



# Applied Artificial Intelligence

## Practical # 8

<b>Name</b>	Ninad Karlekar	<b>Roll Number</b>	22306A1012
<b>Subject/Course:</b>	Applied Artificial Intelligence	<b>Class</b>	M.Sc. IT – Sem III
<b>Topic</b>	Clustering algorithm	<b>Batch</b>	1

### Topic: Clustering algorithm

#### Topic Clustering algorithm

**AIM:** Write an application to implement clustering algorithm.

#### **DESCRIPTION:** **UNSUPERVISED LEARNING:**

- **Clustering:**

Clustering is a machine learning technique that involves grouping similar data points together based on their inherent characteristics or features. It is used to discover patterns, associations, or structures within datasets, helping to identify natural groupings or clusters of data points, which can be valuable for tasks like customer segmentation, anomaly detection, and recommendation systems.

- **Association:**

Association is a data mining technique used to discover relationships, connections, or patterns within large datasets. It identifies associations or frequent co-occurrences of items in transactions, revealing insights into customer behavior, product affinities, and more. Association rule mining, a common application, is often used in market basket analysis and recommendation systems to uncover item associations and improve decision-making processes.

#### **CLUSTERING:**

#### **Clustering Types: -**

##### K-Means Clustering:

K-Means is one of the most popular clustering algorithms. It partitions data into K clusters based on the mean (center) of data points. It's effective for spherical clusters and works well when the number of clusters is known in advance.

##### Hierarchical Clustering

This approach creates a hierarchy of clusters by iteratively merging or splitting clusters. It can be agglomerative (bottom-up) or divisive (top-down) and represents clusters in a tree-like structure called a dendrogram.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN groups together data points that are close to each other while considering density. It can discover clusters of arbitrary shapes and is robust to noise.

Agglomerative Clustering:

This hierarchical clustering method starts with each data point as a separate cluster and then merges them iteratively. It is intuitive and useful when the number of clusters is not predetermined.

### Code: Unsupervised Learning

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import scipy.cluster.hierarchy as shc
from sklearn.cluster import AgglomerativeClustering

# Read the customer data from a CSV file
customer_data = pd.read_csv('/content/mall_customers.csv')

# Display the shape and the first few rows of the data
print(customer_data.shape)
customer_data.head()

# Extract the relevant columns from the data
data = customer_data.iloc[:, 3:5].values

# Create a dendrogram plot
plt.figure(figsize=(10, 7))
plt.title("Customer Dendrograms")
dend = shc.dendrogram(shc.linkage(data, method='ward'))

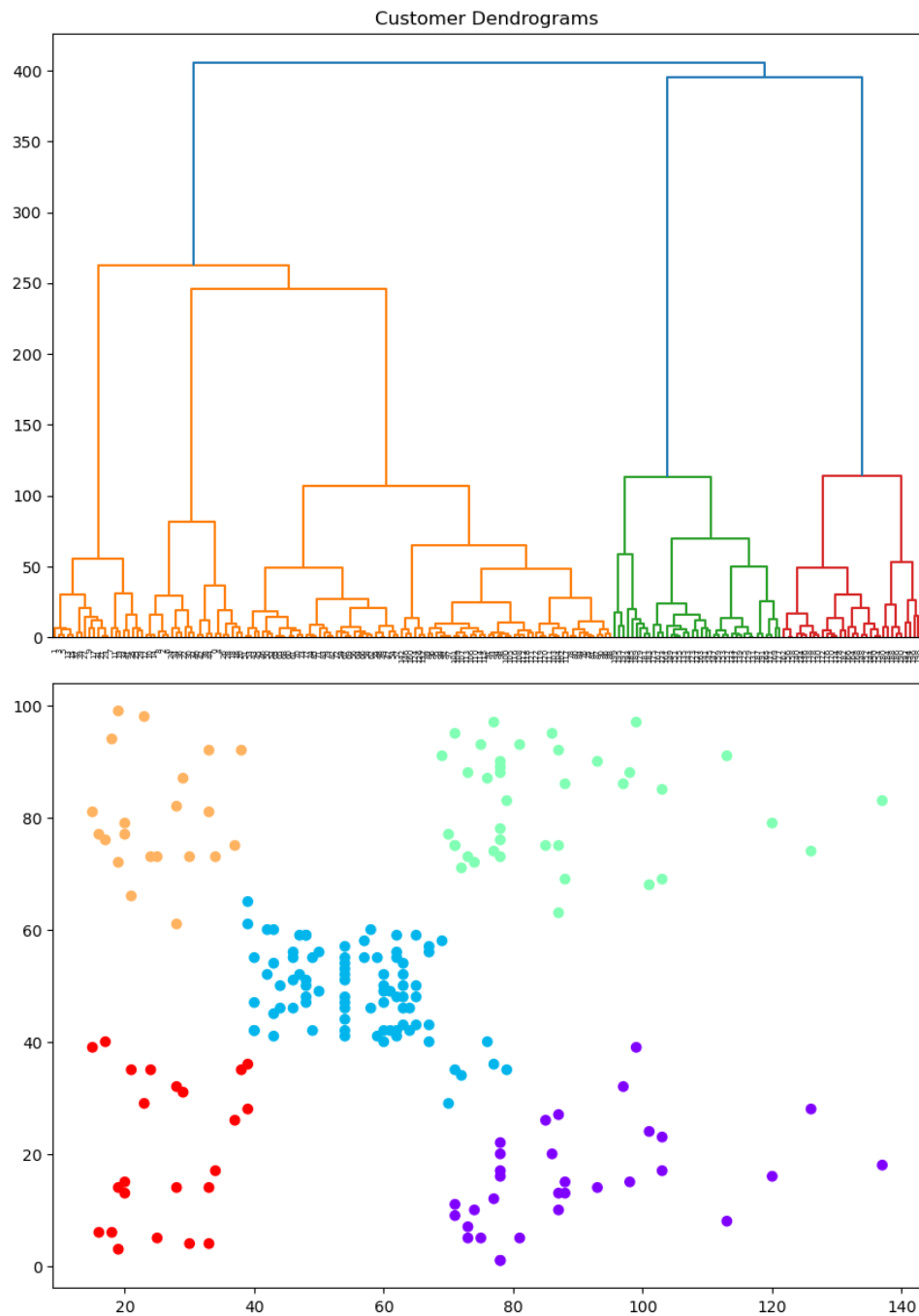
# Perform hierarchical clustering
cluster = AgglomerativeClustering(n_clusters=5, affinity='euclidean', linkage='ward')
cluster_labels = cluster.fit_predict(data)

# Create a scatter plot to visualize the clusters
plt.figure(figsize=(10, 7))
plt.scatter(data[:, 0], data[:, 1], c=cluster_labels, cmap='rainbow')
plt.show()
```

**Output:**

Out[5]:

	customer_id	gender	age	annual_income	spending_score
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40



### Code: Clustering

```
from numpy import where
from sklearn.datasets import make_classification
```

```
from matplotlib import pyplot

x,y =
make_classification(n_samples=1000,n_features=2,n_informative=2,n_redundant=0,n_clusters_per_class=1,ra
ndom_state=4)
for class_value in range(2):
    row_ix=where(y==class_value)
    pyplot.scatter(x[row_ix,0],x[row_ix,1])
pyplot.show()
```

**Output:**

