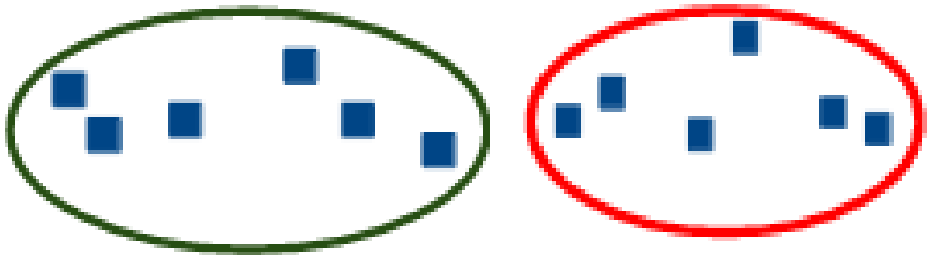
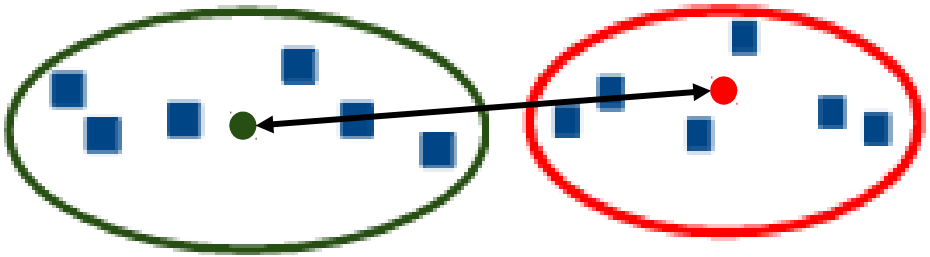
How to define similarity between clusters?

How to define similarity between clusters?

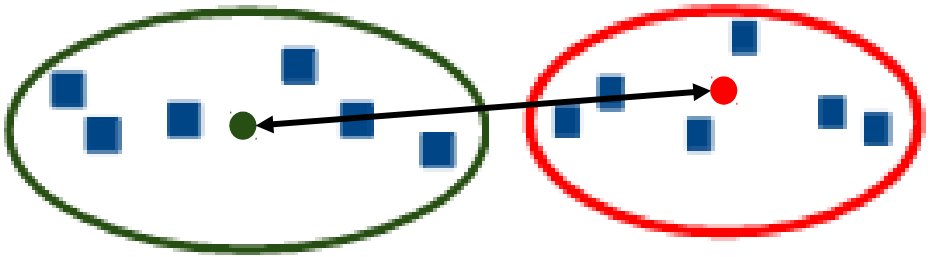


How to define similarity between clusters?



How to define similarity between clusters?

- Distance between Centroids

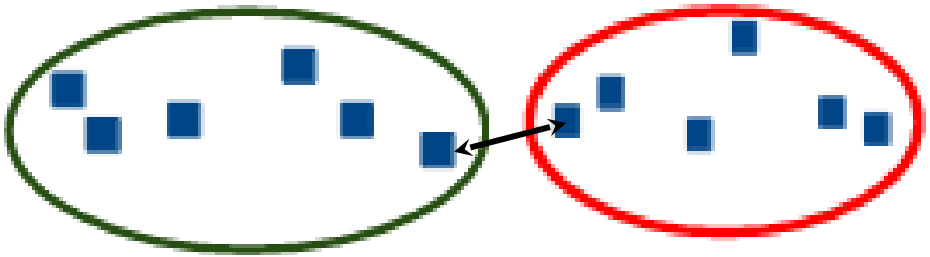


How to define similarity between clusters?

- Distance between Centroids
- MIN

How to define similarity between clusters?

