Pattern Recognition: Introduction

Shahadat Hoshen
Lecturer
Dept. of CSE; NUBTK

Introduction

- A shop manager of electronic goods may use traditional data analysis tools to investigate the sales of air conditioners in different months of a year.
- He/she can also analyze the sales relationship between computers and printers.
- Precisely, a domain expert needs to provide some sort of assumptions to formulate a query in traditional data analysis.

Introduction

- However, with the increase of volume and dimensions of data sets, manual assumptions become more and more complicated and thus traditional approaches become unsuitable in most domains.

What is pattern recognition?

- Pattern recognition is a data analysis method that uses machine learning algorithms to automatically recognize patterns and regularities in data.

- This data can be anything from text and images to sounds or other definable qualities.
- Pattern recognition systems can recognize familiar patterns quickly and accurately.
- They can also recognize and classify unfamiliar objects, recognize shapes and objects from different angles, and identify patterns and objects even if they're partially obscured.

Knowledge Discovery from Data (KDD)

KDD is a process of discovering useful patterns and knowledge from large datasets. Steps of KDD:

- Data Selection
- Data Cleaning
- Data Transformation
- Data Integration
- Data Mining
- Pattern Evaluation
- Knowledge Representation

Application of PR

- Computer Vision
- Business
- Speech Recognition
- Natural Language Processing
- Medical Diagnosis
- Biometrics
- Robotics

