Anomaly Detection

MACHINE LEARNING

Pakarat Musikawan

What Is Anomaly Detection?

Anomaly detection is the process of identifying unusual patterns, data points, or events that significantly deviate from the **norm** in a dataset.

It is also known as **outlier detection**.

The primary objective is to detect patterns that do not align with the typical distribution or behavior of the data.

Nature of anomaly	Anomalies are infrequent events, occurring significantly less often than normal data points.
Challenges	❖ The large amount of data makes detecting anomalies difficult. ❖ Due to their rarity, it is challenging to establish clear patterns for anomalies.

What Is Anomaly Detection?

Statistical-Based Methods

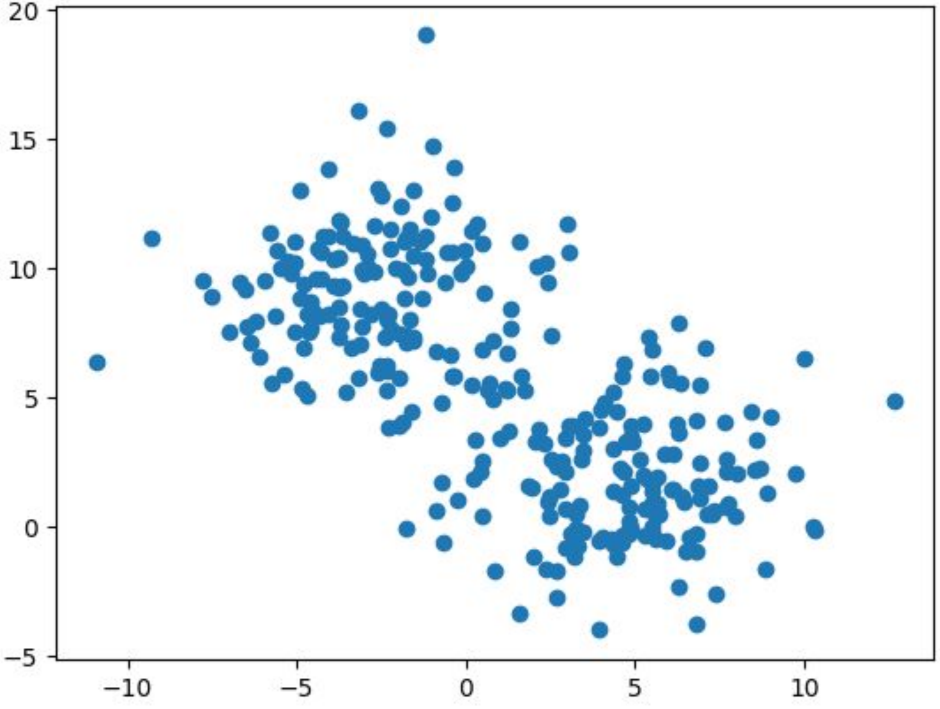
Statistical methods rely on mathematical/statistical models and assumptions about the data distribution to detect anomalies.

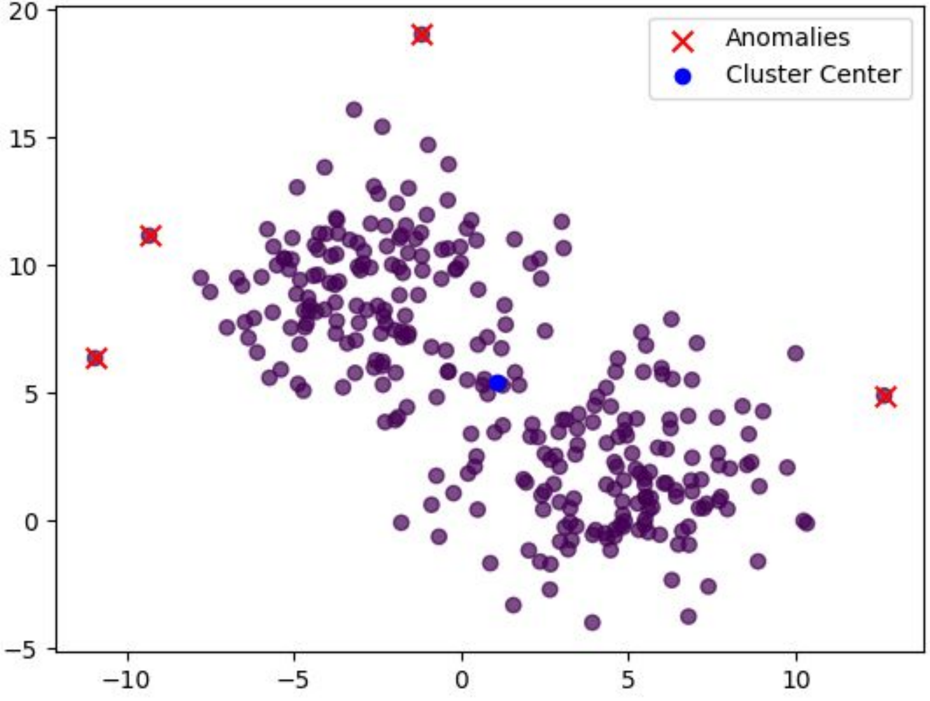
- Z-Score
- Interquartile range (IQR)

