

MACHINE LEARNING ASSIGNMENT

Q.1. d

Q.2. d

Q.3. c

Q.4. b

Q.5. d

Q.6. c

Q.7. d

Q.8. a

Q.9. a

Q.10. b

Q.11. a

Q.12. b

Q.13. What is the importance of Clustering?

Ans. Clustering is the task of dividing the population or data points into number of groups such that data points in the same groups are most similar to other data points in the same group and dissimilar to the data points in other groups. It is basically the collection of objects on the basis of similarity and dissimilarity between them.

Clustering is very much important as it determines the intrinsic grouping among the unlabelled data present. There are no criteria for good clustering. It depends on the user, what is the criteria they may use which satisfy their need.

Importance of Clustering:

1. Having clustering methods helps in restarting the local search procedure and remove the inefficiency. In addition, clustering helps to determine the internal structure of the data.
2. This clustering analysis has been used for model analysis, vector region of attraction.
3. Clustering helps in understanding the natural grouping in a dataset. Their purpose is to make sense to partition the data into some group of logical groupings.
4. Clustering quality depends on the methods and the identification of the hidden patterns.
5. They play a wide role in applications like marketing, economic research and weblogs to identify similarity measures, image processing and spatial research.
6. They are used in outliers' detection to detect credit card fraudulence.

Q.14. How can I improve my clustering performance?

Ans.: There are three main methodologies to improve the clustering performance which are as below

1. Graph based clustering performance can easily be improved by applying ICA blind source separation during the graph Laplacian embedding step.
2. Applying unsupervised feature learning to input data using either RICA or SFT, improves clustering performance.
3. 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 or components needs to be limited to the number of unique classes.