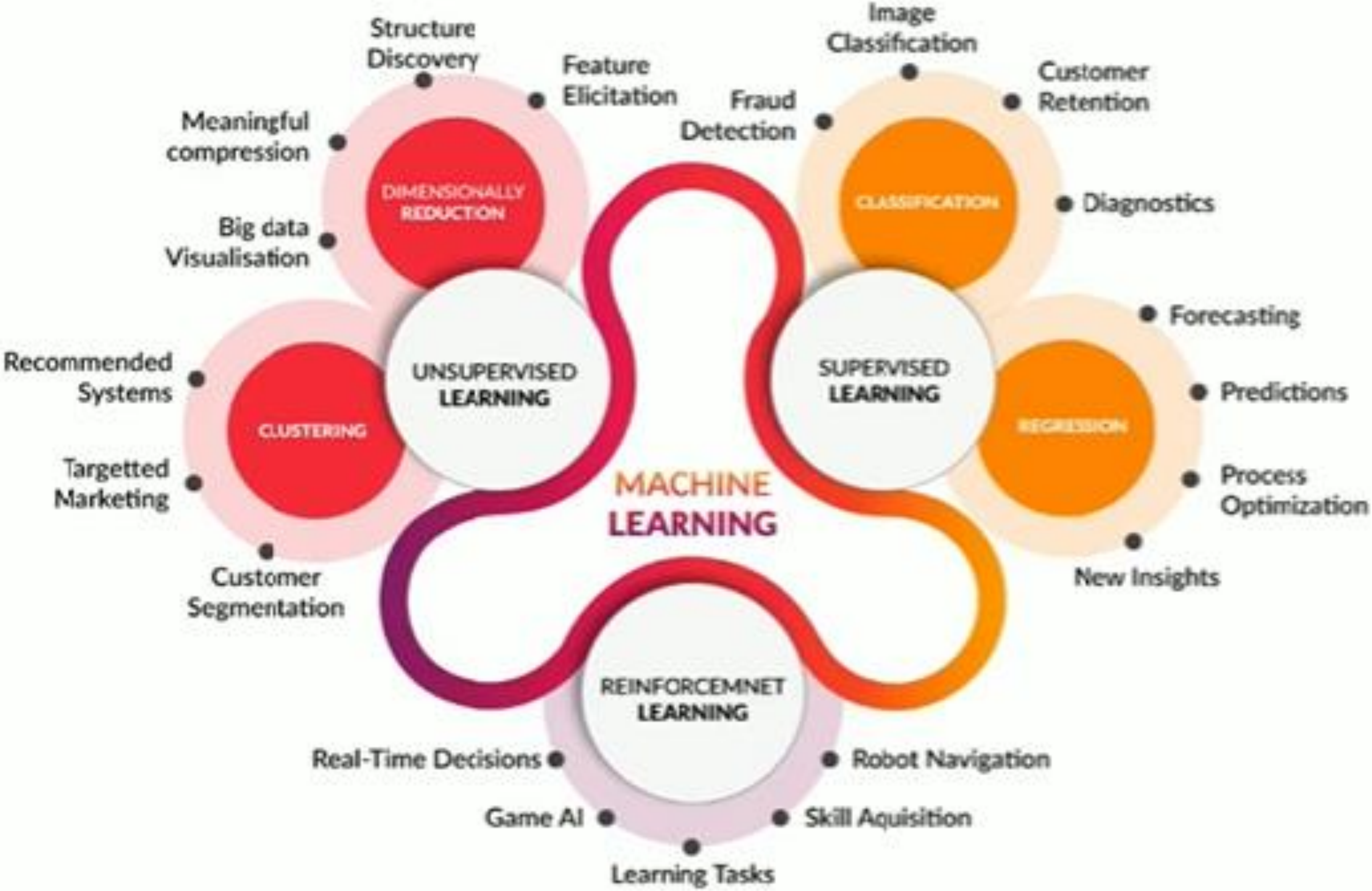
Machine Learning

- Machine learning (ML) is defined as a discipline of artificial intelligence (AI) that provides machines the ability to automatically learn from data and past experiences to identify patterns and make predictions with minimal human intervention.

Self-Study: Application areas of machine learning in modern life

Types ML

- Unsupervised Learning
  - models are trained using unlabeled datasets and are allowed to act on that data without any supervision.
  - types - Clustering and Association
- Supervised Learning
  - train the machines using the “labeled" dataset, and based on the training, the machine predicts the output on new, unseen data.
  - Types- classification, and regression
- Reinforcement Learning
  - works on a feedback-based process, in which an AI agent (A software component) automatically explores its surrounding by hitting & a trail, taking action, learning from experiences, and improving its performance.


MACHINE
LEARNING
UNSUPERVISED
LEARNING
SUPERVISED
LEARNING
REINFORCEMNET
LEARNING
DIMENSIONALLY
REDUCTION
CLUSTERING
CLASSIFICATION
REGRESSION
Structure
Discovery
Feature
Elicitation
Meaningful
compression
Big data
Visualisation
Recommended
Systems
Targetted
Marketing
Customer
Segmentation
Image
Classification
Customer
Retention
Fraud
Detection
Diagnostics
Forecasting
Predictions
Process
Optimization
New Insights
Real-Time Decisions
Game AI
Learning Tasks
Skill Aquisition
Robot Navigation

Association Rules Learning

- Association rules are generally used for discovering interesting relationships in a data set.
- They are often expressed in a rule form {X} ⇒ {Y} to express reasoning that;
  - if X is satisfied then Y is also satisfied.
- Market-Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.
- In a real-world scenario, the following rule can be discovered from a transactional data set of a retail shop. {Milk} ⇒ {Bread}

Support

- Support of an association rule is measured by its frequency in the transactions of a data set.

$$Support(\{X\} \to \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

For example, {Milk} ⇒ {Bread} has a support of 60/100 meaning that 60% of all transactions contain the association rule.

- Thus, a low-support rule is likely to be uninteresting from a business perspective as it may occur simply by chance.

Confidence

- Many customers who buy Milk may buy other goods instead of Bread. Thus, the reliability (confidence) of an association rule needs to be measured.

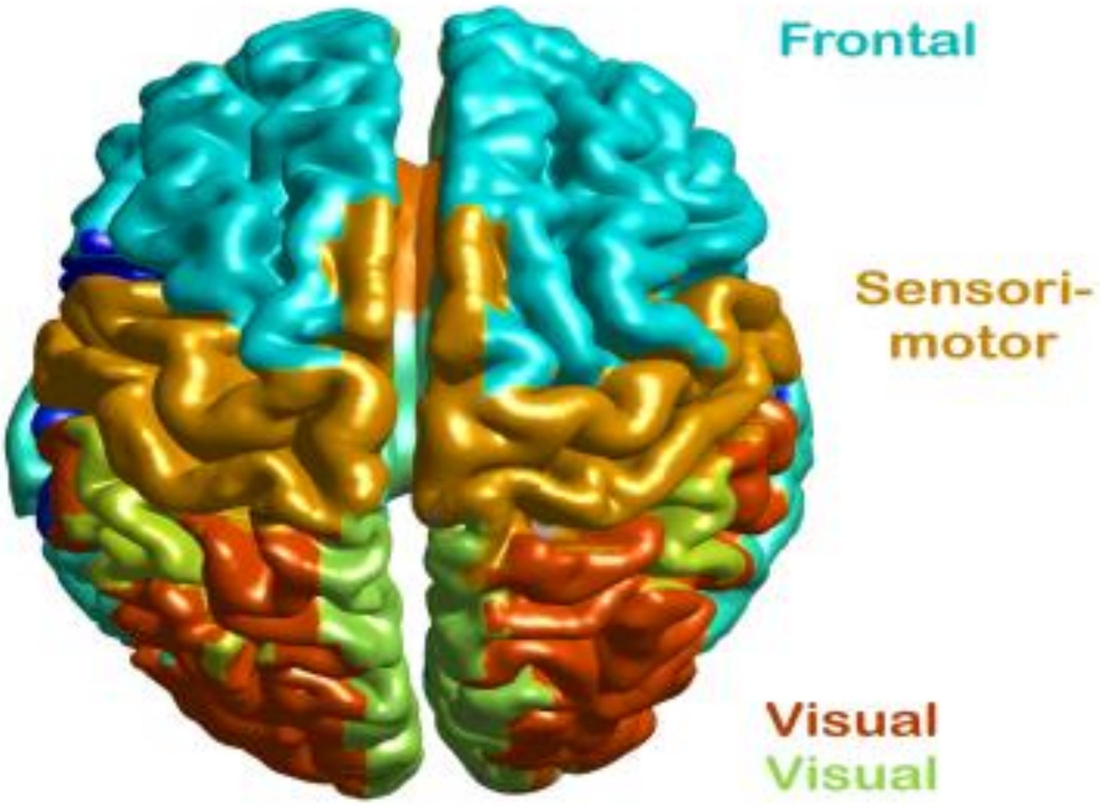
$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Transactions\ containing\ X}$$

