Fakultas Informatika
School of Computing
Telkom University

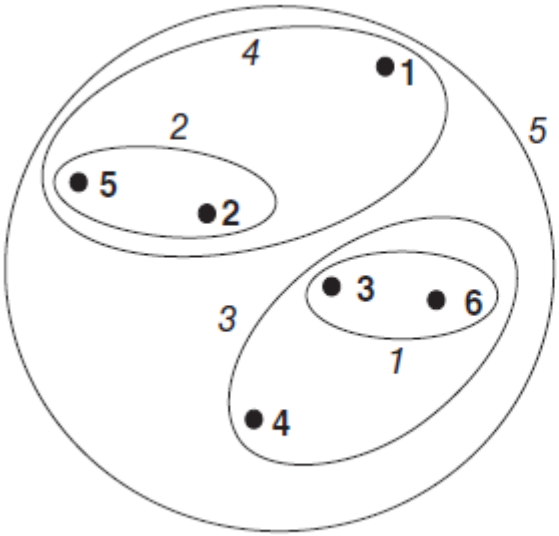
# Machine Learning

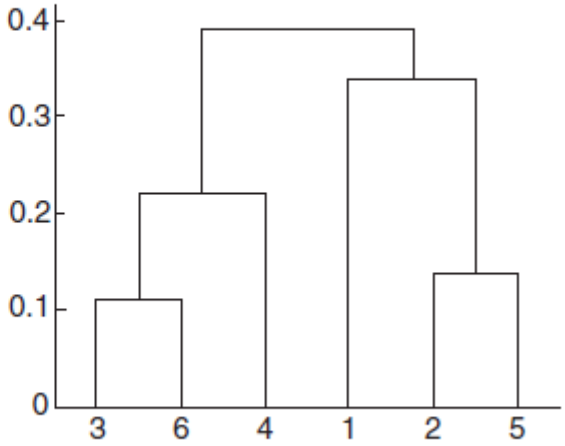
## Unsupervised Learning - Clustering

Adopted from ADF Slides

# Hierarchical Clustering

# Hierarchical Clustering



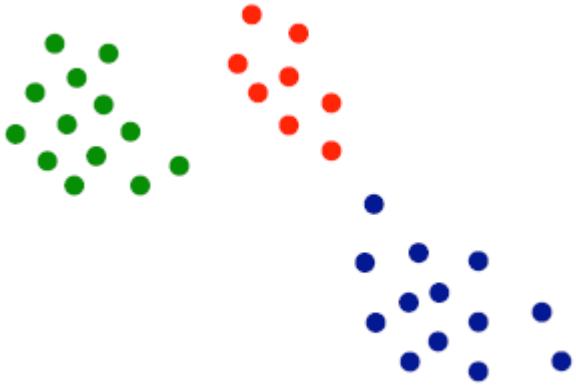

How to combine clusters?
What is the optimum cluster?

# Hierarchical Clustering

- Divisive Hierarchical Clustering (top-down)
  - Start with one cluster of all data
  - Repeat for all non-singleton clusters
    - Divide cluster into two using partitional method (e.g. k-means)
  - Rarely used in practice
- Agglomerative Hierarchical Clustering (bottom-up)
  - Start from an object as a cluster
  - Repeat until remain one cluster
    - Choose a pair of nearest cluster (most similar)
    - Combine into a larger cluster
  - Commonly used in practice

# Agglomerative Hierarchical Clustering

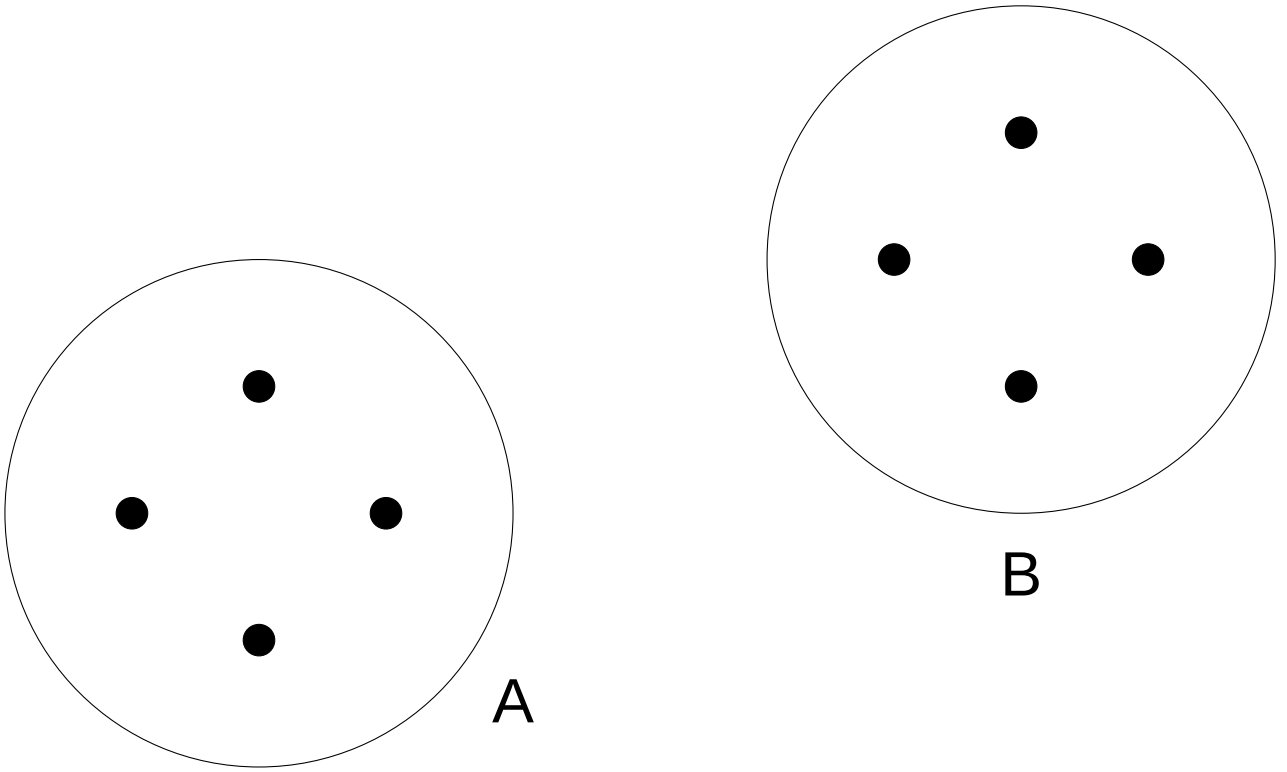
- Start from an object as a cluster
- Repeat until remain one cluster
  - Choose a pair of nearest cluster (most similar)
  - Combine into a larger cluster
- Define what is the similarity/proximity measures



