# Machine Learning

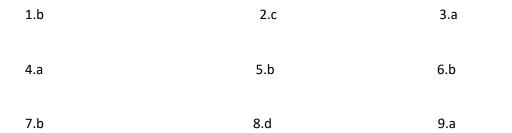| | | |
|---|---|---|
| 1.d | 2.d | 3.c |
| 4.d | 5.d | 6.c |
| 7.d | 8.a | 9.a |
| 10.b | 11.a | 12.b |

13.Important of clustering, (i)Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.

(ii)This clustering analysis has been used for model analysis, vector region of attraction.

(iii)Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.

(iv)They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.

(v)They are used in outlier detections to detect credit card fraudulence.

14. 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

1.Create table customer (customerNumber int primary key,customerName text,contactLastName text,contactFirstName text, phone text,addressLine1 text,addressLine2 text,city text,state text,postalCode int,country text,saleRepEmployeeNumber int,creditLimit float)

2. Create table orders (orderNumber int primary key, orderDate date , requiredDate date , shippedDate date , status text, comments text, customerNumber int)

3. SELECT * FROM Orders;

4. SELECT `comments` FROM Orders;

5. SELECT `orderDate`, COUNT(`orderNumber`) as `Total number of orders` FROM Orders GROUP BY `orderDate`;

6. SELECT `employeeNumber` ,`lastName` ,`firstName` FROM Employees;

7. SELECT `orderNumber` ,`customerName` FROM Orders INNER JOIN customers ON orders.`customerNumber`= customers.`customerNumber`;

8. SELECT `customerName` , CONCAT(`firstName`,`lastName`) FROM employees INNER JOIN customers ON Employees.`employeeNumber`= customers.`salesRepEmployeeNumber`;

9. SELECT `paymentDate`, SUM(amount) FROM payments GROUP BY `paymentDate`;

10. SELECT `productName`,`MSRP`,`productDescription` FROM products;

11. SELECT `productName`,`productDescription` FROM Products INNER JOIN Orderdetails ON Products.`productCode` = Orderdetails.`productCode` GROUP BY Products.`productCode` ORDER BY SUM(`quantityOrdered`) DESC LIMIT 1;

12. SELECT `city` FROM Orders as a INNER JOIN Customers as b ON a.`customerNumber` = b.`customerNumber` GROUP BY `city` ORDER BY COUNT(`orderNumber`) DESC LIMIT 1;

13. SELECT `state` FROM Customers GROUP BY `state` ORDER BY COUNT(`customerNumber`) DESC LIMIT 1;

14. SELECT `employeeNumber`, concat(`firstName`,`lastName`) as `Full name` FROM Employees;

15. SELECT `orderNumber`, `customerName`, `quantityOrdered` * `priceEach` as `total amount paid` FROM OrderDetails as a INNER JOIN Orders as b ON a.`orderNumber` = b.`orderNumber` INNER JOIN Customers as c ON b.`customerNumber` = c.`customerNumber`;

# STATISTICS WORKSHEET-3

1.b                     2.c                     3.a

4.a                     5.b                     6.b

7.b                     8.d                     9.a

10. Bayes' Theorem Bayes' theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of conditional probability. Bayes theorem is also known as the formula for the probability of "causes".

11. A Z-score is a numerical measurement used in statistics of a value's relationship to the mean (average) of a group of values, measured in terms of standard deviations from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score.

12.A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.

13.In statistics percentile is use to indicate the value below which the group of percentage of data fall below.

14.Anova is a statistical method in which the variation in a set of observations is divided into distinct components.

15.Anova is helpful for testing three or more variables. It is similar to multiple two-sample t-test. However, it results in fewer type I error and is appropriate for a range of issues. ANOVA group differences by comparing the means of each group and include spreading out the variance into diverse sources.