- For example, {Milk} ⇒ {Bread} has the confidence of 90/100 meaning that when Milk was purchased, in 90% of cases bread was also purchased.
- Thus, an association rule {Milk} ⇒ {Bread} [Support: 60%; Confidence: 90%] indicates that there exists a strong relationship between the sales of Milk and Bread.

Clustering

- Clustering divides a data set into a set of clusters (groups) where instances of the same cluster have high similarity while instances of different clusters have high dissimilarity.
- It is an unsupervised learning technique, meaning that it does not require labeled data to learn from.
- It is useful in many applications, including market segmentation, customer profiling, image and text analysis, and anomaly detection.

Example: in a real-world scenario, clustering can identify different brain areas (i.e. frontal, sensorimotor, and visual areas) with similar spectral profiles.



Distance functions

"Similarity" can be expressed by different types of distance functions.

Euclidean Distance: Ordinary distance between two points that one would measure with a ruler. It is the straight-line distance between two points. $\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

Manhattan (City Block): Distance between two points is the absolute difference between the points. Absolute value distance gives more robust results. whereas Euclidean was influenced by unusual values.

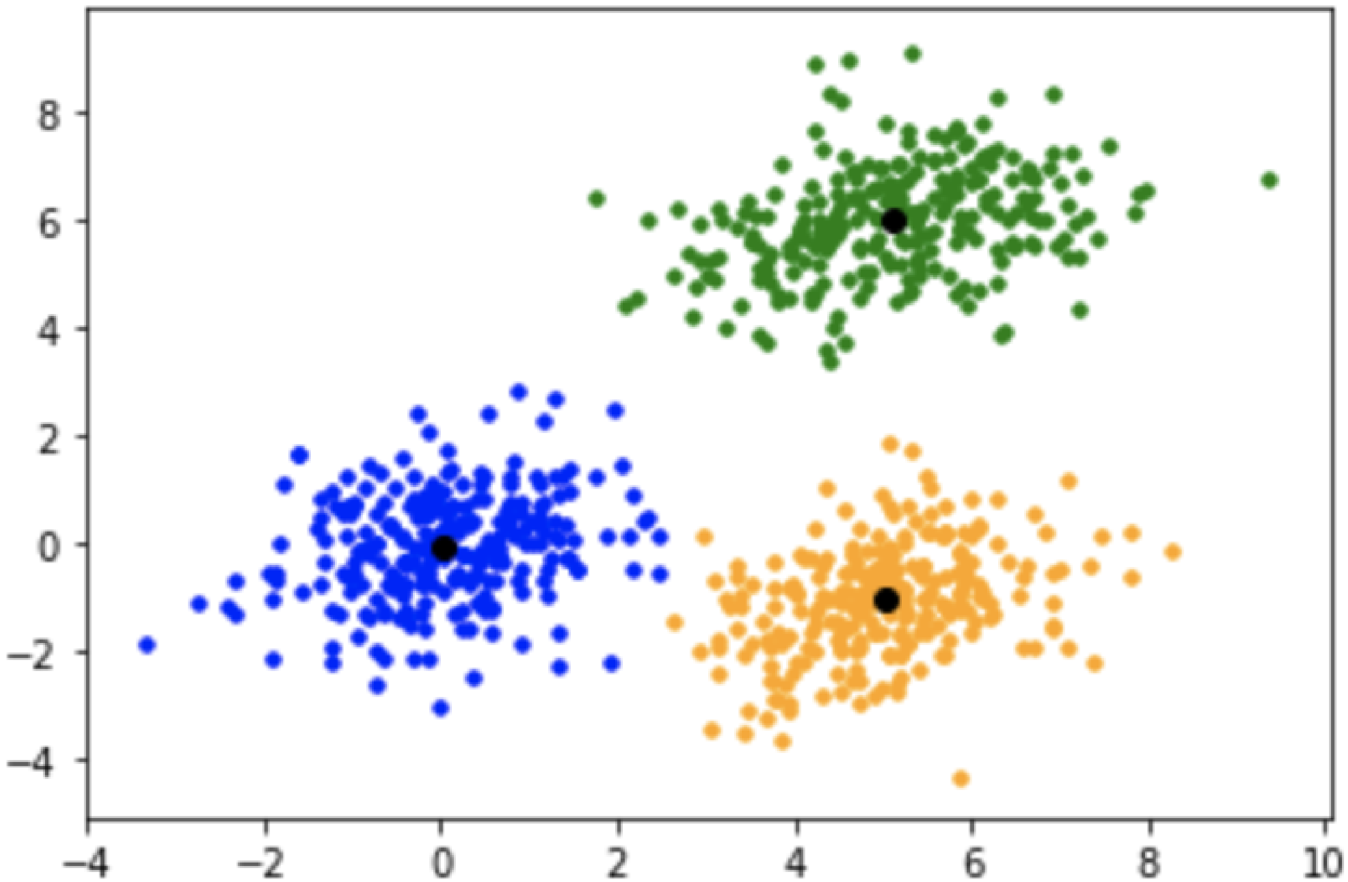
- $\sum_{\mathrm{i}=1}^{\mathrm{n}} |(\mathrm{x_i} - \mathrm{y_i})^2|$