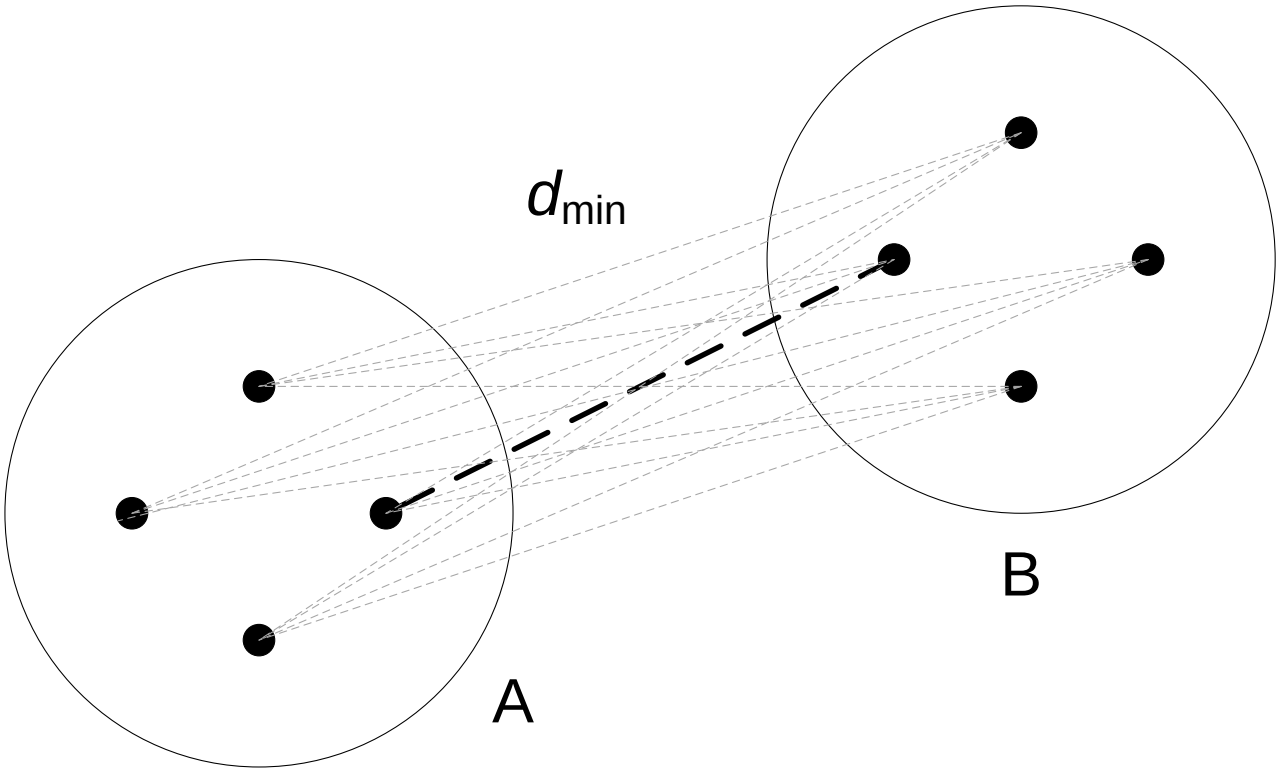
# Proximity Measures

- Single link
  - The lowest distance of a data point from cluster A to cluster B
- Complete link
  - The highest distance of a data point from cluster A to cluster B
- Group average
  - Average all distances from each data in cluster A to every data in cluster B
- Centroid
  - Distance of average data in cluster A to average data in cluster B
- Ward's

