**Himanshu Kumar**

**Assignment Clustering**

Q.1- What is Clustering/ Segmentation?

**ANS:-** Clustering algorithm is an unsupervised machine learning algorithm that discovers groups of data points that are closely related. The fundamental difference between supervised and unsupervised algorithm is that:

* **supervised algorithm:** it requires partitioning the dataset into train and test set, and the algorithm learned based on the output/label of the train dataset and generalize it to unobserved data. For instance, decision tree, regression, neural networks.
* **unsupervised algorithm:** it is used to discover hidden patterns when there isn’t any defined output/label from the dataset. For instance, clustering, association rule mining, dimension reduction

Q.2- How do you determine the "nearness" of clusters?

**ANS :-**

1. intercluster distance = minimum of the distances between any two points, one from each cluster.

(2) Pick a notion of “cohesion” of clusters, e.g., maximum distance from the clustroid.

Merge clusters whose union is most cohesive.

Q.3- What is elbow method in K means Clustering & what is it used for?

**ANS:-** The elbow method runs k-means clustering on the dataset for a range of values for k (say from 1-10) and then for each value of k computes an average score for all clusters. By default, the distortion score is computed, the sum of square distances from each point to its assigned center.

The **Elbow Method** is one of the most popular methods to determine this optimal value of k.

Q.4- Discuss any 2 applications of Clustering?

### ANS:- ****Social Network Analysis (SNA)****

 It is the process of examining qualitative and quantitative social structures by utilizing Graph Theory (a major [branch of discrete mathematics](https://www.analyticssteps.com/blogs/7-major-branches-discrete-mathematics)) and networks. Here the mapping of social networks structure is arranged in terms of nodes (individual personality, people, or other entity inside the network) and the edges or links (relationships, interaction, or communication) that connect them.

Clustering methods are required in such analysis in order to map and measure the relationship and conflicts amid people, groups, companies, compute networks, and other similar connected information/knowledge entities.

### ****Search Result Clustering****

 You must have encountered similar results obtained while searching something particular at Google, these results are a mixture of the similar matches of your original query.

Basically, this is the result of clustering, it makes groups of similar objects in a single cluster and renders to you, i.e provides results of searched data in terms with most closely related objects that are clustered across the data to be searched.

**Better the clustering algorithm deployed, more the possibilities of achieving required outcomes of the leading desk.**

Therefore, the concept of similar objects serves as a backbone in getting searched results. Even though, most of the parameters are taken into consideration for defining the portrait of similar objects.

Check the image below while typing **“search engine” at Google**, we get the more keywords of a similar search, for example, search engine list, top 50 search engine, etc.

Q.5- Which of the following is required by K-means clustering?

a) defined distance metric

b) number of clusters

c) initial guess as to cluster centroids

d) all of the mentioned

**ANS:- (D). all of the mentioned**

**Explanation: k-means clustering follows partitioning approach.**

Q.-6 Point out the wrong statement.

a) k-means clustering is a method of vector quantization

b) k-means clustering aims to partition n observations into k clusters

c) k-nearest neighbor is same as k-means

d) none of the mentioned

**ANS:- (C) k-nearest neighbor is same as k-means**

**Explanation: k-nearest neighbor has nothing to do with k-means.**

Q.7- Which of the following function is used for k-means clustering?

a) k-means

b) k-mean

c) heatmap

d) none of the mentioned

**ANS:- A. k-means**

**Explanation: k-means requires a number of clusters.**

Q.8- Clustering is a…

a) Supervised learning

b) Un Supervised learning

c) Reinforcement learning

d) None

**ANS:- (B) Un Supervised learning**

Q.9- For Clustering we don’t require

a) Labeled data

b) Unlabeled data

c) Numerical data

d) Categorical data

**ANS:- (A) Labeled data**

Q.10- Which of the step is not required for K-means clustering?

a) Distance Metrics

b) Initial no of clusters

c)Initial guess as to no of centroids

d) None

**ANS:- (D) None**