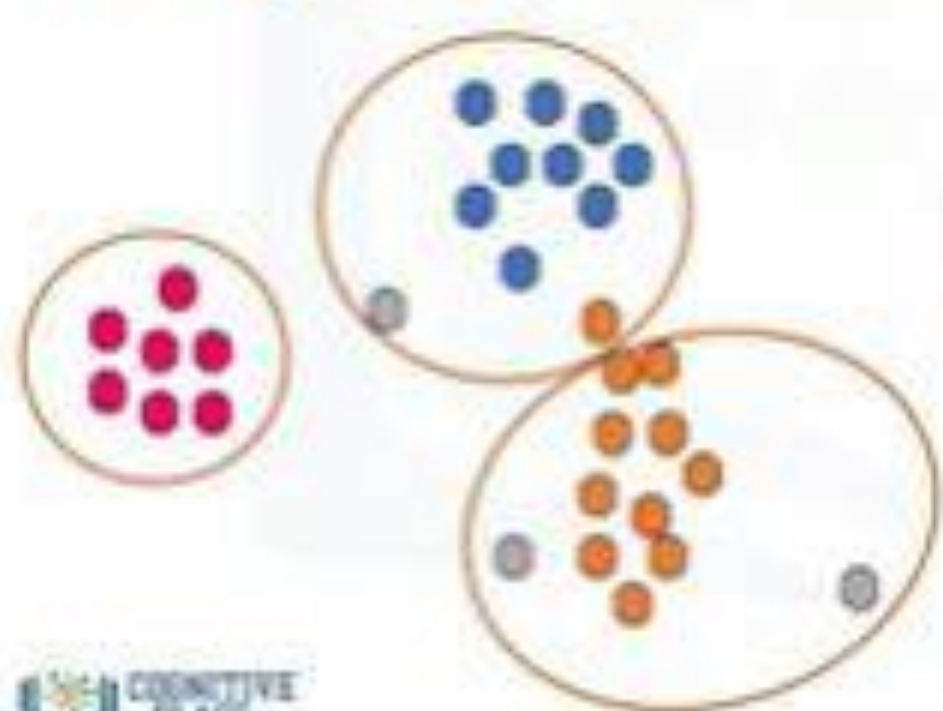


- Arbitrary-shape clusters



k-Means Vs. density-based clustering

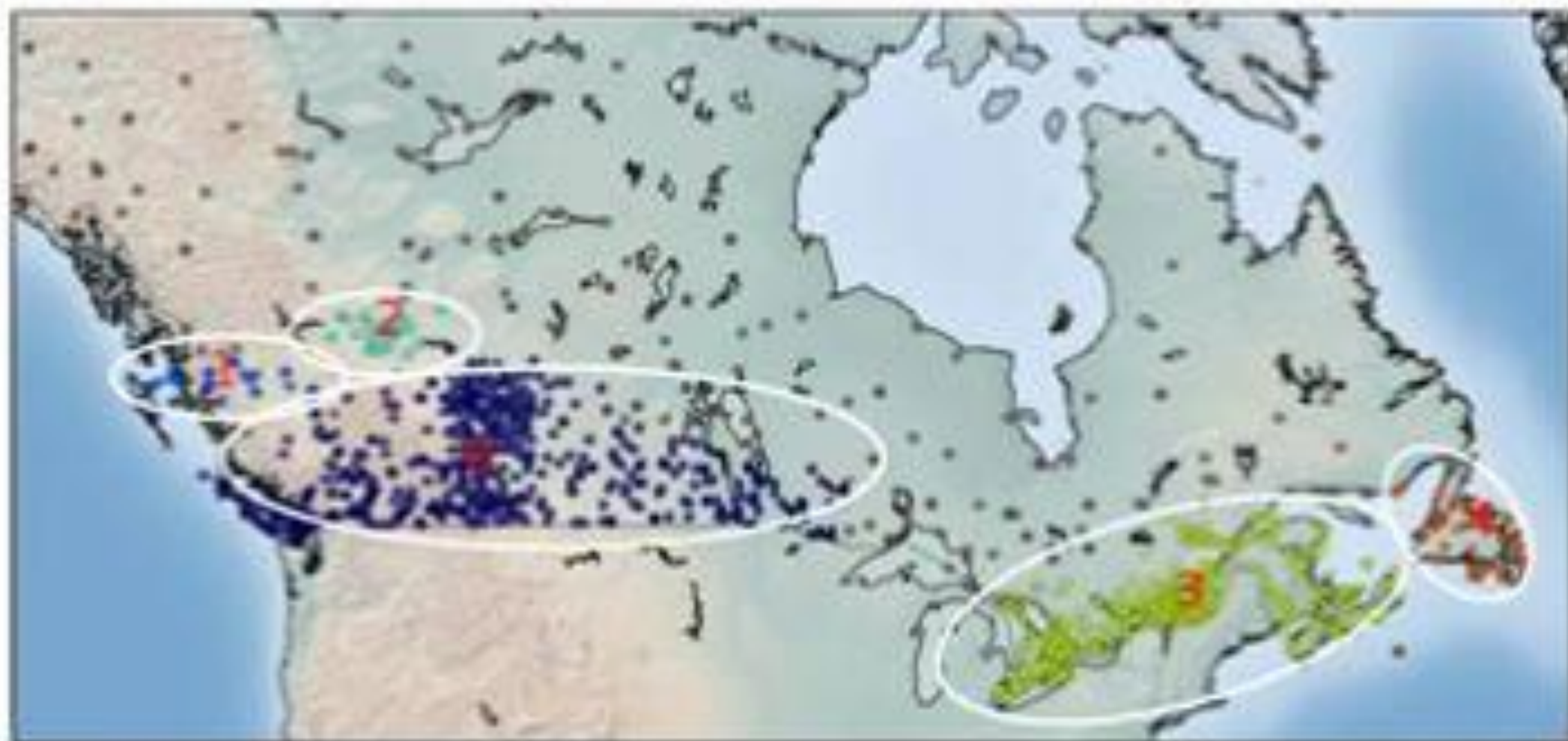
- k-Means assigns all points to a cluster even if they do not belong in any
- Density-based Clustering locates regions of **high density**, and separates outliers



DBSCAN for class identification



DBSCAN for class identification



What is DBSCAN?

