Pattern Classification

EET3053

Lecture 09: Clustering

Dr. Kundan Kumar
Associate Professor
Department of ECE

Faculty of Engineering (ITER)
S'O'A Deemed to be University, Bhubaneswar, India-751030

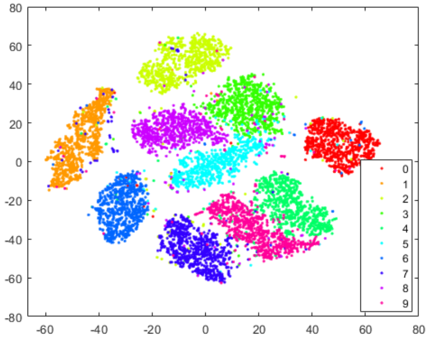
Unsupervised Learning: Clustering

Introduction

- In previous lectures, we have seen that how samples are classified if a training set is available along with class labels to design a classifier. For example, iris data set for classification problem using k-NN.
- In many situations, class labels are not known for the training samples.
- The area of pattern classification which cluster / group the data samples based on some similarity is known as clustering.
- Clustering refers to the process of grouping samples so that the samples are similar within each group. The groups are called *clusters*.
- There are two major issue in clustering:
  - how to measure similarity.
  - the criterion function to be optimized.

Applications

- Marketing: find the group of customers with similar behavior.
- Biology: classification of plants and animals.
- Library: book ordering.
- Medical Imaging: clustering based segmentation.
- Data analysis
- Data Visualization: clustered visualization.



Broad classification

- There are various types of clustering algorithms and they all use different approaches to cluster.
- Broadly classified in three categories:
  - Partitioning algorithm
  - Hierarchical algorithm
  - Graph based algorithm

Hierarchical Clustering

Hierarchical Clustering

- *Hierarchical clustering* refers to a clustering process that organizes the data into large groups, which contain smaller groups, and so on.
- Hierarchical clustering approach can be categorize as
  - Bottom-Up approach (called *Agglomerative* clustering)
  - Top-Down approach (called *Divisive* clustering)
- A hierarchical clustering may be drawn as a *tree* or *dendrogram*.
  - A dendrogram is a diagram that shows the hierarchical relationship between objects.

Agglomerative Clustering

- Consider $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$ are n d-dimensional feature vectors.
- Algorithm:
  1: Begin with n clusters, each consisting of one sample.
  2: Repeat step 3 a total of $n-1$ times
  3: Find the most similar clusters C_i and C_j and merge C_i and C_j into one cluster. If there is a tie, merge the first pair found.

Agglomerative Clustering

- How to find similar clusters?
- Similar clusters can be obtained using these approaches:
  - Single-linkage algorithm

$$D_{SL}(C_i, C_j) = \min_{a \in C_i, b \in C_j} d(a, b)$$

  - Complete-linkage algorithm

$$D_{CL}(C_i, C_j) = \max_{a \in C_i, b \in C_j} d(a, b)$$

  - Average-linkage algorithm

$$D_{AL}(C_i, C_j) = \operatorname*{avg}_{a \in C_i, b \in C_j} d(a, b)$$

Single-linkage Algorithm

- The *Single linkage algorithm* is obtained by defining the distance between two clusters to be the smallest distance between two points such that one point is in each cluster.
- If C_i and C_j are two clusters then distance between cluster is defined as

$$D_{SL}(C_i, C_j) = \min_{a \in C_i, b \in C_j} d(a, b)$$

where $d(a, b)$ – distance between data sample a and b

Example: Single-linkage algorithm

Question 01: Perform a hierarchical clustering of five samples using the single-linkage algorithm and two features x and y.

	x	y
1	4	4
2	8	4
3	15	8
4	24	4
5	24	12

	1	2	3	4	5
1	—	4.0	11.7	20.0	21.5
2	4.0	—	8.1	16.0	17.9
3	11.7	8.1	—	9.8	9.8
4	20.0	16.0	9.8	—	8.0
5	21.5	17.9	9.8	8.0	—