Minkowski Distance: Generalization of both Euclidean and Manhattan distance metric $\sqrt[m]{\sum_{i=1}^{n}(x_i - y_i)^m}$

K-Means Clustering

- K-Means Clustering is an Unsupervised iterative Learning algorithm, which groups the unlabeled dataset into K different clusters in such a way that each dataset belongs to only one group that has similar properties.
- It is a centroid-based algorithm, where each cluster is associated with a centroid.
- The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.
- It can be used in Image Segmentation, Customer Segmentation, Fraud Detection, Bioinformatics, Text Clustering, Recommendation Systems, Environmental Monitoring.

K-Means



K-Means algorithms

1. Choose the number of clusters(K) and obtain the data points
2. Randomly initialize the centroids c_1, c_2, ... C_k
3. for each data point x_i:
   - find the nearest centroid(c_1, c_2 .. c_k)
   - assign the point to that cluster
4. for each cluster j = 1..k
   - new centroid = mean of all points assigned to that cluster
5. Repeat steps 3 and 4 until convergence or until the end of a fixed number of iterations
6. End

Self-study: limitation of K-means algorithm and how can solve?

K-Means: Step-By-Step Example

- Consider the following data set consisting of the scores of two variables on each of seven individuals.
- This data set is to be grouped into two clusters.

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

K-Means Example(cont.)

As a first step in finding a sensible initial partition, let the A & B values of the two individuals furthest apart (using the Euclidean distance measure), define the initial cluster means, giving:

	Individual	Mean Vector (centroid)
Group 1	1	(1.0, 1.0)
Group 2	4	(5.0, 7.0)

Now the initial partition has changed, and the two clusters at this stage having the following characteristics:

	Individual	Mean Vector (centroid)
Cluster 1	1, 2, 3	(1.8, 2.3)
Cluster 2	4, 5, 6, 7	(4.1, 5.4)

K-Means Example(cont.)

Only individual 3 is nearer to the mean of the opposite cluster (Cluster 2) than its own (Cluster 1). Thus, individual 3 is relocated to Cluster 2 resulting in the new partition

	Individual	Mean Vector (centroid)
Cluster 1	1, 2	(1.3, 1.5)
Cluster 2	3, 4, 5, 6, 7	(3.9, 5.1)

K-Means

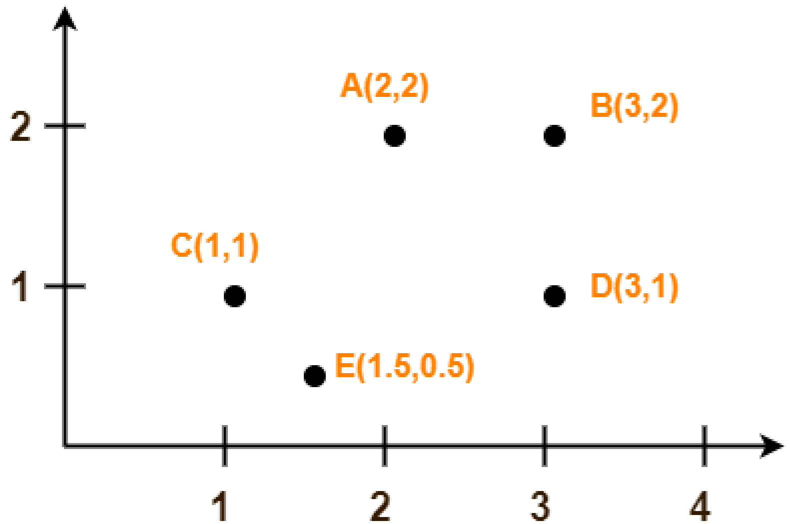
Problem-01:

- Cluster the following eight points (with (x, y) representing locations) into three clusters:
- A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)
- Initial cluster centers are A1(2, 10), A4(5, 8), and A7(1, 2).
- Use K-Means Algorithm to find the three cluster centers after the second iteration.

