IIT ROORKEE

FREE ONLINE EDUCATION swayam शिक्षित भारत, उन्नत भारत

NPTEL ONLINE CERTIFICATION COURSE

Hierarchical method of clustering - I

Dr. A. Ramesh

DEPARTMENT OF MANAGEMENT STUDIES

Agenda

- Introduction to Hierarchical clustering
- Partitioning Vs. Hierarchical

Introduction

- A hierarchical method creates a hierarchical decomposition of the given set of data objects
- A hierarchical clustering method works by grouping data objects into a tree of clusters
- A hierarchical method can be classified as being either agglomerative or divisive, based on how the hierarchical decomposition is formed
- The agglomerative approach, also called the bottom-up approach, starts with each object forming a separate group

Introduction

- It successively merges the objects or groups that are close to one another, until all of the groups are merged into one (the topmost level of the hierarchy), or until a termination condition holds
- The divisive approach, also called the top-down approach, starts with all of the objects in the same cluster
- In each successive iteration, a cluster is split up into smaller clusters, until eventually each object is in one cluster, or until a termination condition holds

Introduction

- Hierarchical methods suffer from the fact that once a step (merge or split) is done, it can never be undone
- This rigidity is useful in that it leads to smaller computation costs by not having to worry about a combinatorial number of different choices
- However, such techniques cannot correct erroneous decisions

Agglomerative and Divisive Hierarchical Clustering

Agglomerative

- This bottom-up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all of the objects are in a single cluster or until certain termination conditions are satisfied
- Most hierarchical clustering methods belong to this category

Divisive Hierarchical

- This top-down strategy does the reverse of agglomerative hierarchical clustering by starting with all objects in one cluster
- It subdivides the cluster into smaller and smaller pieces, until each object forms a cluster on its own or until it satisfies certain termination conditions, such as a desired number of clusters is obtained or the diameter of each cluster is within a certain threshold