- Combine 1 and 2 in single cluster

$$\{1,2\},\{3\},\{4\},\{5\}$$

Example: Single-linkage algorithm

	{1,2}	3	4	5
{1,2}	—	8.1	16.0	17.9
3	8.1	—	9.8	9.8
4	16.0	9.8	—	8.0
5	17.9	9.8	8.0	—

- Minimum value is 8.0 so merge cluster $\{4\}$ and $\{5\}$.

$$\{1,2\},\{3\},\{4,5\}$$

Example: Single-linkage algorithm

	{1,2}	3	{4,5}
{1,2}	—	8.1	16.0
3	8.1	—	9.8
{4,5}	16.0	9.8	—

- Minimum value is 8.1 so merge cluster $\{1,2\}$ and $\{3\}$.

$$\{1,2,3\},\{4,5\}$$

- In next step will merge the two remaining clusters at a distance of 9.8.

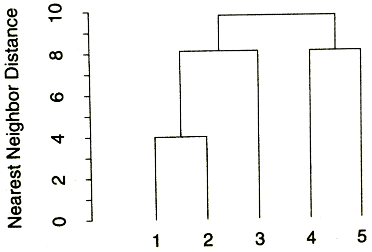


Figure: Dendrogram of Hierarchical clustering using single-linkage algorithm

Complete-linkage algorithm

- The *Complete-linkage algorithm* is obtained by defining the distance between two clusters to be the maximum distance between two points such that one point is in each cluster.
- If C_i and C_j are two clusters then distance between cluster is defined as

$$D_{CL}(C_i, C_j) = \max_{a \in C_i, b \in C_j} d(a, b)$$

where $d(a, b)$ – distance between data sample a and b

Example: Complete-linkage algorithm

Question 02: Perform a hierarchical clustering of five samples using the complete-linkage algorithm and two features x and y.

	x	y
1	4	4
2	8	4
3	15	8
4	24	4
5	24	12

	1	2	3	4	5
1	—	4.0	11.7	20.0	21.5
2	4.0	—	8.1	16.0	17.9
3	11.7	8.1	—	9.8	9.8
4	20.0	16.0	9.8	—	8.0
5	21.5	17.9	9.8	8.0	—

- Combine 1 and 2 in single cluster

$$\{1, 2\}, \{3\}, \{4\}, \{5\}$$

Example: Complete-linkage algorithm

	{1,2}	3	4	5
{1,2}	—	11.7	20.0	21.5
3	11.7	—	9.8	9.8
4	20.0	9.8	—	8.0
5	21.5	9.8	8.0	—

- Minimum value is 8.0 so merge cluster $\{4\}$ and $\{5\}$.

$$\{1,2\},\{3\},\{4,5\}$$

Example: Complete-linkage algorithm

	{1,2}	3	{4,5}
{1,2}	—	11.7	21.5
3	11.7	—	9.8
{4,5}	21.5	9.8	—

- Minimum value is 9.8 so merge cluster $\{4,5\}$ and $\{3\}$.

$$\{1,2\},\{3,4,5\}$$

- In next step will merge the two remaining clusters.

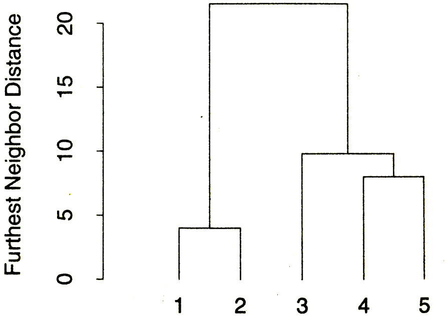


Figure: Dendrogram of Hierarchical clustering using complete-linkage algorithm

Average-linkage algorithm

- The *Average-linkage algorithm* is obtained by taking the average distance of all possible pair of clusters such that one point is in each cluster.
- If C_i and C_j are two clusters then distance between cluster is defined as

$$D_{AL}(C_i, C_j) = \underset{a \in C_i, b \in C_j}{\text{avg}} d(a, b)$$

where $d(a, b)$ – distance between data sample a and b

Example: Average-linkage algorithm

Question 03: Perform a hierarchical clustering of five samples using the average-linkage algorithm and two features x and y.

