

- 1) D
- 2) D
- 3) C
- 4) B
- 5) D
- 6) C
- 7) A
- 8) B
- 9) D
- 10) C
- 11) A
- 12) A

13) **IMPORTANCE OF CLUSTERING:** 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 identify clusters irrespective of their shapes.

14) A solution for this problem is the k-means++ algorithm, which uses a different initialization. The idea is pretty simple:

Instead of random initialization, we only choose the first center randomly. All

following centers are then still sampled, but with a probability that is proportional to their squared distance from all current centers.

Points further

away from current centers get a higher probability to become a center in the next iteration of initialization.

STATISTICS

- 1) B
- 2) C
- 3) A
- 4) A
- 5) A

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". For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black. In this case, the probability of occurrence of an event is calculated depending on other conditions is known as conditional probability. In this article, let us discuss the statement and proof for Bayes theorem, its derivation, formula, and many solved examples.

- 11) **Z_score:** 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) **t_test:** A t-test is a test that is used to compare the of two groups. It is often used to determine whether a process or treatment actually has an effect on the population of interest, or whether two groups are different from one another.

You want to know whether the mean petal length of iris flowers differs according to their species. You find two different species of irises growing in a garden and measure 25 petals of each species. You can test the difference between these two groups using a t-test

- The null hypothesis (H_0) is that the true difference between these group means is zero.
- The alternate hypothesis (H_a) is that the true difference is different from zero.

13) **n statistics,** a percentile (percentile score or centile) is a score below which a given percentage of scores in its frequency distribution falls (exclusive definition) or a score at or below which a given percentage falls (inclusive definition).

- 14) **ANOVA:** What is ANOVA? Developed by Ronald Fisher, ANOVA stands for Analysis of Variance. One-Way Analysis of Variance tells you if there are any statistical differences between the means of three or more independent groups.

MACHINE LEARNING

- 1) CREATE TABLE Customers(customerNumber int,customerName text,contactLastName varchar(255),contactFirstName varchar(255),phone int,addressLine1 int primary key,addressline2 int primary key,city text,state text,country text,saleRepEmployeeNumber int,creditLimit int);
- 2) CREATE TABLE Orders(orderNumber int,productcode int primary key,required date int primary key,shipped date int primary key,status varchar(255),comments text,customerNumber int);
- 3) SELECT* FROM Orders
- 4) DBA_TAB_COMMENTS
- 5) SELECT MAX(orders) FROM City
- 6) SELECT MAX(customers) FROM Customers