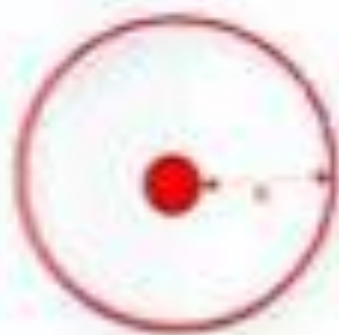
- DBSCAN (**D**ensity-**B**ased **S**patial Clustering of **A**pplications with **N**oise)
- R (Radius of neighborhood)
- M (Min number of neighbors)

What is DBSCAN?

- DBSCAN (**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise)
 - Is one of the most common clustering algorithms
 - Works based on density of objects
- R (Radius of neighborhood)
- M (Min number of neighbors)

What is DBSCAN?

- DBSCAN (**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise)
 - Is one of the most common clustering algorithms
 - Works based on density of objects
- R (Radius of neighborhood)
 - Radius (R) that if includes enough number of points within, we call it a dense area
- M (Min number of neighbors)



What is DBSCAN?

- DBSCAN (**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise)
 - Is one of the most common clustering algorithms
 - Works based on density of objects
- R (Radius of neighborhood)
 - Radius (R) that if includes enough number of points within, we call it a dense area
- M (Min number of neighbors)
 - The minimum number of data points we want in a neighborhood to define a cluster