Machine learning based methods

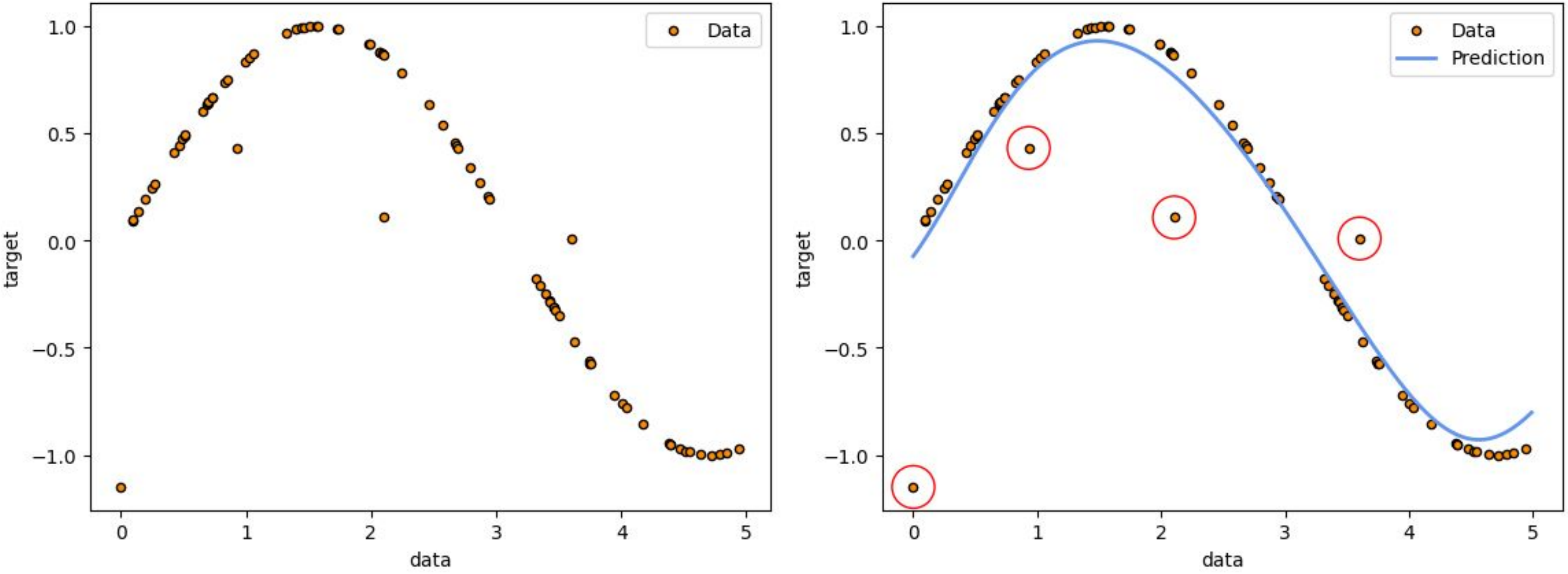
Machine learning approaches learn patterns directly from the data, making them more flexible for complex and high-dimensional datasets.

Anomaly Detection Example





Anomaly Detection Example



What Are the Applications of Anomaly Detection?

Cybersecurity – Network intrusion detection is a key example. An anomaly detection algorithm monitors traffic to establish normal patterns and identifies deviations that may indicate a security breach.

Fraud detection – Commonly used for identifying fraudulent activities, such as unusual credit card transactions.

Social media monitoring – Helps track user activity and engagement, identifying spikes in search terms or trends, allowing advertisers and marketers to optimize budget allocation for specific times.

Machine performance – Utilizes digital twin technologies to detect deviations in performance, signaling potential failures in real-world machines, enabling preventive maintenance and reducing downtime.

Medical monitoring – Detects irregularities in individual health (e.g., abnormal heart rhythms) or public health (e.g., unexpected disease outbreaks in specific regions).

Different Types of Anomalies in Anomaly Detection

Point Anomalies

Point anomalies refer to individual data points that deviate significantly from the expected pattern or norm. These are isolated outliers that don't fit the general pattern of the dataset.

