MACHIN LEARNING

1. Which of the following is an application of clustering?
    - All of the above .
2. On which data type, we cannot perform cluster analysis?
    - None.
3. Netflix's movie recommendation system uses-
    - Reinforcement learning and Unsupervised learning.
4. The final output of Hierarchical clustering is-
    - The tree representing how close the data points are to each other.
5. Which of the step is not required for K-means clustering?
    - None.
6. Which is the following is wrong?
    - k-nearest neighbour is same as k-means.
7. Which of the following metrics, do we have for finding dissimilarity between two clusters in hierarchical clustering?
    - 1,2 and 3
8. Which of the following are true?
    - 1 Only.
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?
    - 2
10. For which of the following tasks might clustering be a suitable approach?
    - Predicting whether stock price of a company will increase tomorrow.

11. Given, six points with the following attributes:

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

- (b)answer

12. Given, six points with the following attributes:

Which of the following clustering representations and dendrogram depicts the use of MAX or Complete link proximity function in hierarchical clustering.

- (c)answer

13. What is the importance of clustering?

- Clustering is used to identify groups of similar objects in datasets with two or more variable quantities.

14. How can I improve my clustering performance?

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

STATISTICS

1. Which of the following is the correct formula for total variation?
    - Total Variation = Residual Variation + Regression Variation.
2. Collection of exchangeable binary outcomes for the same covariate data are called_____ outcomes.
    - Random.
3. How many outcomes are possible with Bernoulli trial?
    - 2
4. If Ho is true and we reject it is called-
    - Type-I error.
5. Level of significance is also called-
    - Level of confidence.
6. The chance of rejecting a true hypothesis decreases when sample size is-
    - Increase.
7. Which of the following testing is concerned with making decisions using data?
    - Hypothesis.
8. What is the purpose of multiple testing in statistical inference?
    - Minimize false positives.

9. Normalized data are centred at _____and have units equal to standard deviations of the original data-
    - 0
10. What Is Bayes' Theorem?
    - 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?
    - 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.
12. What is t-test?
    - The Student's t-test is a statistical hypothesis test that **two independent data** samples known to have a Gaussian distribution, have the same Gaussian distribution,

    The assumption or null hypothesis of the test is that the means of two populations are equal. A rejection of this hypothesis indicates that there is sufficient evidence that the means of the populations are different, and in turn that the distributions are not equal.
13. What is percentile?
    - A p-value is the probability of rejecting a null-hypothesis when the hypothesis is proven true. The null hypothesis is a statement that says that there is no difference between two measures.

14. What is ANOVA?

- Analysis of variance (ANOVA) uses F-tests to statistically assess the equality of means when you have three or more groups.

15. How can ANOVA help?

- An F-statistic is the ratio of two variances, or technically, two mean squares. Mean squares are simply variances that account for the degrees of freedom (DF) used to estimate the variance.