1. What is your definition of clustering? What are a few clustering algorithms you might think of?

What is clustering and its type?

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.

Two different types

There are **two** different types of clustering, which are hierarchical and non- hierarchical methods.

1. What are some of the most popular clustering algorithm applications?

Here, the cluster center i.e. centroid is formed such that the distance of data points is minimum with the center. This problem is basically one of the NP-Hard problems and thus solutions are commonly approximated over a number of trials. For  **Ex- K – means algorithm** is one of the popular examples of this algorithm.

1. When using K-Means, describe two strategies for selecting the appropriate number of clusters.

The optimal number of clusters can be defined as follow: **Compute clustering algorithm (e.g., k-means clustering) for different values of k**. For instance, by varying k from 1 to 10 clusters. For each k, calculate the total within-cluster sum of square (wss).

1. What is mark propagation and how does it work? Why would you do it, and how would you do it?

* While designing a Neural Network, in the beginning, we initialize weights with some random values or any variable for that fact.
* Now obviously, we are not *superhuman.*So, it’s not necessary that whatever weight values we have selected will be correct, or it fits our model the best.
* Okay, fine, we have selected some weight values in the beginning, but our model output is way different than our actual output i.e. the error value is huge.
* Now, how will you reduce the error?
* Basically, what we need to do, we need to somehow explain the model to change the parameters (weights), such that error becomes minimum.

5. Provide two examples of clustering algorithms that can handle large datasets. And two that look

for high-density areas?

6. Can you think of a scenario in which constructive learning will be advantageous? How can you go

about putting it into action?

The Constructivist process works best when **children can compare and share their ideas with others**. At home, this can be a parent, a sibling or even a playmate. Discussions with others can provide learners with the opportunities to express what they have learned and even correct mistakes that may take place.

7. How do you tell the difference between anomaly and novelty detection?

Novelty detection is when you have new data (i.e. OOD) and you want to know whether or not it is in-distribution. You want to know if it looks like the data you trained on. Anomaly detection is when you test your data to see if it is different than what you trained the model.

8. What is a Gaussian mixture, and how does it work? What are some of the things you can do about

it?

Gaussian mixture models (GMMs) are **a type of machine learning algorithm**. They are used to classify data into different categories based on the probability distribution. Gaussian mixture models can be used in many different areas, including finance, marketing and so much more!

Gaussian Mixture Models (GMMs) assume that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Hence, **a Gaussian Mixture Model tends to group the data points belonging to a single distribution together**.

9. When using a Gaussian mixture model, can you name two techniques for determining the correct

number of clusters?

This score, as clearly stated by the SKLearn developers, consider two measures: **The mean distance between a sample and all other points in the same cluster**. The mean distance between a sample and all other points in the next nearest cluster .