**MACHINE-LEARNING-WORKSHEET-1**

Solution:1 – B

Solution:2 – E

Solution:3 – D

Solution:4 – A

Solution:5 – B

Solution:6– D

Solution:7 – A

Solution:8 – B

Solution:9 – D

Solution:10 – A

Solution:11 – D

Solution:12 – A

**Solution:13** – Cluster Analysis is an analysis that tries to identify structures within the data. The cluster analysis Follows three basic steps: 1) calculate the distances, 2) link the clusters, and 3) choose a solution by selecting the right number of clusters.

**Solution:14** – The quality of a cluster ca be measured by using the CrossValidationScore of every cluster. Higher quality cluster consist high score and low quality cluster consist lower score.

**Solution:15** – Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

There are 4 types of clustering analysis

1. **Centroid Clustering**: This is one of the more common methodologies used in cluster analysis. In centroid cluster analysis you choose the number of clusters that you want to classify. For example, if you’re a pet store owner you may choose to segment your customer list by people who bought dog and/or cat products.
2. **Density Clustering**: Density clustering groups data points by how densely populated they are. To group closely related data points, this algorithm leverages the understanding that the more dense the data points...the more related they are. To determine this, the algorithm will select a random point then start measuring the distance between each point around it. For most density algorithms a predetermined distance between data points is selected to benchmark how closely points need to be to one another to be considered related.. Then, the algorithm will identify all other points that are within the allowed distance of relevance. This process will continue to iterate by selecting different random data points to start with until the best clusters can be identified.
3. **Distribution Clustering**: Distribution clustering identifies the probability that a point belongs to a cluster. Around each possible centroid The algorithm defines the density distributions for each cluster, quantifying the probability of belonging based on those distributions The algorithm optimizes the characteristics of the distributions to best represent the data.
4. **Connectivity Clustering**: Unlike the other three techniques of clustering analysis reviewed above, connectivity clustering initially recognizes each data point as its own cluster. The primary premise of this technique is that points closer to each other are more related. The iterative process of this algorithm is to continually incorporate a data point or group of data points with other data points and/or groups until all points are engulfed into one big cluster. The critical input for this type of algorithm is determining where to stop the grouping from getting bigger.