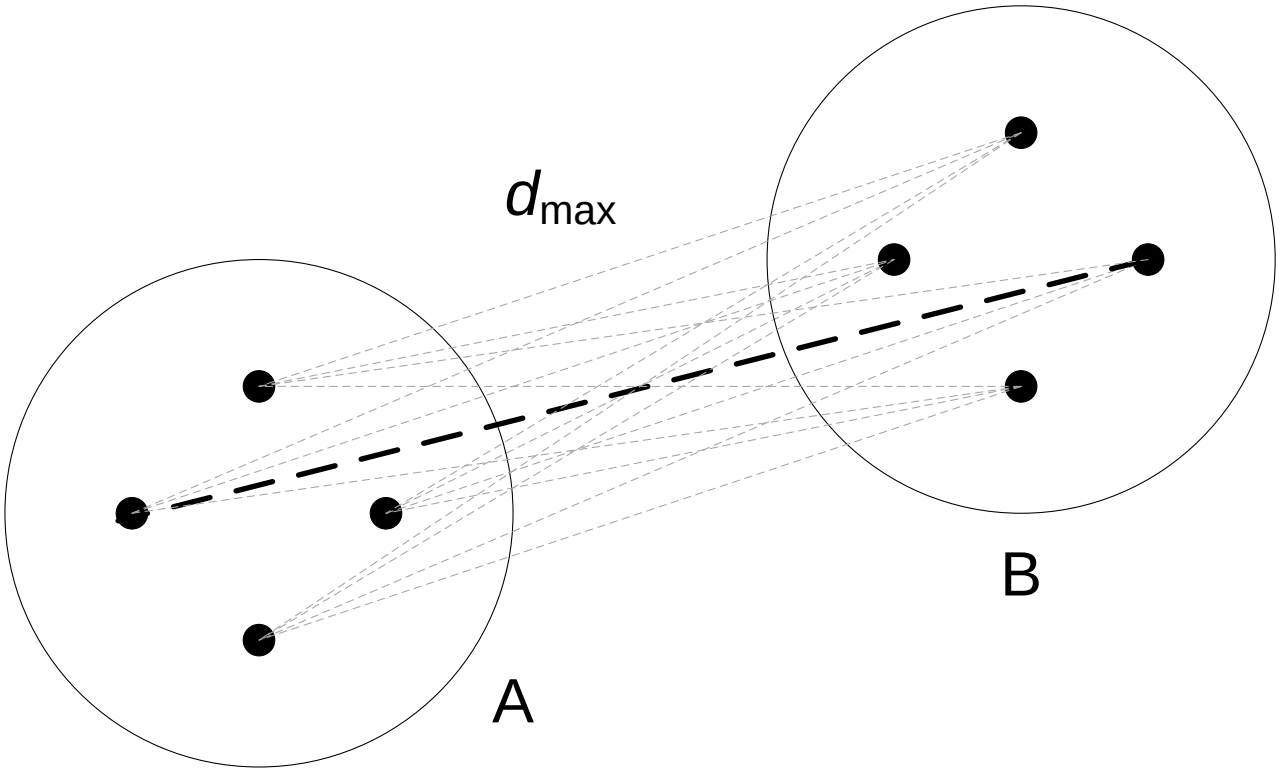
# Cluster distance/similarity



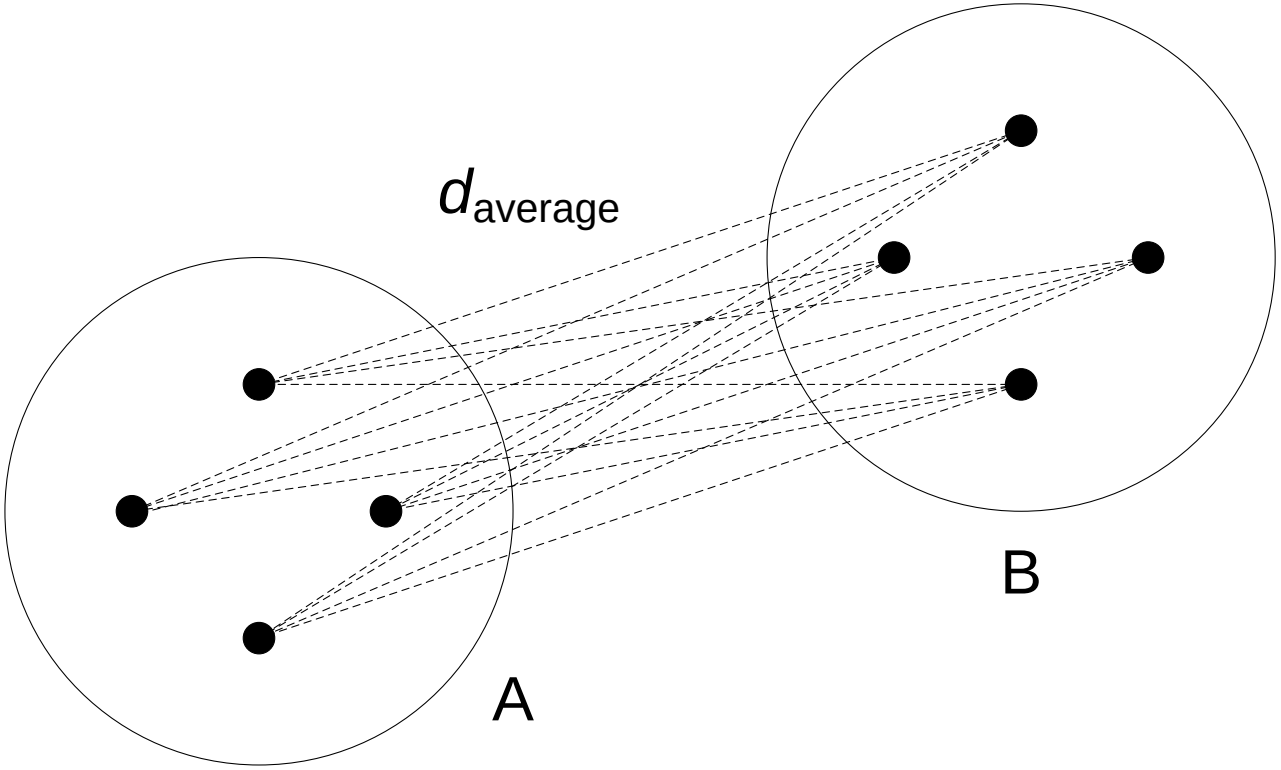
# Single Link (MIN)



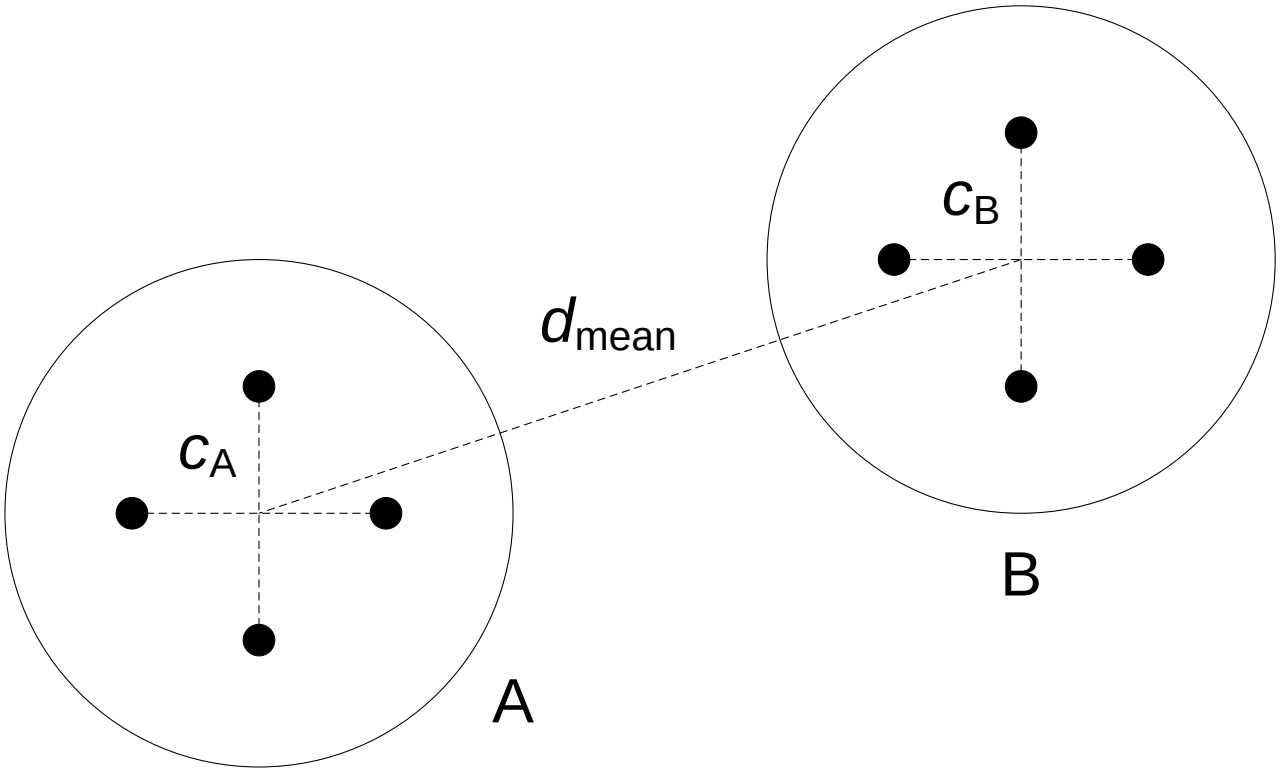
# Complete Link (MAX)



