If you are from a Manufacturing industry, have you ever seen some strange behaviour of one of your newly manufactured devices/products?

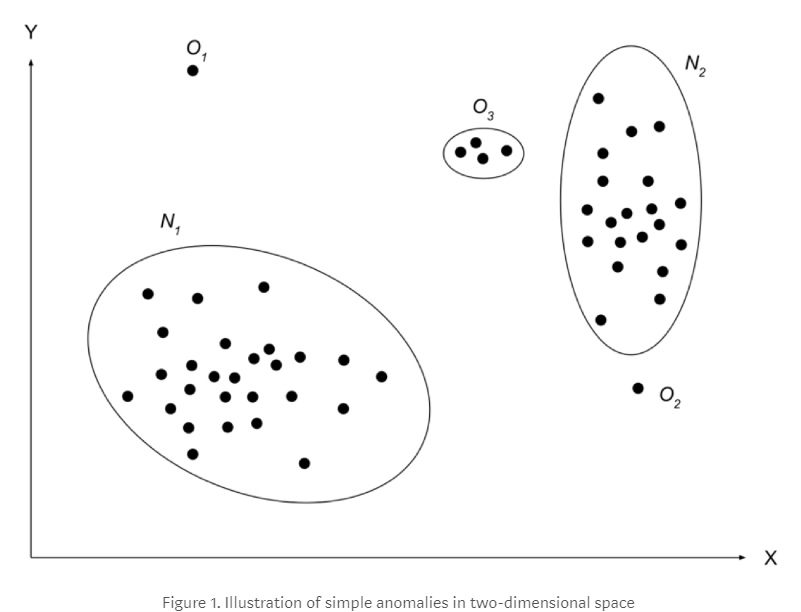
If you belong to a Telecom industry, have you ever noticed some abnormal performance of traffic at some of the nodes in your network?

If you belong to a Healthcare sector, have you ever come across a patient records that are pretty much beyond your estimations?

If you are unsure of the above questions, then this article explains you the methodologies that one can apply in your respective domains to discover those unknown findings. **Anomaly detection** is all about finding what you don’t know to look for.

**What is an Anomaly?**

Unfortunately, there is no clear cut definition for an Anomaly. In simple words, anomalies are the observations that do not match with a well characterized normal behaviour/pattern. Sometimes anomalies are also called as Outliers in a dataset.



For example, in the above plot, the O1 is clearly an anomaly since it does not belong to any of the groups. On the other hand O2 looks like an anomaly, but its characteristics are not too far away from the normal pattern N2.But the probability of O2 being treated as anomaly is less when compared to O1. O3 is considered as a Group Anomaly.

We have 3 types of Anomalies:

1. **Point Anomaly:** If one observation can be observed against other observation as anomaly, it is a point anomaly. O1 and O2 in the above plot are Point anomalies.
2. **Contextual Anomaly:** If an observation is anomalous in some defined context (say - time), then it is contextual Anomaly.

For example, when you get a call from your boss during official hours, then it is normal, but when you receive a call during out of office hours, then it is considered as a Contextual Anomaly.

1. **Collective/Group Anomaly:** If some linked observations can be treated against other observations as anomaly, then it is called Collective anomaly, for e.g. O3 in the above plot.

On the abstract level detection of the anomalies seems like a simple task. But this task can be very challenging. Here are some challenges bellow:

* Defining normal regions is very difficult. In many cases boundaries between anomalies and normal data are not precise. In this case, normal observations could be considered as anomalies and vice-versa
* If action is malicious, like fraud, it is considered as anomaly. Very often attackers try to adapt their actions to the normal behaviour. And again, task to identify anomalies in this case is not so simple
* What is considered normal today can be not normal in the future. Most of the business systems change in time under the influence of the various factors.

**How to detect an Anomaly?**

There are various algorithms and processes developed in the recent past to find a particular kind of Anomaly. Let’s discuss few of them.

**Static Rules:**

This is a simple approach of finding anomaly. Here, based on the list of known anomalies we find in day to day life, the experts in the domain defines static rules that best describes those anomalies. Sometimes, these rules become more complex to understand the behaviour of the new anomaly observation. As discussed before, we are not sure of how many various segments of anomalies that may arise from the incoming new data.

Therefore, statistical or machine learning based approach, which automatically learns the general rules, are preferred to static rules.

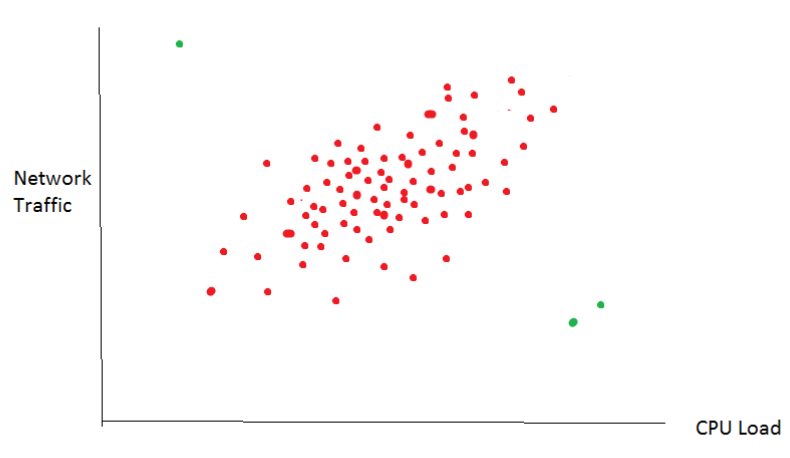
**Statistical Modelling Approach:**

This approach can be used when we have unlabelled data. i.e. (The data is not segmented/classified)

Let’s consider an example of Monitoring computers in a dataset and assume we have below variables of each computer.

* X1 = CPU load
* X2 = network traffic
* X3 = number of disk accesses/sec
* X4 = memory use

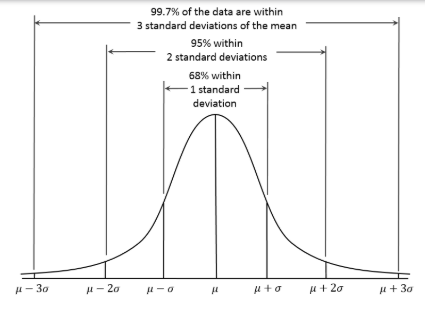
From the above, we need not use all variables for modelling. Based on the expertise in this area, you suspect only few which can take very large values or too small values that may potentially cause an Anomaly. Let us plot these two variables CPU load and network traffic.



The data values of (say 100) computers are which are functioning well are shown in **red**. We consider this as normality in this example. The area where there is more density in red points will be close to the average value and the Probability of these observations will be high. The density decreases on either side which in turn decreases the Probability.

This can be better understood my comparing with the characteristics of a Gaussian density function given below. A data distribution is said to be normal when mean is at the centre and 68% of total data lies between 1 standard deviations. 99.7% of the data will fall between 3 standard deviations.

So those observations that do not fall in this 99.7% of the data are considered as Anomalies. Here the points in Green are anomalies as they violate the normal behaviour.

****

**Fig -** Gaussian (Normal) distribution with mean ( U) and variance (sigma)

**Algorithm – Point Anomaly:**

Let’s consider a dataset having ‘N’ variables and ‘M’ observations. (I.e. M \* N matrix)

X1, X2, X3……XN are the variables.

We will build model that gives the P(X) score which is nothing but the product of Gaussian Probabilities of individual variables. i.e.

P(X) = P(X1) \* P(X2) \* P(X3) \*…..P (XN)

Where P(X1), P(X2) are the individual variable Gaussian Probabilities

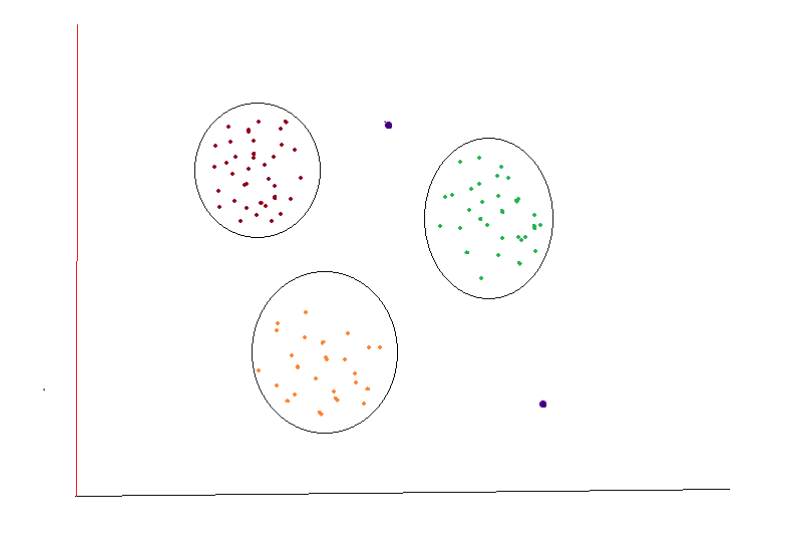
The assumption we make here is that the each variable is normally distributed with different values of mean and standard variation. We also assume some threshold value epsilon € to compare with outcome of P(X).

For any new observation If P(X) > € value, then it is a not an Anomaly. If P(X) < €, then such observations are treated as Anomalies.

**The Machine Learning Way:**

Here, we can make use of the Unsupervised Learning algorithms. For example. Let’s consider K-means clustering.

The data is fed into this model to perform the segmentation and come up with clusters. Behind the scenes, these clusters are formed by grouping the observations that are closer to the centroid of every segment. Distances such as Euclidian distance is being considered to create the borders for each cluster.



Here there are 3 segmentations of the observations each represented by circle. The points in Blue do not belong to any of the above clusters as they are very far from the centroids. So they can be treated as anomalies.

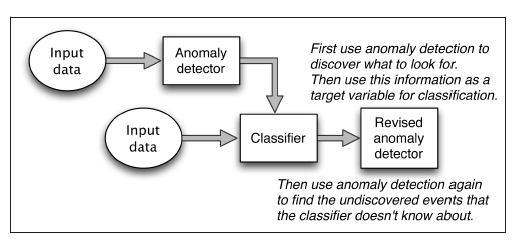
**Anomaly detection VS Supervised Classification:**

In Supervised Classification models, there is labelled data for all the classifications. Here the data of our interest will be very less (say 10 in 1000s) and when we build a model by feeding such data, the accuracy will always come high but the model misclassifies. This is because the dataset has unbalanced classes. One best example would be identifying fraud among 1000s of customers.

This issue can be solved by applying the resampling techniques on the majority class observations to make the classes balanced before feeding into the model. Later by optimizing the Precision and Recall scores, the models will be able to produce good results.

Here, we have a limited set of anomalous data which we re–use frequently to train the model which means the characteristics of these anomalous data are properly defined and well understood. But, when there comes a new type of Anomaly whose characteristics are different to the ones that were just defined, then the trained model will fail to classify accurately. That’s where Anomaly detection algorithms come handy. They are mainly used to find those outliers which are 10 among a million.

So, the bottom line is that the supervised learning algorithm comes after Anomaly Detection.



So far, the above methods helps in identifying the Point anomalies, where a single observation is compared with the remaining observations.

Let’s move onto the other form of anomaly which is Contextual anomaly. This is basically finding anomalies in a time series data. For e.g.. a) Finding abnormalities in a network traffic at a node. b) Finding the abnormal behaviour of a patient’s heart beat etc.

**Contextual Anomaly:**

Contextual anomalies are calculated by focusing on segments of data (e.g. spatial area, graphs, sequences, customer segment) and applying collective anomaly techniques within each segment independently. A practical example and demonstration of identifying this type of anomalies is available at my GitHub Repository.

Anomaly Detection can be explained by many models which form the basis of its Algorithms.

Using Gaussian distribution, we can detect an anomaly by manually creating few features to capture anomalies when the original features form unusual combinations of values. Here the scalability can be increased to many no of variables.

On the flip side, The Multivariate Gaussian distribution automatically captures positive and negative correlations between features and flags the new observation as anomaly if it has unusual combinations of the features.