

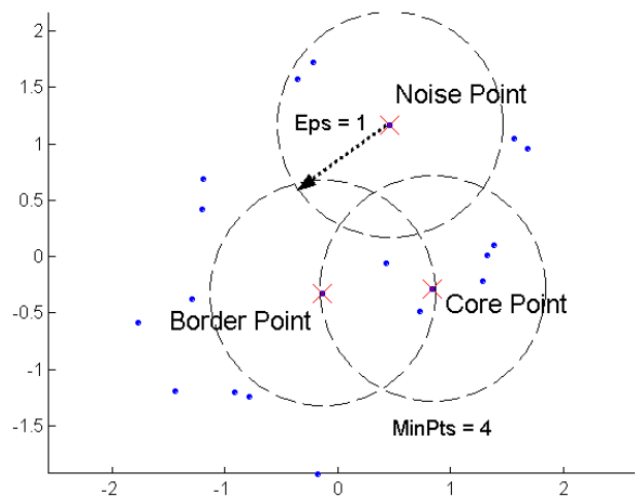
DBSCAN: Density-Based Clustering

- DBSCAN is a Density-Based Clustering algorithm
- Reminder: In density based clustering we partition points into dense regions separated by not-so-dense regions.
- Important Questions:
 - How do we measure density?
 - What is a dense region?
- DBSCAN:
 - Density at point p : number of points within a circle of radius Eps
 - Dense Region: A circle of radius Eps that contains at least $MinPts$ points

DBSCAN

- Characterization of points
 - A point is a core point if it has more than a specified number of points ($MinPts$) within Eps
 - These points belong in a dense region and are at the interior of a cluster
 - A border point has fewer than $MinPts$ within Eps , but is in the neighborhood of a core point.
 - A noise point is any point that is not a core point or a border point.

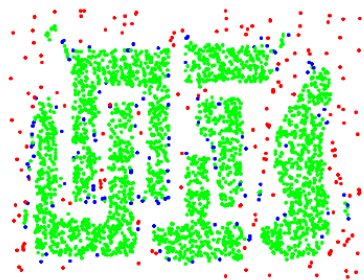
DBSCAN: Core, Border and Noise Points



DBSCAN: Core, Border and Noise Points



Original Points

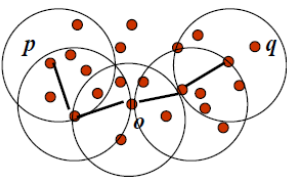
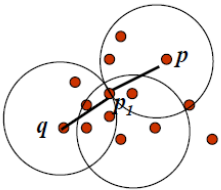


Point types: core,
border and noise

Eps = 10, MinPts = 4

Density-Connected Points

- **Density edge**
 - We place an **edge** between two core points **q** and **p** if they are within distance **Eps**.
- **Density-connected**
 - A point **p** is **density-connected** to a point **q** if there is a **path of edges** from **p** to **q**



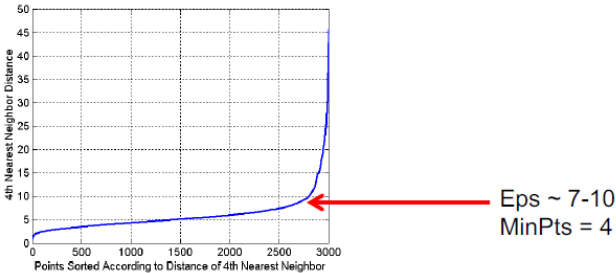
DBSCAN Algorithm

DBSCAN Algorithm

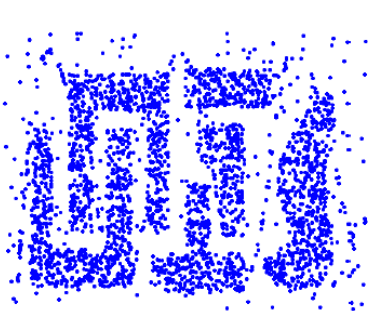
- Label points as **core**, **border** and **noise**
- Eliminate **noise** points
- For every **core** point **p** that has not been assigned to a cluster
 - Create a new cluster with the point **p** and all the points that are **density-connected** to **p**.
- Assign **border** points to the cluster of the closest core point.

DBSCAN: Determining Eps and MinPts

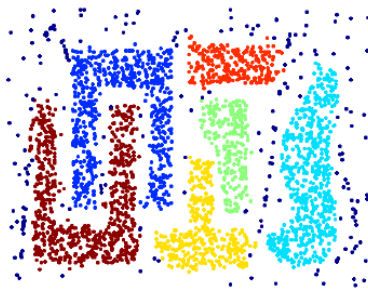
- Idea is that for points in a cluster, their k^{th} nearest neighbors are at roughly the same distance
- Noise points have the k^{th} nearest neighbor at farther distance
- So, plot sorted distance of every point to its k^{th} nearest neighbor
- Find the distance d where there is a “knee” in the curve
 - $\text{Eps} = d$, $\text{MinPts} = k$



When DBSCAN Works Well



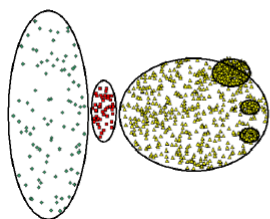
Original Points



Clusters

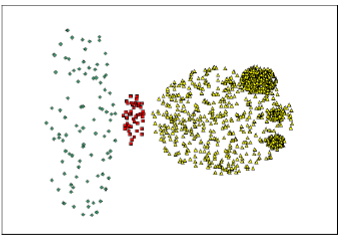
- Resistant to Noise
- Can handle clusters of different shapes and sizes

When DBSCAN Does Not Work Well

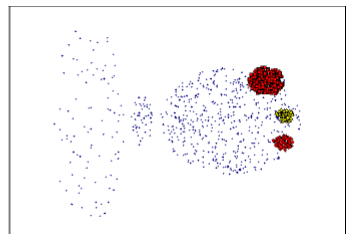


Original Points

- Varying densities
- High-dimensional data



(MinPts=4, Eps=9.75).



(MinPts=4, Eps=9.92)

DBSCAN: Sensitive to Parameters

Figure 8. DBScan results for DS1 with MinPts at 4 and Eps at (a) 0.5 and (b) 0.4.

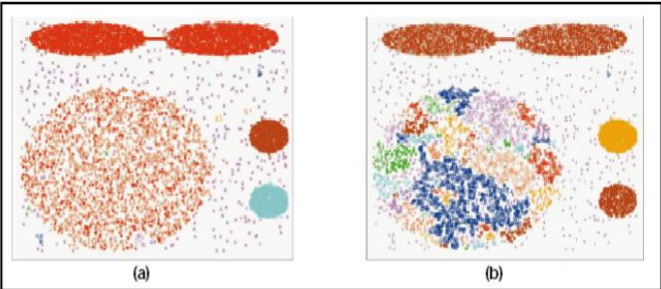


Figure 9. DBScan results for DS2 with MinPts at 4 and Eps at (a) 5.0, (b) 3.5, and (c) 3.0.

