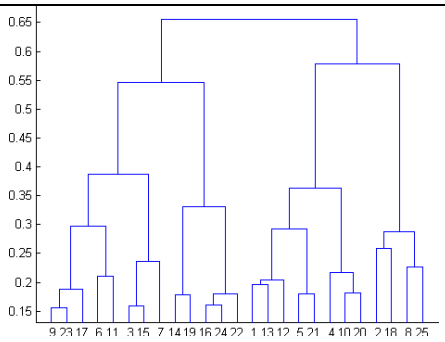
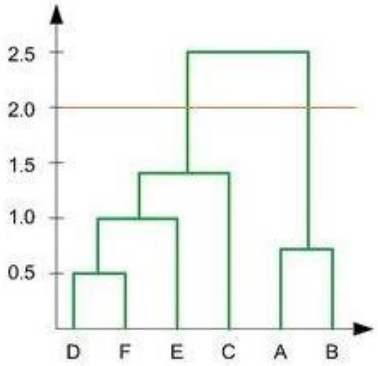
| | | | |
|---|---|---|---|
| **Course/Branch** : B.E / CSE | **Year / Semester** : IVth YR / VII Sem | Format No. | NAC/TLP-07a.13 |
| **Subject Code** : CS8091 | **Subject Name** : Big Data Analytics | Rev. No. | 02 |
| **Unit No** : 02 | **Unit Name** : Clustering and Classification | Date | 30.09.2020 |

**OBJECTIVE TYPE QUESTION BANK**

| S. No. | Objective Questions (MCQ /True or False / Fill up with Choices ) | BTL |
|---|---|---|
| 1. | Movie Recommendation systems are an example of<br>1.Classification 2. Clustering 3. Reinforcement Learning 4. Regression<br>a. 2 Only<br>b. 1 and 2<br>c. 1 and 3<br>d. **2 and 3** | L3 |
| 2. | Sentiment Analysis is an example of<br>1. Regression 2. Classification 3. Clustering 4 Reinforcement Learning<br>**a. 1, 2 and 4**<br>b. 1 and 3<br>c. 1, 2 and 3<br>d. 1 and 2 | L3 |
| 3. | Can decision trees be used for performing clustering?<br>**a. True**<br>**b.** False | L4 |
| 4. | What is the minimum no. of variables/ features required to perform clustering?<br>1. 0<br>**2. 1**<br>3. 2<br>4. 3 | L1 |
| 5. | For two runs of K-Mean clustering is it expected to get same clustering results?<br>1. Yes<br>**2. No** | L3 |
| 6. | Which of the following can act as possible termination conditions in K-Means?<br>1. For a fixed number of iterations.<br>2. Assignment of observations to clusters does not change between iterations. Except for cases with a bad local minimum.<br>3. Centroids do not change between successive iterations. 4.Terminate when RSS falls below a threshold.<br>a. 1, 3 and 4<br>b. 1, 2 and 3<br>c. 1, 2 and 4<br>d**. All of the above** | L1 |
| 7. | Which of the following algorithm is most sensitive to outliers?<br>**1. K-means clustering algorithm**<br>2. K-medians clustering algorithm<br>3. K-modes clustering algorithm<br>4. K-medoids clustering algorithm | L3 |
| 8. | After performing K-Means Clustering analysis on a dataset, you observed the following dendrogram. Which of the following conclusion can be drawn from the dendrogram? | L6 |

| | | | |
|---|---|---|---|
| **Course/Branch** : B.E / CSE | **Year / Semester** : IVth YR / VII Sem | Format No. | NAC/TLP-07a.13 |
| **Subject Code** : CS8091 | **Subject Name** : Big Data Analytics | Rev. No. | 02 |
| **Unit No** : 02 | **Unit Name** : Clustering and Classification | Date | 30.09.2020 |

**OBJECTIVE TYPE QUESTION BANK**

| | | |
|---|---|---|
| |  <br>a. There were 28 data points in clustering analysis<br>b. The best no. of clusters for the analyzed data points is 4<br>c. The proximity function used is Average-link clustering<br>**d. The above dendrogram interpretation is not possible for K-Means clustering analysis** | |
| 9. | In the figure below, if you draw a horizontal line on y- axis for y=2. What will be the number of clusters formed?<br>  <br>1. 1<br>**2. 2**<br>3. 3<br>4. 4 | L6 |
| 10. | In which of the following cases will K-Means clustering fail to give good results?<br>1. Data points with outliers<br>2. Data points with different densities<br>3. Data points with round shapes<br>4. Data points with non-convex shapes<br>a. 1 and 2<br>b. 2 and 3<br>c. 2 and 4<br>d**. 1, 2 and 4** | L4 |
| 11. | The discrete variables and continuous variables are two types of<br>a. Open end classification<br>b. Time series classification<br>c. Qualitative classification<br>**d. Quantitative classification** | L1 |

| | | | |
|---|---|---|---|
| **Course/Branch** : B.E / CSE | **Year / Semester** : IVth YR / VII Sem | Format No. | NAC/TLP-07a.13 |
| **Subject Code** : CS8091 | **Subject Name** : Big Data Analytics | Rev. No. | 02 |
| **Unit No** : 02 | **Unit Name** : Clustering and Classification | Date | 30.09.2020 |

