

Week9 - Kmeans, GMM, DBSCAN

DS3010: Introduction to Machine Learning Lab

Timing: 02:00 PM to 04:45 PM

Max Marks: 5

Instructions

1. Submit one .ipynb file containing all answers. The name should be [student_name]_[rollno]_[lab].ipynb
 2. Write the questions in separate text blocks before the answers.
 3. Outputs for all sub-questions should be given and the code should be executable.
 4. Write justifications for your choices where needed.
 5. Ensure that all plots include clear labels and legends for better interpretation.
 6. Use of generative AI tools (such as ChatGPT, Gemini, etc.) is strictly prohibited. Any submission found to contain AI-generated or plagiarized content will receive a score of zero, and disciplinary action.
-

1. Data Preparation (1 mark)

Generate two synthetic datasets using:

```
from sklearn.datasets import make_blobs, make_moons
X1,y1 = make_blobs(n_samples=500, centers=4, cluster_std=0.8, random_state=42)
X2,y2 = make_moons(n_samples=500, noise=0.07, random_state=42)
```

Visualize both datasets using scatter plots with distinct colors for each cluster.

2. Apply Clustering Algorithms for both the datasets(3 marks)

- K-Means:
 - i. Compute and display inertia (sum of squared distances).
 - ii. Use Elbow Method to determine optimal K using inertia.
 - iii. Plot the clusters and centroids with optimal K.
- Gaussian Mixture Model (GMM):
 - i. Find the Optimal Number of Components using BIC/AIC.
 - ii. Fit and Visualize the clusters with Optimal Component.
- DBSCAN Clustering:
 - i. Find the Optimal eps (k-distance method).
 - ii. Fit and Visualize the clusters.

3. Observation (1 mark)

- Compare the performance of all three methods.
 - Which algorithm worked best for the “moons” dataset and why? Explain
 - What happens if eps is too small or too large in DBSCAN?
-