	x	y
1	4	4
2	8	4
3	15	8
4	24	4
5	24	12

	1	2	3	4	5
1	—	4.0	11.7	20.0	21.5
2	4.0	—	8.1	16.0	17.9
3	11.7	8.1	—	9.8	9.8
4	20.0	16.0	9.8	—	8.0
5	21.5	17.9	9.8	8.0	—

- Combine 1 and 2 in single cluster

$$\{1, 2\}, \{3\}, \{4\}, \{5\}$$

Example: Average-linkage algorithm

	{1,2}	3	4	5
{1,2}	—	9.9	18.0	19.7
3	9.9	—	9.8	9.8
4	18	9.8	—	8.0
5	19.7	9.8	8.0	—

- Minimum value is 8.0 so merge cluster $\{4\}$ and $\{5\}$.

$$\{1,2\},\{3\},\{4,5\}$$

Example: Average-linkage algorithm

	{1,2}	3	{4,5}
{1,2}	—	9.9	18.9
3	9.9	—	9.8
{4,5}	18.9	9.8	—

- Minimum value is 9.8 so merge cluster $\{4,5\}$ and $\{3\}$.

$$\{1,2\},\{3,4,5\}$$

- In next step will merge the two remaining clusters.

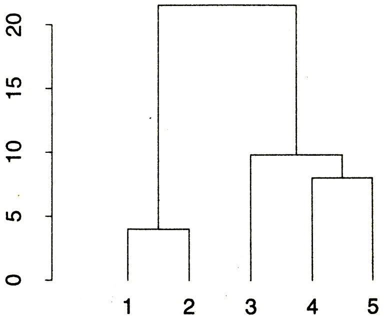


Figure: Dendrogram of Hierarchical clustering using average-linkage algorithm

Ward's Algorithm

- The hierarchical clustering based on variance.
- also called minimum-variance method
- Consider C_j class has k no. of feature vectors $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_k$
- Error within j^{th} cluster

$$E_j = \sum_{i=1}^{k} \|\mathrm{x}_i - \mu\|^2 = k\sigma^2$$

  we need to minimize the variance

- Total error

$$E = \sum_{j=1}^{c} E_j$$

- Computationally expensive because need to check all combination of samples.

Example: Ward's algorithm

Question: Perform Ward's algorithm to cluster the following data samples.

	x	y
1	4	4
2	8	4
3	15	8
4	24	4
5	24	12

Clusters	Squared Error, E
{1,2},{3},{4},{5}	8.0
{1,3},{2},{4},{5}	68.5
{1,4},{2},{3},{5}	200.0
{1,5},{2},{3},{4}	232.0
{2,3},{1},{4},{5}	32.5
{2,4},{1},{3},{5}	128.0
{2,5},{1},{3},{4}	160.0
{3,4},{1},{2},{5}	48.5
{3,5},{1},{2},{4}	48.5
{4,5},{1},{2},{3}	32.0

- Minimum Squared Error is 8.0 so form the cluster $\{1,2\},\{3\},\{4\},\{5\}$

Example: Ward's algorithm

Clusters	Squared Error, E
{1,2,3},{4},{5}	72.7
{1,2,4},{3},{5}	224.0
{1,2,5},{3},{4}	266.7
{1,2},{3,4},{5}	56.5
{1,2},{3,5},{4}	56.5
{1,2},{4,5},{3}	40.0

- Minimum Squared Error is 40.0 so form the cluster $\{1, 2\}, \{3\}, \{4, 5\}$

Example: Ward's algorithm

Clusters	Squared Error, E
{1,2,3},{4,5}	104.7
{1,2,4,5},{3}	380.0
{1,2},{3,4,5}	94.0

- Minimum value is 94.0 so form the cluster $\{1,2\},\{3,4,5\}$
- In next step will merge the two remaining clusters.

