Machine Learning Assignment-3 Answers :-

Ans.1 :- All of the above

Ans.2 :- None

Ans.3 :- Reinforcement learning and Unsupervised learning

Ans.4 :- The tree representing how close the data points are to each other

Ans.5 :- None

Ans.6 :- k-nearest neighbour is same as k-means

Ans.7 :- 1, 2 and 3

Ans.8 :- 1 only

Ans.9 :- 2

Ans.10 :- Given a database of information about your users, automatically group them into different market segments.

Ans.11 :- a

Ans.12 :- b

Ans.13 :-

Importance of clustering :-

- 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.

- This clustering analysis has been used for model analysis, vector region of attraction.

- 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.

- Clustering quality depends on the methods and the identification of hidden patterns.

- They play a wide role in applications like marketing economic research and weblogs to identify similarity measures, Image processing, and spatial research.

- They are used in outlier detections to detect credit card fraudulence.

Ans.14 :-

How to improve my clustering performance :-

- Improving clustering performance using independent component analysis and unsupervised feature learning.

- Clustering was performed on six benchmark datasets, consisting of five image datasets used in object, face, digit recognition tasks (COIL20, COIL100, CMU-PIE, USPS, and MNIST) and one text document dataset (REUTERS-10K) used in topic recognition.

- K-means, spectral clustering, Graph Regularized Non-negative Matrix Factorization, and K-means with principal components analysis algorithms were used for clustering.

- For each clustering algorithm, blind source separation (BSS) using Independent Component Analysis (ICA) was applied. Unsupervised feature learning (UFL) using reconstruction cost ICA (RICA) and sparse filtering (SFT) was also performed for feature extraction prior to the cluster algorithms.

- Clustering performance was assessed using the normalized mutual information and unsupervised clustering accuracy metrics.