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

Using a clustering algorithm means you're going to give the algorithm a lot of input data with no labels and let it find any groupings in the data it can. Those groupings are called clusters. A cluster is a group of data points that are similar to each other based on their relation to surrounding data points

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

Affinity Propagation.

Agglomerative Hierarchical Clustering.

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Gaussian Mixture Models (GMM)

K-Means.

Mean Shift Clustering.

3. 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).

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

Backpropagation, or backward propagation of errors, is an algorithm that is designed to test for errors working back from output nodes to input nodes. It is an important mathematical tool for improving the accuracy of predictions in data mining and machine learning.

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

for high-density areas?

DBSCAN

Hierarchical

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

about putting it into action?

In Constructivist learning, the traditional classroom learning procedure is flipped. Instead of a teacher informing a child about a subject and constructing a meaning for them, in constructivist learning, children construct their own meanings. Now, this may not be the best approach for every learning environment, but it can be tremendously helpful at home. The home environment is where children primarily develop a sense of self and is the perfect setting for children to practice taking initiative in their learning via a Constructivist approach.

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

Both novelty detection and outlier detection are used to detect anomalies. Outlier detection is an unsupervised anomaly detection algorithm. Novelty detection is a semi-supervised anomaly detection algorithm.

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

it?

A Gaussian Mixture is a function that is comprised of several Gaussians, each identified by k ∈ {1,…, K}, where K is the number of clusters of our dataset. Each Gaussian k in the mixture is comprised of the following parameters: A mean μ that defines its centre. A covariance Σ that defines its width.

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

number of clusters?

Often, the best way to find an appropriate cluster number is to try different cluster numbers and see which fits your data appropriately. The two most popular evaluation metrics for picking cluster numbers for fitting Gaussian Mixture models are BIC and AIC.