Agglomerative versus divisive hierarchical clustering

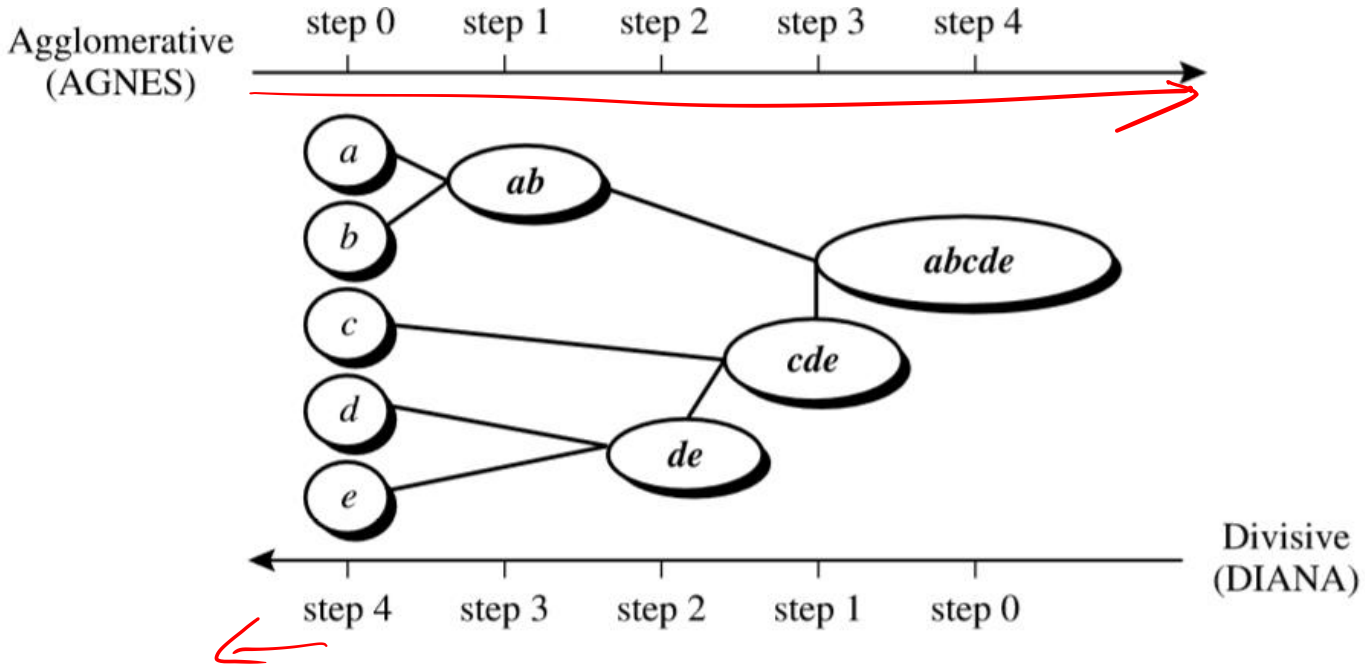


Figure: 1 Agglomerative and divisive hierarchical clustering on data objects{a,b,c,d,e}

Interpretation

- Figure: 1 shows the application of AGNES (AGglomerative NESting), an agglomerative hierarchical clustering method, and DIANA (DIvisive ANAlysis), a divisive hierarchical clustering method, to a data set of five objects, {a,b,c,d,e}
- Initially, AGNES places each object into a cluster of its own
- The clusters are then merged step-by-step according to some criterion
- Let's say for example, clusters C_1 and C_2 may be merged if an object in C_1 and an object in C_2 form the minimum Euclidean distance between any two objects from different clusters

Interpretation

- This is a single-linkage approach in that each cluster is represented by all of the objects in the cluster, and the similarity between two clusters is measured by the similarity of the closest pair of data points belonging to different clusters
- The cluster merging process repeats until all of the objects are eventually merged to form one cluster

Interpretation

- In DIANA, all of the objects are used to form one initial cluster
- The cluster is split according to some principle, such as the maximum Euclidean distance between the closest neighboring objects in the cluster
- The cluster splitting process repeats until, eventually, each new cluster contains only a single object
- In either agglomerative or divisive hierarchical clustering, the user can specify the desired number of clusters as a termination condition

Dendrogram

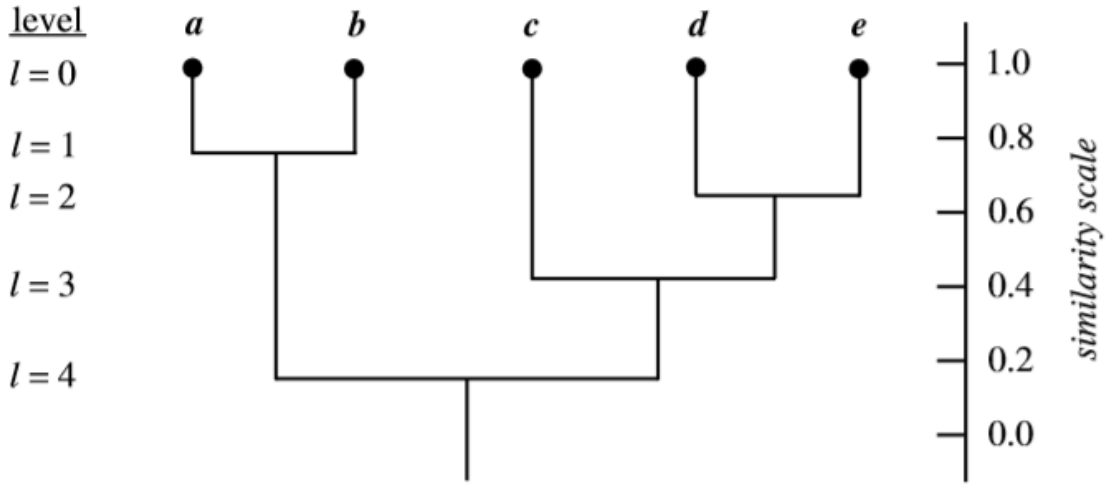


Figure 2: Dendrogram representation for hierarchical clustering of data objects{a,b,c,d,e}

Dendrogram

- A tree structure called a dendrogram is commonly used to represent the process of hierarchical clustering
- It shows how objects are grouped together step by step
- Figure: 2 shows a dendrogram for the five objects presented in Figure:1 , where l =0 shows the five objects as singleton clusters at level 0
- At l =1, objects a and b are grouped together to form the first cluster, and they stay together at all subsequent levels

