

Q1)

ANS.

What do you mean by clustering?

Clustering is an Un-supervised Machine Learning (Algorithm) technique used to group similar data points into clusters without any pre-defined Labels.

- Data points in the same cluster are similar to each other.
- Data points in different clusters are dissimilar.
- It does not require <sup>prior</sup> training data with labels.
- The Algorithm finds patterns within data and group them accordingly.
- Each group formed is known as 'cluster'.

Example:-

① Customer Segmentation:

- Group customers based on buying behaviour.

② Grouping Articles by Topic.

Application:-

1. Pattern Discovery.
2. Data Analysis.
3. Market Segmentation.

Common Algorithms:-

1. K-Means
2. Hierarchical Clustering.
3. DBSCAN.

Q2.)

ANS.

What is a Hierarchical Clustering Algorithm?

Hierarchical clustering builds a tree like structure (dendrogram) of clusters, to represent data grouping.

TYPES:-

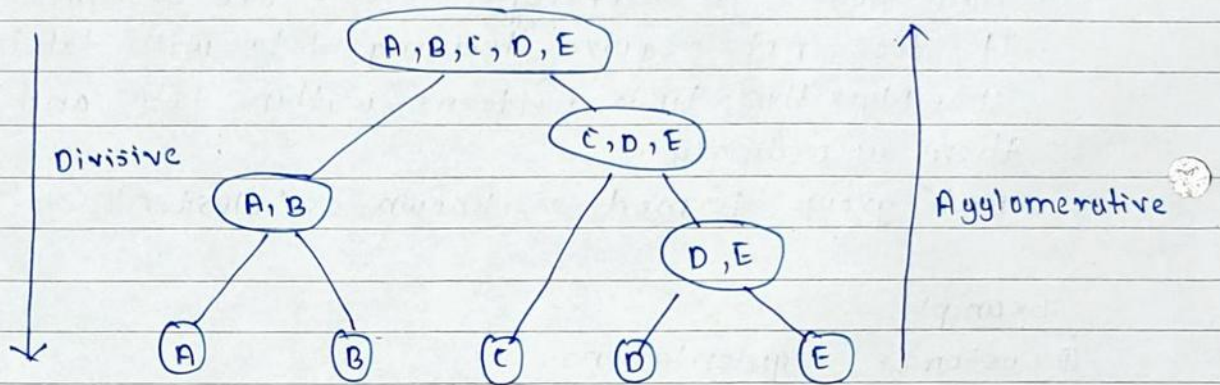
1. Agglomerative (Bottom Up)



- Start with Individual points.
- Merge clusters step by step.

## 2. Divisive (Top-Down)

- Start with One cluster.
- Split into smaller clusters.



### Advantages :-

- ① No need to specify number of cluster initially.
- ② Easy to visualize.

### Dis-Advantages :-

- ① Computationally Expensive.
- ② Not suitable for Very Large Datasets.

Hierarchical Clustering creates needed clusters using a tree structure.

Q4) What is Classification?

Ans. Classification is a supervised Machine Learning (ML) technique used to assign data into pre-defined categories or classes based on training data.

- The Model Learn From Labelled Data.
- Then predicts the class of new data.



For Eg:- Spam Detection (Yes/No)

o Working:- (Process involved)

1. Collect and prepare data.
2. Split Data into training data and testing data.
3. Test the Model on new data.
4. Evaluate Performance.

TYPES:-

1. Binary Classification: Two Classes (Yes/No)
2. Multi-class Classification: More than two classes.
3. Multi-label Classification: One Data point may belong to Multiple classes.

→ Algorithms:-

- ① Logistic Regression, ② Decision Tree, ③ SVM.

Hence, Classification predicts label based on training data.

Q5.) How to Evaluate a Classification Model?

ANS. A Classification Model is evaluated using different performance Metrics:-

1. Accuracy.

- It represents the percentage of correct predictions.

- Formula:-

$$\text{Accuracy} = \frac{\text{Correct Prediction}}{\text{Total Prediction}} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision:-

• Measures how many predicted positives are actually correct.

• Important when false positive must be minimized.

• Formula:-



$$\text{Precision} = \frac{TP}{TP + FP}$$

### 3. Recall.

• Measure how many actual positives are identified.

• Formula:-  $\text{Recall} = \frac{TP}{TP + FN}$

### 4. F1-Score

- Balance between precision and recall.

- Formula :-  $\text{F1-Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$

### 5. Confusion Matrix:-

A Table that shows:-

- |                      |                       |
|----------------------|-----------------------|
| ① True positive (TP) | ③ False positive (FP) |
| ② True Negative (TN) | ④ False Negative (FN) |

Q3.) What is the Difference between Classification & clustering?

NO.	<u>Feature:-</u>	<u>Classification:-</u>	<u>Clustering:-</u>
1.	Type	Supervised.	Un-supervised.
2.	Data	Labelled	Un-labelled.
3.	Goal.	Predict class labels.	Group similar data.
4.	Output.	Predefined classes.	New clusters.
5.	Example.	Spam Detection	Customer Segmentation.
6.	—	Categories are known (Yes/No)	Unknown group formed Automatically.