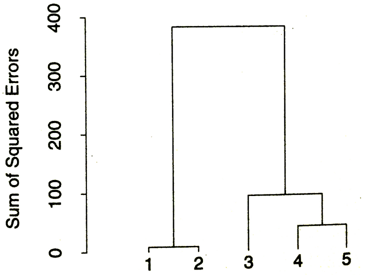
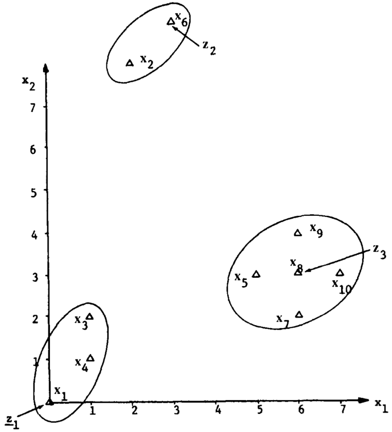


Figure: Dendrogram of Ward's algorithm

Batchelor and Wilkins' Algorithm

- Heuristic procedure for clustering also called *maximum distance algorithm*



Example: Batchelor and Wilkins' Algorithm

Question:

Perform the Batchelor and Wilkins' algorithm to cluster the following data samples.

$A \to (1,1)$
$B \to (1,2)$
$C \to (3,2)$
$D \to (3,3)$
$E \to (6,6)$
$F \to (6,7)$
$G \to (7,6)$
$H \to (7,6)$
$I \to (7,1)$
$J \to (8,2)$
$K \to (9,1)$

Distance matrix

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B		0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C			0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D				0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E					0.0	1.0	1.0	1.4	5.1	4.5	5.8
F						0.0	1.4	1.0	6.1	5.4	6.7
G							0.0	1.0	5.0	4.1	5.4
H								0.0	6.0	5.1	6.3
I									0.0	1.4	2.0
J										0.0	1.4
K											0.0

Batchelor and Wilkins' Algorithm

Solution

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B		0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C			0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D				0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E					0.0	1.0	1.0	1.4	5.1	4.5	5.8
F						0.0	1.4	1.0	6.1	5.4	6.7
G							0.0	1.0	5.0	4.1	5.4
H								0.0	6.0	5.1	6.3
I									0.0	1.4	2.0
J										0.0	1.4
K											0.0

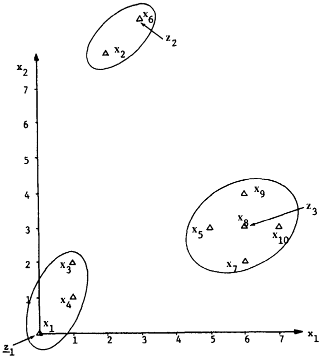
Batchelor and Wilkins' Algorithm

Solution

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B		0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C			0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D				0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E					0.0	1.0	1.0	1.4	5.1	4.5	5.8
F						0.0	1.4	1.0	6.1	5.4	6.7
G							0.0	1.0	5.0	4.1	5.4
H								0.0	6.0	5.1	6.3
I									0.0	1.4	2.0
J										0.0	1.4
K											0.0

Batchelor and Wilkins' Algorithm

Step 1: Arbitrarily, let x_1 be the first cluster center, designated by z_1.
Step 2: Determine the pattern sample farthest from x_1, which is x_6. Call it cluster center z_2.
Step 3: Compute the distance from each remaining pattern sample to z_1 and z_2
Step 4: Save the minimum distance for each pair of these computations.
Step 5: Select the maximum of these minimum distances.
Step 6: If the distance is appreciably greater than a fraction of the distance $d(z_1, z_2)$, call the corresponding sample cluster center z_3. Otherwise, the algorithm is terminated.
Step 7: If this distance from each of the three established cluster centers to the remaining samples and save the minimum of every group of three distances. Again select the maximum of these minimum distances. If this distance is an appreciable fraction of the "typical" previous maximum distances, the corresponding sample becomes cluster center z_4. Otherwise, the algorithm is terminated.
Step 8: Repeat until the new maximum distance at a particular step fails to satisfy the condition for the creation of a new cluster center.
Step 9: Assign each sample to its nearest cluster center.



