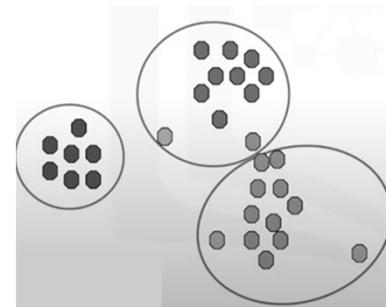


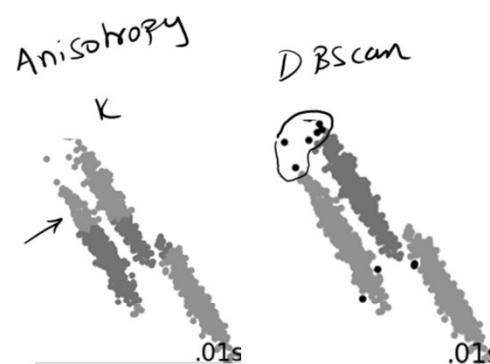
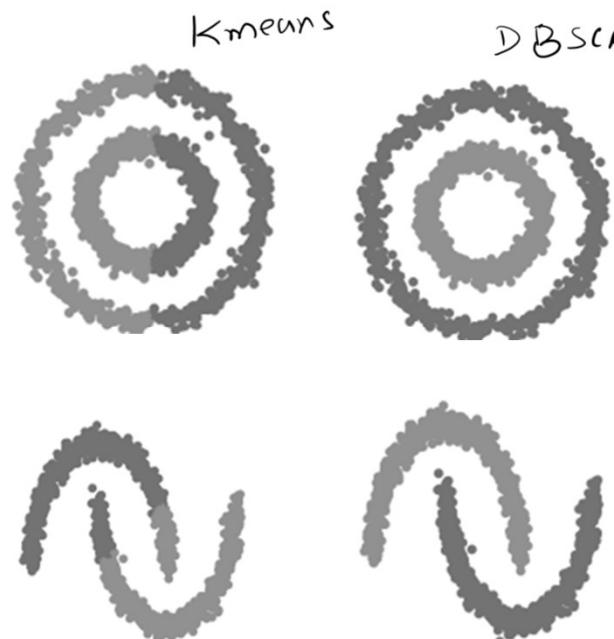
Problem with K-means

- Convex shape
- Isotropic variance
- outliers

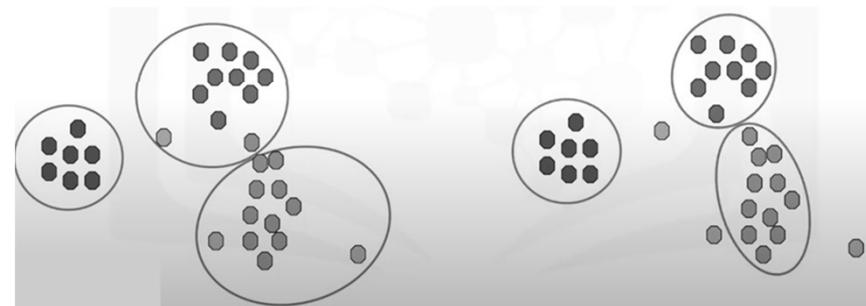


DBSCAN

- Density Bases Spatial Clustering of Applications with Noise



Better at outlier handling



DBSCAN

- Density Bases Spatial Clustering of Applications with Noise
- Eps: Radius of neighborhood
- n: Number of neighbors

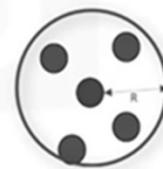
(Radius of neighborhood)

- Radius (R) that if includes enough number of points within, we call it a dense area

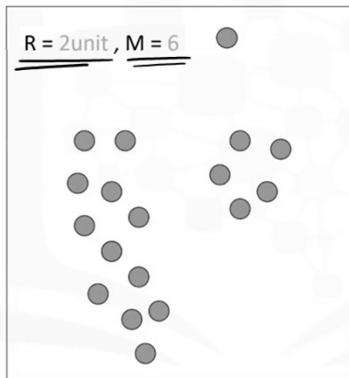


| (Min number of neighbors)

