MACHINE LEARNING ASSIGNMENT 3

Q1 TO 12 OBJECTIVE ANSWERS

1. A) Biological network analysis
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 neighbor is same as k-means
7. D) 1,2 and 3[single-link, complete-link, average-link]
8. B) 2 only [clustering analysis is negatively affected by heteroscedasticity]
9. A) 2 [two clusters are formed because the number of vertical lines intersecting the red horizontal line at y=2 in the dendrogram are 2.]
10. B) Given a database of information about your users, automatically group them into different market segments.
11. A)
12. B)

Q 13 TO 14 SUBJECTIVE ANSWERS

13. Clustering is a powerful machine learning method involving data point grouping. It is an unsupervised learning method and is a popular technique among data scientists to obtain statistical data analysis in data science.
Importance of clustering:

- Relatively fast and efficient.
- They analyze the groups each data point falls into when applying clustering algorithms.
- They are essential for data scientists to discover innate groupings among unlabeled and labeled data presets.
- Only has one parameter to tune and can easily see the direct impact of adjusting the value of parameter k.
- It comes down to individual preferences, requirements, and what a data scientist utilizes to fulfill their need.
- Easy to implement.
- There are different types of clustering method which help them to approach different titles; Hierarchical clustering, Exclusive clustering, Overlapping clustering.

14. For improving the clustering performance by merging neighboring clusters if the resulting cluster's variance is below the threshold isolating elements that are "far" if a cluster's variance is above the threshold or moving some elements between neighboring clusters if it decreases the sum of squared errors. There is an improved method used, k-means++, which can help to improve the performance. There are various methods proposed for improving performance of the k-means clustering algorithm. For example PCA is an important approach to unsupervised dimensionality reduction technique.

THANK YOU