Batchelor and Wilkins' Algorithm

	C_1 (A)	C_2 (H)
B	1.0	7.8
C	2.2	6.4
D	2.8	5.6
E	7.07	1.41
F	7.81	1.0
G	7.81	1.0
I	6.0	6.0
J	7.07	5.09
K (C_3)	8.0	6.32

6.32 → maximum among all minimum.

Batchelor and Wilkins' Algorithm

	C_1 A	C_2 H	C_3 K
B	1.0	7.8	8.06
C	2.2	6.4	6.08
C_4 D	2.8	5.6	6.32
E	7.07	1.41	5.83
F	7.81	1.0	6.70
G	7.81	1.0	5.38
I	6.0	6.0	2.0
J	7.07	5.09	1.41

Final clusters are $\{A, B, C, D\}, \{E, F, G, H\}, \{I, J, K\}$

Graph Based Clustering

Graph Based Clustering

- Graph can be represented as
$$G = \langle V, E \rangle$$
where V is set of nodes or vertices and E is set of Edges.
- There are many ways to represent a graph and one of them is adjacency matrix also called similarity matrix.
- For n number of nodes, similarity matrix will be of size $n \times n$.
- Similarity matrix
  - The elements of similarity matrix will be either 0 or 1.
  - Similarity matrix: $S(i, j) = 1$, if V_i and V_j are connected in some sense.
- Types of graph: Complete graph, Tree, Spanning tree, Weighted graph, etc.
- Graph based algorithms:
  - Similarity Matrix based clustering
  - Minimal Spanning Tree based clustering

Similarity Matrix based clustering

Question:

Perform the Similarity Matrix based clustering technique to cluster the following data samples.

$A \rightarrow (1,1)$
$B \rightarrow (1,2)$
$C \rightarrow (3,2)$
$D \rightarrow (3,3)$
$E \rightarrow (6,6)$
$F \rightarrow (6,7)$
$G \rightarrow (7,6)$
$H \rightarrow (7,7)$
$I \rightarrow (7,1)$
$J \rightarrow (8,2)$
$K \rightarrow (9,1)$

Distance matrix

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B		0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C			0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D				0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E					0.0	1.0	1.0	1.4	5.1	4.5	5.8
F						0.0	1.4	1.0	6.1	5.4	6.7
G							0.0	1.0	5.0	4.1	5.4
H								0.0	6.0	5.1	6.3
I									0.0	1.4	2.0
J										0.0	1.4
K											0.0

Similarity Matrix based clustering

Question:

Perform the Similarity Matrix based clustering technique to cluster the following data samples.

$A \rightarrow (1,1)$
$B \rightarrow (1,2)$
$C \rightarrow (3,2)$
$D \rightarrow (3,3)$
$E \rightarrow (6,6)$
$F \rightarrow (6,7)$
$G \rightarrow (7,6)$
$H \rightarrow (7,7)$
$I \rightarrow (7,1)$
$J \rightarrow (8,2)$
$K \rightarrow (9,1)$

Distance matrix

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B	1.0	0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C	2.2	2.0	0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D	2.8	2.2	1.0	0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E	7.0	6.4	5.0	4.2	0.0	1.0	1.0	1.4	5.1	4.5	5.8
F	7.8	7.0	4.9	5.0	1.0	0.0	1.4	1.0	6.1	5.4	6.7
G	7.8	7.0	5.6	5.0	1.0	1.4	0.0	1.0	5.0	4.1	5.4
H	8.5	7.8	6.4	5.6	1.4	1.0	1.0	0.0	6.0	5.1	6.3
I	6.0	6.0	4.1	4.5	5.1	6.1	5.0	6.0	0.0	1.4	2.0
J	7.0	7.0	5.0	5.1	4.5	5.4	4.1	5.1	1.4	0.0	1.4
K	8.0	8.1	6.1	6.3	5.8	6.7	5.4	6.3	2.0	1.4	0.0

