

Density-Based Clustering

Isaiah Shaefer

Overview

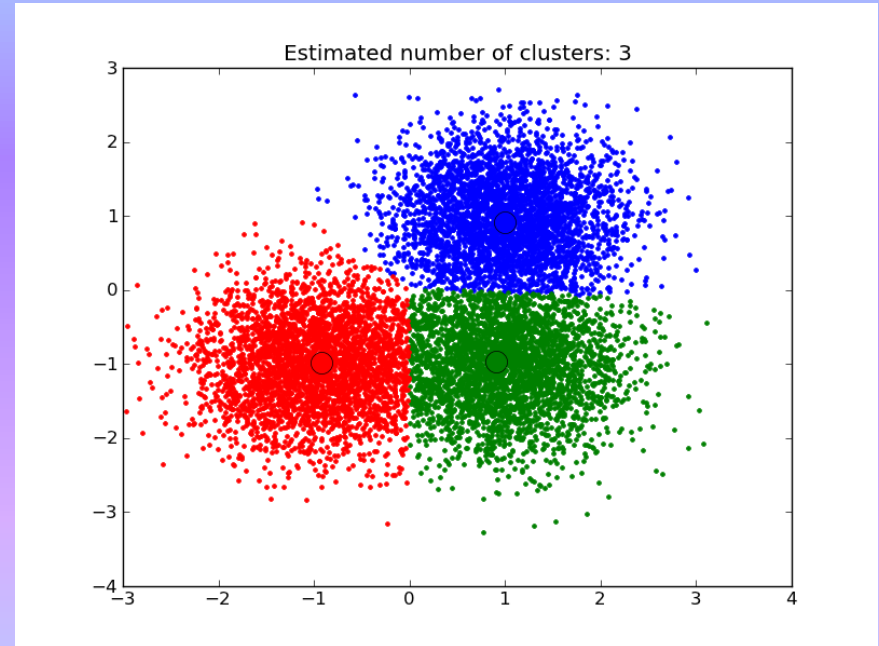
- Clustering in general
- Different types of clustering
- DBSCAN algorithm
- DBSCAN benefits
- DBSCAN problems
- OPTICS glaze-over.

Clustering

- Type of data mining technique.
- Separates data into specific groups.
- Algorithm not precisely defined.
- Many models for clustering.

Clustering Models

- Connectivity Models
- Centroid Models
- Distribution Models
- Density Models
- Subspace Models

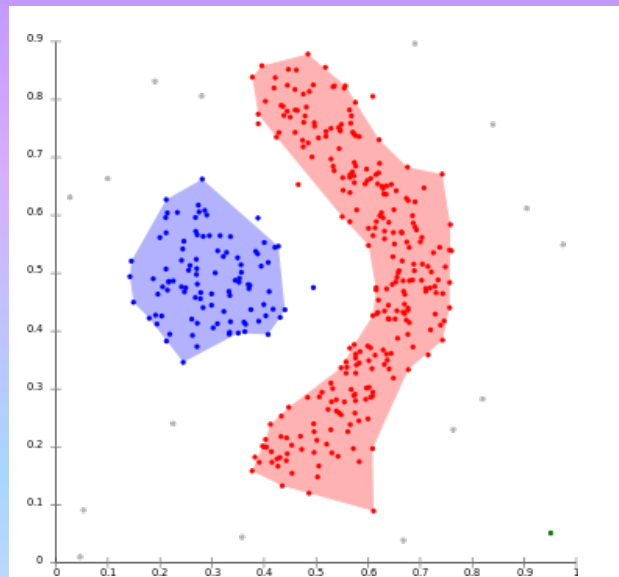


DBSCAN Algorithm

- Uses neighborhood (ϵ) and minimum points (minPts) parameters.
- A point's ϵ is scanned for other points.
 - If ϵ contains minPts, the center point and all points within ϵ become a cluster.
 - Else, the scanned point becomes noise.
- This model is called Density Reachability.
- Works well for unknown data.

DBSCAN benefits

- Given number of clusters not needed.
- Can form unusual shapes.
- Mostly insensitive to order.

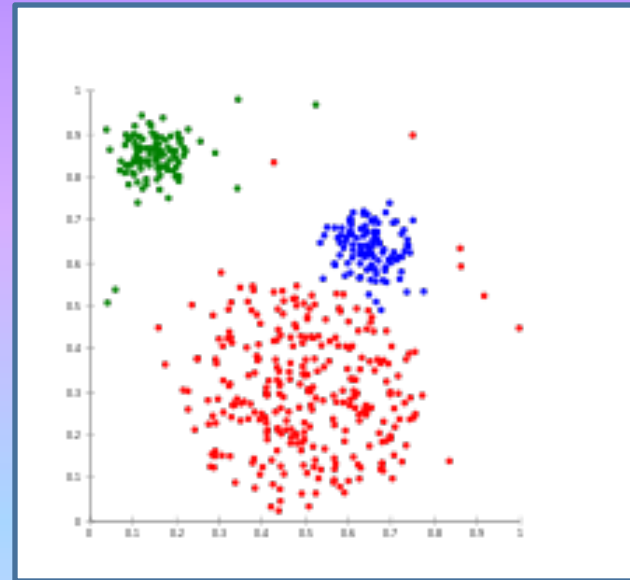
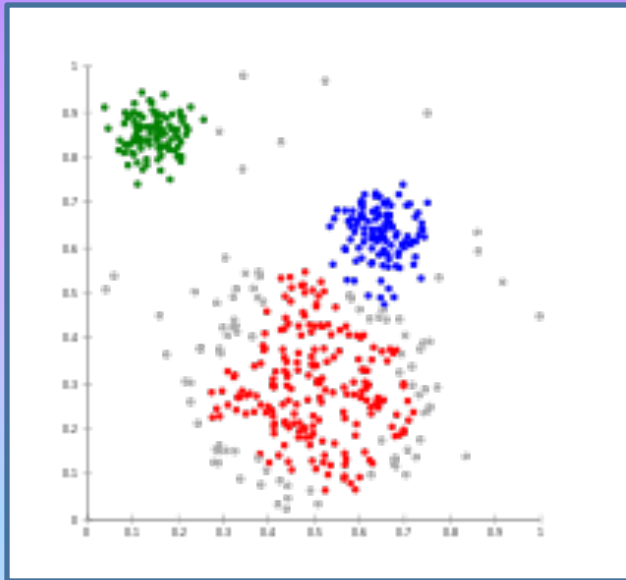


DBSCAN problems

- It relies on distance
 - Curse of Dimensionality.
- It has issues with varying densities.
- Sometimes has issues separating nearby clusters properly.

OPTICS algorithm

- Improved form of DBSCAN
- Look it up if you are interested.



Summary

- Clustering used to group data.
- Many models for clustering.
- Density-Based Clustering is unique.
- DBSCAN has its pros and cons.
- OPTICS expands on DBSCAN.

Works Cited

- Dunham, M. H. (2002). *Data Mining - Introductory and Advanced Topics*. Pearson Education, Inc.
- Bellman, R. E. (1957). *Dynamic Programming*. Princeton University Press.