

# Machine Learning Assignment 1 – Answers

Q. 1

Ans – b) 4

Q. 2

Ans - d) 1, 2 and 4

Q. 3

Ans - d) formulating the clustering problem

Q. 4

Ans - a) Euclidean distance

Q. 5

Ans - b) Divisive clustering

Q. 6

Ans - d) All answers are correct

Q. 7

Ans - d) All of the above

Q. 8

Ans - b) Unsupervised learning

Q. 9

Ans - a) K- Means clustering

**Q. 10**

**Ans - a) K-means clustering algorithm**

**Q. 11**

**Ans - d) All of the above**

**Q. 12**

**Ans - a) Labeled data**

**Q. 13**

**Ans -** For instance, by varying k from 1 to 10 **clusters**. For each k, **calculate** the total within-**cluster** sum of square (wss). Plot the curve of wss according to the number of **clusters** k. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of **clusters**.

**Q. 14**

**Ans -** To measure a cluster's fitness within a clustering, we can compute the average silhouette coefficient value of all objects in the cluster. To measure the quality of a clustering, we can use the average silhouette coefficient value of all objects in the data set. The silhouette coefficient and other intrinsic measures can also be used in the elbow method to heuristically derive the number of clusters in a data set by replacing the sum of within-cluster variances.

**Q.15**

**Ans -** Cluster analysis is the task of grouping a set of data points in such a way that they can be characterized by their relevancy to one another. These techniques create clusters that allow us to understand how our data is related. The most common applications of cluster analysis in a business setting is to segment customers or activities.

## Types of clustering -

1. Centroid Clustering
2. Density Clustering
3. Distribution Clustering
4. Connectivity Clustering