

DBSCAN

While K-means generates a fixed quantity of k variably-sized, globular clusters from a dataset, **Density-based spatial clustering of applications with noise** (DBSCAN) fits the data to some number of clusters given a threshold `min_pts` for what defines a cluster.

1. Find the ϵ -neighborhood of each point
2. Label the point as **core** if it contains at least `min_pts`
3. For each **core** point, assign to the same cluster all **core** points in its neighborhood
4. Label points in its neighborhood that are not core as **border**
5. Label points as noise if they are neither **core** nor **border**
6. Assign border points to nearby clusters

Benefits of DBSCAN

- Can identify clusters of different shapes and sizes
- Resistant to noise

Limitations of DBSCAN

Since the density that constitutes a cluster is predefined by ϵ , DBSCAN is subject to the following limitations:

- Can fail to identify clusters of varying densities
- Tends to create clusters of the same density
- Notion of density doesn't translate well to high-dimensional spaces