**Machine Learning**

1. Which of the following is an application of clustering?

Ans: d) All of the above

2. On which data type, we cannot perform cluster analysis?

Ans: d) None

3. Netflix’s movie recommendation system uses-

Ans: d) All of the above

4. The final output of Hierarchical clustering is-

Ans: b) The tree representing how close the data points are to each other

5. Which of the step is not required for K-means clustering?

Ans: d) None

6. Which is the following is wrong?

Ans: c) 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?

Ans: d) 1,2 and 3

8. Which of the following are true?

Ans: a) 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?

Ans: b) 2

10. For which of the following tasks might clustering be a suitable approach?

Ans: b) Given a database of information about your users, automatically group them into different market segments.

11. Given, six points with the following attributes:

Ans: a)

12. Given, six points with the following attributes:

Ans: b)

13. What is the importance of clustering?

Ans: Cluster analysis groups objects (observations, events) based on the information found in the data describing the objects or their relationships. The goal is that the objects in a group will be similar (or related) to one other and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group, and the greater the difference between groups, the ―better‖ or more distinct the clustering.

14. How can I improve my clustering performance?

Ans: K-means [clustering algorithm](https://www.sciencedirect.com/topics/computer-science/clustering-algorithm) can be significantly improved by using a better initialization technique, and by repeating (re-starting) the algorithm.

•When the data has overlapping clusters, k-means can improve the results of the initialization technique.

•When the data has well [separated clusters](https://www.sciencedirect.com/topics/computer-science/separated-cluster), the performance of k-means depends completely on the goodness of the initialization.

•Initialization using simple furthest point heuristic (Maxmin) reduces the clustering error of k-means from 15% to 6%, on average.

**Statistics**

1. Which of the following is the correct formula for total variation?

Ans: b) Total Variation = Residual Variation + Regression Variation

2. Collection of exchangeable binary outcomes for the same covariate data are called outcomes.

Ans: c) Binomial

3. How many outcomes are possible with Bernoulli trial?

Ans: a) 2

4. If Ho is true and we reject it is called

Ans: a) Type 1 Error

5. Level of significance is also called:

Ans: c) Level of Confidence

6. The chance of rejecting a true hypothesis decreases when sample size is:

Ans: b) Increase

7. Which of the following testing is concerned with making decisions using data?

Ans: b) Hypothesis

8. What is the purpose of multiple testing in statistical inference?

Ans: d) All of the mentioned

9. Normalized data are centred at and have units equal to standard deviations of the original data

Ans: 1

10. What Is Bayes' Theorem?

Ans: Bayes’ theorem describes the probability of occurrence of an event related to any condition. It is also considered for the case of [conditional probability](https://byjus.com/maths/conditional-probability-and-conditional-probability-examples/). Bayes theorem is also known as the formula for the probability of “causes”.

11. What is z-score?

Ans:  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.

12. What is t-test?

Ans: T-tests are statistical [hypothesis tests](https://statisticsbyjim.com/glossary/hypothesis-tests/) that you use to analyze one or two [sample](https://statisticsbyjim.com/glossary/sample/) means. Depending on the t-test that you use, you can compare a sample [mean](https://statisticsbyjim.com/glossary/mean/) to a hypothesized value, the means of two independent samples, or the difference between paired samples. he term “t-test” refers to the fact that these hypothesis tests use t-values to evaluate your sample data. T-values are a type of test statistic. [Hypothesis tests](https://statisticsbyjim.com/glossary/hypothesis-tests/) use the test statistic that is calculated from your sample to compare your sample to the [null hypothesis](https://statisticsbyjim.com/glossary/null-hypothesis/). If the test statistic is extreme enough, this indicates that your data are so incompatible with the null hypothesis that you can reject the null.

13. What is percentile?

Ans: A percentile (or a centile) is a score below which a given percentage of scores in its [frequency distribution](https://en.wikipedia.org/wiki/Frequency_distribution) falls (exclusive definition) or a score at or below which a given percentage falls (inclusive definition). For example, the 50th percentile (the [median](https://en.wikipedia.org/wiki/Median)) is the score below which (exclusive) or at or below which (inclusive) 50% of the scores in the distribution may be found.

14. What is ANOVA?

Ans: ANOVA is a statistical method that stands for analysis of variance. Before the use of ANOVA, the t-test and z-test were commonly used. But the problem with the T-test is that it cannot be applied for more than two groups. This test is also called the Fisher analysis of variance, which is used to do the analysis of variance between and within the groups whenever the groups are more than two. In its simplest form, ANOVA provides a [statistical test](https://en.wikipedia.org/wiki/Statistical_test) of whether two or more population [means](https://en.wikipedia.org/wiki/Mean) are equal,

15. How can ANOVA help?

Ans: The one-way ANOVA can help you know whether or not there are significant differences between the means of your independent variables (such as: age, sex, income). When you understand how each independent variable’s mean is different from the others, you can begin to understand which of them has a connection to your dependent variable for example: buying behaviour, and begin to learn what is driving that behaviour.