Dendrogram

- We can also use a vertical axis to show the similarity scale between clusters
- For example, when the similarity of two groups of objects, {a,b} and {c,d,e}, is roughly 0.16, they are merged together to form a single cluster

Measures for distance between clusters

- Four widely used measures for distance between clusters are as follows, where $|p-p'|$ is the distance between two objects or points, p and p' , m_i is the mean for cluster, C_i and n_i is the number of objects in C_i
- Minimum distance: $d_{min}(C_i,C_j)= min_{p\in Ci,\ p'\in Cj} |p-p'|$
- Maximum distance: $d_{max}(C_i,C_j)= max_{p\in Ci,\ p'\in Cj} |p-p'|$
- Mean distance: $d_{mean}(C_i,C_j)= |m_i-m_j|$
- Average distance: $d_{avg}(C_i,C_j) = \frac{1}{n_i n_j} \sum_{p\in C_i} \sum_{p'\in C_j} |p-p'|$

Measures for distance between clusters

- When an algorithm uses the minimum distance, $d_{min}(C_i,C_j)$, to measure the distance between clusters, it is sometimes called a nearest-neighbor clustering algorithm
- Moreover, if the clustering process is terminated when the distance between nearest clusters exceeds an arbitrary threshold, it is called a single-linkage algorithm
- If we view the data points as nodes of a graph, with edges forming a path between the nodes in a cluster, then the merging of two clusters, C_i and C_j, corresponds to adding an edge between the nearest pair of nodes in C_i and C_j

Measures for distance between clusters

- Because edges linking clusters always go between distinct clusters, the resulting graph will generate a tree
- Thus, an agglomerative hierarchical clustering algorithm that uses the minimum distance measure is also called a minimal spanning tree algorithm
- When an algorithm uses the maximum distance, $d_{max}(C_i,C_j)$, to measure the distance between clusters, it is sometimes called a farthest-neighbor clustering algorithm
- If the clustering process is terminated when the maximum distance between nearest clusters exceeds an arbitrary threshold, it is called a complete-linkage algorithm

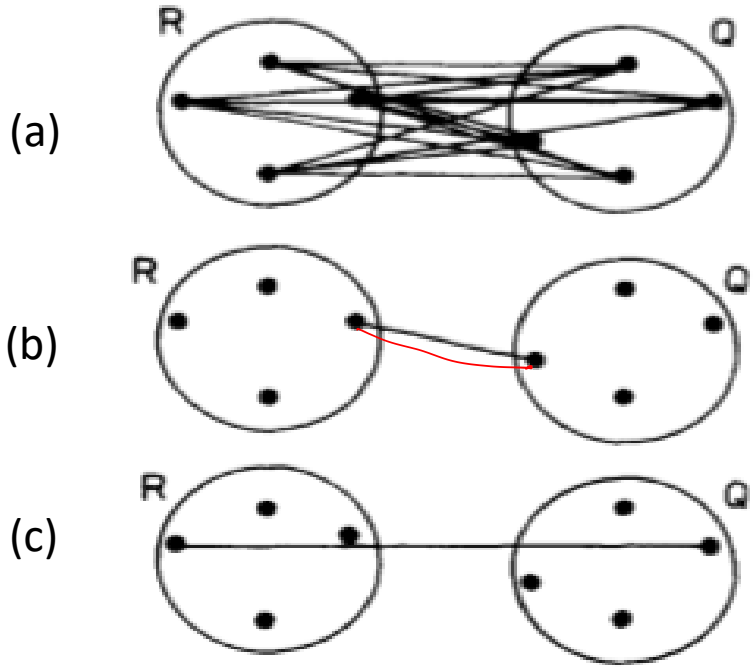
Measures for distance between clusters

- By viewing data points as nodes of a graph, with edges linking nodes, we can think of each cluster as a complete sub graph, that is, with edges connecting all of the nodes in the clusters
- The distance between two clusters is determined by the most distant nodes in the two clusters
- Farthest-neighbor algorithms tend to minimize the increase in diameter of the clusters at each iteration as little as possible
- If the true clusters are rather compact and approximately equal in size, the method will produce high-quality clusters
- Otherwise, the clusters produced can be meaningless

