Density-based Clusters

A cluster is a dense region of objects surrounded by a region of low density.

DBSCAN

Density-Based Spatial Clustering of Applications with Noise

Classification of Points

Given a radius $\varepsilon$ and the minimum number of points $\text{MinPts}$ within a radius of $\varepsilon$ ($\varepsilon$-neighborhood)
- Core point
  - Points in its $\varepsilon$-neighborhood $\geq \text{MinPts}$
- Border points
  - Within the $\varepsilon$-neighborhood of a core point
- Noise points

Point Examples

The DBSCAN Algorithm

- Label all points as core, border, or noise
- Remove all noise points
- Put an edge between all core points that are within $\varepsilon$ of each other
- Make each connected group of core points a cluster
- Assign border points to one of the clusters of their associated core points
DBSCAN Example

Select DBSCAN Parameters

- \( k \)-dist: distance to the \( k \)th nearest neighbor
- \( k = 4 \) is usually reasonable for most 2-D datasets

More DBSCAN Examples

About DBSCAN

- Handle clusters with arbitrary shapes and sizes
- Limitations
  - Clusters with varying densities
  - High dimensional data
- Could be expensive because of nearest neighbor computation
  - Use a spatial index structure like R tree or k-d tree

Readings

- Textbook 10.4.1