# Group Average

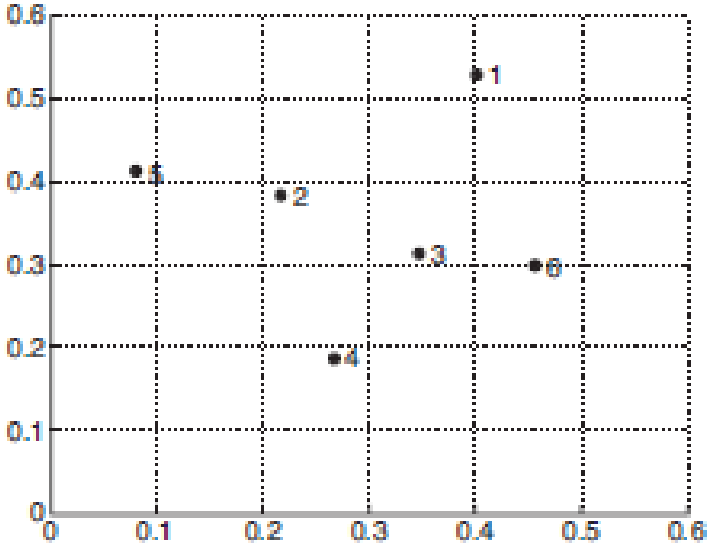


# Centroid (Mean distance)



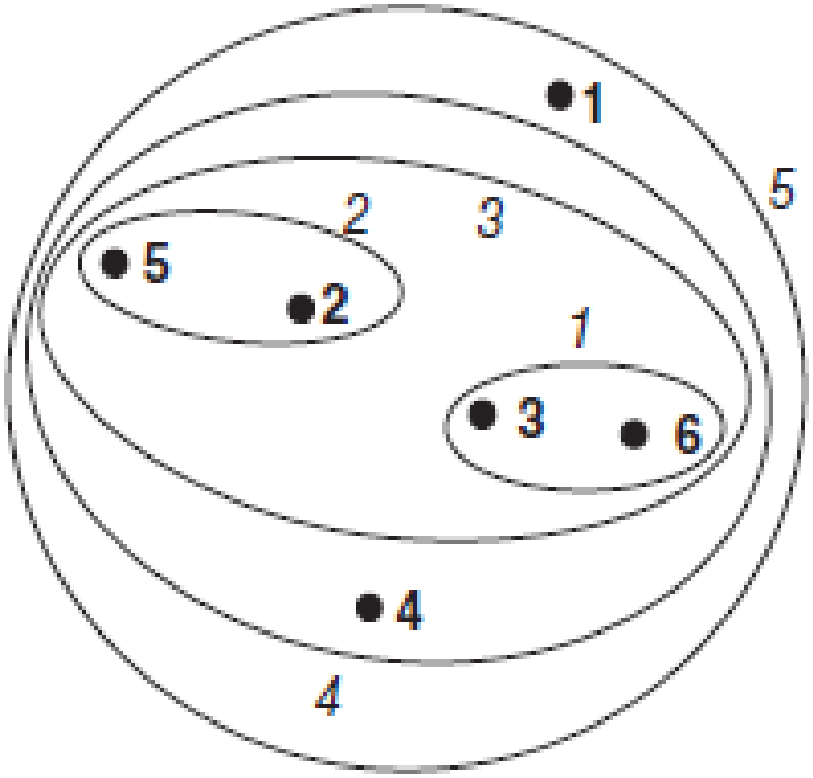
# Agglomerative Hierarchical Clustering Example

