# Dbscan 101

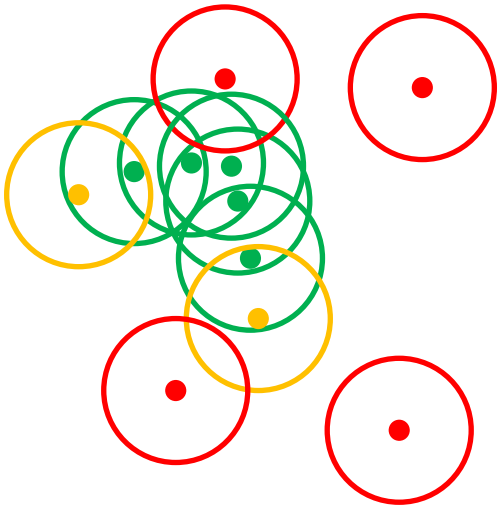
Bert Gollnick

# Dbscan

Introduction

- Density based Spatial Clustering of Applications with Noise
- identifies points in crowded regions
- Core points
- Reachable points
- Non-reachable points

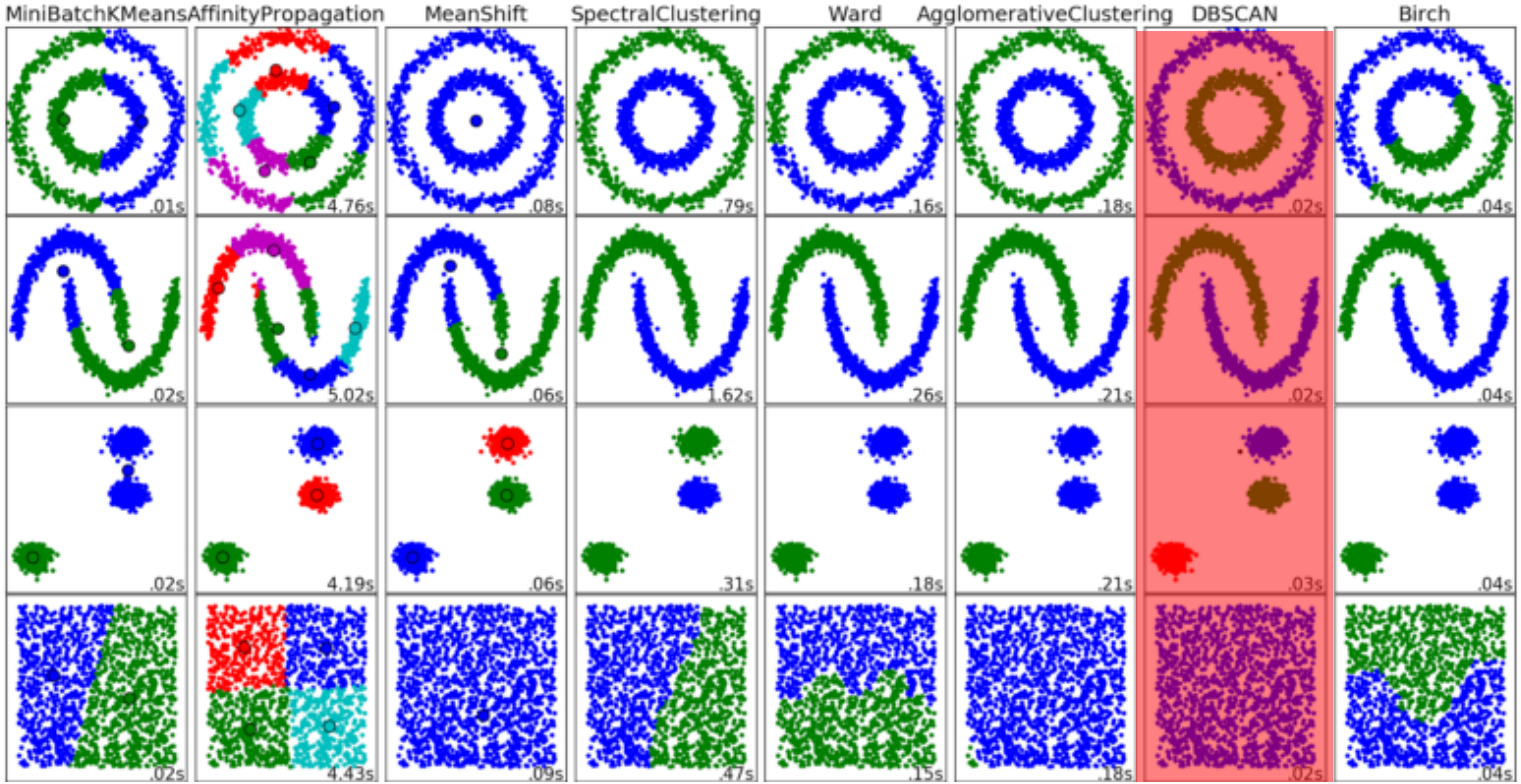
- Two parameters
  - min points
  - epsilon



