1)What is the role of feature selection in anomaly detection?

Ans- Feature selection plays an important role when it comes to improve outlier detection in terms of identifying noisy data that contain irrelevant or redundant features. State-of-the-art work either focuses on unsupervised feature selection for data streams or (offline) outlier detection.

2) What are some common evaluation metrics for anomaly detection algorithms and how are they computed?

Ans- Generally, in order to evaluate the quality of an anomaly detection technique, the confusion matrix and its derived metrics such as precision and recall are used. These metrics, however, do not take this temporal dimension into consideration.

3) What is DBSCAN and how does it work for clustering?

Ans- DBSCAN is a density-based clustering algorithm that works on the assumption that clusters are dense regions in space separated by regions of lower density. It groups 'densely grouped' data points into a single cluster.

4) How does the epsilon parameter affect the performance of DBSCAN in detecting anomalies?

Ans- It is the distance that DBSCAN uses to determine if two points are similar and belong together. A larger epsilon will produce broader clusters (encompassing more data points) and a smaller epsilon will build smaller clusters.

5) What are the differences between the core, border, and noise points in DBSCAN, and how do they relate to anomaly detection?

Ans-The Core Points, as the name suggests, lie usually within the interior of a cluster. A Border Point has fewer than MinPts within its ϵ-neighborhood (N), but it lies in the neighborhood of another core point. Noise is any data point that is neither core nor border point.

6) How does DBSCAN detect anomalies and what are the key parameters involved in the process?

Ans- The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm. The main principle of this algorithm is that it finds core samples in a dense area and groups the samples around those core samples to create clusters. The samples in a low-density area become the outliers.

7) What is the make\_circles package in scikit-learn used for?

Ans- Make a large circle containing a smaller circle in 2d. A simple toy dataset to visualize clustering and classification algorithms.

8) What are local outliers and global outliers, and how do they differ from each other?

Ans- There are two general types of outlier detection: global and local. Global outliers fall outside the normal range for an entire dataset, whereas local outliers may fall within the normal range for the entire dataset, but outside the normal range for the surrounding data points.

9) How can local outliers be detected using the Local Outlier Factor (LOF) algorithm?

Ans- The Local Outlier Factor (LOF) algorithm is an unsupervised anomaly detection method which computes the local density deviation of a given data point with respect to its neighbors. It considers as outliers the samples that have a substantially lower density than their neighbors.

10) How can global outliers be detected using the Isolation Forest algorithm?

Ans- Isolation Forest is based on the Decision Tree algorithm. It isolates the outliers by randomly selecting a feature from the given set of features and then randomly selecting a split value between the max and min values of that feature.

11) What are some real-world applications where local outlier detection is more appropriate than global outlier detection, and vice versa?

Ans- Catching and identifying anomalies is what we call anomaly or outlier detection. For example, if large sums of money are spent one after another within one day and it is not your typical behavior, a bank can block your card. They will see an unusual pattern in your daily transactions.