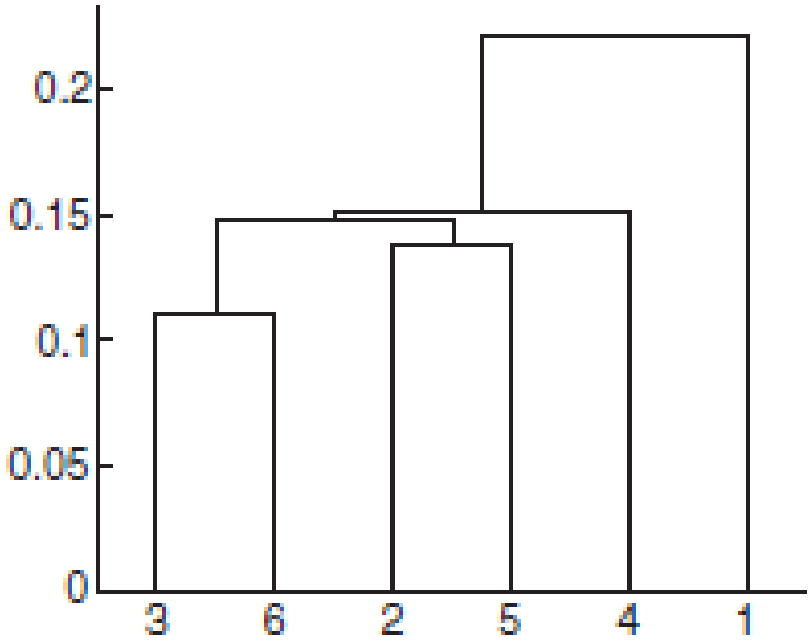
# Example



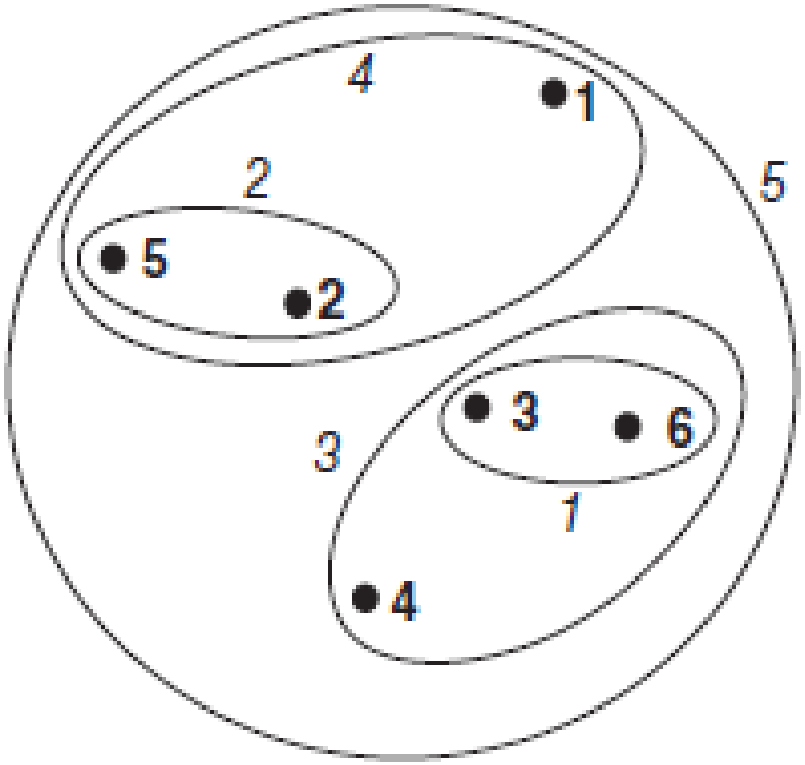

| | p1 | p2 | p3 | p4 | p5 | p6 |
|---|---|---|---|---|---|---|
| p1 | 0.00 | 0.24 | 0.22 | 0.37 | 0.34 | 0.23 |
| p2 | 0.24 | 0.00 | 0.15 | 0.20 | 0.14 | 0.25 |
| p3 | 0.22 | 0.15 | 0.00 | 0.15 | 0.28 | 0.11 |
| p4 | 0.37 | 0.20 | 0.15 | 0.00 | 0.29 | 0.22 |
| p5 | 0.34 | 0.14 | 0.28 | 0.29 | 0.00 | 0.39 |
| p6 | 0.23 | 0.25 | 0.11 | 0.22 | 0.39 | 0.00 |

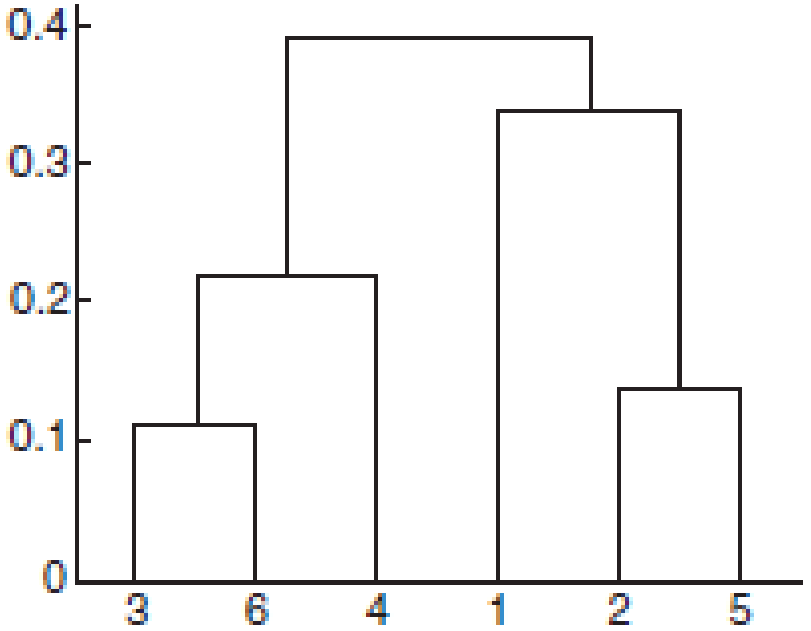
# Single Link (MIN)



