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

K-means clustering is the most commonly used clustering algorithm. It's a centroid-based algorithm and the simplest unsupervised learning algorithm. This algorithm tries to minimize the variance of data points within a cluster. It's also how most people are introduced to unsupervised machine learning.

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

* K-Means Algorithm. The most commonly used algorithm, K-means clustering, is a centroid-based algorithm.
* Mean-Shift Algorithm.
* DBSCAN Algorithm.
* Expectation-Maximization Clustering using Gaussian Mixture Models.
* Agglomerative Hierarchical Algorithm.

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

K-means clustering is an unsupervised algorithm. In an unsupervised algorithm, we are not interested in making predictions (since we don’t have a target/output variable). The objective is to discover interesting patterns in the data, e.g., are there any subgroups or ‘clusters’ among the bank’s customers.

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

The algorithm is used to effectively train a neural network through a method called chain rule. In simple terms, after each forward pass through a network, backpropagation performs a backward pass while adjusting the model’s parameters (weights and biases).

5. Provide two examples of clustering algorithms that can handle large datasets. And two that look for high-density areas?

What are the two main types of clustering methods?

There are two different types of clustering, which are hierarchical and non-hierarchical methods. Non-hierarchical Clustering In this method, the dataset containing N objects is divided into M clusters. In business intelligence, the most widely used non-hierarchical clustering technique is K-means.

6. Can you think of a scenario in which constructive learning will be advantageous? How can you go about putting it into action?

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

In "novelty detection", you have a data set that contains only good data, and you're trying to determine whether new observations fit within the existing data set. In "outlier detection", the data may contain outliers, which you want to identify

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!

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

The two most popular evaluation metrics for picking cluster numbers for fitting Gaussian Mixture models are BIC and AIC. BIC stands for Bayesian information criterion and AIC stands for Akaike information criterion.