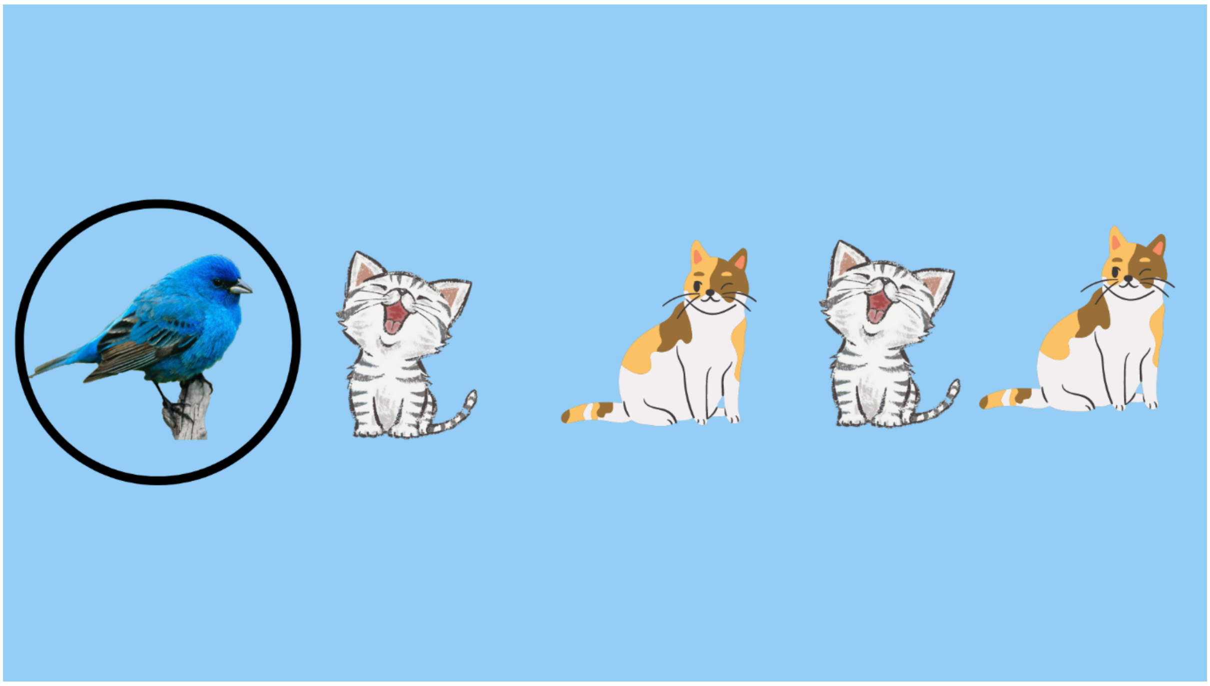
What Is Anomaly Detection in Machine Learning?

Anomaly Detection is the **technique of identifying rare events** or observations which can raise suspicions by being **statistically different from the rest of the observations.**



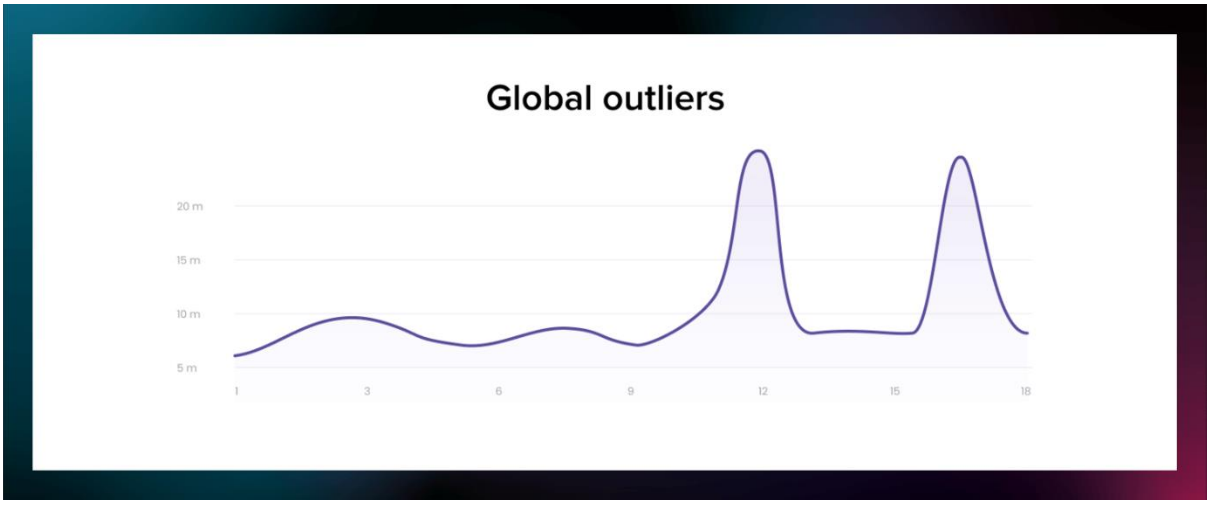
Anomaly detection is a process of finding those rare items, data points, events, or observations that make suspicions by being different from the rest data points or observations. Anomaly detection is also known as **outlier detection**.

Types of anomalies

Global outliers

When a data point assumes a value that is far outside all the other data point value ranges in the dataset, it can be considered a **global anomaly**. In other words, it's a rare event.

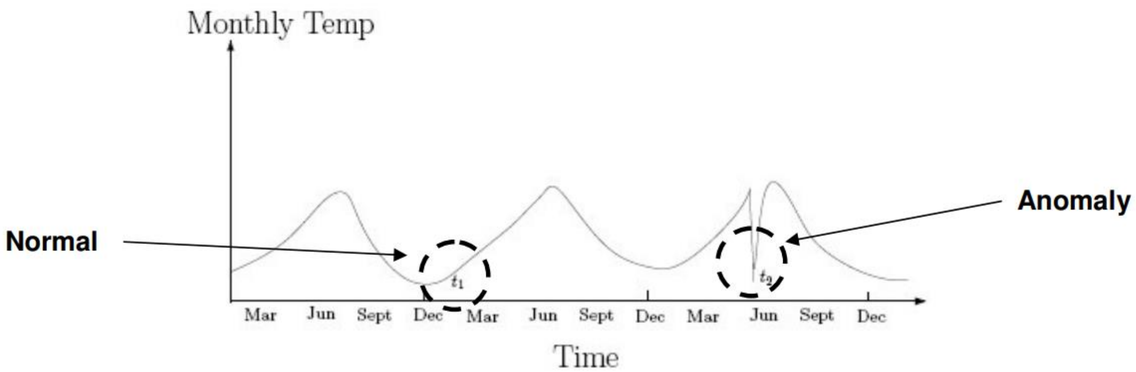
For example, if you receive an average American salary to your bank accounts each month but **one day get a million dollars,** that would look like a global anomaly to the bank's analytics team.



Contextual Anomaly

Contextual anomaly is also known as conditional outliers. If a particular observation is different from other data points, then it is known as a **contextual Anomaly**

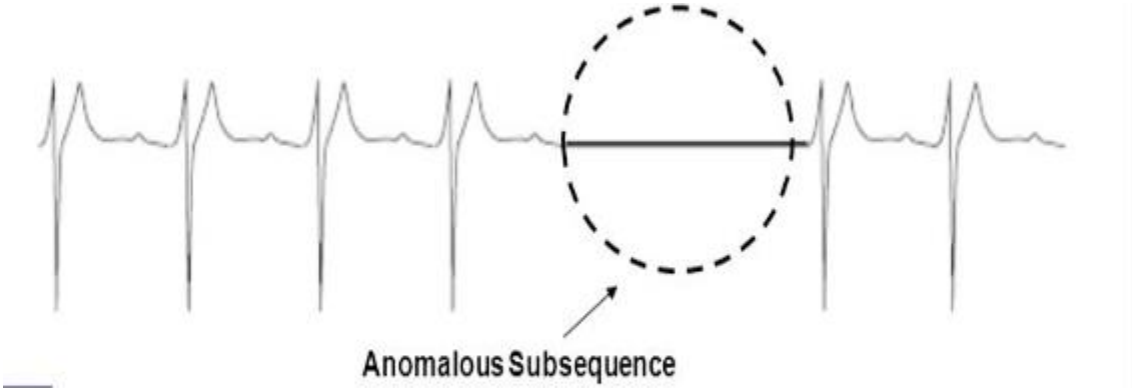
- **Contextual anomalies** – The **abnormality is based on the specific context**. We can find this kind of anomaly in **time-series based data**. Use case – Amount spent on petrol on a daily basis throughout the working days is normal but will be found odd when spent on vacation.



Collective Anomaly

Collective anomalies occur when a data point within a set is anomalous for the whole dataset, and such values are known as collective outliers.

- **Collective anomalies** – A set of data instances put together helps in detecting anomalies. Use case – Someone attempting to copy data from a remote machine to a local host unexpectedly such an anomaly would be flagged as possible cyber-attack.



Anomaly Detection Approaches

- ***Model-based***

- ***Distance-based***

This approach is based on the proximities. Consider a 2D or 3D scatter plot all the data objects are in one proximity. But Anomalous objects are away from them.

- ***Distance-based***

This approach is based on the proximities. Consider a 2D or 3D scatter plot all the data objects are in one proximity. But Anomalous objects are away from them.

DBSCAN

DBSCAN (Density-based Spatial Clustering of Applications with Noise) is one of the extensively used density-based clustering algorithms which is used in the applications of outlier detection.

The basic idea behind DBSCAN is that a cluster has to contain minimum number of points within the specified radius. DBSCAN algorithm uses two parameters, minimum points ("minpts") and epsilon("eps") to perform clustering.

The Model-based Approach (also called Statistical Method)

Time Series Analysis and Forecasting | Data-Driven Insights

What Is Time Series Analysis?

Time series analysis is a specific way of analyzing a sequence of data points collected over time. In TSA, analysts record data points at consistent intervals over a set period rather than just recording the data points intermittently or randomly.

Introduction

What Is Time Series Analysis?

Time series analysis is a specific way of analyzing a sequence of data points collected over time. In TSA, analysts record data points at consistent intervals over a set period rather than just recording the data points intermittently or randomly.

Objectives of Time Series Analysis

How to Analyze Time Series?

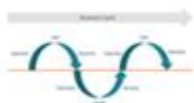
To perform the time series analysis, we have to follow the following steps:

- Collecting the data and cleaning it
- Preparing Visualization with respect to time vs key feature
- Observing the stationarity of the series
- Developing charts to understand its nature.
- Model building – AR, MA, ARMA and ARIMA
- Extracting insights from prediction

Components of Time Series Analysis

Let's look at the various components of Time Series Analysis:

- **Trend**: In **which there is no fixed interval** and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend
- **Seasonality**: In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth
- **Cyclical**: In which there is no fixed interval, uncertainty in movement and its pattern
- **Irregularity**: Unexpected situations/events/scenarios and spikes in a short time span.

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

Designed by Author (Shanthababu)

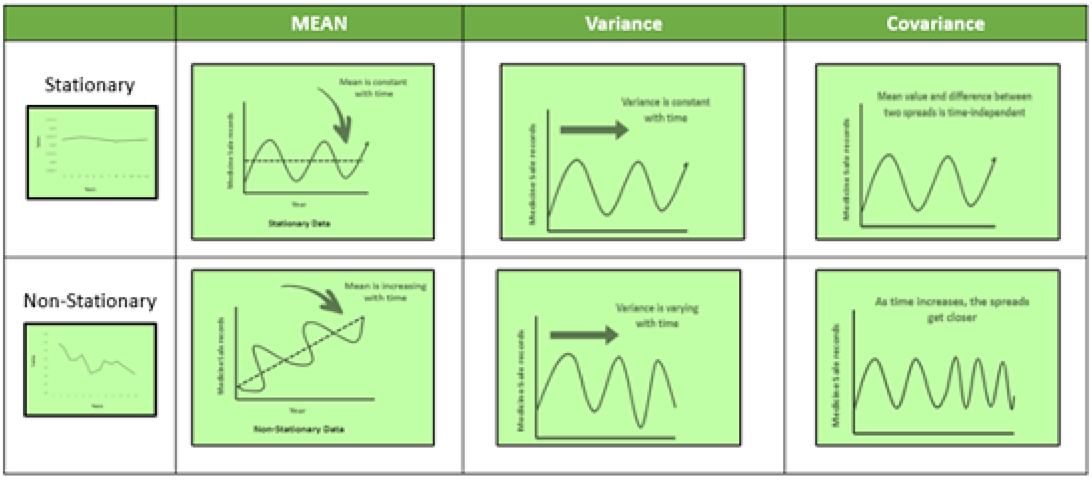
Data Types of Time Series

Let's discuss the time series' data types and their influence. While discussing TS data types, there are two major types – stationary and non-stationary.

Stationary: A dataset should follow the below thumb rules without having Trend, Seasonality, Cyclical, and Irregularity components of the time series.

- The **mean** value of them should be completely constant in the data during the analysis.
- The **variance** should be constant with respect to the time-frame
- **Covariance** measures the relationship between two variables.

Non- Stationary: If either the mean-variance or covariance is changing with respect to time, the dataset is called non-stationary.



Designed by Author (Shanthababu)

Methods to Check Stationarity

During the TSA model preparation workflow, we must assess whether the dataset is stationary or not. This is done using **Statistical Tests.** There are two tests available to test if the dataset is stationary:

- **Augmented Dickey-Fuller (ADF) Test**
- Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Augmented Dickey-Fuller (ADF) Test or Unit Root Test

The ADF test is the most popular statistical test. It is done with the following assumptions:

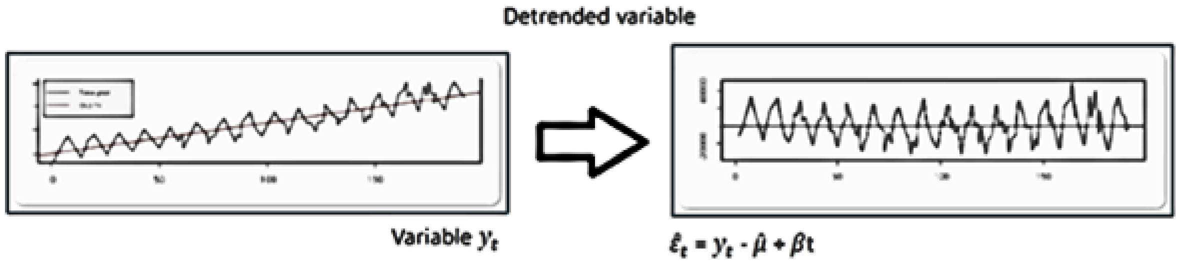
- Null Hypothesis (H0): Series is non-stationary
- Alternate Hypothesis (HA): Series is stationary
    - p-value >0.05 Fail to reject (H0)
    - p-value <= 0.05 Accept (H1)

Converting Non-Stationary Into Stationary

Let’s discuss quickly how to convert non-stationary to stationary for effective time series modeling. There are three methods available for this conversion – detrending, differencing, and transformation.

Detrending

It involves removing the trend effects from the given dataset and showing only the differences in values from the trend. It always allows cyclical patterns to be identified.

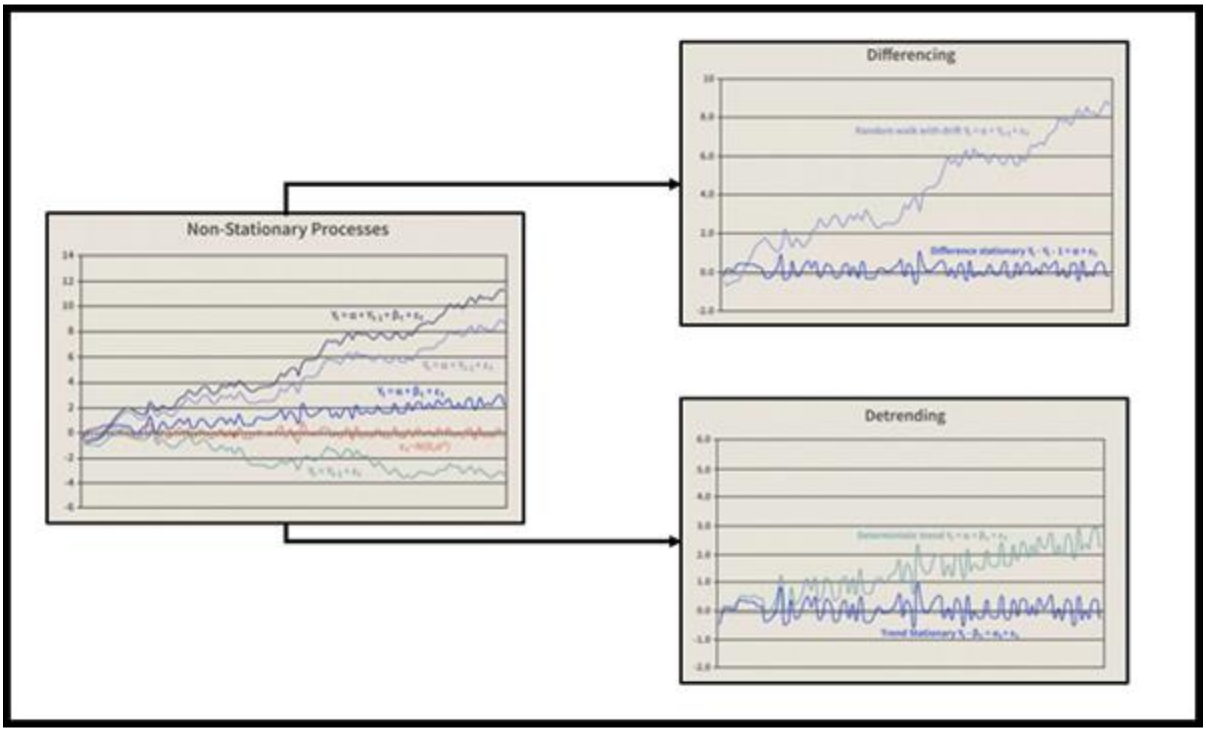


Designed by Author (Shanthababu)

Differencing

This is a simple transformation of the series into a new time series, which **we use to remove the series dependence on time and stabilize the mean of the time series, so trend and seasonality are reduced during this transformation.**

- Yt= Yt – Yt-1
- Yt=Value with time



Transformation

This includes three different methods they are **Power Transform, Square Root, and Log Transfer.** The most commonly used one is **Log Transfer.**

Moving Average Methodology

The commonly used time series method is the Moving Average. This method is slick with random short-term variations. Relatively associated with the components of time series.

The Moving Average (MA) (or) Rolling Mean: The value of MA is calculated by taking average data of the time-series within k periods.

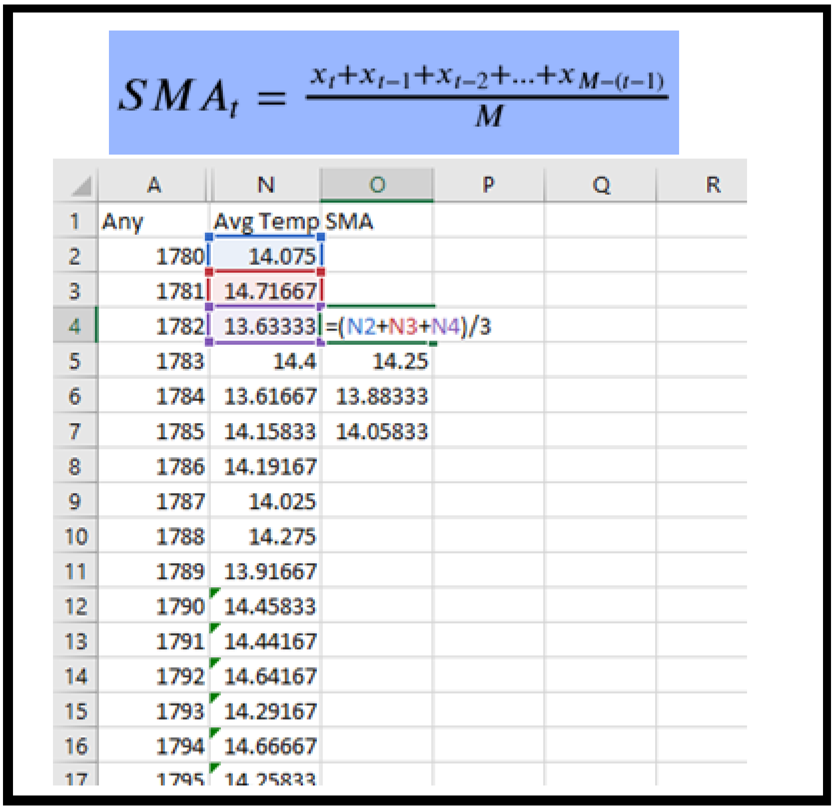
Let’s see the types of moving averages:

- Simple Moving Average (SMA),
- Cumulative Moving Average (CMA)
- Exponential Moving Average (EMA)

Simple Moving Average (SMA)

The Simple Moving Average (SMA) **calculates the unweighted mean of the previous M or N points.** We **prefer selecting sliding window data points based** on the **amount of smoothing,** as increasing the value of M or N improves smoothing but reduces accuracy.

To understand better, I will use the air temperature dataset.

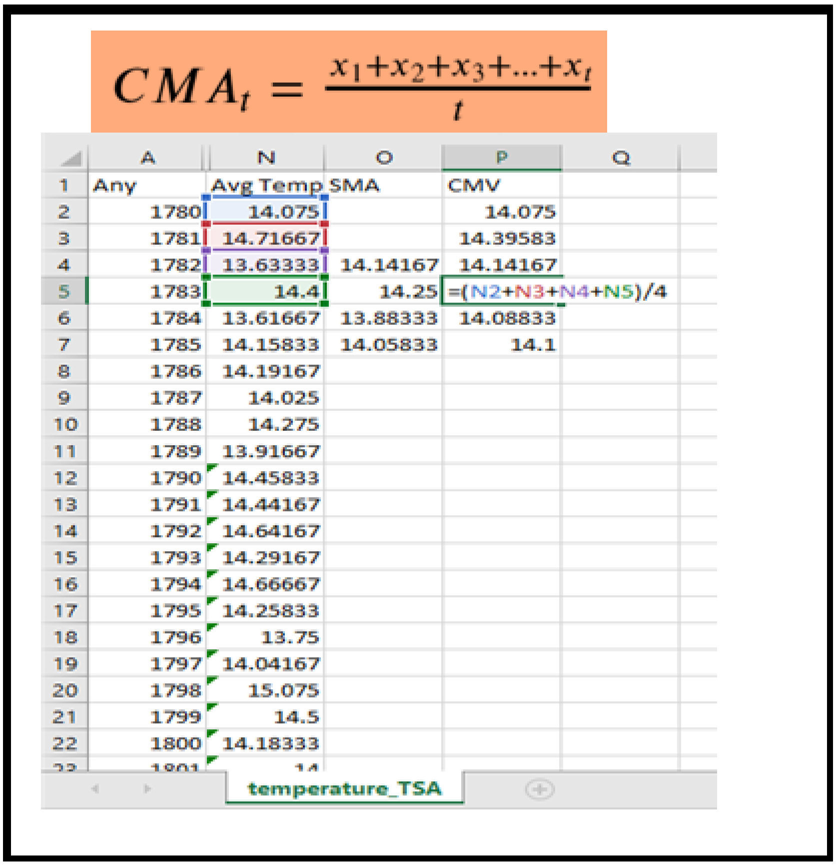


$$SMA_t = \frac{x_t + x_{t-1} + x_{t-2} + \ldots + x_{M-(t-1)}}{M}$$

	A	N	O	P	Q	R
1	Any	Avg Temp	SMA			
2	1780	14.075				
3	1781	14.71667				
4	1782	13.63333	=(N2+N3+N4)/3			
5	1783	14.4	14.25			
6	1784	13.61667	13.88333			
7	1785	14.15833	14.05833			
8	1786	14.19167				
9	1787	14.025				
10	1788	14.275				
11	1789	13.91667				
12	1790	14.45833				
13	1791	14.44167				
14	1792	14.64167				
15	1793	14.29167				
16	1794	14.66667				
17	1795	14.25833				

Cumulative Moving Average (CMA)

The CMA is the unweighted mean of past values till the current time.



$$CMA_t = \frac{x_1+x_2+x_3+...+x_t}{t}$$

	A	N	O	P	Q
1	Any	Avg Temp	SMA	CMV	
2	1780	14.075		14.075	
3	1781	14.71667		14.39583	
4	1782	13.63333	14.14167	14.14167	
5	1783	14.4	14.25	=(N2+N3+N4+N5)/4	
6	1784	13.61667	13.88333	14.08833	
7	1785	14.15833	14.05833	14.1	
8	1786	14.19167			
9	1787	14.025			
10	1788	14.275			
11	1789	13.91667			
12	1790	14.45833			
13	1791	14.44167			
14	1792	14.64167			
15	1793	14.29167			
16	1794	14.66667			
17	1795	14.25833			
18	1796	13.75			
19	1797	14.04167			
20	1798	15.075			
21	1799	14.5			
22	1800	14.18333			
23	1801	14			

temperature_TSA

Exponential Moving Average (EMA)

EMA is **mainly used to identify trends and filter out noise**. The **weight of elements is decreased gradually over time.** This means **It gives weight to recent data points, not historical ones.** Compared with SMA, the EMA is faster to change and more sensitive.

