MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question.

Answer- d. All of the above

2. On which data type, we cannot perform cluster analysis?

Answer- d. None

3. Netflix's movie recommendation system uses

Answer- c. Reinforcement learning and Unsupervised learning

4. The final output of Hierarchical clustering is

Answer- c. A map defining the similar data points into individual groups

5. Which of the step is not required for K-means clustering?

Answer- d. None

6. Which is the following is wrong?

Answer- d. None

7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?

Answer- i. Single-link ii. Complete-link iii. Average-link Options:

d- 1,2 and 3

8. Which of the following are true?

Answer- i. Clustering analysis is negatively affected by multicollinearity of features ii. Clustering analysis is negatively affected by heteroscedasticity Options:

a. 1 only

9. 9. In the figure above, if you draw a horizontal line on y-axis for y=2. What will be the number of clusters formed?

Answer- b 4

10. For which of the following tasks might clustering be a suitable approach?

Answer- b.Given a database of information about your users, automatically group them into different market

11. Given, six points with the following attributes:

Answer-

points	X coordinate	Y coordinate
P1	0.4005	0.5306
P2	0.2148	0.3854
P3	0.3457	0.3156
P4	0.2652	0.1875
P5	0.0789	0.4139
P6	0.4548	0.3022

Table : X-Y coordinate of six points

P1	0.0000	0.2357	0.2213	0.3688	0.3421	0.2347
P2	0.2357	0.0000	0.1483	0.2042	0.1388	0.2540
P3	0.2218	0.1483	0.0000	0.1513	0.2843	0.1100
P4	0.3688	0.2042	0.1513	0.0000	0.2932	0.2216
P5	0.3421	0.1388	0.2843	0.2932	0.0000	0.3921
P6	0.2347	0.2540	0.1100	0.2216	0.3921	0.0000

Table : Distance Matrix for Six Points.

12. Which of the following clustering representations and dendrogram depicts the use of MIN or Single link proximity function in hierarchical clustering:

Answer- B

13. 13. What is the importance of clustering?

Answer- Clustering is important in data analysis and data mining applications. It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups. A good clustering algorithm is able to identity clusters irrespective of their shapes.

14.. How can I improve my clustering performance ?

Answer- Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance. Surprisingly for some cases, high clustering performance can be achieved by simply performing K-means clustering on the ICA components after PCA dimension reduction on the input data.

Worksheet 3 SQL

Refer the following ERD and answer all the questions in this worksheet. You have to write the queries using mysql for the required Operation.

- Customers: stores customer's data.
- Products: stores a list of scale model cars.
- ProductLines: stores a list of product line categories.
- Orders: stores sales orders placed by customers.
- OrderDetails: stores sales order line items for each sales order.
- Payments: stores payments made by customers based on their accounts.
- Employees: stores all employee information as well as the organization structure such as who reports to whom.
- Offices: stores sales office data.

1. Write SQL query to create table Customers

   Answer- create table customers (customerName varchar(20) not null, contactFirstName varchar(20), contactLastName varchar(25), customersNumber varchar(300), phone varchar(300) not null, addressLine1 text, addressLine2 text, city varchar(20) not null, state varchar(20) not null,postalCode int(11) not null,country varchar(30) not null, salesRepEmployeeNumber int not null,creditLimit int not null) ;

2. **Write SQL query to create table Orders.**

Answer- create table orders (orderNumber int PRIMARY KEY NOT NULL, orderDate DATE ,requiredDate DATE, shippedDate DATE, status varchar(20), comments text ,customersNumber int);