K-Means

Problem-02:

- Use K-Means Algorithm to create two clusters-



- Assume A(2, 2) and C(1, 1) are centers of the two clusters.

Hierarchical clustering

- It is a type of unsupervised machine learning algorithm that groups similar objects into clusters based on their similarity or dissimilarity.
- The algorithm builds a hierarchy of clusters by successively merging or dividing them until a stopping criterion is met.
- To measure similarity or dissimilarity between two clusters or data points, hierarchical clustering algorithms use a distance metric such as Euclidean distance
- We develop the hierarchy of clusters in the form of a tree, and this tree-shaped structure is known as the **dendrogram** that shows the hierarchical relationships between clusters.

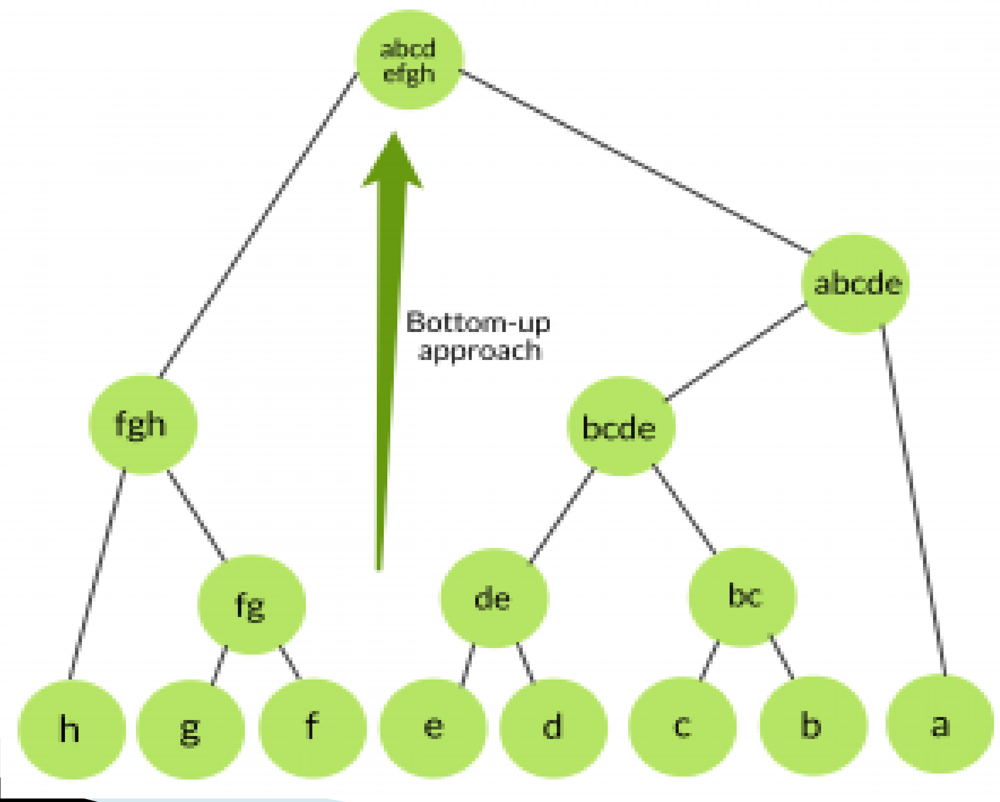
Hierarchical clustering(cont.)

There are two types of hierarchical clustering:

1. **Agglomerative (**(bottom-up)
2. **Divisive(**(top-down**)**

Agglomerative Clustering

- Agglomerative clustering is the most common type of hierarchical clustering that starts with each data point as a separate cluster and iteratively merges clusters until a stopping criterion is met.



Steps to agglomerative clustering

- Initialization
- Compute pairwise distances
- Find the closest pair of clusters
- Merge the closest pair of clusters
- Update the distance matrix: Recalculate the distance between the new cluster and each of the remaining clusters using a linkage method such as single linkage, complete linkage, or average linkage.
- Repeat steps 3-5
- Output the dendrogram

Example-1

For the given dataset find the clusters using a single link technique. Use Euclidean distance and draw the Dendrogram.