Similarity Matrix based clustering

	A	B	C	D	E	F	G	H	I	J	K
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5	6.0	7.0	8.0
B	1.0	0.0	2.0	2.2	6.4	7.0	7.0	7.8	6.0	7.0	8.1
C	2.2	2.0	0.0	1.0	5.0	4.9	5.6	6.4	4.1	5.0	6.1
D	2.8	2.2	1.0	0.0	4.2	5.0	5.0	5.6	4.5	5.1	6.3
E	7.0	6.4	5.0	4.2	0.0	1.0	1.0	1.4	5.1	4.5	5.8
F	7.8	7.0	4.9	5.0	1.0	0.0	1.4	1.0	6.1	5.4	6.7
G	7.8	7.0	5.6	5.0	1.0	1.4	0.0	1.0	5.0	4.1	5.4
H	8.5	7.8	6.4	5.6	1.4	1.0	1.0	0.0	6.0	5.1	6.3
I	6.0	6.0	4.1	4.5	5.1	6.1	5.0	6.0	0.0	1.4	2.0
J	7.0	7.0	5.0	5.1	4.5	5.4	4.1	5.1	1.4	0.0	1.4
K	8.0	8.1	6.1	6.3	5.8	6.7	5.4	6.3	2.0	1.4	0.0

	A	B	C	D	E	F	G	H	I	J	K
A											
B											
C											
D											
E											
F											
G											
H											
I											
J											
K											

Minimal Spanning Tree based clustering

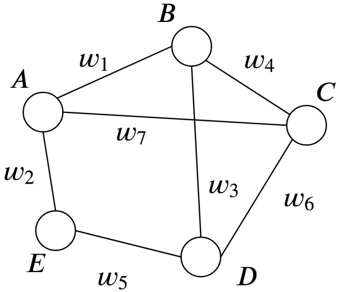
- Spanning tree is the tree representation of graph which contains all nodes present in the graph, i.e., subset of the complete connected weighted graph.
- Difference between tree and graph $\rightarrow$ Graph can have cycle but tree cannot have cycle.
- Weighted Graph have weighted edge which can be represented as

$$G = \langle V, E, W \rangle$$

  $W \rightarrow$ weight or cost
- Weight is the distance (various distance measure) between two feature points.
- The minimal spanning tree is a spanning tree having minimum cost (sum of the weights of edges).

Minimal Spanning Tree based clustering

- Many ways to represent a graph
  - Adjacency matrix
  - Edge list



$$A - B \to w_1$$
$$B - C \to w_4$$
$$A - E \to w_2$$
$$E - D \to w_5$$
$$D - C \to w_6$$
$$A - C \to w_7$$
$$B - D \to w_3$$

Minimal Spanning Tree based clustering

Question:

Perform the Minimal Spanning Tree based clustering technique to cluster the following data samples.

$A \rightarrow (1,1)$
$B \rightarrow (1,2)$
$C \rightarrow (3,2)$
$D \rightarrow (3,3)$
$E \rightarrow (6,6)$
$F \rightarrow (6,7)$
$G \rightarrow (7,6)$
$H \rightarrow (7,7)$

Distance matrix

	A	B	C	D	E	F	G	H
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5
B	1.0	0.0	2.0	2.2	6.4	7.0	7.0	7.8
C	2.2	2.0	0.0	1.0	5.0	4.9	5.6	6.4
D	2.8	2.2	1.0	0.0	4.2	5.0	5.0	5.6
E	7.0	6.4	5.0	4.2	0.0	1.0	1.0	1.4
F	7.8	7.0	4.9	5.0	1.0	0.0	1.4	1.0
G	7.8	7.0	5.6	5.0	1.0	1.4	0.0	1.0
H	8.5	7.8	6.4	5.6	1.4	1.0	1.0	0.0