- Distance between Centroids
- MIN

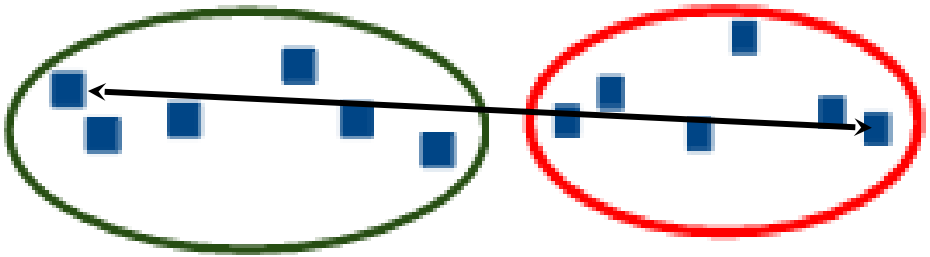


How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX

How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX

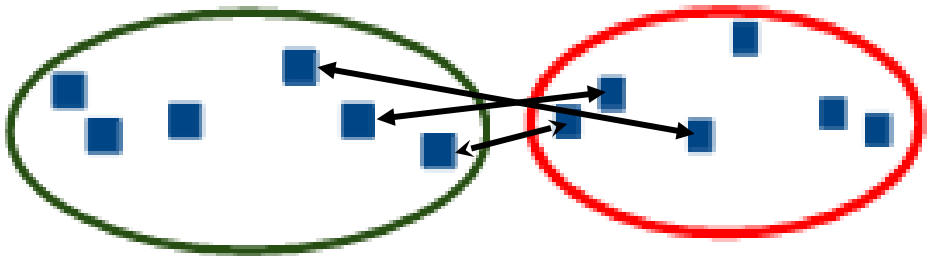


How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX
- K-Linkage

How to define similarity between clusters?

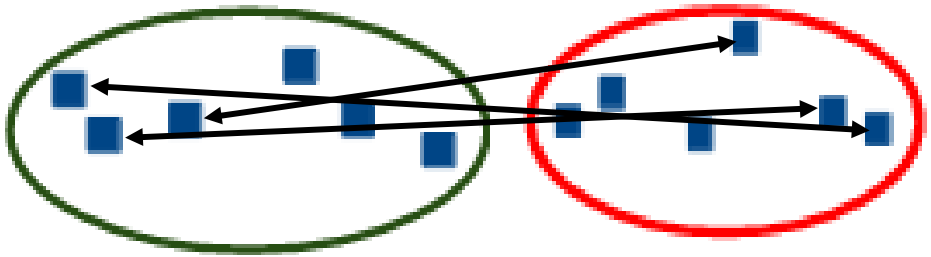
- Distance between Centroids
- MIN
- MAX
- K-Linkage
  - K-min Linkage



K = 3

How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX
- K-Linkage
  - K-min Linkage
  - K-max Linkage



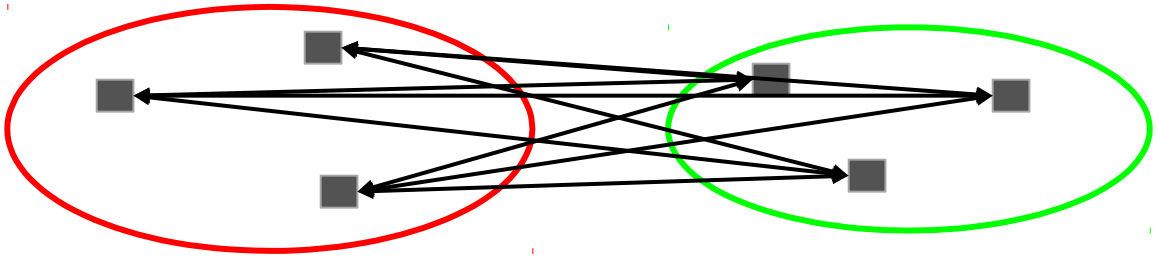
K = 3

How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX
- K-Linkage
  - K-min Linkage
  - K-max Linkage
- Group Average

How to define similarity between clusters?