Choice of measurement

- The above minimum and maximum measures represent two extremes in measuring the distance between clusters
- They tend to be overly sensitive to outliers or noisy data
- The use of mean or average distance is a compromise between the minimum and maximum distances and overcomes the outlier sensitivity problem
- Whereas the mean distance is the simplest to compute, the average distance is advantageous in that it can handle categorical as well as numeric data

Illustration



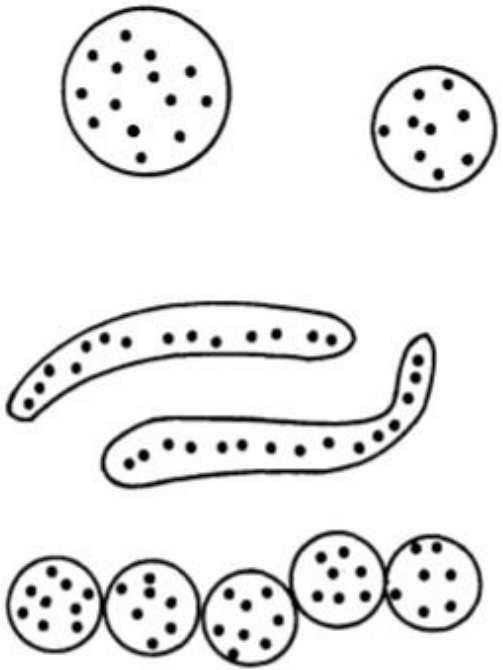
Representation of some definitions of inter-cluster dissimilarity: (a) Group average
(b) Nearest neighbor
(c) Furthest neighbor

Illustration

(a)

(b)

(c)



Some types of clusters:

(a) Ball-shaped
(b) Elongated
(c) Compact but not well separated

Difficulties with hierarchical clustering

- The hierarchical clustering method, though simple, often encounters difficulties regarding the selection of merge or split points
- Such a decision is critical because once a group of objects is merged or split, the process at the next step will operate on the newly generated clusters
- It will neither undo what was done previously nor perform object swapping between clusters

Difficulties with hierarchical clustering

- Thus merge or split decisions, if not well chosen at some step, may lead to low-quality clusters
- Moreover, the method does not scale well, because each decision to merge or split requires the examination and evaluation of a good number of objects or cluster
- For improving the clustering quality of hierarchical methods is to integrate hierarchical clustering with other clustering techniques, resulting in multiple-phase clustering

Partitioning Vs. Hierarchical

Method	General Characteristics
Partitioning methods	– Find mutually exclusive clusters of spherical shape – Distance-based – May use mean or medoid (etc.) to represent cluster center – Effective for small- to medium-size data sets
Hierarchical methods	– Clustering is a hierarchical decomposition (i.e., multiple levels) – Cannot correct erroneous merges or splits – May incorporate other techniques like microclustering or consider object "linkages"

K-means versus hierarchical clustering

K means versus hierarchical clustering

K- means clustering

- Non-hierarchical methods, such as k-means, using a pre-specified number of clusters, the method assigns records to each cluster to find the mutually exclusive cluster of spherical shape based on distance
- In this case, one can use mean or median as a cluster centre to represent each cluster

Hierarchical clustering

- Hierarchical methods can be either agglomerative or divisive
- Agglomerative methods begin with ‘n’ clusters and sequentially merge similar clusters until a single cluster is obtained

K means versus hierarchical clustering

K- means clustering

- These methods are generally less computationally intensive and are therefore preferred with very large datasets

Hierarchical clustering

- Divisive methods work in the opposite direction, starting with one cluster that includes all records
- Hierarchical methods are especially useful when the goal is to arrange the clusters into a natural hierarchy

