DBSCAN: Density-Based Spatial Clustering of Applications with Noise.

Detects noise in the data and uses it for clustering. We can have outliers

Kmeans big O = O(n³)

DBSCAN big O = O(n²)

Afbeelding met tekst, schermopname, Lettertype, diagram

Automatisch gegenereerde beschrijving

Afbeelding met tekst, schermopname, Lettertype, Perceel

Automatisch gegenereerde beschrijving

DBSCAN sees what density the points have.

DBSCAN parameters:

* Epsilon
  + Distance extended from a point
* Minimum number of points
  + minimum number of points in an epsilon distance

3 types of points:

* core
* border
* outlier

outliers cannot be reached by points in a cluster assignment

everything directly reachable is a core point

if it satisfies both conditions 🡪 core

1 condition 🡪 border

0 🡪 outlier

How to know the optimal number of samples 🡪 for loop

Epsilon intuiton:

* decreasing it causes more point to not be in range