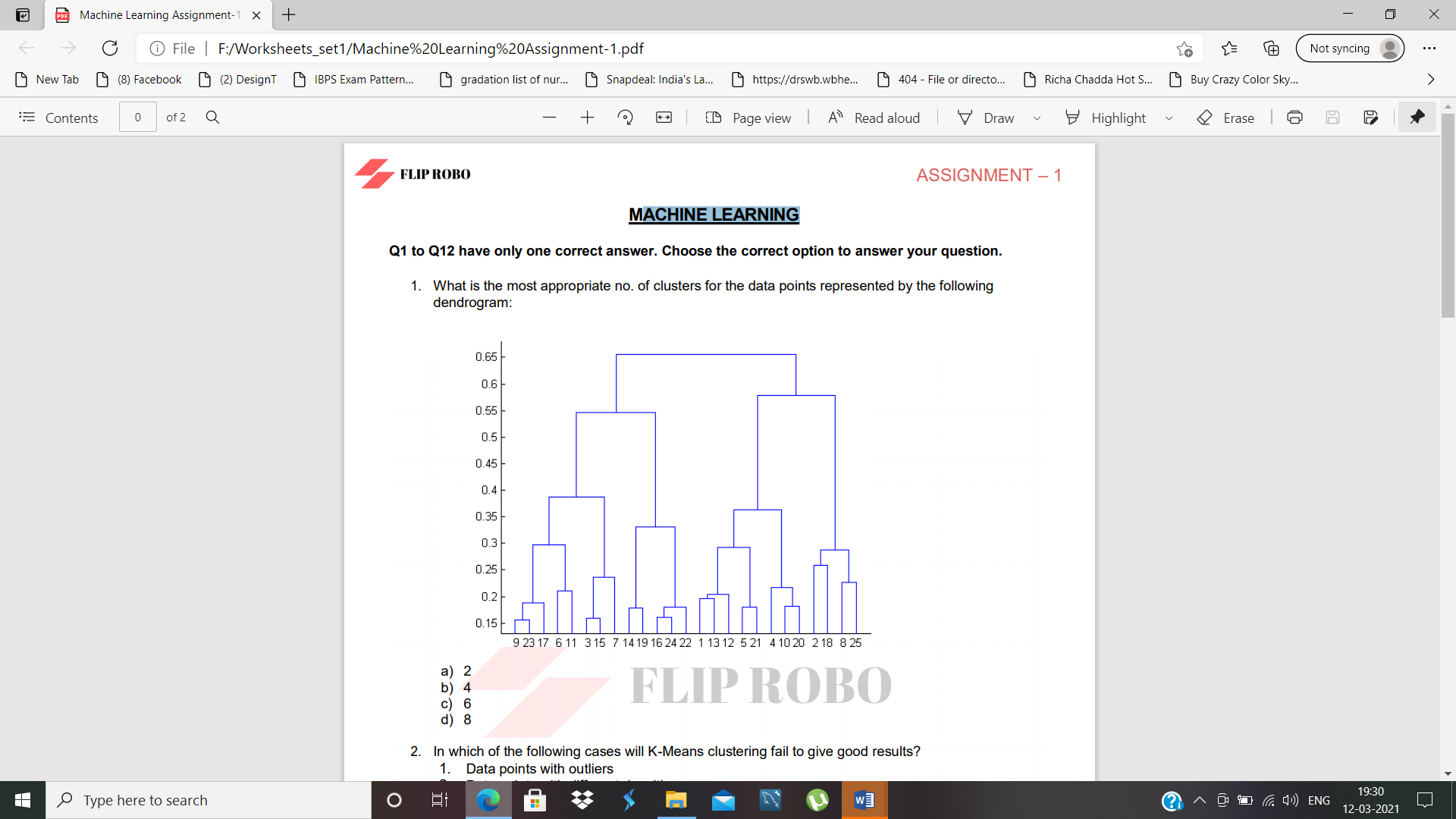
MACHINE LEARNING

Q1 to Q12 have only one correct answer. Choose the correct option to answer your question

. 1. What is the most appropriate no. of clusters for the data points represented by the following dendrogram:



Ans) 4.

2. In which of the following cases will K-Means clustering fail to give good results?

1. Data points with outliers

2. Data points with different densities

3. Data points with round shapes

4. Data points with non-convex shapes

Ans) d) 1, 2 and 4

3. The most important part of is selecting the variables on which clustering is based.

Ans) d) formulating the clustering problem

4. The most commonly used measure of similarity is the or its square.

Ans) a) Euclidean distance

5. is a clustering procedure where all objects start out in one giant cluster. Clusters are formed by dividing this cluster into smaller and smaller cluster.

Ans) b) Divisive clustering

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

Ans) d) All answers are correct

7. The goal of clustering is to

Ans) d) All of the above

8. Clustering is a

Ans) b) Unsupervised learning

9. Which of the following clustering algorithms suffers from the problem of convergence at local optima?

Ans) a) K- Means clustering

10. Which version of the clustering algorithm is most sensitive to outliers?

Ans) a) K-means clustering algorithm

11. Which of the following is a bad characteristic of a dataset for clustering analysis

Ans) a) Data points with outliers

12. For clustering, we do not require

Ans) a) Labeled data

Q13 to Q15 are subjective answers type questions, Answers them in their own words briefly.

13. How is cluster analysis calculated?

Ans) The hierarchical 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.

14. How is cluster quality measured?

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.

15. What is cluster analysis and its types?

Ans) Clustering is a type of unsupervised learning method of machine learning. Clustering is a task of dividing the data sets into a certain number of clusters in such a manner that the data points belonging to a cluster have similar characteristics. Clusters are nothing but the grouping of data points such that the distance between the data points within the clusters is minimal.

The various types of clustering are:

* Connectivity-based Clustering (Hierarchical clustering)
* Centroids-based Clustering (Partitioning methods)
* Distribution-based Clustering.
* Density-based Clustering (Model-based methods)
* Fuzzy Clustering.
* Constraint-based (Supervised Clustering)