Sample No.	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

$$d(p_1, p_2) = \sqrt{(0.22-0.40)^2 + (0.38-0.53)^2}$$

$$= 0.23$$

$$d(p_1, p_3) = \sqrt{(0.35-0.40)^2 + (0.32-0.53)^2}$$

$$= 0.22$$

$$d(p_2, p_3) = \sqrt{(0.35-0.22)^2 + (0.32-0.38)^2}$$

$$= 0.14$$

$$\begin{array}{cc}
 & \begin{array}{cccccc} P1 & P2 & P3 & P4 & P5 & P6 \end{array} \\
\begin{array}{c} P1 \\ P2 \\ P3 \\ P4 \\ P5 \\ P6 \end{array} &
\begin{pmatrix}
0 & & & & & \\
0.23 & 0 & & & & \\
0.22 & 0.14 & 0 & & & \\
0.37 & 0.19 & 0.13 & 0 & & \\
0.34 & 0.14 & 0.28 & 0.23 & 0 & \\
0.24 & 0.24 & 0.10 & 0.22 & 0.39 & 0
\end{pmatrix}
\end{array}$$

$$\begin{pmatrix} & P1 & P2 & P3, P6 & P4 & P5 \\ P1 & 0 & & & & \\ P2 & 0.23 & 0 & & & \\ P3, P6 & 0.22 & 0.14 & 0 & & \\ P4 & 0.37 & 0.19 & 0.13 & 0 & \\ P5 & 0.34 & 0.14 & 0.28 & 0.23 & 0 \end{pmatrix}$$

(P3, P6)

$$\begin{pmatrix} & P1 & P2 & P3, P6, P4 & P5 \\ P1 & 0 & & & \\ P2 & 0.23 & 0 & & \\ P3, P6, P4 & 0.22 & 0.14 & 0 & \\ P5 & 0.34 & 0.14 & 0.28 & 0 \end{pmatrix}$$

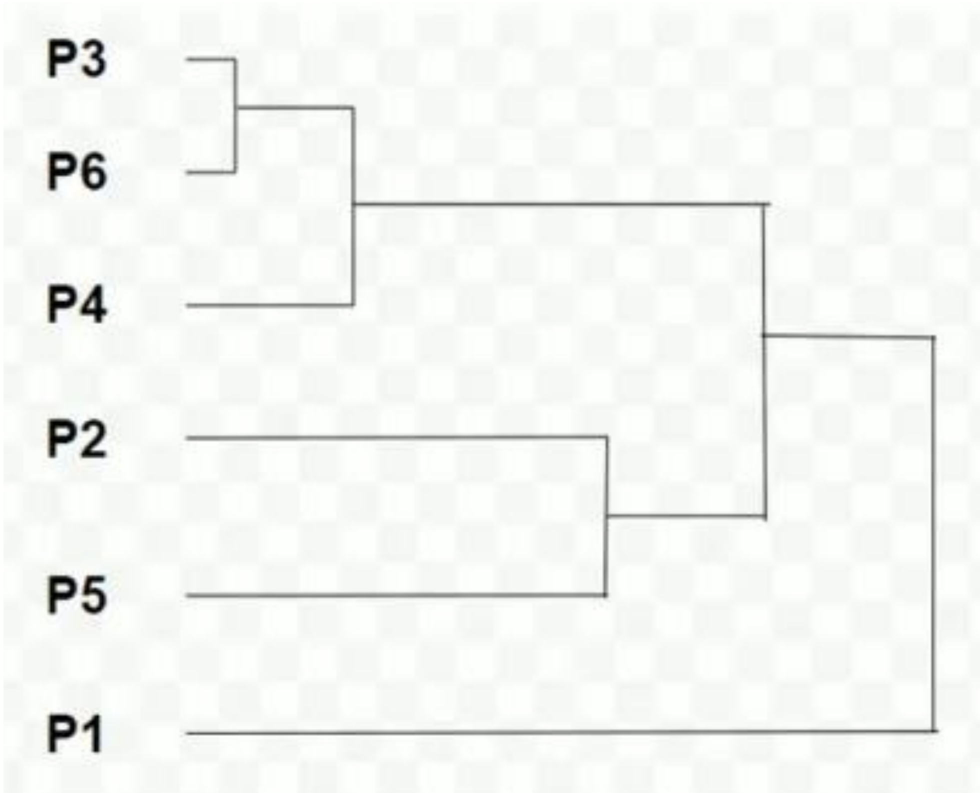
{(P3, P6), P4}

$$\begin{pmatrix} & P1 & P2, P5 & P3, P6, P4 \\ P1 & 0 & & \\ P2, P5 & 0.23 & 0 & \\ P3, P6, P4 & 0.22 & 0.14 & 0 \end{pmatrix}$$

{(P3, P6), P4} and **(P2, P5)**

$$\begin{pmatrix} & P1 & P2, P5, P3, P6, P4 \\ P1 & 0 & \\ P2, P5, P3, P6, P4 & 0.22 & 0 \end{pmatrix}$$