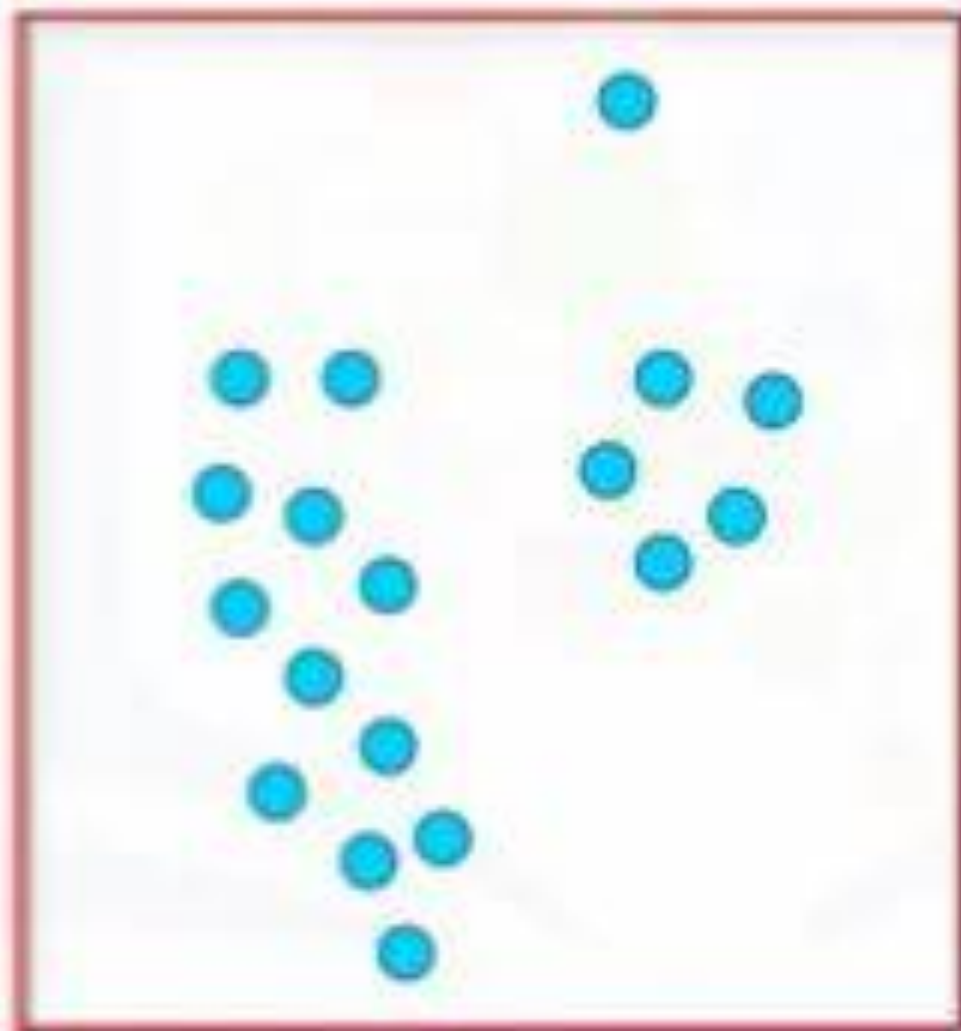


How DBSCAN works



$R = 2 \text{ unit}$, $M = 6$

How DBSCAN works

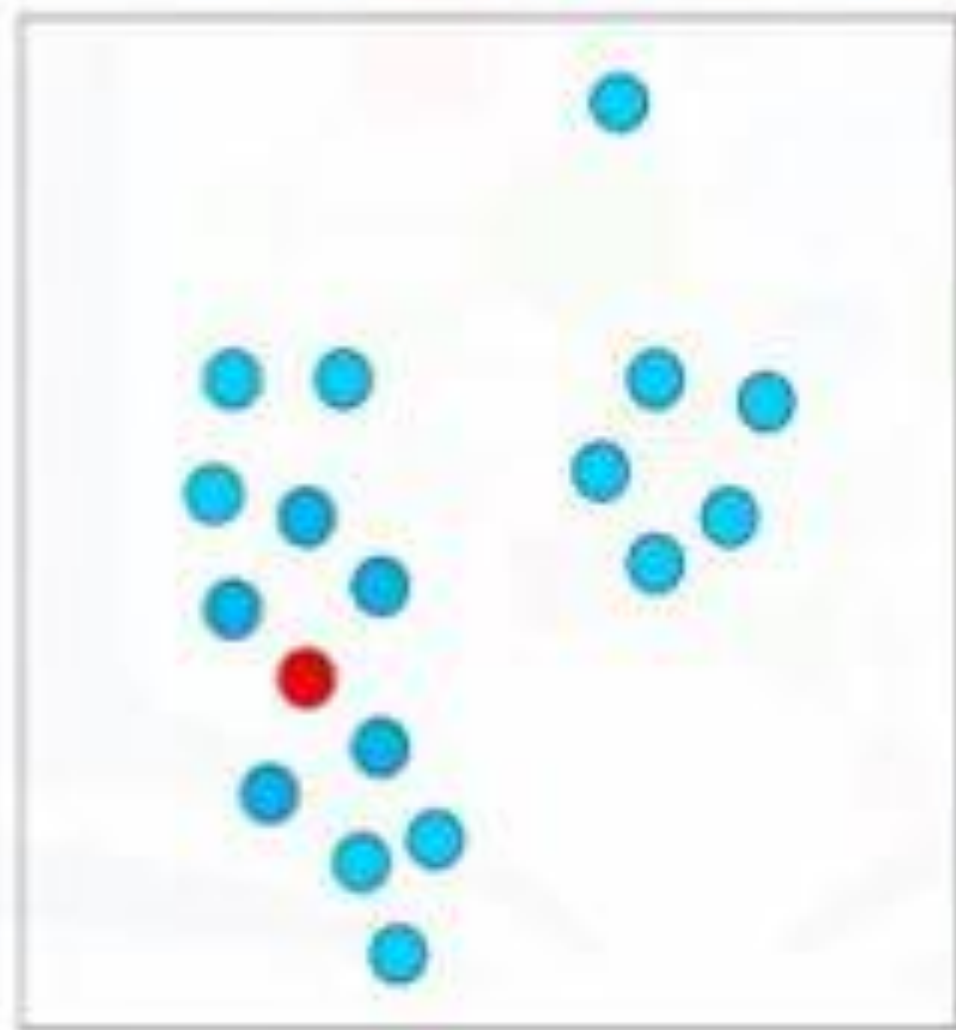


Each point is either:

- *core point*
- *border point*
- *outlier point*

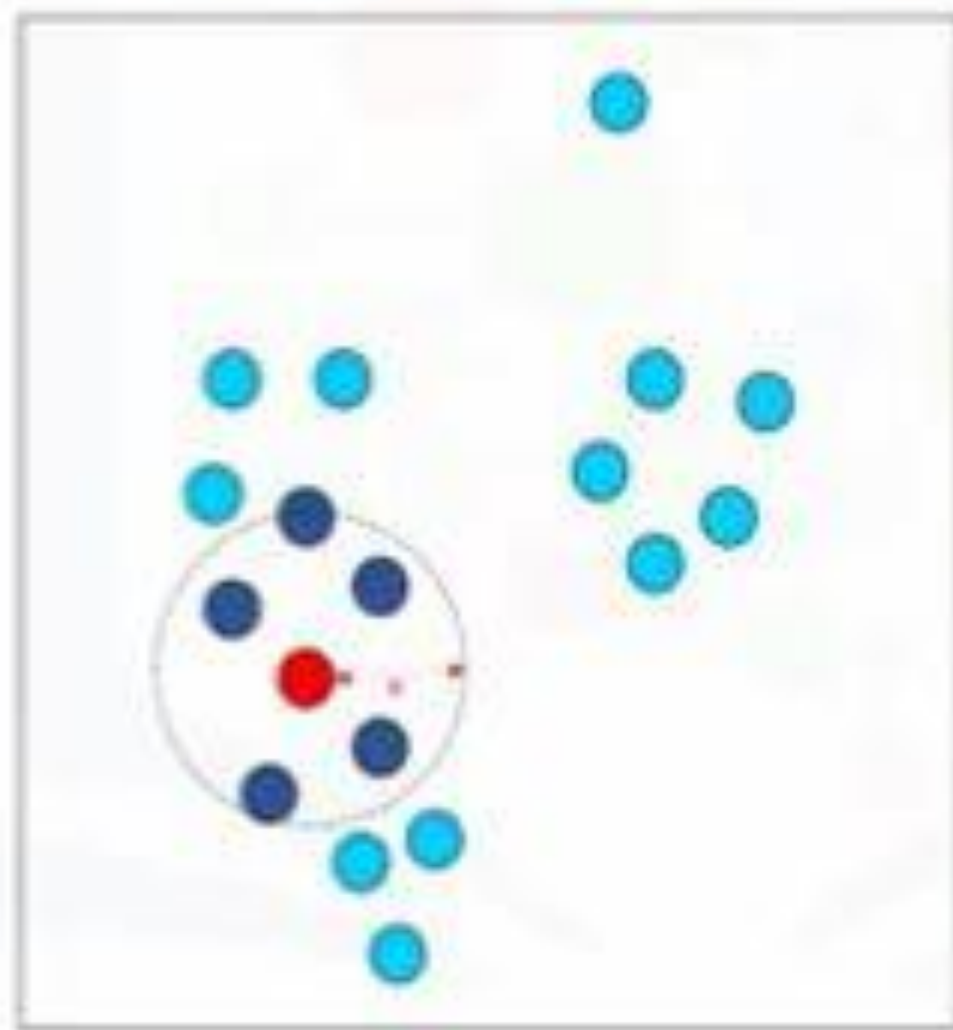
$R = 2 \text{ unit}$, $M = 5$

DBSCAN algorithm – core point?



