

## MACHINE LEARNING WORKSHEET\_SET3

1)D

2)D

3)D

4)B

5)D

6)C

7)D

8)A

9)A

10B

11)A

12)B

13) What is the importance of clustering?

Ans: Clustering methods (like Hierarchical method, Partitioning, Density-based method, Model-based clustering, and Grid-based model) help in grouping the data points into clusters, using the different techniques are used to pick the appropriate result for the problem, these clustering techniques helps in grouping the data points into similar categories, and each of these subcategories is further divided into subcategories to assist the exploration of the queries output.

### **Clustering Methods?**

This clustering method helps grouping valuable data into clusters and picks appropriate results based on different techniques. For example, in information retrieval, the results of the query are grouped into small clusters, and each cluster has irrelevant results. By Clustering techniques, they are

grouped into similar categories, and each category is subdivided into sub-categories to assist in the exploration of queries output.

There are various [types of clustering](#) methods; they are

- Hierarchical methods
- Partitioning methods
- Density-based
- Model-based clustering
- Grid-based model

14) How can I improve my clustering performance?

Ans: From the new cluster processing pipeline presented in this study, the three main methodological findings are:

- Graph-based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step.
- Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.
- Surprisingly for some cases, high clustering performance can be achieved by simply performing K-means clustering on the ICA components after PCA dimension reduction on the input data. However, the number of PCA and ICA signals/components needs to be limited to the number of unique classes.

The main clustering results from the new processing pipeline compared to other clustering studies are:

- Compared to state-of-the-art non-deep learning clustering methods, ICA BSS and/or UFL with graph-based clustering algorithms outperformed all other methods.
- With respect to deep learning-based clustering algorithms, our new pipeline obtained the following rankings on the six different datasets: (1) COIL20, 2nd out of 5; (2) COIL100, 2nd out of 5; (3) CMU-PIE, 2nd out of 5; (4) USPS, 3rd out of 9; (5) MNIST, 8th out of 15; and (6) REUTERS-10K, 4th out of 5.