**OBJECTIVE TYPE QUESTION BANK**

| | | |
|---|---|---|
| 12. | Bayesian classifiers is<br>**1. A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.**<br>2. Any mechanism employed by a learning system to constrain the search space of a hypothesis<br>3. An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.<br>4. None of these | L1 |
| 13. | Classification accuracy is<br>1. A subdivision of a set of examples into a number of classes<br>**2. Measure of the accuracy, of the classification of a concept that is given by a certain theory**<br>3. The task of assigning a classification to a set of examples<br>4. None of these | L1 |
| 14. | Classification task referred to<br>1. A subdivision of a set of examples into a number of classes<br>2. A measure of the accuracy, of the classification of a concept that is given by a certain theory<br>**3. The task of assigning a classification to a set of examples**<br>4. None of these | L1 |
| 15. | Euclidean distance measure is<br>1. A stage of the KDD process in which new data is added to the existing selection.<br>2. The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them<br>3. The distance between two points as calculated using the Pythagoras theorem<br>**4. None of these** | L1 |
| 16. | _____ is good at handle missing data and support both the kind of attributes ( i.e Categorial and Continuous attributes )<br>a. ID3.<br>**b. C4.5.**<br>c. CART.<br>d. Naïve Bayes. | L4 |
| 17. | Decision trees use _____, in that they always choose the option that seems the best available at that moment.<br>**a. Greedy Algorithms.**<br>b. Divide and Conquer.<br>c. Backtracking.<br>d. Shortest Path Method. | L2 |
| 18. | Decision trees cannot handle categorical attributes with many distinct values, such as country codes for telephone numbers.<br>a. TRUE<br>b. **FALSE** | L4 |
| 19. | _____are easy to implement and can execute efficiently even without | L2 |

| | | | |
|---|---|---|---|
| **Course/Branch** : B.E / CSE | **Year / Semester** : IVth YR / VII Sem | Format No. | NAC/TLP-07a.13 |
| **Subject Code** : CS8091 | **Subject Name** : Big Data Analytics | Rev. No. | 02 |
| **Unit No** : 02 | **Unit Name** : Clustering and Classification | Date | 30.09.2020 |

**OBJECTIVE TYPE QUESTION BANK**

| | | |
|---|---|---|
| | prior knowledge of the data, they are among the most popular algorithms for classifying text documents.<br>a. ID3<br>**b. Naïve Bayes classifiers**<br>c. CART<br>d. None of these. | |
| 20. | High entropy means that the partitions in classification are<br>a. Pure<br>**b. Not pure**<br>c. Useful<br>d. Useless | L2 |
| 21. | Which of the following statements about Naive Bayes is incorrect?<br>a. Attributes are equally important.<br>**b. Attributes are statistically dependent of one another given the class value.**<br>c. Attributes are statistically independent of one another given the class value.<br>d. Attributes can be nominal or numeric | L4 |
| 22. | The maximum value for entropy depends on the number of classes so if we have 8 Classes what will be the max entropy.<br>$Entropy(x) = -\sum(P(x=k) * log_2(P(x=k))$<br>a. Max Entropy is 1<br>b. Max Entropy is 2<br>**c. Max Entropy is 3**<br>d. Max Entropy is 4 | L6 |
| 23. | John flies frequently and likes to upgrade his seat to first class. He has determined that if he checks in for his flight at least two hours early, the probability that he will get an upgrade is 0.75; otherwise, the probability that he will get an upgrade is 0.35. With his busy schedule, he checks in at least two hours before his flight only 40% of the time. Suppose John did not receive an upgrade on his most recent attempt. What is the probability that he did not arrive two hours early?<br>a. 0.892<br>**b. 0.796**<br>c. 0.685<br>d. 0.999 | L6 |
| 24. | Point out the wrong statement.<br>**a. k-nearest neighbor is same as k-means**<br>b. k-means clustering is a method of vector quantization<br>c. k-means clustering aims to partition n observations into k clusters<br>d. none of the mentioned | L4 |
| 25. | Consider the following example "How we can divide set of articles such that those articles have the same theme (we do not know the theme of the articles ahead of time) " is this:<br>**1. Clustering**<br>2. Classification | L3 |

| | | | |
|---|---|---|---|
| **Course/Branch** : B.E / CSE | **Year / Semester** : IVth YR / VII Sem | Format No. | NAC/TLP-07a.13 |
| **Subject Code** : CS8091 | **Subject Name** : Big Data Analytics | Rev. No. | 02 |
| **Unit No** : 02 | **Unit Name :** Clustering and Classification | Date | 30.09.2020 |

**OBJECTIVE TYPE QUESTION BANK**

| | | |
|---|---|---|
| | 3. Regression<br>4. None of These | |
| 26. | Can we use K Mean Clustering to identify the objects in video?<br>**1. Yes**<br>2. No | L4 |
| 27. | Clustering techniques are _____ in the sense that the data scientist does not determine, in advance, the labels to apply to the clusters.<br>**1. Unsupervised**<br>2. Supervised<br>3. Reinforcement<br>4. Neural network | L2 |