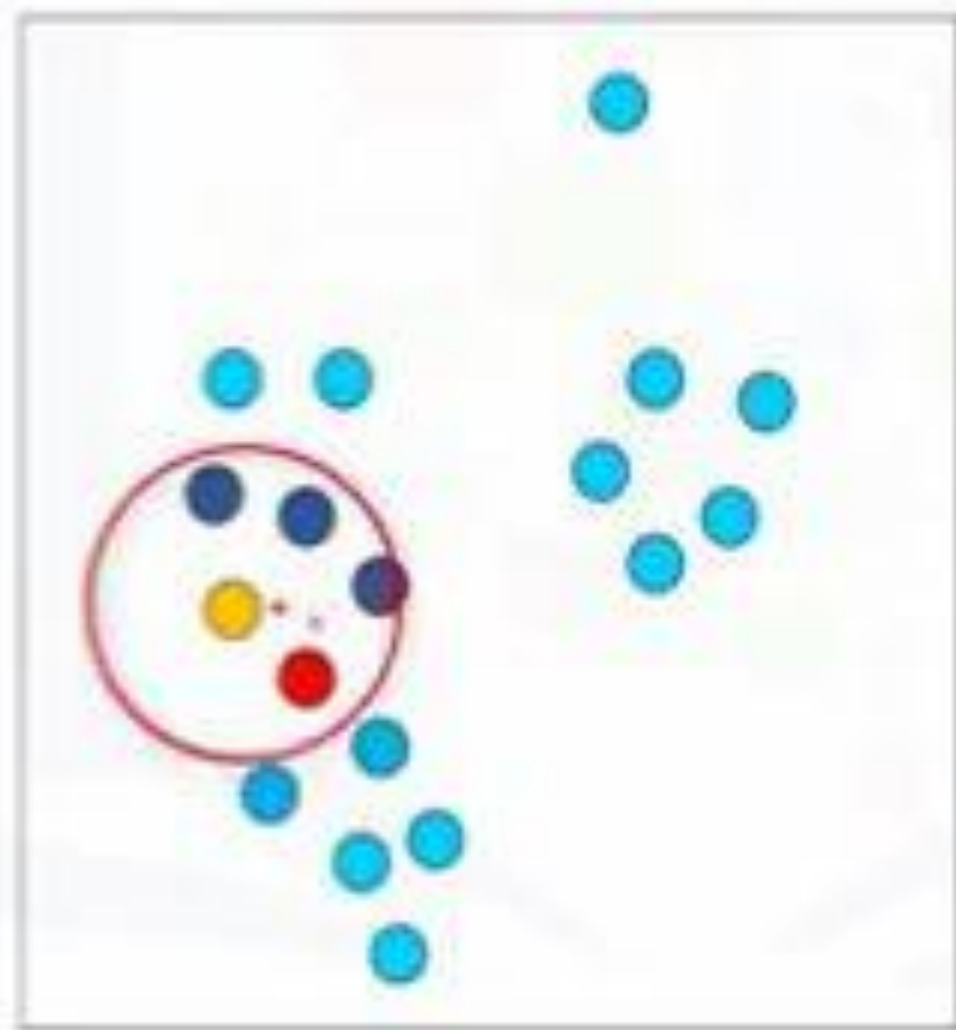
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – core point



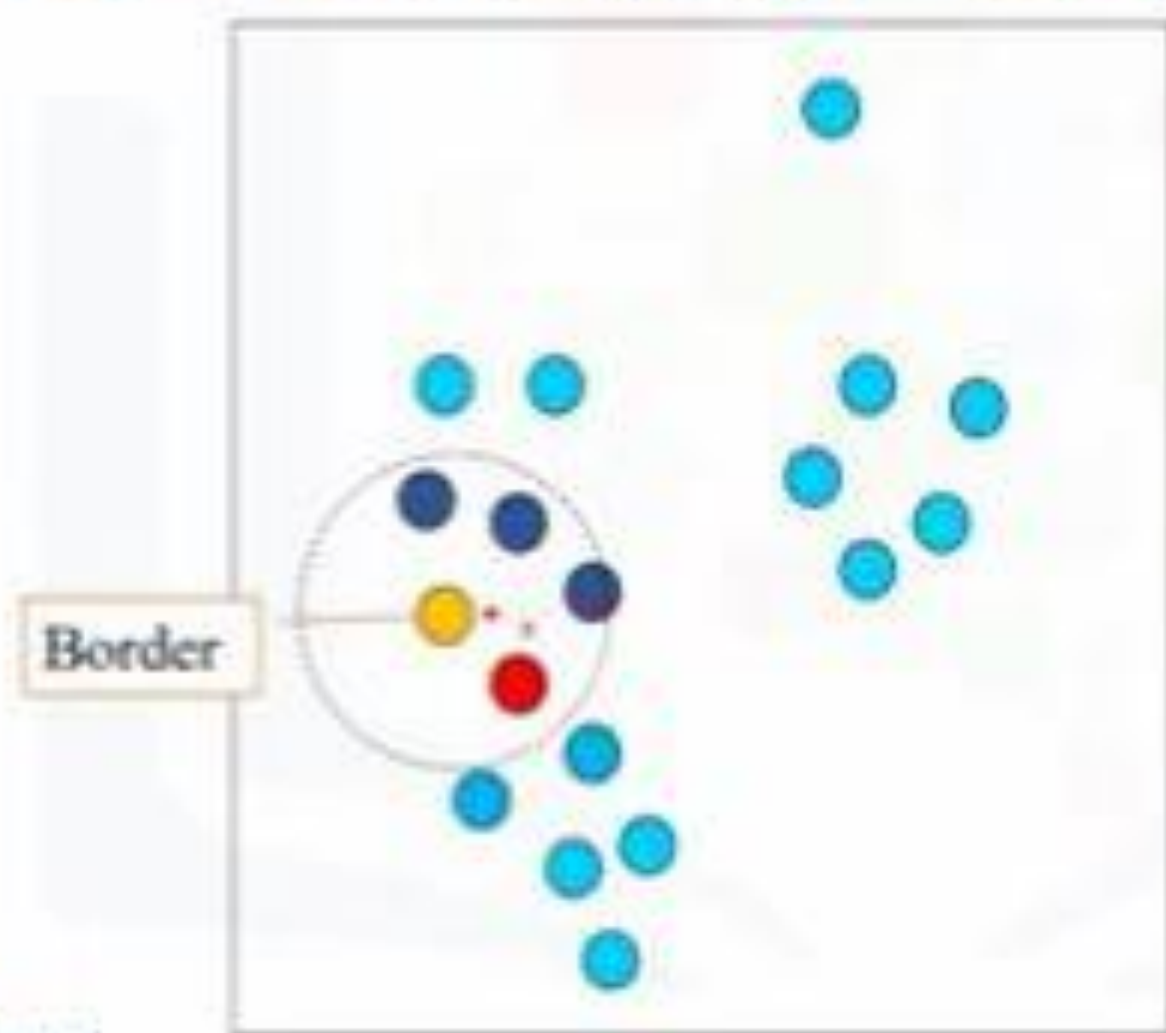
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – border points?



