# UNIT-V

<table>
<tr><td>1</td><td>Apply K-means clustering method for the following dataset with K=2 and initial cluster centroids are object 1 and object 2
<table>
<tr><th>Object</th><th>A</th><th>B</th></tr>
<tr><td>1</td><td>1.0</td><td>1.0</td></tr>
<tr><td>2</td><td>1.5</td><td>2.0</td></tr>
<tr><td>3</td><td>3.0</td><td>4.0</td></tr>
<tr><td>4</td><td>5.0</td><td>7.0</td></tr>
<tr><td>5</td><td>3.5</td><td>5.0</td></tr>
<tr><td>6</td><td>4.5</td><td>5.0</td></tr>
<tr><td>7</td><td>3.5</td><td>4.5</td></tr>
</table></td><td>10</td></tr>
<tr><td>2</td><td>Consider the data set shown in the table below. Apply agglomerative clustering techniques for the data with single linkage. Construct the dendrogram and show the updation of distance matrix at every step.
<table>
<tr><th>Point</th><th>X</th><th>Y</th></tr>
<tr><td>P1</td><td>0.40</td><td>0.53</td></tr>
<tr><td>P2</td><td>0.22</td><td>0.38</td></tr>
<tr><td>P3</td><td>0.35</td><td>0.32</td></tr>
<tr><td>P4</td><td>0.26</td><td>0.19</td></tr>
<tr><td>P5</td><td>0.08</td><td>0.41</td></tr>
<tr><td>P6</td><td>0.45</td><td>0.30</td></tr>
</table></td><td>10</td></tr>
<tr><td>3</td><td>Above question can be asked with complete linkage</td><td>10</td></tr>
<tr><td>4</td><td>Illustrate the steps involved in the Expectation-Maximization algorithm.</td><td>10</td></tr>
<tr><td>5</td><td>Illustrate the working of the K-means clustering algorithm.</td><td>10</td></tr>
<tr><td>6</td><td>Explain different types of clustering methods.</td><td>10</td></tr>
<tr><td>7</td><td>Explain various linkage methods in hierarchical clustering.</td><td>8</td></tr>
</table>

<table>
<tr><td>8</td><td>Explain the convergence process in EM algorithm.</td><td>6</td></tr>
<tr><td>9</td><td>For the given intial set of three clusters (k=3)<br>C1 = {(1,3),(3,6),(3,5)}<br>C2 = {(5,3),(6,7),(2,2)}<br>C3 = {(6,5),(3,1),(2,3)}<br>Apply k-means algorithm and show when the clusters converge.<br>Note: Use Euclidean Distance<br><br>NOTE: Can be asked for Manhattan distance also</td><td>10</td></tr>
<tr><td>10</td><td>Discuss the use of clustering in supervised learning with suitable examples.</td><td>5</td></tr>
<tr><td>11</td><td>Illustrate the Hierarchical clustering with an example.</td><td>10</td></tr>
<tr><td>12</td><td>Difference between agglomerative and divisive</td><td>5</td></tr>
<tr><td>13</td><td>Explain the process of Gaussian Mixture Models and their advantages.</td><td>8</td></tr>
<tr><td>14</td><td>Identify any two clustering techniques commonly used in data analysis and provide a brief description.</td><td>2</td></tr>
<tr><td>15</td><td>Outline the significance of the distance calculation between two clusters and its impact on clustering rules. Discuss the concept of Linkage methods.</td><td>2</td></tr>
<tr><td>16</td><td>Consider the research context and explain how Gaussian Mixture Models (GMMs) could be utilized to identify distinct customer segments based on purchasing behavior. Highlight the benefits of using GMMs over simpler clustering methods in this specific scenario.</td><td>2</td></tr>
<tr><td>17</td><td>Analysis of the dataset given below using the K-means clustering method for the following dataset with K=2 and initial cluster centroids are object 1 and object 2.<br><br>(Show till the second iteration and the centroids of the two clusters )<br><br><table><tr><th>Object</th><th>A</th><th>B</th></tr><tr><td>1</td><td>1.0</td><td>1.0</td></tr><tr><td>2</td><td>1.5</td><td>2.0</td></tr><tr><td>3</td><td>3.0</td><td>4.0</td></tr><tr><td>4</td><td>5.0</td><td>7.0</td></tr><tr><td>5</td><td>3.5</td><td>5.0</td></tr><tr><td>6</td><td>4.5</td><td>5.0</td></tr><tr><td>7</td><td>3.5</td><td>4.5</td></tr></table></td><td>7</td></tr>
<tr><td>18</td><td>Describe the basic working principles behind the functioning of Agglomerative Hierarchical</td><td>3</td></tr>
</table>

| | | |
|---|---|---|
| | clustering. | |
| 19 | Analyze and critically evaluate the process applied to the one-dimensional data set {7, 10, 20, 28, 35}. Apply the Complete Linkage - Agglomerative Hierarchical Clustering Euclidean. | 7 |
| 20 | Apply your understanding of the k-means algorithm by describing two specific scenarios where it is effectively utilized?" | 3 |
| 21 | Demonstrate the concept of convergence within the Expectation-Maximization (EM) algorithm and explain the four key steps integral to the EM algorithm. | 5 |
| 22 | Imagine you are working with a dataset that contains information about various measurements related to smartphones, such as screen size, battery life, camera quality, and storage capacity. Your goal is to identify the most critical factors influencing the overall performance of smartphones.<br>Which dimensionality reduction technique could you apply to extract the most important features from the dataset and reduce the complexity of the information? How would this technique help you better understand the key factors contributing to smartphone performance? | 5 |
| 23 | Classify the different types of anomaly detection methods and explain their applications in specific scenarios. | 5 |
| 24 | Analyze and critically evaluate the process applied to the one-dimensional data set {18, 22, 25, 42, 27,43}. Apply the Single link -Agglomerative Hierarchical Clustering. | 5 |
| 25 | Cluster the following eight points (with (x, y) representing locations) into three clusters:<br>A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)<br>Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). | 7 |
| 26 | What do you mean by Dimensionality reduction? Explain Principal component analysis working principle. | 7 |
| 27 | Explain Outlier detection / anomaly detection in clustering. | 5 |
| 28 | Describe the basic working principles behind the functioning of Divisive Hierarchical clustering. | 5 |
| 29 | Describe Principal Component Analysis Algorithm. Highlight its significance when the number of dimensions in the dataset in high. | 5 |
| 30 | For the given initial set of three clusters<br>C1 = {(1,3)(3,6)(3,5)} | 8 |

| | | |
|---|---|---|
| | C2 = {(5,3),(6,7),(2,2)}<br><br>C3 = {(6,5),(3,1),(2,3)}<br><br>Apply the agglomerative algorithm with a single linkage cluster distance measure to produce a dendrogram tree.<br><br>Note: Use Manhattan distance to compute the distance matrix. | |
| 31 | Examine the critical factors in evaluating and validating unsupervised learning models for anomaly detection. | 4 |

NOTE: Problems on K-means, Hierarchical clustering using Euclidean distance or Manhattan distance may be asked.