Among the following problems, a clustering algorithm is mostly appropriate for:

(A) Predicting the rainfall in HYD in 2022.

(B) For a robot to decide the direction to travel to Himalaya 105 class room.

(C) **[Ans]** To detect Credit Card transactions that are Frauds

(D) To translate a sentence from Hindo to Telugu

(E) All the above

Consider a Divisive Clustering Algorithm with two steps:

Create an MST

Successively remove the longest or largest edges.

Assume there are 100 samples, and all edges are of unique length. If we have removed 5 edges, the number of clusters is:

(A) 5

(B) 2^5

(C) 5^2

(D) **[Ans]** 6

(E) None of the above.

Consider a Divisive Clustering Algorithm with two steps:

Create an MST

Successively remove the longest or largest edges.

Assume there are 100 samples, and all edges are of unique length.

(A) **[Ans]** This algorithm is yielding a globally optimal solution to a specific objective.

(B) Every run of this algorithm can give different solution and therefore, this is sensitive to the ordering/indices of the samples in the set.

(C) Since this is a global optima, there can not exist a better clustering algorithm.

(D) Since there are better/other clustering algorithms, the final solution is only locally optimal.

(E) **[Ans]** The objective function that this algorithm minimizes is the following:

$$\sum_{l} \sum_{x_i, x_j \in C_l} d(x_i, x_j)^2$$

Assume there are N samples in a data set, the number of distinct ways in which we can cluster this set is:

(A) N

(B) 2^N

(C) $N!$

(D) ${}_N C_2$

(E) **[Ans]** None of the above ‘

Consider a Divisive Clustering Algorithm with two steps:

Create an MST

Successively remove the longest or largest edges.

Assume there are 100 samples, and all edges are unique length.

What can be a **bad** termination criteria?

(A) Stop when all the left out edges are less than p

(B) **[Ans]** Stop when there are no more edges to remove.

(C) Stop when the length of the next largest is less than half of the edge removed in the previous step?

(D) **[Ans]** Stop when the length of the next largest is more than half of the edge removed in the previous step?

(E) **[Ans]** Average length of leftout edges is more than average length of removed edges.