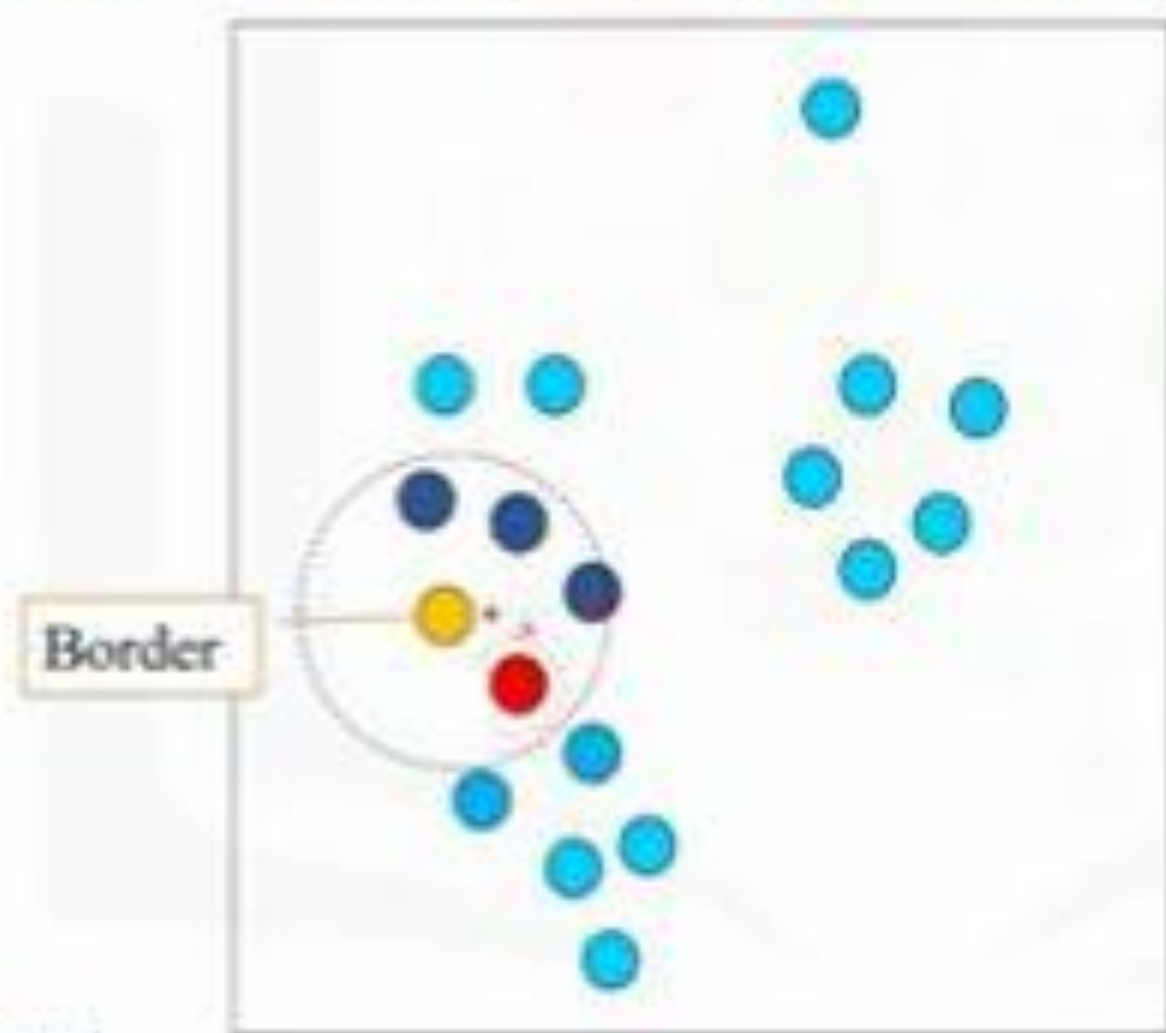
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – border points?



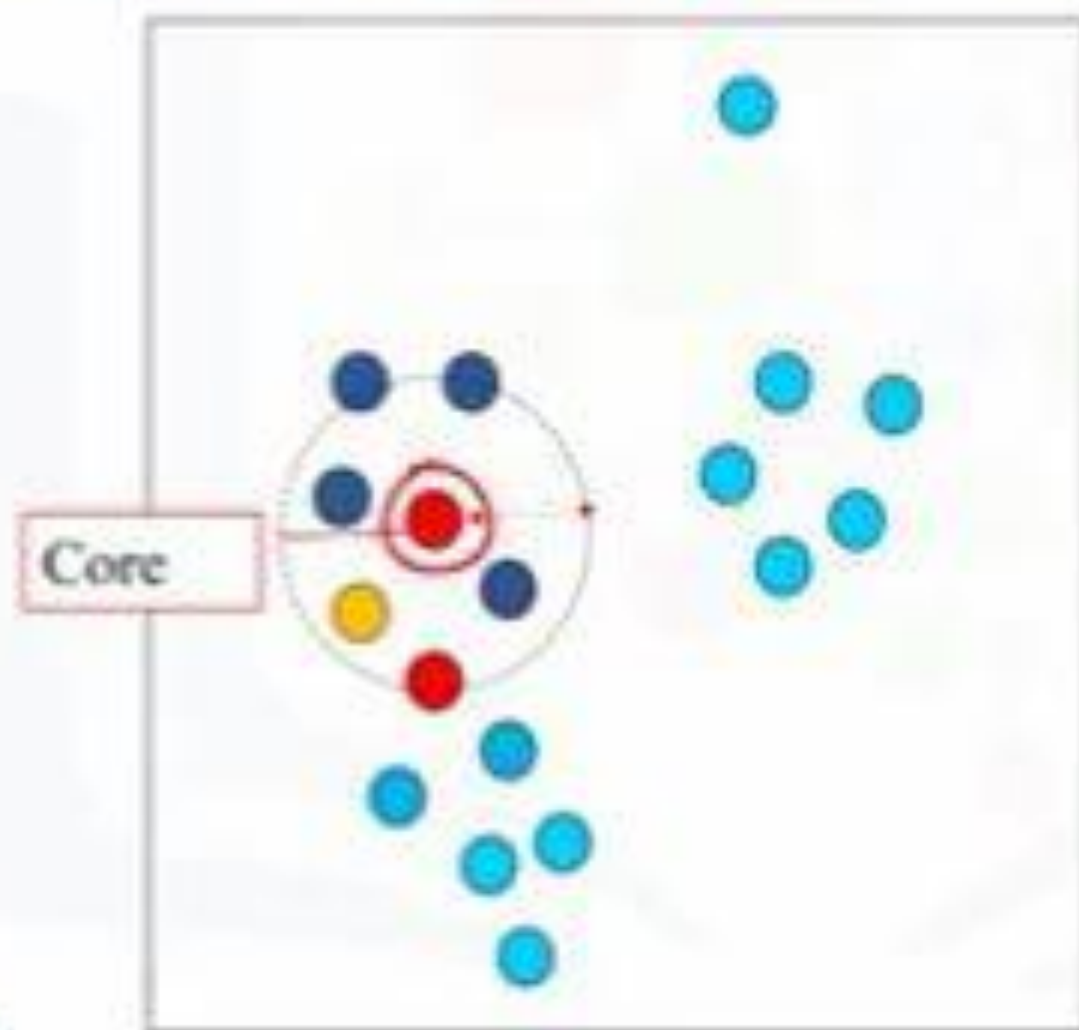
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – border points?