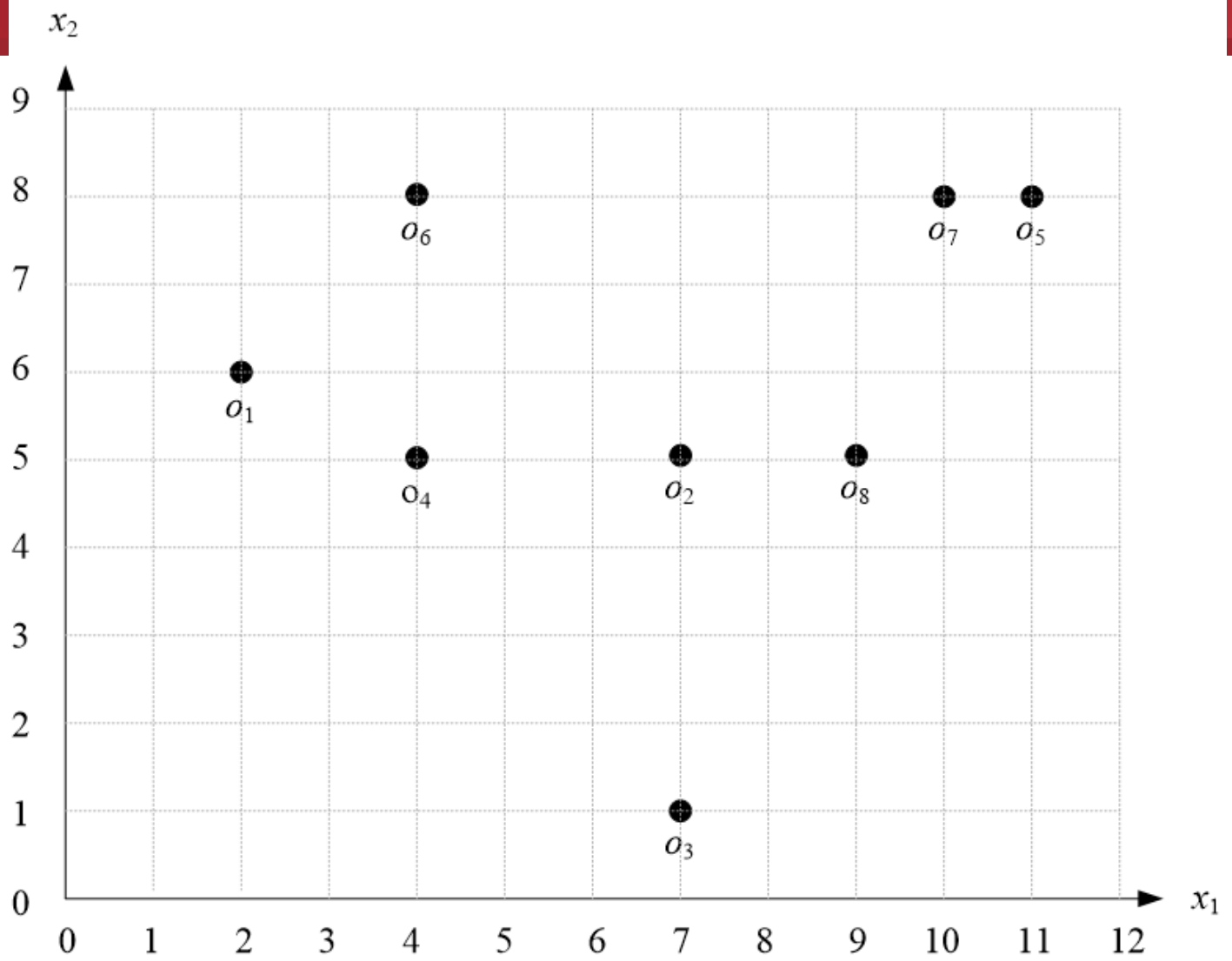



# Complete Link (MAX)



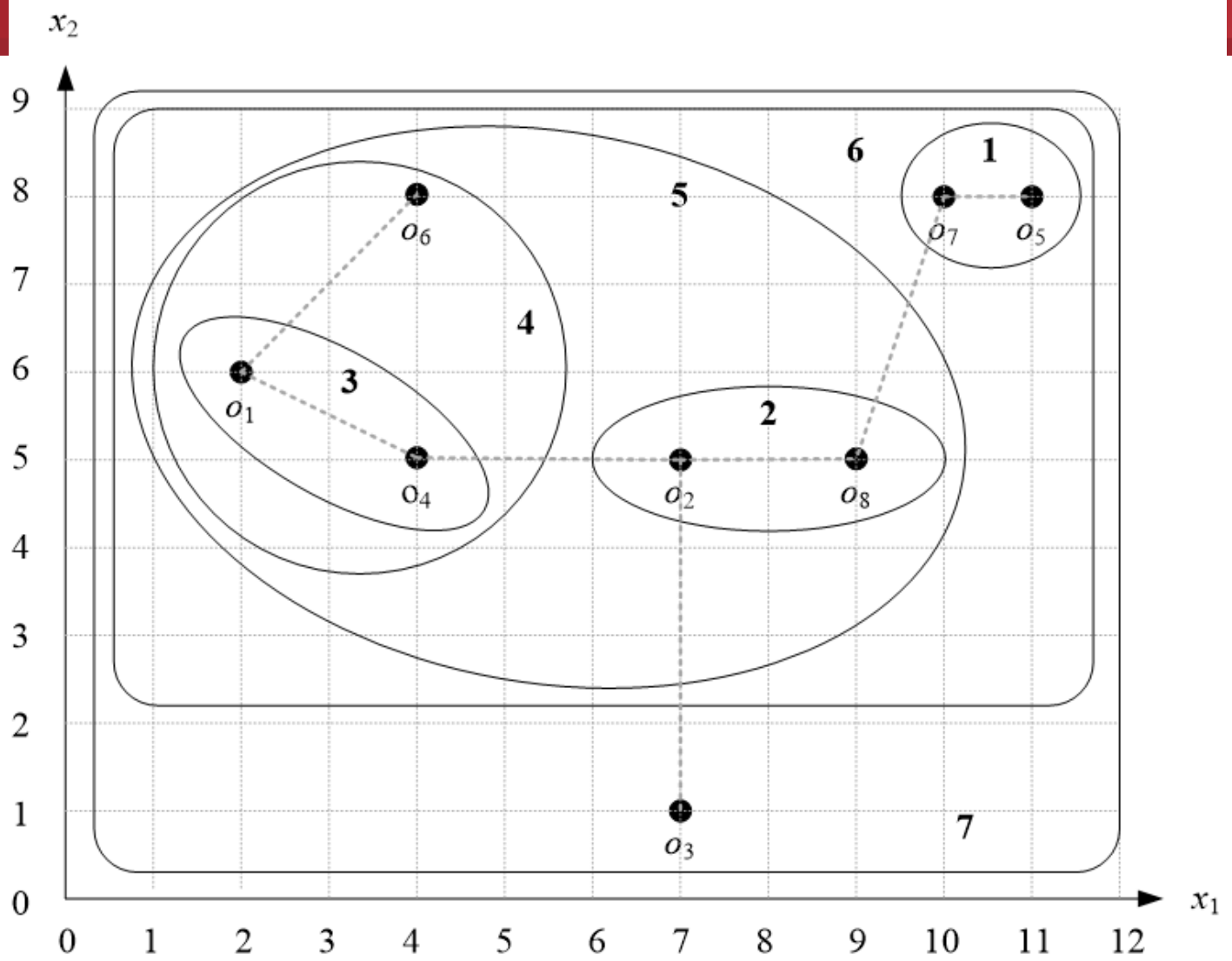



