Answers for MACHINE LEARNING

(CLUSTERING) Worksheet-2

1. d. 2 and 3
2. e. 1, 2 and 4
3. a. True
4. c. 1 and 2
5. b. 1
6. b. No
7. a. Yes
8. d. All of the above
9. a. 1 only
10. a. K-means clustering algorithm
11. f. All of the above
12. e. All of the above
13. Yes, K-means algorithm is sensitive to outliers, as in the step of the algorithm while updating the centroid value involves taking mean of all the data points assigned to that centroid, which further depends on the distribution of the data points, therefore, the presence of outliers/extreme data points may distort the actual mean and the centroid may gravitate towards the outlier leading to wrong cluster formation.

14. K-means algorithm can be a better choice due to following reasons:
    - Easy implementation for quick cluster analysis.
    - Less complex than the other clustering algorithms which translates to better interpretation of the results.
    - Very efficient algorithm in terms of computational cost.
    - This algorithm can be applied as a pre-clustering solution for forming initial clusters, which can then be analysed further by using other clustering algorithms.

15. K-means algorithm is basically non-deterministic algorithm due to its nature of random selection of the centroid as one of the data points. This method of randomly initializing the centroid may lead to different cluster formation.