$R = 2\text{unit}$, $M = 6$

DBSCAN algorithm – core point



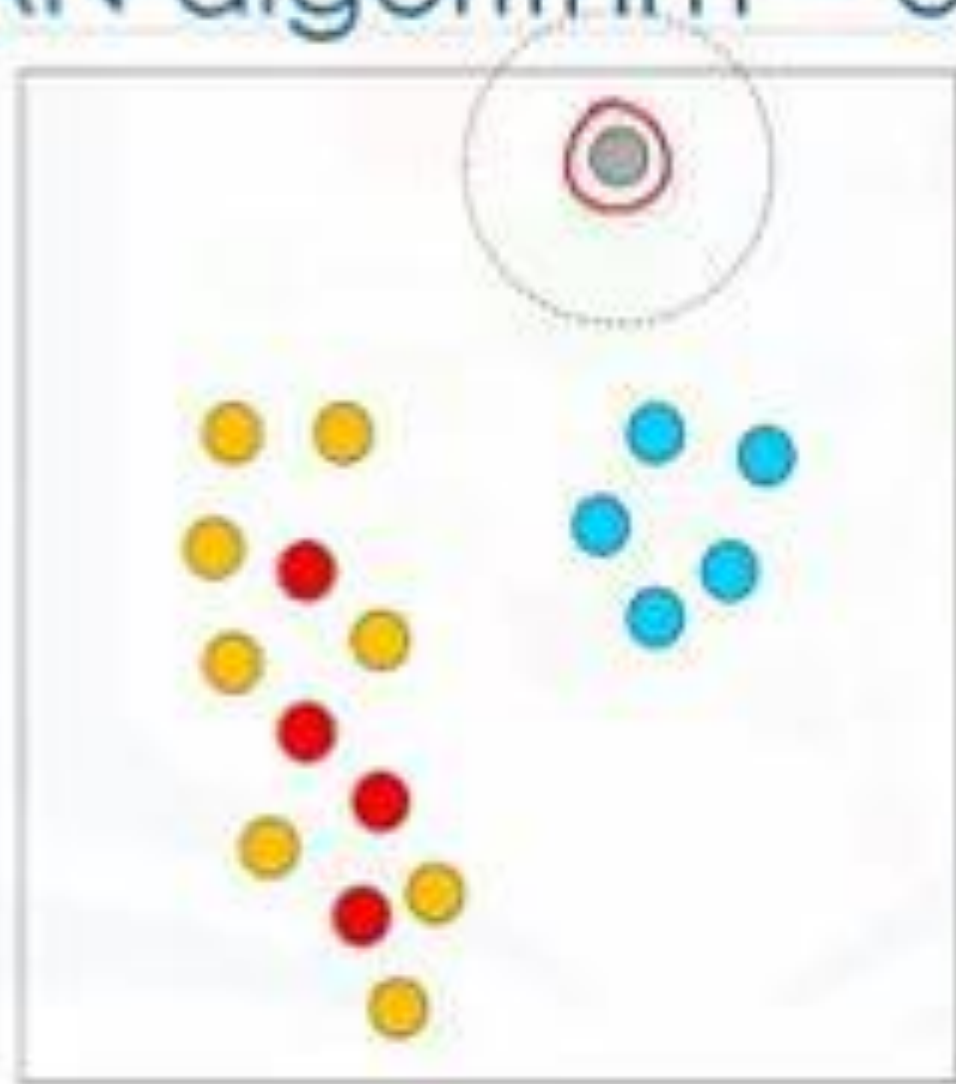
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – next core point



