MACHINE LEARNING

Ans 1. d. All of the above

Ans2. d. None

Ans3. C. Reinforcement learning and Unsupervised learning

Ans4. b. The tree representing how close the data points are to each other

Ans5. d. None

Ans6. c. k-nearest neighbour is same as k-means

Ans7. d. 1, 2 and 3

Ans8. a. 1 only

Ans9. a. 2

Ans10. A. Given sales data from a large number of products in a supermarket, estimate future sales for each of these products.

Ans11. (a)

Ans12. (b)

Ans13. The goal of cluster analysis is to discover natural grouping of a set of patterns, points or objects."
Clustering can be defined on the basis of similarity, such that the intraclass variation is low while the interclass variation is high. Clusters differ in terms of shape, size and density. If there is noise in the data, then detection of cluster becomes even more difficult.

Ans14. Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step.
Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.

WORKSHEET 3 SQL

STATISTICS WORKSHEET-3

Ans1. b) Total Variation = Residual Variation + Regression Variation

Ans2. c) binomial

Ans3. a) 2

Ans4. a) Type I error

Ans5. C) Level of confidence

Ans6. B) Increase

Ans7. B) Hypothesis

Ans8. d) All of the mentioned

Ans9. a) 0

Ans10. Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

Ans11. A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is 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.

Ans12. A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related. T-tests are used when the data sets follow a normal distribution and have unknown variances, like the data set recorded from flipping a coin 100 times.

Ans13. a value on a scale of 100 that indicates the percent of a distribution that is equal to or below it.

Ans 14. ANOVA, which stands for **Analysis of Variance**, is a statistical test used to analyze the difference between the means of more than two groups. A one-way ANOVA uses one independent variable, while a two-way ANOVA uses two independent variables.

Ans15. ANOVA is helpful for **testing three or more variables**. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues.