**Unsupervised Learning**

There are situations where the desired output class/event is unknown for historical data. The objective in such cases would be to study the patterns in the input dataset to get better understanding and identify similar patterns that can be grouped into specific classes or events. As these types of algorithms do not require any intervention from the subject matter experts beforehand, they are called unsupervised learning. Let’s look at some examples of unsupervised learning.

**Clustering**

Assume that the classes are not known beforehand for a given dataset. The goal here is to divide the input dataset into logical groups of related items. Some examples are grouping similar news articles, grouping similar customers based on their profile, etc.

**Dimension Reduction**

Here the goal is to simplify a large input dataset by mapping them to a lower dimensional space. For example, carrying analysis on a large dimension dataset is very computational intensive, so to simplify you may want to find the key variables that hold a significant percentage (say 95%) of information and only use them for analysis.

**Anomaly Detection**

Anomaly detection is also commonly known as outlier detection is the identification of items, events or observations which do not conform to an expected pattern or behavior in comparison with other items in a given dataset. It has applicability in a variety of domains, such as machine or system health monitoring, event detection, fraud/intrusion detection etc. In the recent days, anomaly detection has seen a big area of interest in the word of Internet of Things (IoT) to enable detection of abnormal behavior in a given context. A data point is termed anomaly if it is distant from other data points in a given context, so calculating standard deviation or clustering are the most commonly used techniques for detection of anomaly alongside a whole lot of other techniques. I’ll not be covering the topic in this edition.