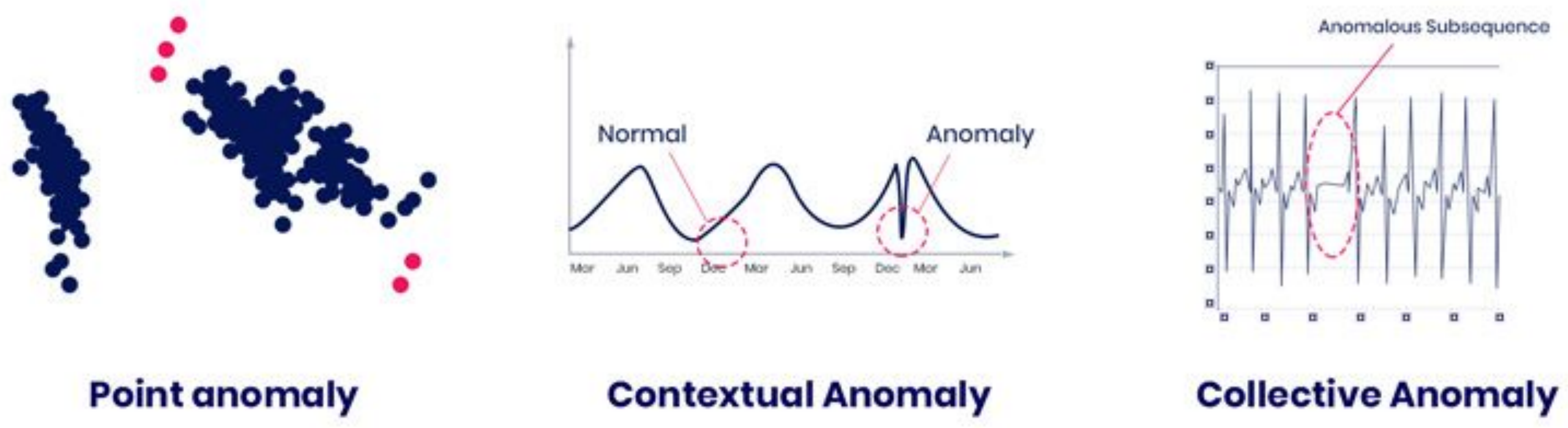
Collective Anomalies

Collective anomalies occur when a group of data points, considered normal individually, together exhibit unusual behavior as a collective pattern.

Contextual Anomalies

Contextual anomalies are data points that are anomalous only in a specific context or under certain conditions. These anomalies are normal in one context but abnormal in another.

What Are the Applications of Anomaly Detection?



Hand on

Assume that $\mathbf{X}_{\text{Train}}$ represents the normal data and $\mathbf{Y}_{\text{Train}}$ represents the reconstruction data. To identify anomalies in $\mathbf{X}_{\text{Test}}$, calculate the maximum Mean Squared Error (MSE) between $\mathbf{X}_{\text{Train}}$ and $\mathbf{Y}_{\text{Train}}$, and compare it with the MSE of $\mathbf{X}_{\text{Test}}$ and $\mathbf{Y}_{\text{Test}}$. If the MSE of the test data exceeds the maximum MSE, it will be classified as an anomaly.

$$\mathbf{X}_{\text{Train}} = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 0.5 & 0.3 & 0.2 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 0.7 & 0.7 & 0.7 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 0.4 & 0.4 & 0.4 \\ 1.0 & 1.0 & 1.0 \end{bmatrix} \mathbf{Y}_{\text{Train}} = \begin{bmatrix} 1.0 & 0.9 & 1.1 \\ 0.5 & 0.3 & 0.3 \\ 0.9 & 1.1 & 1.0 \\ 1.0 & 1.0 & 0.9 \\ 0.6 & 0.8 & 0.7 \\ 1.1 & 0.9 & 1.1 \\ 1.0 & 1.1 & 0.9 \\ 0.9 & 1.0 & 1.0 \\ 0.4 & 0.3 & 0.5 \\ 1.0 & 0.9 & 1.1 \end{bmatrix}$$

$$\mathbf{X}_{\text{Test}} = \begin{bmatrix} 1.0 & 1.0 & 1.0 \\ 0.6 & 0.5 & 0.4 \\ 0.9 & 0.8 & 1.0 \\ 0.7 & 0.6 & 0.7 \end{bmatrix}$$

$$\mathbf{Y}_{\text{Test}} = \begin{bmatrix} 1.1 & 1.0 & 0.9 \\ 0.6 & 0.4 & 0.3 \\ 0.9 & 1.1 & 0.9 \\ 0.7 & 0.6 & 0.8 \end{bmatrix}$$