K means versus hierarchical clustering

K- means clustering

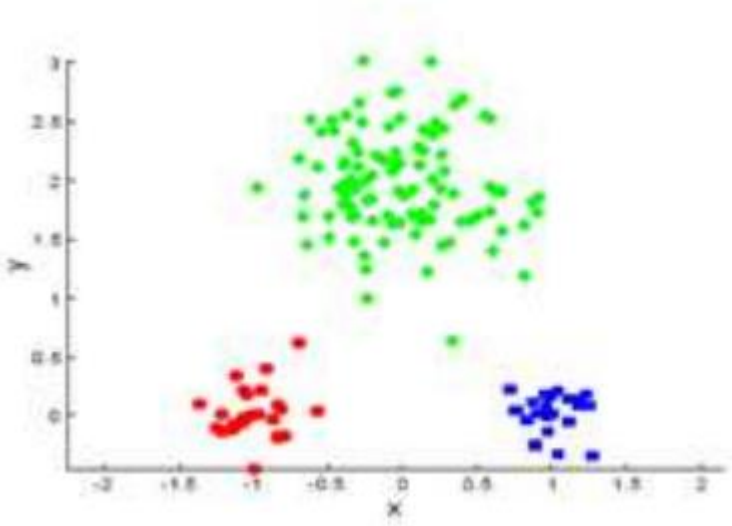
- A partitioning (K- means) clustering a simply a division of the set of data objects into non-overlapping subsets (clusters) such that each data object is in exactly one subset)

Hierarchical clustering

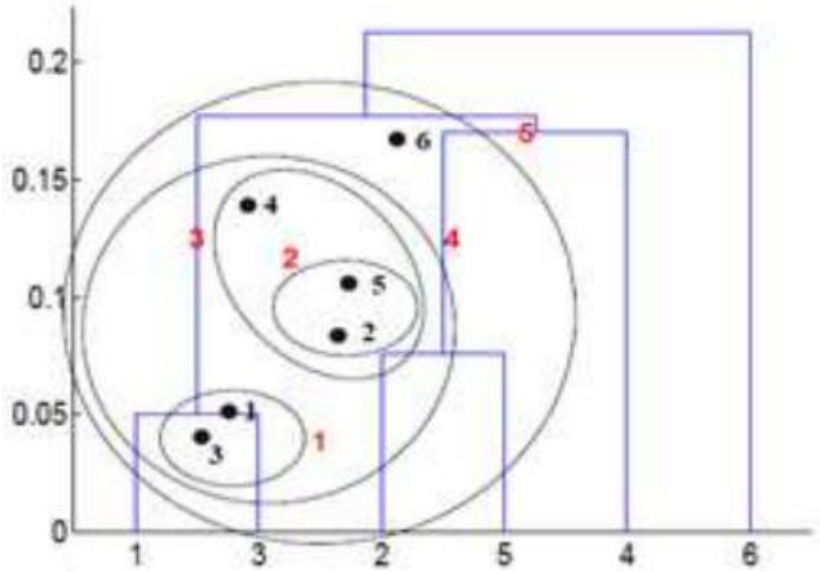
- A hierarchical clustering is a set of nested clusters that are organized as a tree

K means versus hierarchical clustering

Un-nested cluster



Nested cluster



Ashok, A.R., Prabhakar, C.R. and Dyaneshwar, P.A., Comparative Study on Hierarchical and Partitioning Data Mining Methods.

K means versus hierarchical clustering

- Hierarchical clustering does not assume a particular value of 'k', as needed by k-means clustering
- The generated tree may correspond to a meaningful taxonomy
- Only a distance or "proximity" matrix is needed to compute the hierarchical clustering

	a	b	c	d	e	f
a	0	184	222	177	216	231
b	184	0	45	123	128	200
c	222	45	0	129	121	203
d	177	123	129	0	46	83
e	216	128	121	46	0	83
f	231	200	203	83	83	0

Proximity matrix

K means versus hierarchical clustering

K Means clustering

- In K Means clustering, since one start with random choice of clusters, the results produced by running the algorithm multiple times might differ
- K Means is found to work well when the shape of the clusters is hyper spherical (like circle in 2D, sphere in 3D)

Hierarchical clustering

- Results are reproducible in Hierarchical clustering
- Hierarchical clustering don't work as well as, k means when the shape of the clusters is hyper spherical