3. **Write SQL query to show all the columns data from the Orders Table.**
   Ans. select * from orders;
4. **Write SQL query to show all the comments from the Orders Table.**
   Ans. select comments from orders
5. **Write a SQL query to show the order date and the Total number of orders placed on that date, from the Orders table.**
   Ans. select orderDate ,count(*) as Total_number_of_orders from orders GROUP BY orderDate
6. **Write a SQL query to show employeeNumber, lastName, firstName of all the employees from employees Table.**
   Ans. select employeesNumber,lastName,firstName from employees
7. **Write a SQL query to show all orderNumber, customerName of the person who placed the respective order.**
   Ans. select orders.orderNumber ,customers.customerName from customers INNER JOIN orders ON orders.customersNumber = customers.customersNumber
8. **Write a SQL query to show name of all the customers in one column and salerepemployee name in another column.**
   Ans. select customers.customerName ,employees.firstName as salerepemployee from employees INNER JOIN customers ON customers.salesRepEmployeeNumber = employeesNumber
9. Write a SQL query to show Date in one column and total payment amount of the payments made on that date from the payments table
   Ans. select paymentDate,sum(amount) from payments GROUP BY paymentDate
10. **Write a SQL query to show all the products productName, MSRP, productDescription from the products table.**

Ans. select productName, MSRP, productDescription from products

11. **Write a SQL query to print the productName, productDescription of the most ordered product.**
    Ans. select productName, productDescription, count(productName) from products GROUP BY productCode limit 1
12. Write a SQL query to get the name of the state having maximum number of customers
    Ans. select state ,count(state) from customers GROUP BY state limit 1
13. **Write a SQL query to print the employee number in one column and Full name of the employee in the second column for all the employees.**
    Ans. select concat(firstName,lastName) as Full_Name,employeesNumber from employees;
14. **Write a SQL query to print the orderNumber, customer Name and total amount paid by the customer for that order (quantityOrdered × priceEach).**
    Ans. select orders.orderNumber ,payments.amount ,customers.customerName from orders,payments,customers where payments.customersNumber = customers.customersNumber and customers.customersNumber = orders.customersNumber

STATISTICS WORKSHEET-3

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following is the correct formula for total variation?
   Ans- b) Total Variation = Residual Variation + Regression Variation
2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.
   Ans- c) binomial
3. How many outcomes are possible with Bernoulli trial?
   Ans- a) 2
4. If Ho is true and we reject it is called
   Ans- a) Type-I error
5. Level of significance is also called:
   Ans- b) Size of the test
6. The chance of rejecting a true hypothesis decreases when sample size is:
   Ans- b) Increase
7. Which of the following testing is concerned with making decisions using data?
   Ans- b) Hypothesis
8. What is the purpose of multiple testing in statistical inference?
   Ans- d) All of the mentioned
9. Normalized data are centred at and have units equal to standard deviations of the original data
   Ans- a) 0
10. What Is Bayes' Theorem?

Ans- . In probability, the Bayes theorem is a mathematical formula, which is used to determine the conditional probability of a given event. Conditional probability is defined as the likelihood that an event will occur, based on the occurrence of a previous outcome.

P(A|B) is the probability of event A occurring given that B is true. P(B|A) is the probability of event B occurring given that A is true. P(A) and P(B) are the probabilities of observing A andB respectively without any given conditions

11. **What is the z-score?**

Ans- In statistics, Z-score is the method to find out the outliers present in the data, and also z-score shows how much the particular point is away from the standard deviation. Z-scores range from -3 standard deviations up to +3 standard deviations.
Formula for find out the z-score is : $z = (x – \mu) / \sigma$ where , x = data point μ = Mean value σ =Standard deviation

12. **What is a t-test?**
    Ans- The independent sample t-test or 2 samples t-test compares the mean of two independent groups in order to determine whether the mean of two different variables is identical or not.
13. **What is a percentile?**

Ans- In statistics, the percentile is used to indicate the value below which the group the percentage of data fall. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.

14. What is ANOVA?

Ans- ANOVA test is a type of statical test that allows a comparison of more than two groups at thesame time it helps to determine whether a relationship exists between them or not.

15. **How can ANOVA help?**

    Ans- The one-way ANOVA can help you to determine whether or not there are significant differences between the means of your independent variables(for ex- Age, Sex, Position). When you understand how each independent variables are different from others, you can begin to understand which of them has a connection to your dependent variables and begin to learn what is driving that behavior