$R = 2 \text{ unit}, M = 6$

DBSCAN algorithm – outliers



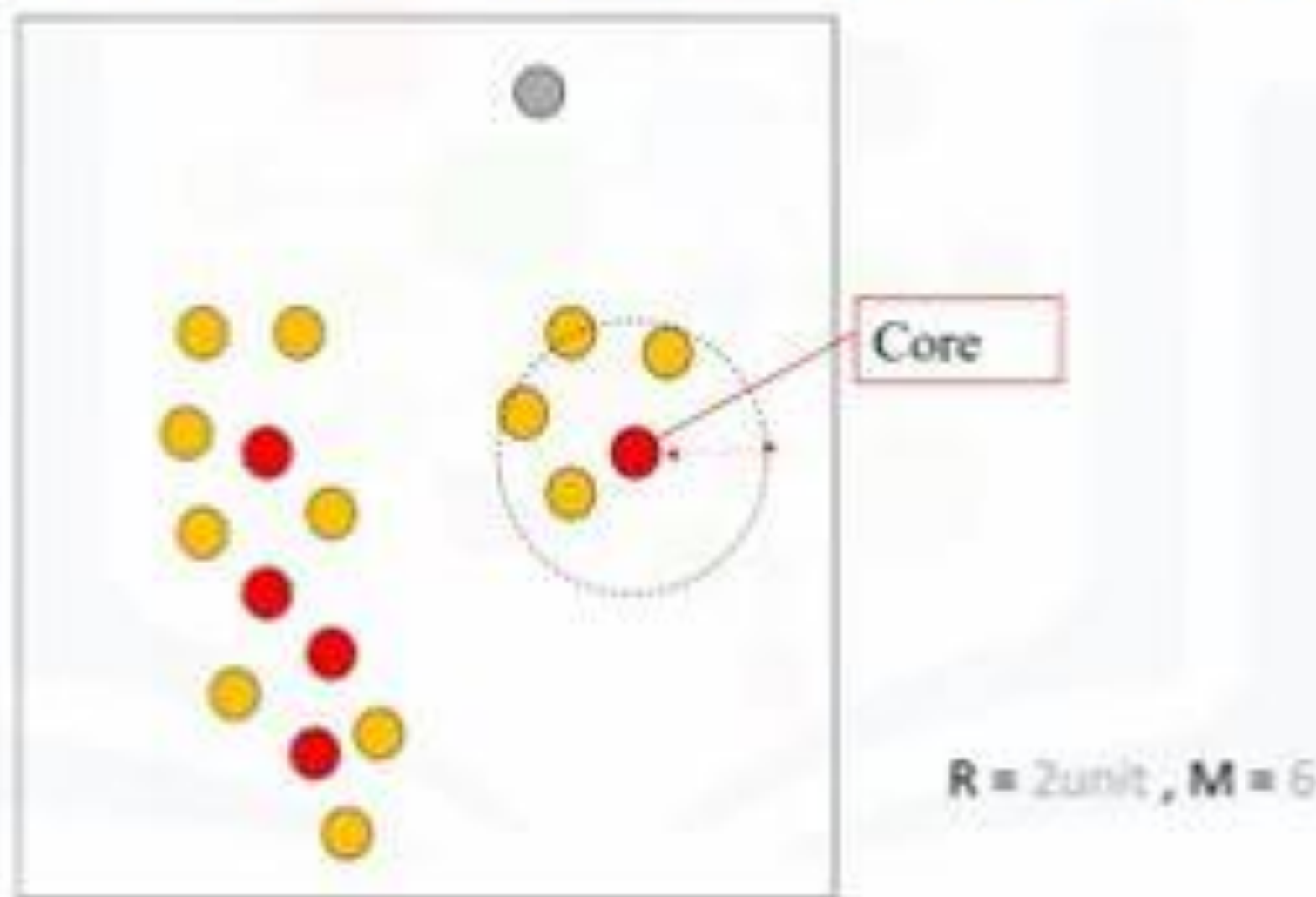
$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – outliers