- Distance between Centroids
- MIN
- MAX
- K-Linkage
  - K-min Linkage
  - K-max Linkage
- Group Average



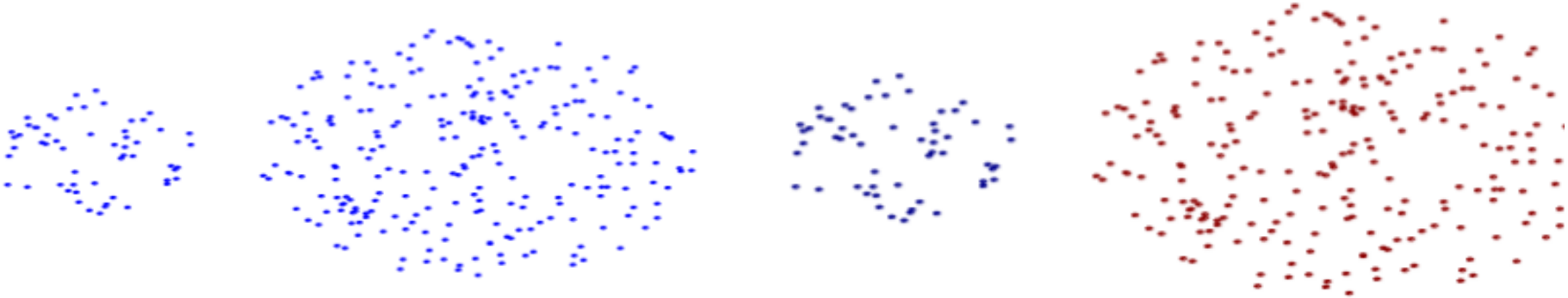
Strengths of using MIN distance

Strengths of using MIN distance

Original
Points

Strengths of using MIN distance

- It can handle non-elliptical shapes



Original Points

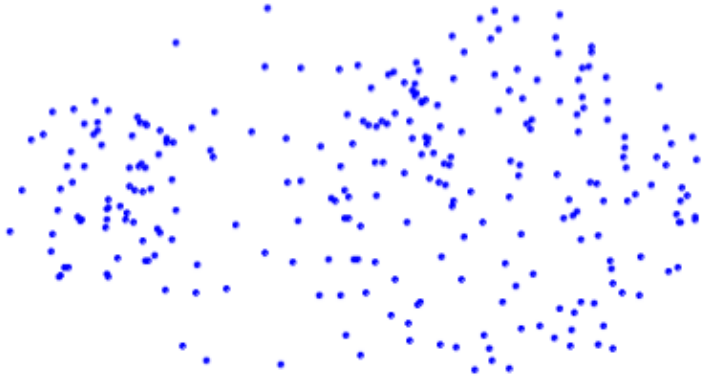
Clusters = 2

Weakness of using MIN distance

- Can not handle noise and outliers

Weakness of using MIN distance

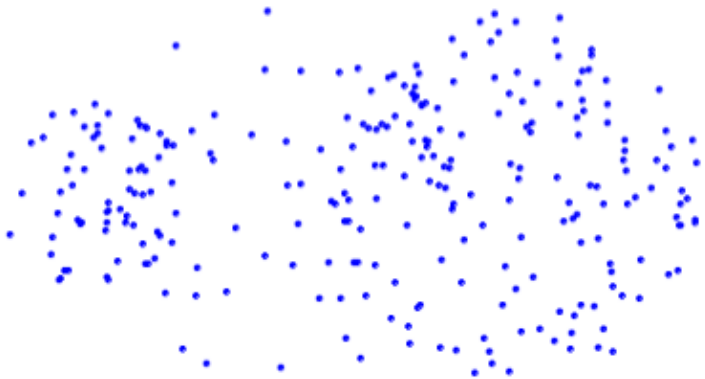
