

Unsupervised Learning — Concept Document

Name: Sambathini Paul Jonathan

Day: 2

Topic: Unsupervised Learning – K-Means Clustering

1. **Definition:**

Unsupervised learning finds patterns in **unlabeled data** without predefined outputs.

2. **Goal:**

To group similar data points or discover hidden structures in datasets.

3. **Data Used:**

Synthetic data generated with `make_blobs()` containing 300 samples and 4 clusters.

4. **Algorithm:**

K-Means Clustering divides data into K groups where each point belongs to the nearest centroid.

5. **Process:**

Initialize K centers → assign points → update centers → repeat until stable.

6. **Implementation Steps:**

- Imported sklearn, numpy, and matplotlib.
- Created data and plotted before clustering.
- Applied `KMeans(n_clusters=4)` and predicted cluster labels.
- Visualized clusters and centroids on a scatter plot.

7. **Output:**

Displayed 4 distinct clusters with their center coordinates and inertia value.

8. **Observation:**

K-Means successfully separated data into clear groups; thread warning fixed using `OMP_NUM_THREADS=2`.

9. **Applications:**

Customer segmentation, image compression, fraud detection, document grouping.

10. **Conclusion:**

K-Means is a simple and effective unsupervised learning method for pattern discovery in unlabeled data.