

## **Essay 1**

### **Global Outlier**

A global outlier is a data point that significantly deviates from the rest of the dataset. For example, consider a dataset containing the annual incomes of individuals in a small town. If most incomes range from \$30,000 to \$80,000 but there is one income of \$1,000,000, this would be considered a global outlier. Methods such as the Z-score method can detect global outliers effectively. In this method, the Z-score for each data point is calculated, and points with a Z-score beyond a certain threshold (e.g.,  $|Z| > 3$ ) are classified as outliers.

### **Contextual Outlier**

A contextual outlier, also known as a conditional outlier, is a data point that is an outlier within a specific context but not necessarily in the overall dataset. For instance, in a temperature dataset, a temperature of 25°C might be normal in summer but abnormal in winter. Contextual outliers can be detected using methods such as the Seasonal Hybrid Extreme Studentized Deviate (S-H-ESD) test, which takes into account the context, such as seasonal patterns or time series trends.

### **Collective Outlier**

A collective outlier occurs when a group of data points deviates significantly from the rest of the dataset, even though individual points within the group might not be outliers. For example, a sequence of consecutive zeros in a sensor reading that typically fluctuates between 10 and 20 can be considered a collective outlier. Techniques such as clustering algorithms (e.g., DBSCAN) can identify collective outliers by recognizing unusual patterns or clusters in the data.

## **Essay 2**

### **Global Outlier**

Global outliers are single data points that stand out significantly from the rest. In a dataset of student exam scores ranging from 50 to 90, a score of 20 would be a global outlier. One effective method for detecting global outliers is the interquartile range (IQR) method. This involves calculating the IQR (the range between the first quartile,  $Q1$ , and the third quartile,  $Q3$ ) and identifying data points that lie beyond 1.5 times the IQR above  $Q3$  or below  $Q1$ .

### **Contextual Outlier**

Contextual outliers are data points that deviate significantly within a specific context. For example, a daily water usage of 500 liters may be normal for a household, but the same usage for an office building during the weekend might be a contextual outlier. Contextual outliers can be detected using machine learning techniques like time series anomaly detection algorithms, which consider the temporal context of the data.

### **Collective Outlier**

Collective outliers are groups of data points that together form an anomaly. For example, in network traffic data, a sudden surge of identical requests to a server can indicate a DDoS attack, forming a collective outlier. Methods such as Principal Component Analysis (PCA) can help in detecting collective outliers by transforming the data into a space where the normal patterns are more distinguishable from the anomalous collective patterns.

### **Essay 3**

#### **Global Outlier**

A global outlier is an individual data point that deviates significantly from the rest of the data. In a dataset of house prices in a city where most houses are priced between \$100,000 and \$500,000, a house priced at \$10,000,000 would be a global outlier. Detection methods like the Modified Z-score can be used, where a score greater than 3.5 indicates an outlier.

#### **Contextual Outlier**

A contextual outlier is a data point that is only an outlier when considered within a specific context. For example, in a series of monthly sales figures, a spike in sales during December might be normal due to the holiday season, but the same spike in August would be a contextual outlier. Methods like ARIMA (AutoRegressive Integrated Moving Average) models can be used to detect these outliers by considering the context of the data.

#### **Collective Outlier**

Collective outliers are sets of data points that together form an anomaly. For example, in a time series of stock prices, a sudden, consistent drop over several days could be a collective outlier, indicating an unusual market event. Methods like sequence mining or Hidden Markov Models (HMM) can be used to detect these patterns by analyzing the sequences and identifying deviations from the normal patterns.

### **Essay 4**

#### **Global Outlier**

Global outliers are single points that stand out significantly from the rest of the data. For example, in a dataset of body temperatures where the normal range is 36°C to 37.5°C, a reading of 40°C would be a global outlier. The Grubbs' Test is a statistical test that can be used to detect global outliers by identifying the maximum or minimum value that deviates significantly from the rest of the data.

#### **Contextual Outlier**

A contextual outlier deviates within a specific context but may be normal in another context. For example, a social media post frequency of 50 posts per day might be normal for a news organization but abnormal for an individual user. Contextual outliers can be detected using contextual anomaly detection methods like Bayesian Networks, which consider the context of the data to determine the likelihood of an outlier.

#### **Collective Outlier**

Collective outliers are groups of data points that collectively represent an anomaly. For example, in an e-commerce website's traffic data, a sudden burst of traffic from a single IP address over a short period could indicate a bot attack, forming a collective outlier. Methods such as K-means clustering can help detect these outliers by identifying unusual clusters of data points that deviate from the norm.

## **Essay 5**

### **Global Outlier**

A global outlier is an individual data point that differs significantly from the rest of the dataset. In a dataset of daily step counts where most values range from 5,000 to 10,000 steps, a count of 30,000 steps would be a global outlier. The boxplot method, which visually represents the distribution of the data and highlights points beyond the whiskers, can be used to detect such outliers.

### **Contextual Outlier**

Contextual outliers are data points that are only outliers within a specific context. For instance, in a dataset tracking hourly electricity consumption, a high usage of 10 kWh might be normal during the day but abnormal at midnight. Contextual outliers can be detected using techniques like LOF (Local Outlier Factor), which takes into account the local density of the data points and identifies anomalies based on their context.

### **Collective Outlier**

Collective outliers are sets of data points that together form an unusual pattern. For example, a sequence of unusually high sales figures over several consecutive days in an otherwise stable sales period can indicate a promotional event or a data error, forming a collective outlier. Methods like the Mahalanobis distance can be used to detect collective outliers by measuring the distance between data points in a multidimensional space and identifying unusual clusters.