Worksheet_set_3

Machine Learning

1. d. All of the above
2. d. None
3. c. Reinforcement learning and Unsupervised learning
4. b. The tree representing how close the data points are to each other
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. b. Given a database of information about your users, automatically group them into different market segments.
11. Option (B) depicts the use of MIN or Single link proximity function
12. Option (B) depicts the use of MAX or Complete link proximity function
13. Clustering is a way that classifies the raw data reasonably and searches the hidden patterns that may exist in datasets. Cluster analysis is done to classify objects into groups where objects in one group are more similar to each other and different from objects in other groups. It is normally used for exploratory data analysis and as a method of discovery by solving classification issues. Clustering helps to simplify processing of large datasets.
14. 0

Statistics

1. b) Total Variation = Residual Variation + Regression Variation
2. c) binomial
3. a) 2
4. a) Type-I error
5. b) Size of the test
6. a) Decrease
7. b) Hypothesis
8. d) All of the mentioned
9. a) 0
10. 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. A z score is simply defined as the number of standard deviation from the mean. The z-score can be calculated by subtracting mean by test value and dividing it by standard value. Z-score indicates how much a given value differs from the standard deviation.
12. The t-test is a test that is mainly used to compare the mean of two groups of samples. It is meant for evaluating whether the means of the two sets of data are statistically significantly different from each other.
13. Percentile means: A value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. A percentile score of 95 is a score equal to or better than 95 percent of the scores.
14. Analysis of variance or ANOVA: is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among means. ANOVA was developed by the statistician Ronald Fisher.
15. 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. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources.