In [11]:

```
%run 00_basic.ipynb
```

What is an Anomaly

Anomaly is something that is not normally observed. It does not mean fault or failure. Just something different.

A **fault** in a system is a negative behaviour and has the potential to cause **Failure** which is the inability of the system to execute the intended operation.

Anomalies are also referred to as discordants, deviations, outliers, novelty - something that stands out when compared to normal population.

Hawkins [1] defines:

> "An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism."

Applications of Anomaly Detection

System operations is to detemine if the system is operating normally or were there any deviations from is nomrla beharviour. For example a spacecraft operations can be monitored using a system in place.

Intrusion Detection Systems : is to check the behavior of system calls, Network traffic, activity during normal and holidays are not unsual. An unusual slowness or elevated activitied may indicate an attack

In **Fraud Detection** a typical usage pattern and buyng behavior of the individuals can be modelled and compared with each future transaction to detect an anomalous transaction. The model can take into account the charge amount, location, speed and frequency of purchases etc. to compare to detect normal or fradualant transactions.

Medical Diagnosis, the authors have applied to predict Ashtma triggers in patients

Weather predictions or **aging predictions** to see if there are unsual patterns that lead to environmental trends.

Determine and set normal Operating range of values for a measured entity. For example, in spacecraft operations one could ask, what is the normal operating temperature of thermal heaters. In medical application, one migth want to know normal heart rate during rest and activity periods among different segments of the population.

Other applications one can imagine are **Cyber Security Money Laundering Banking and Financial applications** etc.

Concerns of Anomaly Detection System

Challenges, Methods and Approaches to Anomaly Detection Algorithms

Anomaly is out of the norm behaviour, therefore usually due to inherant nature of it, one might often enounter:

- significant class imbalance.
- Concept Drift - the behavior evolves and drifts its dynamic
- New Anomalies - it is unlikely to enumerate all the anomalies
- Lack of supervised set (especially during new product development)
- Class Overlap - anamlous data can overlap on non-anomalous datasets.
- Anomaly data in very high dimenstions - the spacecraft data sets range from 5000 to 150000 sensors

Methods

The methods to detect Anomaly could be

> - Supervised
> - Unsupervised
> - Semi Supervised

- Denisity Based methods (AKS Procimity based Methods)

  > - DBSAN -LOF

- Distance Baed

  > - Clustering
  >
  >   > - K-NN
  >   > - K-Means
  >
  > - Regression Hyperplane distane

- Parametric

  > - Gaussian mixture Models
  > - Single class SVMs
  > - Extreme Value Theorem
  > - Hidden Markov Models
  > - Isolation Forests
  > - Extreme Value Analysis
  > - Linear Models
  > - Spectral Models

- System Invariance Models

  > - System Invariant Analysis Technology (SIAT) [2][3]

- Deep Neural networks (numerous articles)

  - Sequence models
  - Long Short Term Memory (LSTM)
  - Convolutional Neural networks (CNN)

Short sight on Detection and need for extentions'

Almost all methods consists of developing a model by observing the data in normal time. Compare the model to remaining dataset and classify the new observations as abnormal or normal.

It is just not sufficient to detect an anomaly. The complex models such as LSTM's can detect anomalies, however just merely detecting anomaly does not solve real world issues.

The other concerns are:

- When is the time to retrain the model? or how long does a trained model can predict anomalies - model warranty
- **Most critical** are how to explain the anomaly; the cause of the anomaly; It is importance to note that the LSTM models can detect anomalies with high fidelity. However it is very **critical** to explain the anomaly; most methods fail to explain the cause. the models such as SIAT are built to show the cause of anomaly and trouble shoot the root cause and offer remedial recommendations.
- How to capture and visualize the **Logical mapping of the system**. SIAT models inherently show the dynamics or logical-model of the system. This is especially important to compare the constllation of similar systems. In case of satellites, how does constellation of spacecrafts behave - an important dynamics to understand the space weather effecs on spacecraft operations that undergo frequent **Elecro static discharge (ESD)** events that cause what is known as **Single Event Failures SEF** that causes intermittent failure of the spacecraft; these intermittent failures have tremondous isses when it occurs in communication spacecrafts or Weather or GPS satellites.
- Are there seasonality in system behavior? Is it diffent during night and day time. In case of spacecrafts, the battery behaviors and temperature change rapidly. Spacecrafts can show various dynamics as seasons change (summer, winter, fall, etc) depending upon its exposure to sun which is the primary means of charging batteries.
- How reduce False positives? In case of spacecraft operations, there are operations such as **Station keeping**, **North South** or **East West** station keeping operations (AKA maneuvers) that are done to align the spacecraft pointing to antennas to enable **Communication Sub System (CSS)** to operate normally. These maneuvers are difficult to capture in general dynamics - how shouls the system handle these normal and yet difficult to capture dynamics of the system.
- How should data be preprocessed for various algorithsms? For LSTM, it is critical to normalize the data; for SIAT, critical to eliminate categorical variables. In all cases, the data must be numeric and quality tested.
- How to capture the **"false positive"** patterns and to apply post-processing to reduce the false alarms.
- How to customize **automated actions** upon known anomalous patterns.
- How not to miss **True Positives**
- How to conduct feature engineering to detemine most critical sensors
- How to trigger anomaly if if it occurs in sub-space (this is especially true in case higher dimentional space of 100k sensors where anomaly may be caused due to small deviations in sensors)
- How to evolve the model; Use this to show how the system is aging and offer insights to robust design and operations.

- What is the **Concept of Operations** (CONOPS). When to build the model and how to deploy. The computation power requirements for developing the model and power requirements for inferences.
- Where does the computation run. This is especially important when the anomaly detection framework needs to be deployed in Space environments. In addition, power, memory requirements plays a vital role in deplying system where it is expected to discover abd self-calibrate.
- How to handle diffent types of data? Most algorithms are not capable of handling categorical or binary type of data. (For example, switches ON/OFF or status reporting switches). How to holistically handle all types of data.
- How to augment (or enrich) the data set and remove irrelvant data. (For ex. a sensor that is high correlated with time is rather not have much signals and thus can be omitted)
- How to handle "Log data", text data or Image data that have charateristics of time series.
- How to develop of user interface for vairous stake holders.
- Determine the sufficient data quantity to capture the system dynamics.
- How to patch or handle missing data. This is especailly true in spacecraft (or any high dimentional dataset) there will be missing or bad-quality data (anomaly); how to detect and patch before capturing the model to reduce false positives.

  > Data patching technics such as "forward fill", backward-fill are not necesarily effective always.

- How to predict future anomalies - i.e. capture the trending signals that leads to anamoly
- How to use it for maintenance

Type *Markdown* and LaTeX: α^2