# Exercise



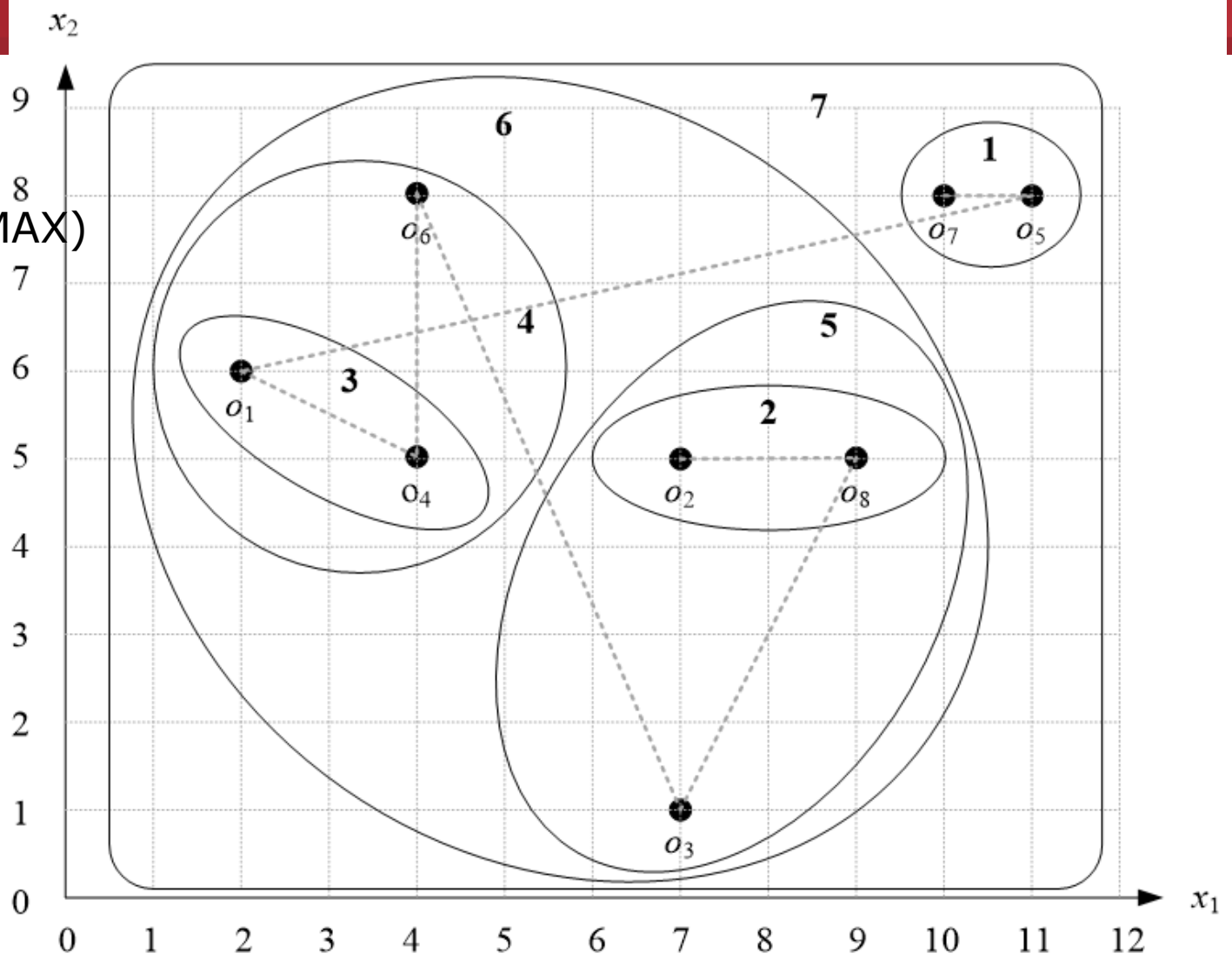
# Exercise

Single Link (MIN)



# Exercise

Complete Link (MAX)