[{(P3, P6), P4}, (P2, P5)] , P1

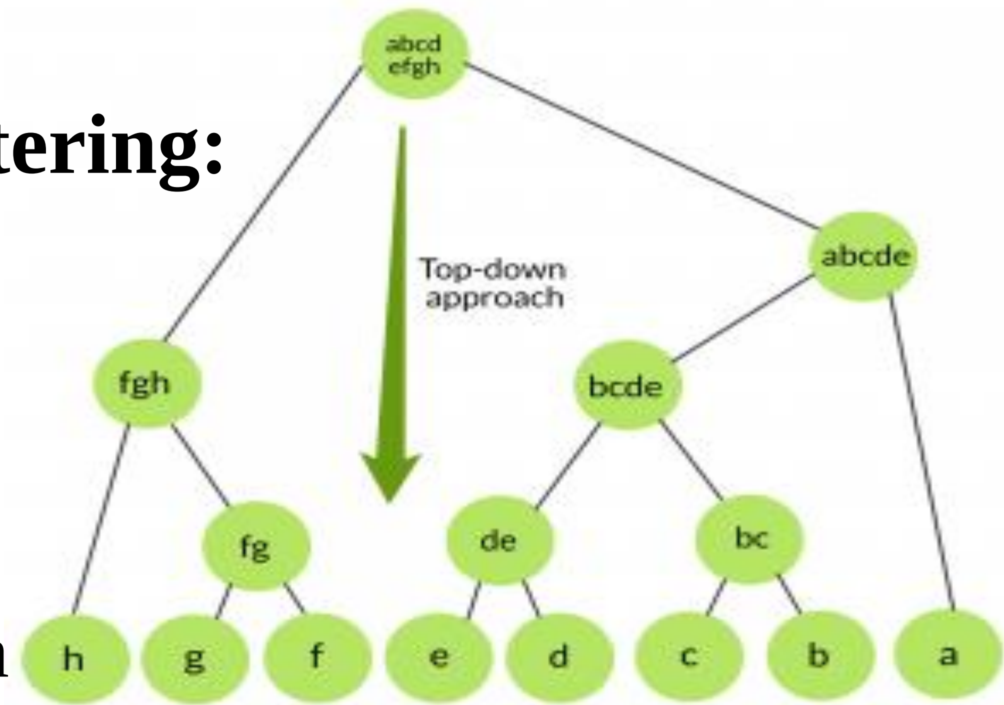


Divisive hierarchical clustering

All the data points are considered an individual cluster, and recursively dividing the data into smaller subsets until each subset contains only one data point or satisfies a stopping criterion.

Steps of Divisive Clustering:

- Initialize the algorithm
- Select a cluster to split
- Split the cluster
- Recurse on the subsets
- Generate the dendrogram



Hierarchical divisive clustering

Fuzzy c-means (FCM)

- fuzzy c-means clustering, is a soft clustering technique in machine learning in which each data point is separated into different clusters and then assigned a probability score for being in that cluster that provides one element of data or image belonging to two or more clusters.

Self-study: soft vs hard clustering

Fuzzy c-means (FCM)

- It starts with a random initial guess for the cluster centers; that is the mean location of each cluster.
- Next, FCM assigns every data point a random membership grade for each cluster.
- By iteratively updating the cluster centers and the membership grades for each data point, FCM moves the cluster centers to the correct location within a data set and, for each data point, finds the degree of membership in each cluster.

K-Means vs FCM

- k-means clustering algorithm gives the values of any point lying in some particular cluster to be either 0 or 1 i.e., either true or false. But the fuzzy logic gives the fuzzy values of any particular data point to be lying in either of the clusters.

K Mean

Data	Cluster-1	Cluster-2
(3,4)	100%	0%
(4,5)	100%	0%
(6,7)	0%	100%
(5,7)	0%	100%

Fuzzy C Mean

Data	Cluster-1	Cluster-2
(3,4)	91%	9%
(4,5)	82%	18%
(6,7)	30%	70%
(5,7)	10%	90%

Steps of FCM

Step -1: Randomly initialize the membership matrix $U^{(0)}$

Step -2: Calculate the centroid.

Step -3: Update membership matrix.

$$\mathrm{U}_{ij} = \left[\sum_{k=1}^{c} \left(\frac{\|\mathbf{x}_i - \mathbf{c}_j\|}{\|\mathbf{x}_i - \mathbf{c}_k\|} \right)^{\frac{2}{m-1}} \right]^{-1}$$

Step -4: Check for convergence

If $| U^{(1)} - U^{(0)} | < \varepsilon$ **Stop**

Else return **Step 2**

Example-1:

Cluster data (3, 4) (6, 7) and (4, 6) into two. Use degree of fuzziness, **m=2** and **ε=0.3**

Answer:

Step 1:

Initialize $U^{(0)}$

	Clus-1	Clus-2	
	0.8	0.2	X1 (3, 4)
	0.3	0.7	X2 (6, 7)
	0.9	0.1	X3 (4, 6)

❖ summation of membership of each data point should be equal to one.

