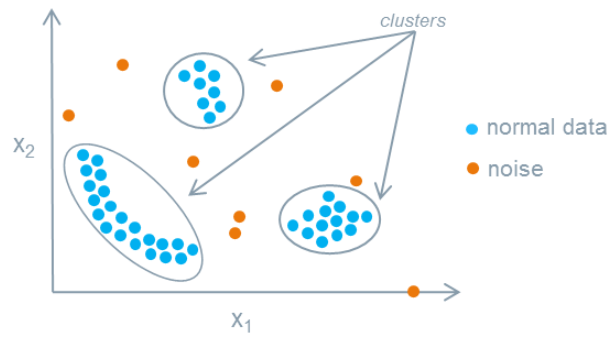
**ANAMOLY DETECTION**

Anamoly Detection is also known as outlier detection. Let us understand the Outlier in the Laymen language. For instance, you are asked to remove the rotten tomatoes from bucket because if not separated it will also spoil the other good tomatoes.

Similarly, there are variable/features/data points which are of no use or making no difference but could be responsible for greater loss. Thus we need to find the Outliers and remove them for better accuracy.



The noise are the data points which are detected as the outliers.

Anamoly Detection is categorized into three broad categories -

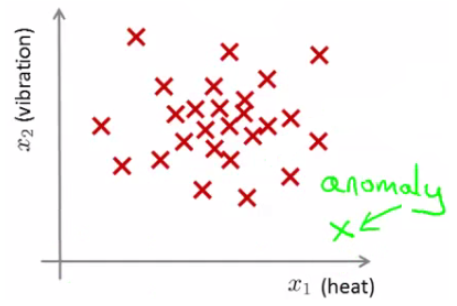
1. **Supervised Anamoly Detection** In Supervised Detection, there is a classifier which classifies whether the data points is Normal or Abnormal.
2. **Unsupervised Anamoly Detection** It detects the anomalies in the given dataset by assuming that the testing dataset contains the least fit to the remainder of the data set.
3. **Semi-Supervised Anamoly Detection** The training data set to construct the normal behaviour to the model and it checks the test data for the likelihood by the experience the model generated.

***Anamolies and is classifications***

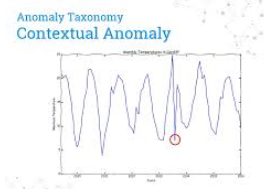
Anamoly is use to identify the rare items, suspcious items, events and outcomes which can raise a harm to the model.

The anamolies have several classifications -

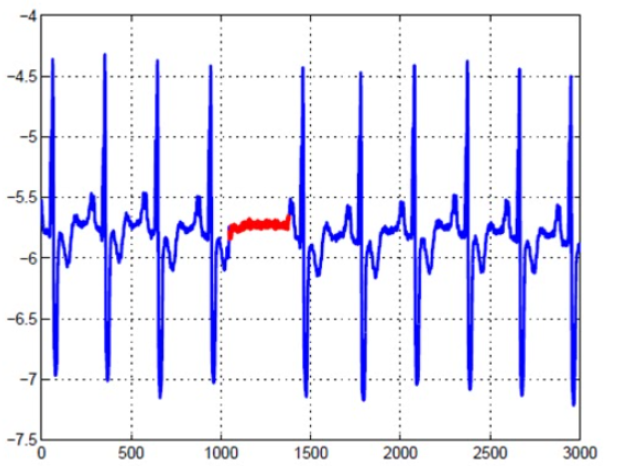
1. **Point anamolies** When a single data point is too far from the rest data points which makes it merely impossible to make the cluster or map it to the data points or cluster then we simply remove such data points. This is called the point anamolies.



1. **Contextual anamolies** If the abnormality is context specific, For instance investing 1000 rupee everyday on buying shoe since you play football is normal, but odd anyway.



1. **Collective anamolies** A set of data instance is responsible to track this anamoly. If someone is remotely using a machine and extracting the information to the local host. It gives the sign of the cyber attack.



Anamoly Detection is similar to Novelty detection but not completly similar. Novelty Detection is mainly concerned of identifying the unobseverd pattern in the observations.

Anamoly Detection Techniques

Simple Statistical Methods - The simple methods to find the irregularities in data points that deviate from common statistical properties of distribution including mean, median, standard deviation, etc.

**Anamoly detection techniques**

Isolation Forest Anomaly Detection Algorithm

Density-Based Anomaly Detection (Local Outlier Factor)Algorithm

Support Vector Machine Anomaly Detection Algorithm

## **Applications of Anamoly detection**

1. Intrusion Detection
2. Fraud Detection
3. Fault Detection
4. System Health Monitoring
5. Event Detection in networks
6. Detecting Natural disturbances.

# Why Data processing and visualization is important ?

It is very important to clean the data(preprocess) before using it to fit the model. The method helps in removing the outliers and make the data standardized.

To understand the data more easily and widely we visualize the data

Now, let us preprocess the data, visualize the data and fit the data into the model.

1. Pandas - It is an open-source library which we can use to manipulate, create or wrangle the data.
2. Numpy - NumPy stands for 'Numerical Python'. It is a python package used to perform scientific computations like performing linear algebra, arranging the data, dropping the data, etc.

