**1. What is the difference between supervised and unsupervised learning? Give some examples to illustrate your point.**

The main difference between supervised and unsupervised learning lies in the presence or absence of labeled data. In supervised learning, the algorithm is trained using labeled examples, where the input data is paired with corresponding desired output. Examples include image classification, spam detection, and sentiment analysis. In unsupervised learning, there are no labeled examples, and the algorithm learns patterns and relationships directly from the input data. Examples include clustering, dimensionality reduction, and anomaly detection.

**2. Mention a few unsupervised learning applications.**

Unsupervised learning has various applications, including:

* Clustering
* Dimensionality reduction
* Anomaly detection
* Recommendation systems
* Generative modeling

**3. What are the three main types of clustering methods? Briefly describe the characteristics of each.**

The three main types of clustering methods are:

* K-means clustering: Divides data into a predefined number of clusters (k) based on minimizing the distance between data points and cluster centroids. It assigns each data point to the nearest centroid, creating compact and spherical clusters.
* Hierarchical clustering: Builds a hierarchy of clusters by either merging or splitting clusters based on their similarity. It can be agglomerative (bottom-up) or divisive (top-down), resulting in a dendrogram that shows the cluster relationships.
* Density-based clustering: Identifies clusters based on the density of data points. It groups together data points that are within a specified distance threshold, forming areas of high density separated by areas of low density. Examples include DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

**4. Explain how the k-means algorithm determines the consistency of clustering.**

The k-means algorithm determines the consistency of clustering by minimizing the within-cluster sum of squared errors (SSE). It iteratively updates the cluster centroids and assigns data points to the nearest centroid, aiming to minimize the distance between data points and their respective centroids. The algorithm converges when the SSE no longer significantly decreases with each iteration.

**5. With a simple illustration, explain the key difference between the k-means and k-medoids algorithms.**

The key difference between the k-means and k-medoids algorithms is the choice of representatives for each cluster. In k-means, the centroid of a cluster represents the cluster, while in k-medoids, an actual data point from the cluster is chosen as the medoid. This difference makes k-medoids more robust to outliers and able to handle non-Euclidean distances.

**6. What is a dendrogram, and how does it work? Explain how to do it.**

A dendrogram is a visual representation of hierarchical clustering results. It illustrates the relationships between data points and clusters in a tree-like structure. The dendrogram is built by merging or splitting clusters based on their similarity, and the vertical axis represents the dissimilarity between clusters. It helps in identifying the optimal number of clusters and understanding the hierarchical structure of the data.

**7. What exactly is SSE? What role does it play in the k-means algorithm?**

SSE stands for Sum of Squared Errors, which is a measure of the within-cluster variability or compactness of clusters in the k-means algorithm. It calculates the squared Euclidean distance between each data point and its assigned centroid within a cluster. The goal of the k-means algorithm is to minimize the SSE, as smaller SSE values indicate more cohesive and well-separated clusters.

**8. With a step-by-step algorithm, explain the k-means procedure.**

The k-means algorithm follows a step-by-step procedure:

* Initialize: Randomly choose k centroids.
* Assign: Assign each data point to the nearest centroid based on distance.
* Update: Recalculate the centroids by taking the mean of data points assigned to each cluster.
* Repeat: Iteratively reassign and update the centroids until convergence or a maximum number of iterations is reached.

**9. In the sense of hierarchical clustering, define the terms single link and complete link.**

In hierarchical clustering, single link and complete link are two linkage criteria used to measure the dissimilarity between clusters:

* Single link (also known as minimum or nearest neighbor): Measures the dissimilarity between clusters by considering the shortest distance between any two points in different clusters. It tends to form long, chain-like clusters.
* Complete link (also known as maximum or farthest neighbor): Measures the dissimilarity between clusters by considering the longest distance between any two points in different clusters. It tends to form compact, spherical clusters.

**10. How does the apriori concept aid in the reduction of measurement overhead in a business basket analysis? Give an example to demonstrate your point.**

The apriori concept helps in reducing measurement overhead in a business basket analysis by focusing on frequent itemsets. It uses the principle that if an itemset is frequent, then all of its subsets must also be frequent. By identifying frequent itemsets and their associations, it enables more efficient analysis and recommendations. For example, in a retail setting, if "bread" and "milk" are frequently purchased together, the retailer can place them close to each other to improve customer convenience.