Step 2:

C1(Centroid1) Calculation

X Dimension

$$C1=\frac{0.8*3+0.3*6+0.9*4}{0.8+0.3+0.9} = 3.9$$

Y Dimension

$$C1=\frac{0.8*4+0.3*7+0.9*6}{0.8+0.3+0.9} = 5.35$$

So, C1 ≡(3.9, 5.35)

$$U^{(0)} \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \\ 0.9 & 0.1 \end{bmatrix}$$

Given Data
(3, 4) (6, 7) and (4, 6)

$$V_{ij} = (\sum_{1}^{n}(\gamma_{ik}^{m} * x_k) / \sum_{1}^{n} \gamma_{ik}^{m}$$

Step 3:

$$U^{(1)} = \begin{array}{|cc|} U11 & U12 \\ U21 & U22 \\ U31 & U32 \end{array}$$

$$U11 = \frac{1}{\sum_{k=1}^{2}\left(\frac{||X1-C1||}{||X1-Ck||}\right)^{\frac{2}{2-1}}}$$

$$= \frac{1}{\left(\frac{||X1-C1||}{||X1-C1||}\right)^2+\left(\frac{||X1-C1||}{||X1-C2||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{||(3,4)-(3.9,5.35)||}{||(3,4)-(5.2,6.3)||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{\sqrt{(3-3.9)^2+(4-5.35)^2}}{\sqrt{(3-5.2)^2+(4-6.3)^2}}\right)^2}$$

$$= \frac{1}{1+\left(\frac{(3-3.9)^2+(4-5.35)^2}{(3-5.2)^2+(4-6.3)^2}\right)}$$

= **0.794**

So, **U12** = 1-U11 = 1-0.794 = **0.206**

$$U21 = \frac{1}{\sum_{k=1}^{2}\left(\frac{||X2-C1||}{||X2-Ck||}\right)^{\frac{2}{2-1}}}$$

$$= \frac{1}{\left(\frac{||X2-C1||}{||X2-C1||}\right)^2+\left(\frac{||X2-C1||}{||X2-C2||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{||(6,7)-(3.9,5.35)||}{||(6,7)-(5.2,6.3)||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{\sqrt{(6-3.9)^2+(7-5.35)^2}}{\sqrt{(6-5.2)^2+(7-6.3)^2}}\right)^2}$$

$$= \frac{1}{1+\left(\frac{(6-3.9)^2+(7-5.35)^2}{(6-5.2)^2+(7-6.3)^2}\right)}$$

$$= \frac{1}{1+\left(\frac{4.41+2.722}{0.64+0.49}\right)}$$

$$=0.142$$

So, U22=1-U12=1-0.142=**0.858**

$$U31 = \frac{1}{\sum_{k=1}^{2}\left(\frac{||X3-C1||}{||X3-Ck||}\right)^{\frac{2}{2-1}}}$$

$$= \frac{1}{\left(\frac{||X3-C1||}{||X3-C1||}\right)^2+\left(\frac{||X3-C1||}{||X3-C2||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{||(4.6)-(3.9,5.35)||}{||(4.6)-(5.2,6.3)||}\right)^2}$$

$$= \frac{1}{1+\left(\frac{\sqrt{(4-3.9)^2+(6-5.35)^2}}{\sqrt{(4-5.2)^2+(6-6.3)^2}}\right)^2}$$

$$= \frac{1}{1+\left(\frac{(4-3.9)^2+(6-5.35)^2}{(4-5.2)^2+(6-6.3)^2}\right)}$$

$$= \frac{1}{1+\left(\frac{0.01+0.4225}{1.44+0.09}\right)}$$

=0.779

So, **U23**=1-U13=1-0.779=**0.221**

$U^{(1)}=$

0.794	0.206
0.142	0.858
0.779	0.221

Step-4

$|U11^1 - U11^0|$= 0.794-0.8=0.006

$|U12^1 - U12^0|$=0.206-0.2=0.006

$|U21^1 - U21^0|$=0.142-0.3=0.158

$|U22^1 - U22^0|$=0.858-0.7=0.158

$|U31^1 - U31^0|$=0.779-0.9=0.121

$|U32^1 - U32^0|$=0.221-0.1=0.121

Max (0.006, 0.158, 0.121) =0.158< ε (True)

Example-2:

Suppose the given data points are **{(1, 3), (2, 5), (6, 8), (7, 9)}**, use fuzzy-c means clustering these data.