Ordered Edge List

	A	B	C	D	E	F	G	H
A	0.0	1.0	2.2	2.8	7.0	7.8	7.8	8.5
B	1.0	0.0	2.0	2.2	6.4	7.0	7.0	7.8
C	2.2	2.0	0.0	1.0	5.0	4.9	5.6	6.4
D	2.8	2.2	1.0	0.0	4.2	5.0	5.0	5.6
E	7.0	6.4	5.0	4.2	0.0	1.0	1.0	1.4
F	7.8	7.0	4.9	5.0	1.0	0.0	1.4	1.0
G	7.8	7.0	5.6	5.0	1.0	1.4	0.0	1.0
H	8.5	7.8	6.4	5.6	1.4	1.0	1.0	0.0

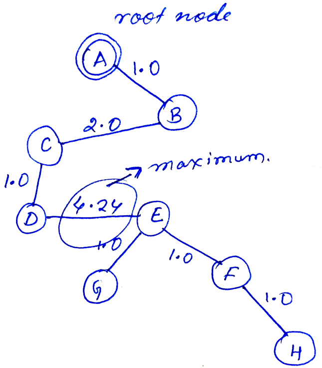
Ordered Edge List

A - B	1.0
C - D	1.0
E - F	1.0
E - G	1.0
F - H	1.0
G - H	1.0
E - H	1.41
F - G	1.41
B - C	2.0
A - C	2.23
B - D	2.23
A - D	2.82
D - E	4.24
C - F	4.89
C - E	5.0
D - F	5.0
D - G	5.0
C - G	5.65
D - H	5.65
B - E	6.40
C - H	6.40
A - E	7.07
B - F	7.07
B - G	7.07
A - F	7.81
A - G	7.81
B - H	7.81
A - H	8.48

- From this ordered edge list find the minimal spanning tree.

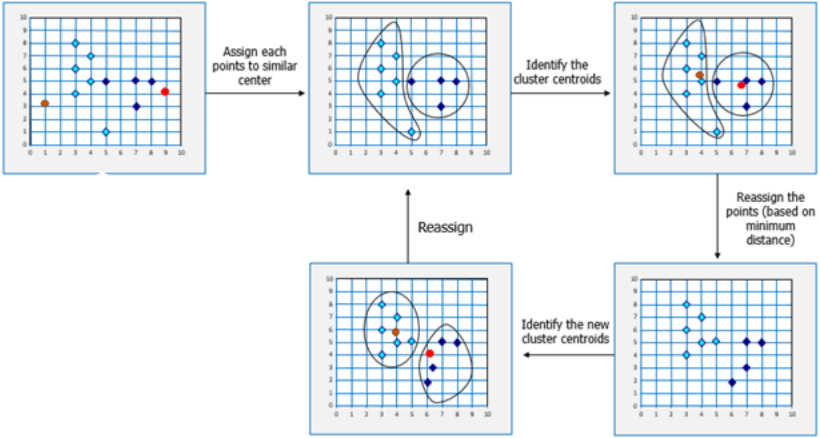
Minimal Spanning Tree based clustering

- There exist only one path between each pair of nodes in the tree.
- Minimal spanning tree is not unique.
- If root node is same than minimal spanning tree will be unique.



Partitioning Clustering

K-means clustering



[1]Source: https://www.edureka.co/blog/k-means-clustering/

K-mean clustering

Let $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots, \mathrm{x}_n$ be the set of data samples.

- Algorithm:
  1. Randomly select 'K' cluster centers.
  2. Calculate the distance of each data samples from each cluster centers and assign the data samples to cluster center whose distance from the cluster center is minimum of all the cluster centers.
  4. Recalculate the new cluster center using:

$$\mu_i = \frac{1}{n_i} \sum_{i=1}^{n_i} \mathrm{x}_i$$

     where, 'n_i' represents the number of data samples in ith cluster.
  5. If the stopping criteria satisfied then stop, otherwise repeat from step 2.

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik} ||\mathrm{x}_i - \mu_k||^2$$

where $w_{ik} = 1$ for data point x_i if it belongs to cluster k; otherwise, $w_{ik} = 0$.

References

[1] Hart, P. E., Stork, D. G., & Duda, R. O. (2000). Pattern classification. Hoboken: Wiley.

[2] Gose, E. (1997). Pattern recognition and image analysis.

Thank you!