- Can not handle noise and outliers



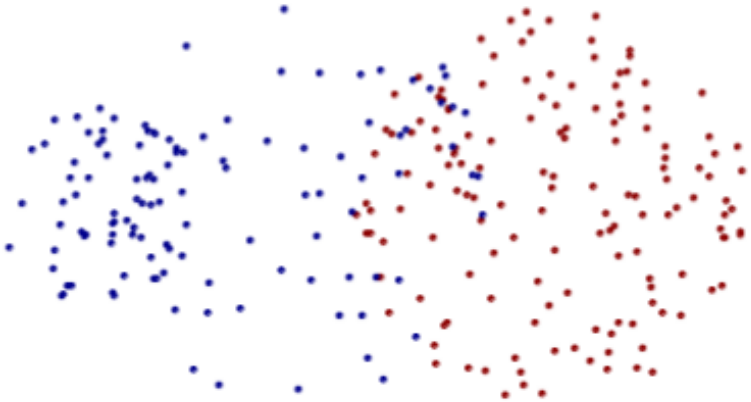
Original
Points

Weakness of using MIN distance

- Can not handle noise and outliers



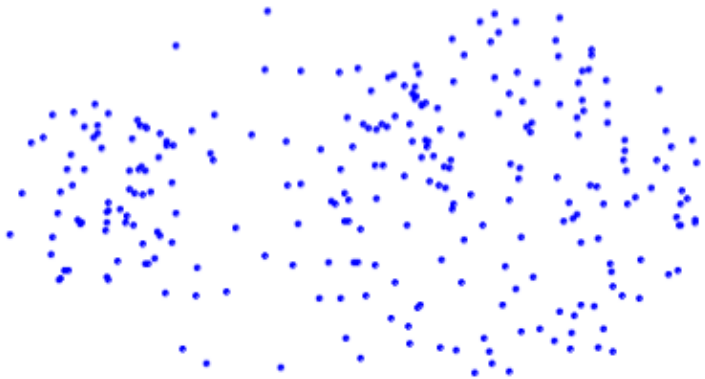
Original Points



Clusters = 2

Strengths of using MAX distance

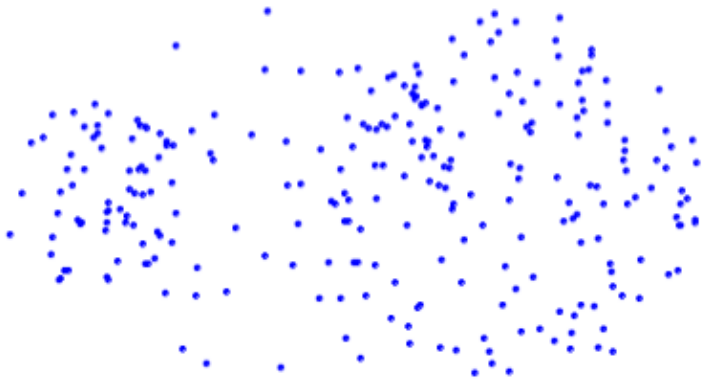
Strengths of using MAX distance



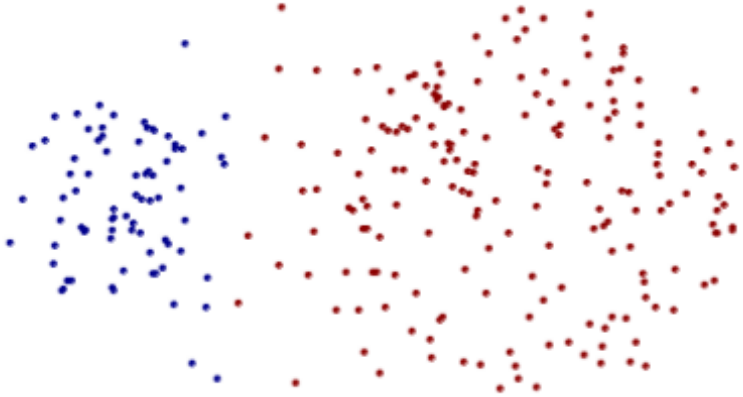
Original
Points

Strengths of using MAX distance

- It can handle noise and outliers



Original Points



