**Machine Learning Worksheet-I**

1. B)
2. A)
3. D)
4. A)
5. B)
6. D)
7. A)
8. B)
9. A)
10. A)
11. D)
12. A)
13. In cluster analysis, we will initially focus on clustering procedures that result in the assignment of each subject to one, and only one, class. Subjects within a class are usually assumed to be indistinguishable from one another. Thus, we assume that the underlying structure of the data involves an unordered set of discrete classes. In some cases we may also view these classes as hierarchical in nature, with some classes divided into subclasses.  Given that no information on group definition is formally evaluated in advance, the major problems of cluster analysis will be discussed as follows:
14. What measure of inter-subject similarity is to be used and how is each variable to be “weighted” in the construction of such a summary measure?
15. After inter-subject similarities are obtained, how are the classes to be formed?
16. After the classes have been formed, what summary measures of each cluster are appropriate in a descriptive sense; that is, how are the clusters to be defined?
17. Assuming that adequate descriptions of the clusters can be obtained, what inferences can be drawn regarding their statistical significance?
18. Clustering quality means evaluation of goodness of clustering results. There are 3 categorization of measures: External, Internal and Relative.
19. External: Supervised, employ criteria not inherent to the dataset. Here, we compare a clustering against prior or expert-specified knowledge
20. Internal: Unsupervised, criteria derived from data itself. Here, we evaluate the goodness of a clustering by considering ow well the clusters are separated and how compact the clusters are.
21. Relative: Here, we directly compare different clusterings, usually those obtained via different parameter settings for the same algorithm.
22. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). Cluster analysis can be a powerful data-mining tool for any organisation that needs to identify discrete groups of customers, sales transactions, or other types of behaviors and things.

Following are the different types of cluster analysis:

* Partitioning methods
* Hierarchical clustering
* Fuzzy clustering
* Density-based clustering
* Model-based clustering