MACHINE LEARNING

1. a. biological network analysis
2. b. Text data
3. a. Supervised learning
4. a. The number of cluster centroids
5. d. None
6. c. k-nearest neighbour is same as k-means
7. d. 1, 2 and 3
8. a. 1 only
9. a. 2
10. c. Predicting whether stock price of a company will increase tomorrow.
11. D.
12. D.

13. What is important of clustering?

Asn:
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. Clustering quality depends on the methods and the identification of hidden patterns.

14. How can I improve my clustering performance?

Ans:
Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step.

WORKSHEET 3 SQL

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial

3. a) 2
4. a) Type-I error
5. a) Power of the test
6. b) Increase
7. b) Hypothesis
8. d) All of the mentioned
9. a) 0

Theory section

10. What Is Bayes' Theorem?

Ans:
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.

11. What is z-score?

Ans:

Z-score indicates how much a given value differs from the standard deviation. The Z-score, or standard score, is the number of standard deviations a given data point lies above or below mean. Standard deviation is essentially a reflection of the amount of variability within a given data set.

12. What is t-test?

Ans:

The t test estimates the true difference between two group means using the ratio of the difference in group means over the pooled standard error of both groups. You can calculate it manually using a formula or use statistical analysis software.

13. What is percentile?

Ans:

A percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed.

14. . What is ANOVA?

Ans:

Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests.

15. How can ANOVA help?

Ans:

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