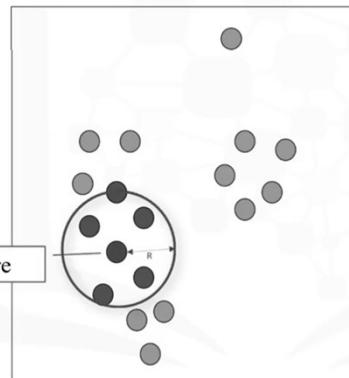
- The minimum number of data points we want in a neighborhood to define a cluster



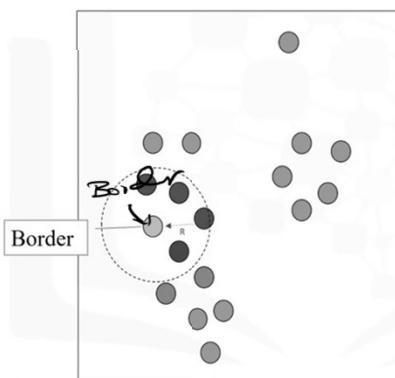
DBSCAN



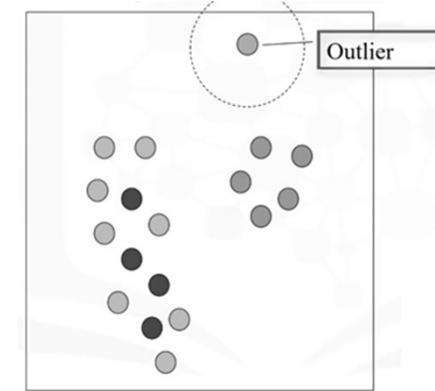
Each point is either:
→ *core point*
→ *border point*
→ *outlier point*



Number of points = M in R



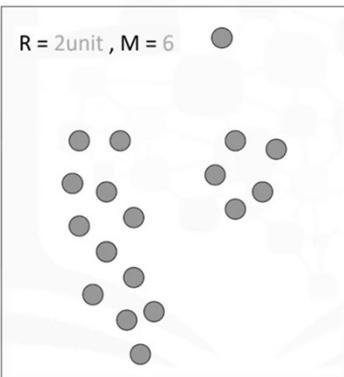
Number of points < M
or Reachable from some
core point



Number of points in $R=0$
or not Reachable from some
core point

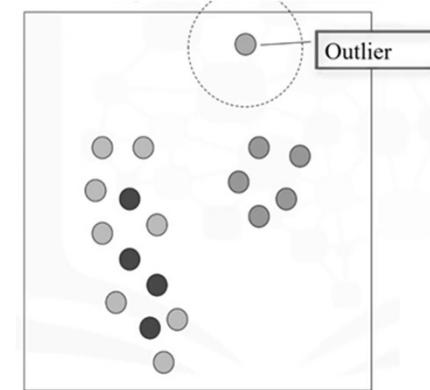
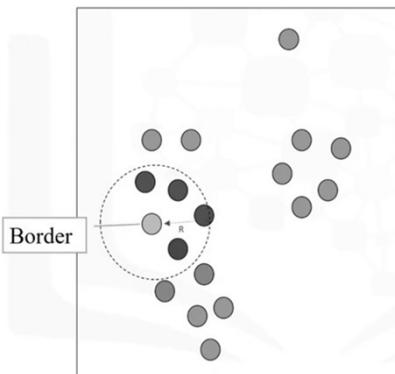
Group points as clusters based on there types

DBSCAN

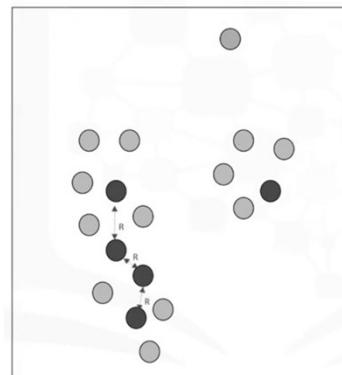


Each point is either:

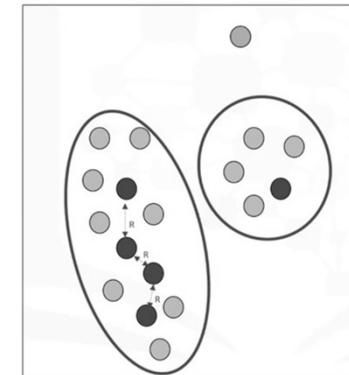
- core point
- border point
- outlier point



Group points as clusters based on there types



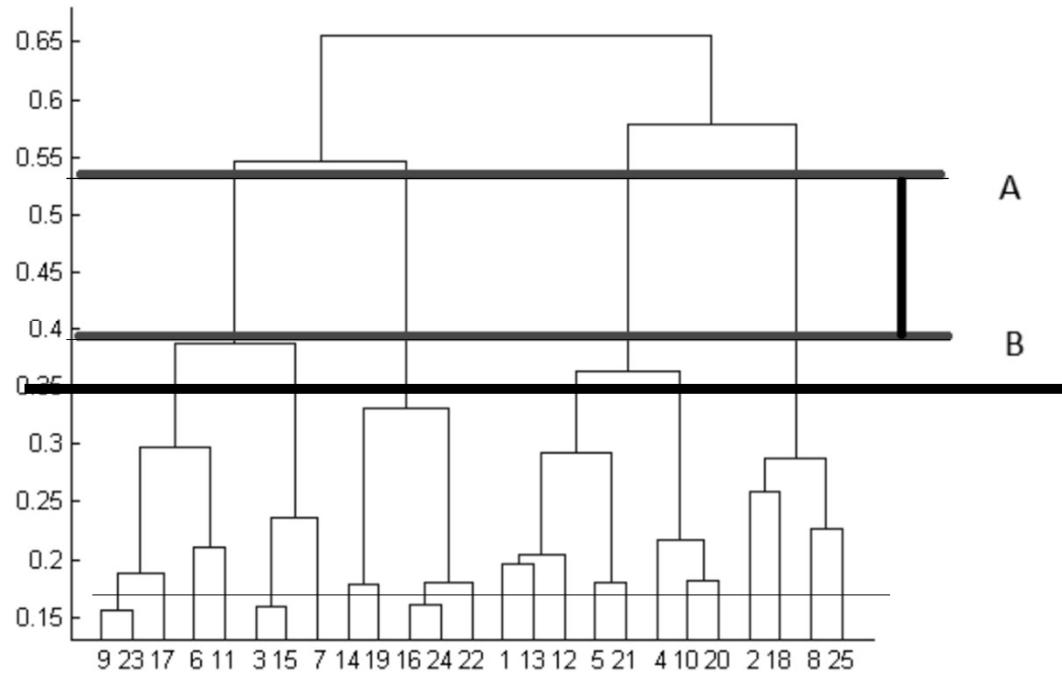
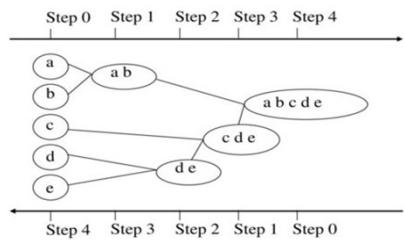
All core reachable core points in a cluster

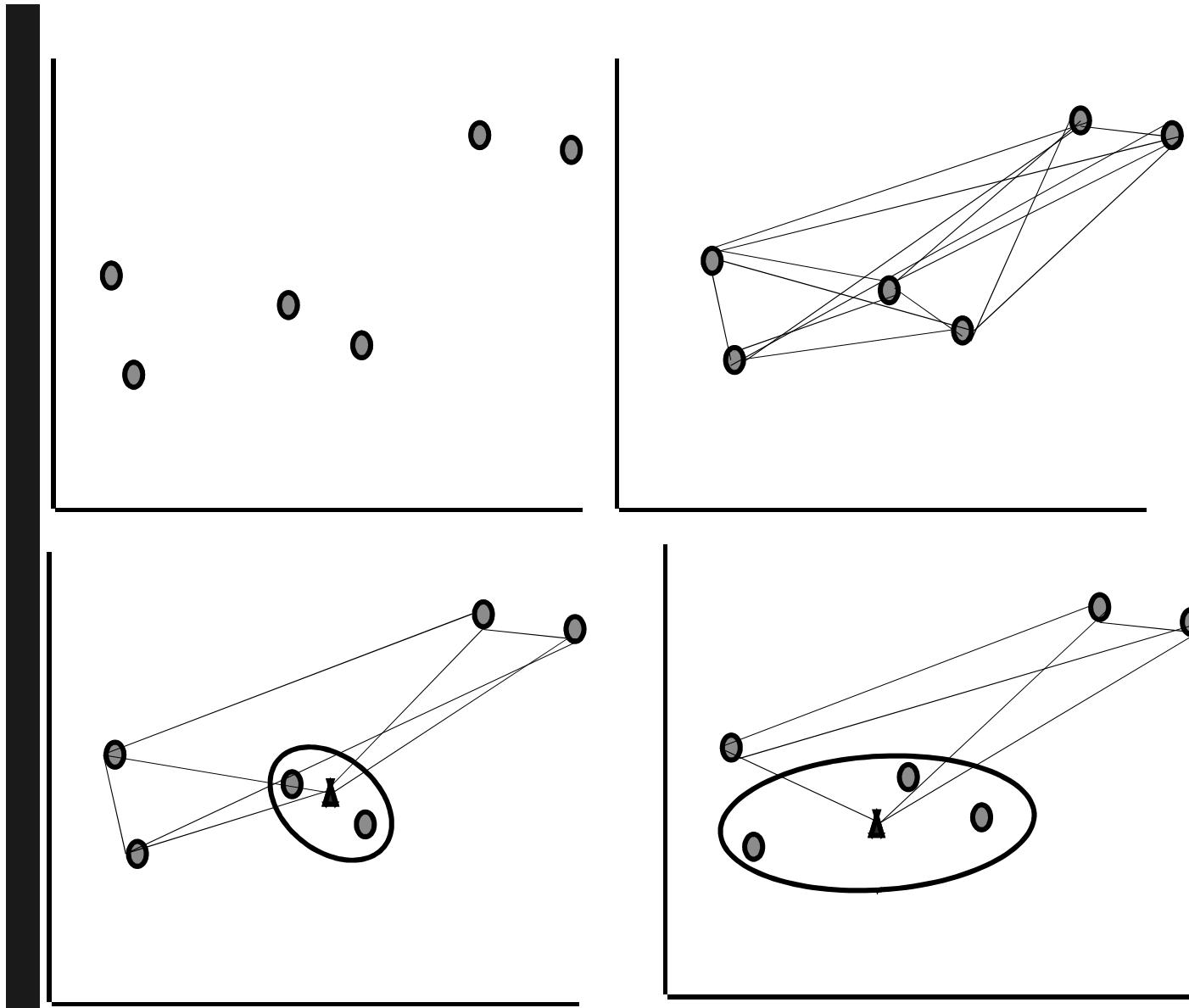


1 core point + border points= 1 cluster

Agglomerative

- Use distance matrix as clustering criteria. This method does not require the number of clusters k as an input, but needs a termination condition





Linkage

Agglomerative Clustering Linkage Algorithms

- Single linkage – Minimum distance or Nearest neighbour rule
- Complete linkage – Maximum distance or Farthest distance
- Average linkage – Average of the distances between all pairs
- Centroid method – combine cluster with minimum distance between the centroids of the two clusters
- Ward's method – Combine clusters with which the increase in within cluster variance is to the smallest degree