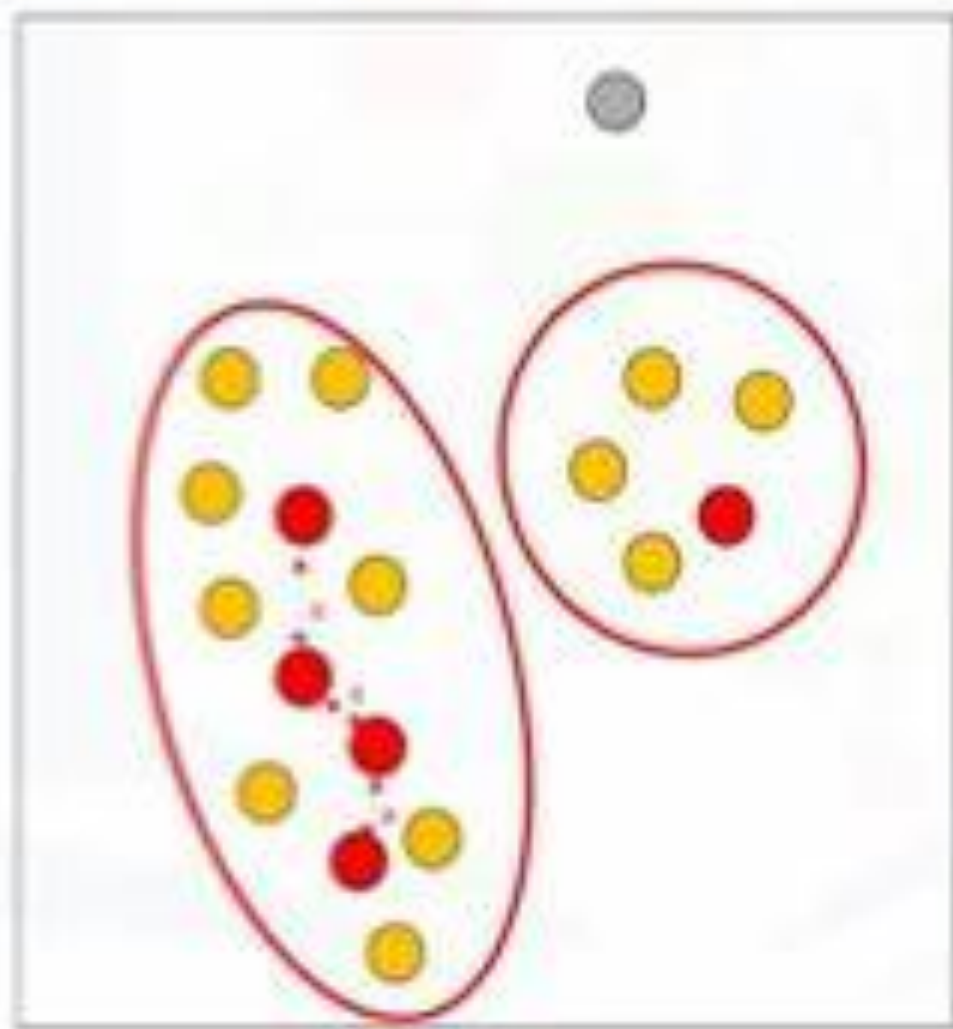


$R = 2 \text{ unit}$, $M = 6$

DBSCAN algorithm – identify all points

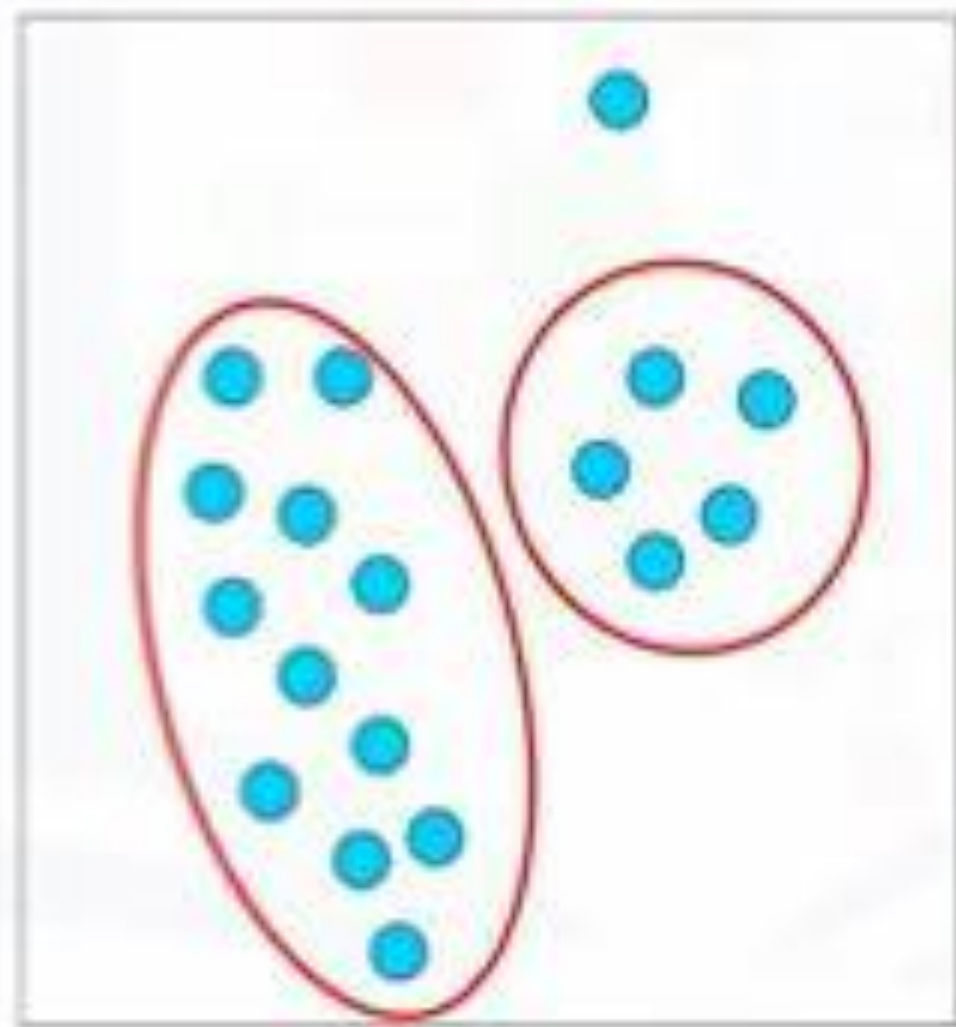


DBSCAN algorithm – clusters?

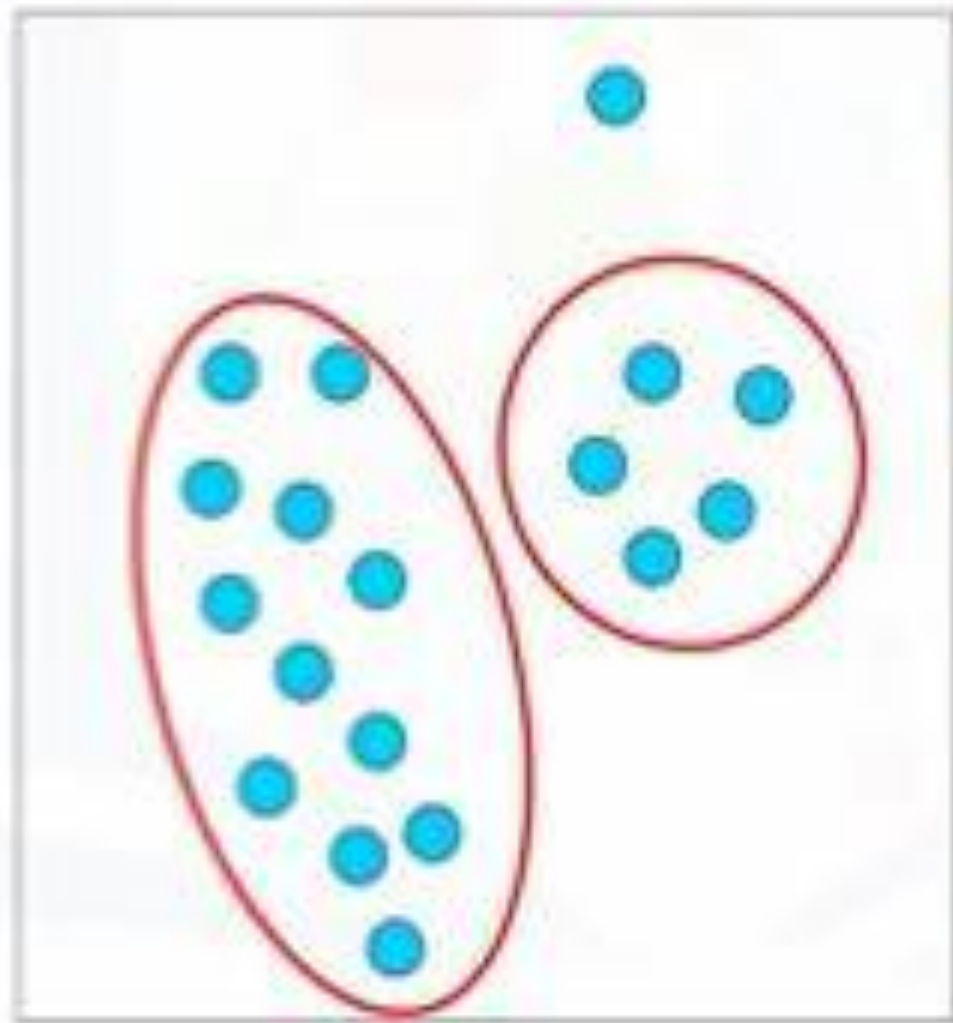


$R = 2\text{unit}$, $M = 6$

Advantages of DBSCAN



Advantages of DBSCAN



1. Arbitrarily shaped clusters
2. Robust to outliers
3. Does not require specification of the number of clusters