K means versus hierarchical clustering

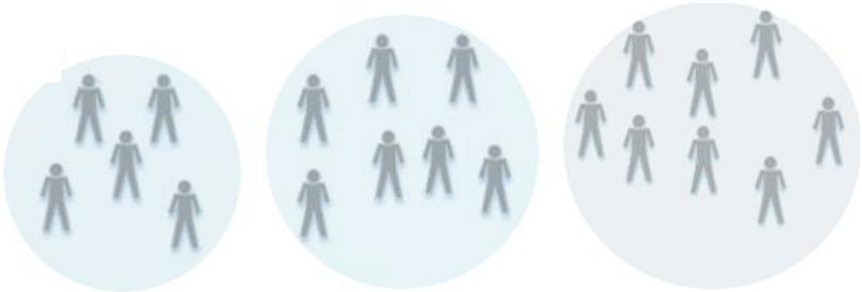
K Means clustering

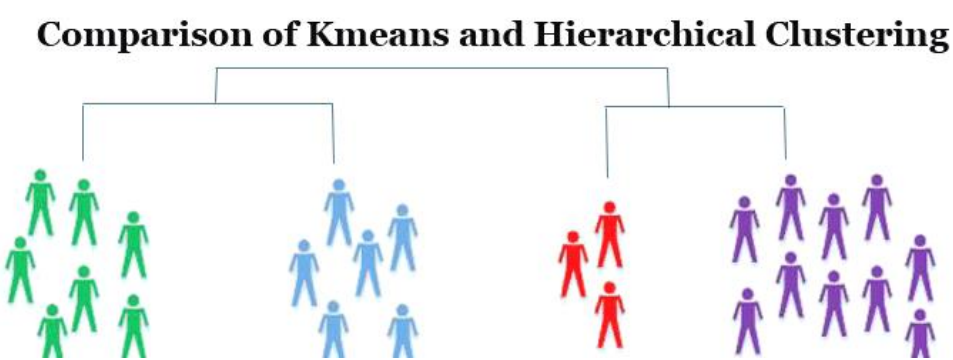
- K Means clustering requires prior knowledge of K i.e. no. of clusters one want to divide your data into

Hierarchical clustering

- In hierarchical clustering one can stop at any number of clusters, one find appropriate by interpreting the dendrogram

K means versus hierarchical clustering





https://stepupanalytics.com/difference-between-k-means-clustering-and-hierarchical-clustering/

Hierarchical clustering

Advantages

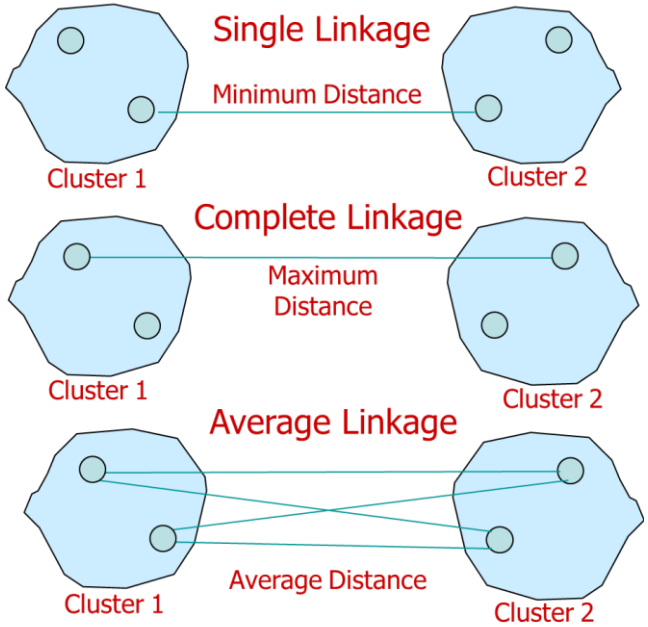
- Ease of handling of any forms of similarity or distance
- Consequently, applicability to any attributes types

Limitations of Hierarchical Clustering

- Hierarchical clustering requires the computation and storage of an n×n distance matrix. For very large datasets, this can be expensive and slow
- The hierarchical algorithm makes only one pass through the data. This means that records that are allocated incorrectly early in the process cannot be reallocated subsequently
- Hierarchical clustering also tends to have low stability. Reordering data or dropping a few records can lead to a different solution

