UNIT I

1. Define unsupervised learning. How does it differ from supervised learning?
2. Explain any three applications of unsupervised learning in real-world domains.
3. Discuss clustering as a machine learning task with suitable examples.
4. Differentiate between partitioning methods and hierarchical clustering.
5. Explain density-based clustering. How does DBSCAN work?
6. Compare hierarchical clustering and k-means clustering.
7. Discuss advantages and limitations of clustering in machine learning.
8. Explain the concept of biclustering.
9. What is spectral co-clustering? Give an application.
10. Define spectral biclustering and explain its importance.
11. What is an association rule? Provide an example.
12. Define support, confidence, and lift in association rule mining.
13. Explain the Apriori algorithm with its steps.
14. Differentiate between clustering and association rule mining.
15. Discuss challenges in applying clustering to high-dimensional data.

Numerical/Problem-Solving Questions

16. Apply k-means clustering (k=2) to the dataset: {(1,1), (2,1), (4,3), (5,4)}. Show steps.
17. Perform hierarchical clustering (single linkage) on data points: {1, 2, 6, 7, 9}.
18. Using DBSCAN with parameters ε=2, minPts=2, classify the dataset: {1, 2, 2.5, 8, 9, 10}.
19. Calculate the support and confidence for the rule {Milk} → {Bread} using a transaction dataset.
20. Given transactions:
    T1 = {Milk, Bread, Butter}
    T2 = {Milk, Bread}
    T3 = {Bread, Butter}
    T4 = {Milk, Butter}
    Find frequent itemsets with minimum support = 50%.
21. Apply Apriori algorithm step by step to generate association rules from the above dataset.
22. For dataset {(2,10), (2,5), (8,4), (5,8), (7,5), (6,4)}, perform k-means clustering with k=2.
23. Compute similarity matrix for points {(0,0), (0,2), (2,2)} using Euclidean distance.

24. Use hierarchical clustering (complete linkage) to cluster {(1,1), (1.5,1.5), (5,5), (3,4)}.
25. Find the clusters formed by k-means (k=2) after 2 iterations on points: {(1,2), (2,1), (4,5), (5,4)}.
26. Derive the time complexity of the Apriori algorithm.
27. A dataset has 10,000 transactions. Rule {A} → {B} occurs 500 times, support(A)=0.1. Compute confidence and lift.
28. For DBSCAN with ε=3, minPts=3, check if point (5,5) is a core, border, or noise given neighbors: {(5,6), (4,5), (7,5)}.
29. Demonstrate how biclustering can be applied in gene expression data.
30. Design a real-world case study of clustering for customer segmentation.

UNIT II

1. Define Gaussian Mixture Models (GMM).
2. Differentiate between Gaussian mixture and k-means clustering.
3. Explain Variational Bayesian Gaussian mixture.
4. What are the advantages of GMM over hard clustering methods?
5. Define manifold learning with examples.
6. Explain the working of Isomap algorithm.
7. Discuss Locally Linear Embedding (LLE).
8. What is Modified Locally Linear Embedding (MLLE)?
9. Explain Spectral Embedding in manifold learning.
10. Define Multi-Dimensional Scaling (MDS).
11. Explain t-distributed Stochastic Neighbor Embedding (t-SNE).
12. Compare PCA and t-SNE for dimensionality reduction.
13. What is factor analysis? How is it different from PCA?
14. Explain Kernel PCA and its advantage over standard PCA.
15. Discuss Latent Dirichlet Allocation (LDA) with applications.

Numerical/Problem-Solving Questions

16. Given data {1, 2, 5, 6}, fit a Gaussian mixture with 2 components. Show steps.
17. For a dataset {(2,3), (3,4), (4,5), (8,7)}, perform PCA and find first principal component.

18. Compute eigenvalues and eigenvectors of covariance matrix [[2,1],[1,2]] for PCA.
19. Perform truncated SVD on matrix [[1,0],[0,1],[1,1]].
20. Apply MDS on points {(0,0), (1,0), (0,1)} using Euclidean distance.
21. Use Isomap to reduce dataset {(0,0), (1,1), (2,2), (3,3)} to 1D.
22. Given a 3×3 covariance matrix, perform factor analysis (extract 2 factors).
23. Apply LLE on dataset {(0,0), (1,0), (0,1)} with k=2 neighbors.
24. Perform independent component analysis (ICA) on mixed signals X1 = S1 + S2, X2 = S1 - S2.
25. Decompose matrix [[4,2],[2,3]] using Non-negative Matrix Factorization (NMF).
26. A dataset has 1000 documents. Apply Latent Semantic Analysis (LSA) using SVD to extract 2 topics.
27. Perform kPCA using RBF kernel on data {(0,0), (1,0), (0,1)}.
28. Given covariance matrix [[3,1],[1,3]], reduce dimension using PCA to 1D.
29. Compute perplexity in t-SNE for dataset of 4 points with probabilities given.
30. Case study: Show how GMM can be used for speaker recognition.