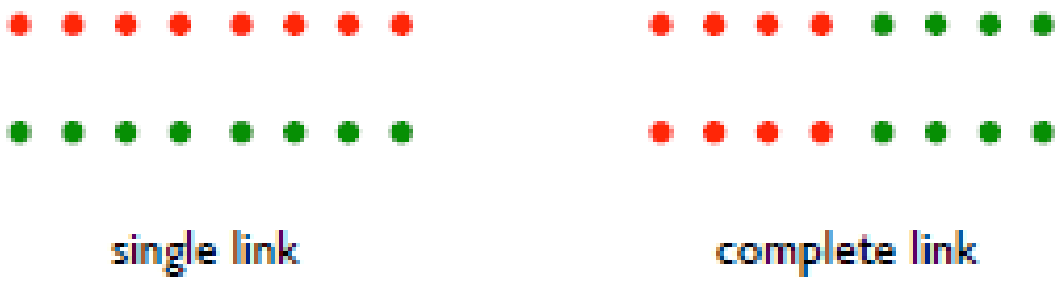
# Agglomerative Pros and Cons

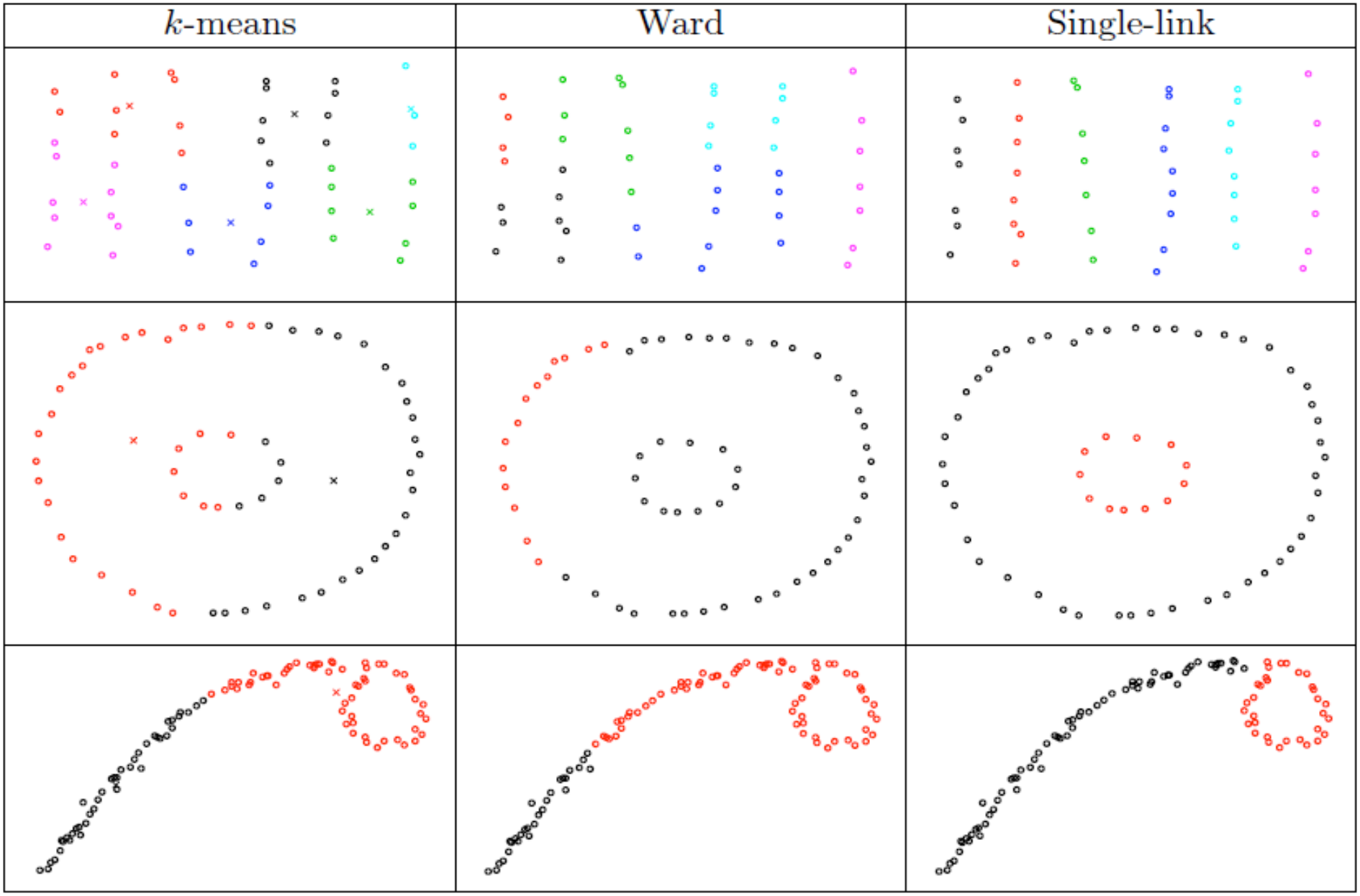
# Pros and Cons

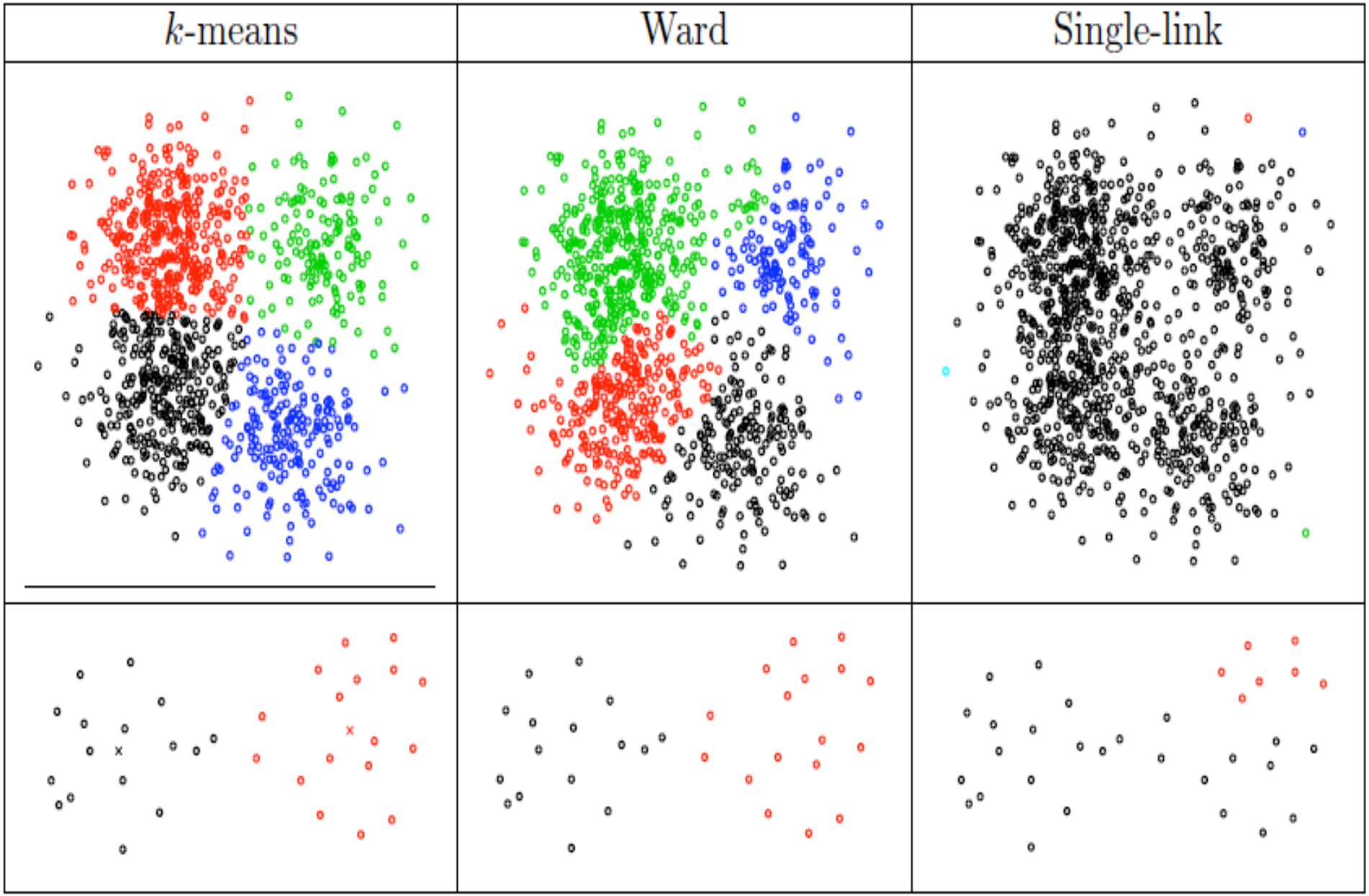
- Pros:
  - do not have a natural objective function that is being optimized (in contrast to K-means)
  - Monotonicity:
    - the dissimilarity between a pair clusters merged at any point in the algorithm is always at least as large as the dissimilarity of the pair of clusters merged in the previous step,
    - Only for Single-Link, Complete-Link, and Group Average
- Cons
  - Space Complexity = $O(N^2)$
  - Time Complexity = $O(N^3)$
    - reduced by various techniques to range of $O(N^2)$ to $O(N^2 \log N)$

# Cluster shapes

- Single-link can produce arbitrarily shaped clusters (joining quite different objects which have some intermediate links that connect them)
- Complete-link tends to produce fairly compact, globular clusters. Problems with clusters of different sizes.
- Group average is a compromise between the two




k-means
Ward
Single-link


k-means
Ward
Single-link

# Question?

Fakultas Informatika
School of Computing
Telkom University

# THANK YOU