Limitations of Hierarchical Clustering

- With respect to the choice of distance between clusters, single and complete linkage are robust to changes in the distance metric (e.g., Euclidean, statistical distance) as long as the relative ordering is kept.
- In contrast, average linkage is more influenced by the choice of distance metric, and might lead to completely different clusters when the metric is changed
- Hierarchical clustering is sensitive to outlier



Average-linkage clustering

- Compromise between Single and Complete Link
- Strengths
  - Less susceptible to noise and outliers
- Limitations
  - Biased towards globular clusters

Distance between two clusters

- **Ward's distance** between clusters C_i and C_j is the ***difference*** between the ***total within cluster sum of squares for the two clusters separately***, and the ***within cluster sum of squares resulting from merging the two clusters*** in cluster C_{ij}

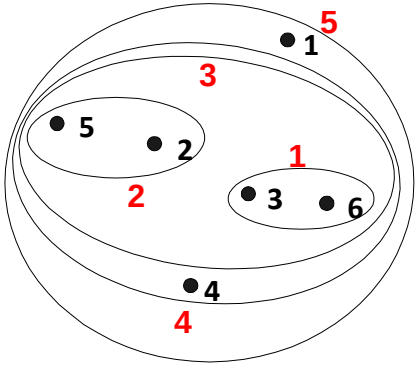
$$D_w(C_i, C_j) = \sum_{x \in C_i} (x - r_i)^2 + \sum_{x \in C_j} (x - r_j)^2 - \sum_{x \in C_{ij}} (x - r_{ij})^2$$

- r_i: centroid of C_i
- r_j: centroid of C_j
- r_{ij}: centroid of C_{ij}

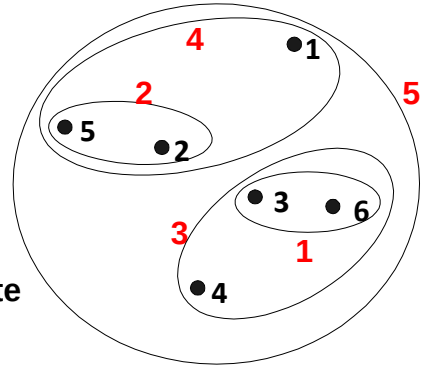
Ward's distance for clusters

- Similar to group average and centroid distance
- Less susceptible to noise and outliers
- Biased towards globular clusters
- Hierarchical analogue of k-means
  - Can be used to initialize k-means

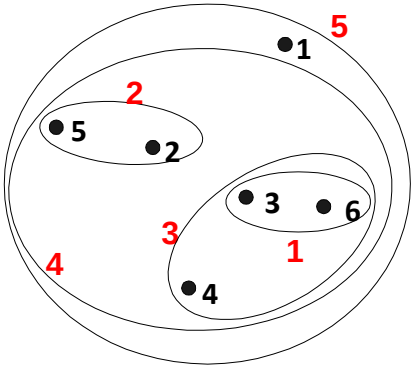
Hierarchical Clustering: Comparison



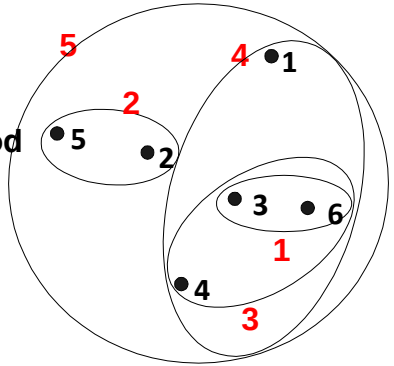
Simple linkage



Complete linkage



Group Average



Ward's Method

K- means clustering

Advantages

- The center of mass can be found efficiently by finding the mean value of each co-ordinate
- This leads to an efficient algorithm to compute the new centroids with a single scan of the data

Disadvantages

- K-means has problems when clusters are of differing sizes, densities, non-globular shapes and when the data contains outliers

Similarity

- Two most popular methods: hierarchical agglomerative clustering and k-means clustering
- In both cases, we need to define two types of distances: distance between two records and distance between two cluster
- In both cases, there is a variety of metrics that can be used

Thank You