Machine learning

answers

1. 4
2. 1,2 and 4
3. Formulating the clustering problem
4. Euclidean distance
5. Divisine clustering
6. All answer are correct
7. Divide the data point into group
8. Unsupervised learning
9. k-means clustering
10. K- means clustering alogrithum
11. All of the above
12. Labeled data
13. It is calculated by measuing the distance between each data point and its squaring across the distance, and summing these squares across one cluster

14. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set.

15.In this techinuque which group the unlabbled data set. Cluster analysis is data analysis technique .cluster analysis dosen’t need to group data points into any predefined group which means that is a unsupervised learning

Types of Clustering

Centroid-based Clustering.

Density-based Clustering.

Distribution-based Clustering.

Hierarchical Clustering.

Statistics worsheet

1. True
2. Central limit theorem
3. Moudeling bounded count data
4. All the mentioned
5. Poisson
6. False
7. Hypothesis
8. 0

9. None of the mentioned
10. Normal distribution – normal distribution is a type of continous distributionin which most data point cluster toward the middle of the range.the middle of the range is also known as the mean of the distribution

11.Handling missing values is the deletion of the rows and column having null values .if any column have more than half of the value as null then you can drop the entire column In the same way, rows can also be dropped if having one or more columns values as null.

And the imputation techniques The simplest imputation method is replacing missing values with the mean or median values of the dataset at large, or some similar summary statistic.

12. A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) although the concept can be also extended to multiple variants of the same variable.

13. The process of replacing null values in a data collection with the data's mean is known as mean imputation.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable.

15. There are three real branches of statistics: data collection, descriptive statistics and inferential statistics.