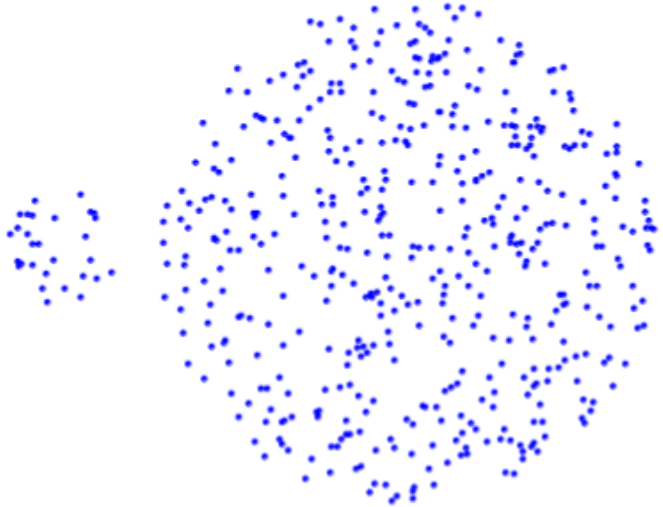
Clusters = 2

Weakness of using MAX distance

- Breaks large clusters
- Can not handle globular clusters

Weakness of using MAX distance

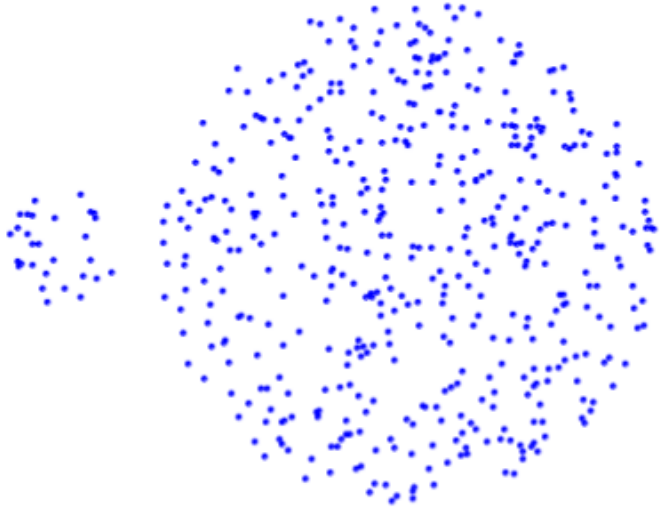
- Breaks large clusters
- Can not handle globular clusters



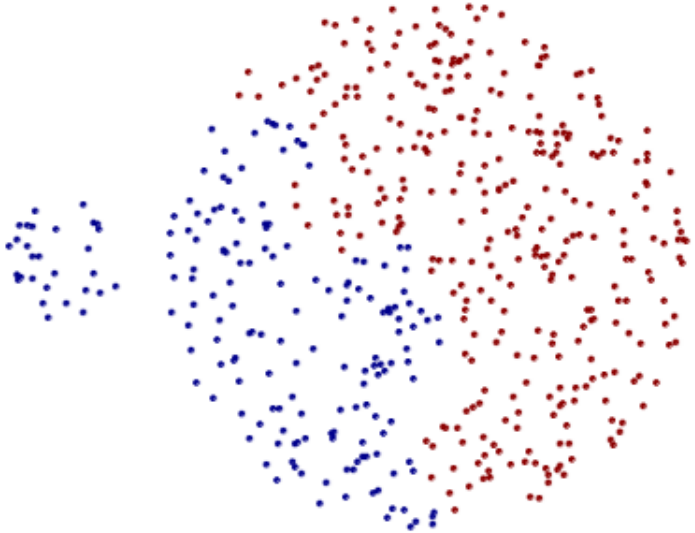
Original
Points

Weakness of using MAX distance

- Breaks large clusters
- Can not handle globular clusters



Original Points



Clusters = 2

Thank
You!

Analytics Vidhya
Learn everything about analytics