Example:
minPoints=3
eps defines radius

# Dbscan

Example

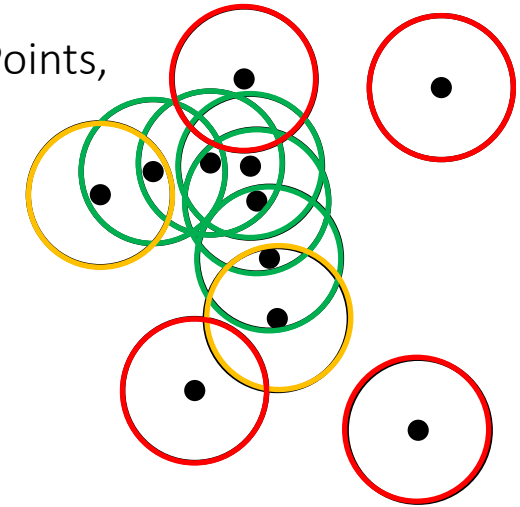



Source: https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80

# Dbscan

## Principle

1. Draw Circles of given eps-radius
2. If Circle contains >= minPoints
   → core point
3. If Circle contains < minPoints, but still reachable
   → Reachable point
4. Non-reachable points
   → Outliers or part of other clusters

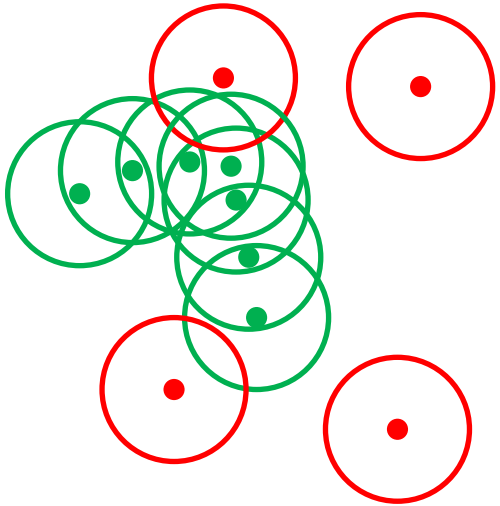


Example:
minPoints=3
eps defines radius

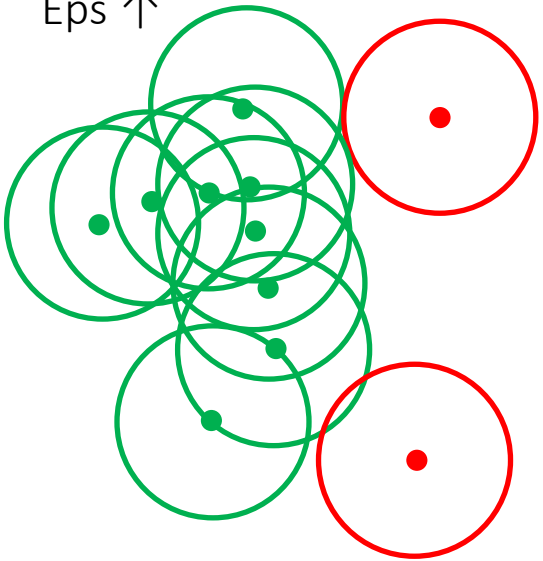
# Dbscan

Parameter Impact: epsilon increased

Baseline:
minPoints=2
Eps defines radius



Parameter Study:
minPoints=2
Eps ↑

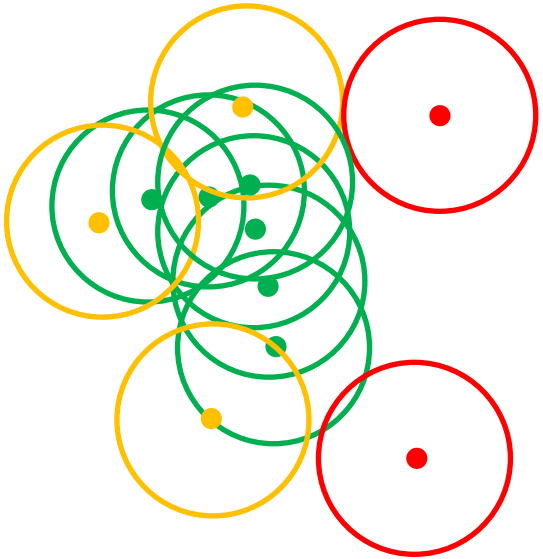


More Points are added to cluster

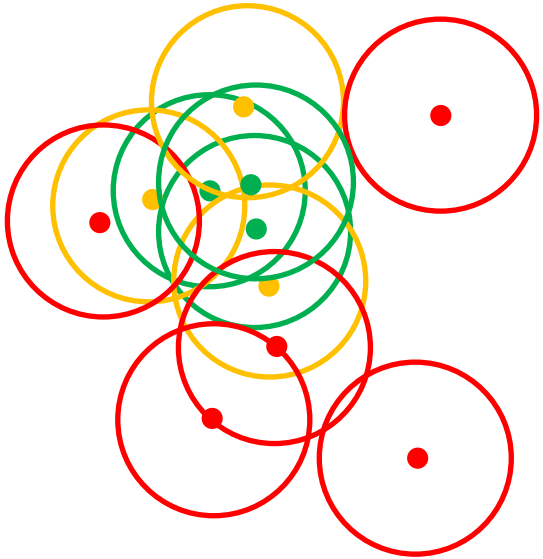
# Dbscan

Parameter Impact: minPoints increased

Baseline:
minPoints = 3
Eps defines radius



Parameter Study:
minPoints = 4
Eps constant



Cluster size decreases

# Dbscan

## Advantages / Disadvantages

\+

- No pre-knowledge or assumption on cluster number
- Can detect complex shapes
- Applicable for outlier detection

\-

- Border points can be part of either one (not completely deterministic)
- Quality relies on distance measure