α –>Smoothing Factor.

- It has a value between 0,1.

- Represents the weighting applied to the very recent period.

Time Series Analysis in Data Science and Machine Learning

When dealing with TSA in Data Science and Machine Learning, there are multiple model options are available. In which the **Autoregressive–Moving-Average (ARMA) models with [p, d, and q].**

- **P==> autoregressive lags**
- **q== moving average lags**
- **d==> difference in the order**

Before we get to know about Arima, first, you should understand the below terms better.

- **Auto-Correlation Function (ACF)**
- **Partial Auto-Correlation Function (PACF)**

Auto-Correlation Function (ACF)

ACF indicates how similar a value is within a given time series and the previous value. (OR) It measures the degree of the similarity between a given time series and the lagged version of that time series at the various intervals we observed.

Python Statsmodels library calculates autocorrelation. It identifies a set of trends in the given dataset and the influence of former observed values on the currently observed values.

Partial Auto-Correlation (PACF)

PACF is similar to Auto-Correlation Function and is a little challenging to understand. It always shows the correlation of the sequence with itself with some number of time units per sequence order in which only the direct effect has been shown, and all other intermediary effects are removed from the given time series.

What Is an Auto-Regressive Model?

An **auto-regressive model is a simple model that predicts future performance based on past performance**. It is mainly used for forecasting when there is some correlation between values in a given time series and those that precede and succeed (back and forth).

An AR is a Linear Regression model **that uses lagged variables as input**. By indicating the input, the Linear Regression model can be easily built using the scikit-learn library.

Statsmodels library provides autoregression model-specific functions where you must specify an appropriate lag value and train the model. It is provided in the AutoTeg class to get the results using simple steps.

- Creating the model AutoReg()
- Call fit() to train it on our dataset.
- Returns an AutoRegResults object.
- Once fit, make a prediction by calling the predict () function

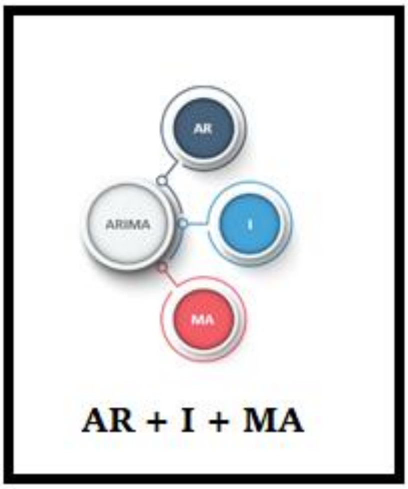
The equation for the AR model (Let's compare Y=mX+c)

Understanding ARMA and ARIMA

ARMA is a combination of the **Auto-Regressive and Moving Average models for forecasting**. This model provides a **weakly stationary stochastic** process in terms of two polynomials, one for the Auto-Regressive and the second for the Moving Average.

$$\boxed{Y_t = \underbrace{\mu + \sum_{i=1}^{q} \gamma_i Y_{t-i} + \varepsilon_t}_{\text{Auto-Regressive}} + \underbrace{\sum_{i=1}^{q} \theta_i \varepsilon_{t-i}}_{\text{Moving Average}}}$$

ARMA is best for predicting stationary series. **ARIMA** was thus developed to support both stationary as well as non-stationary series.



- AR ==> Uses past values to predict the future.
- MA ==> Uses past error terms in the given series to predict the future.
- I==> Uses the differencing of observation and makes the stationary data.

AR+I+MA= ARIMA

1. Univariate data –

This type of data consists of **only one variable**. The analysis of univariate data is thus the simplest form of analysis since the information deals with only one quantity that changes. It does not deal with causes or relationships and the main purpose of the analysis is to describe the data and find patterns that exist within it. The example of a univariate data can be height.

Heights (in cm)	164	167.3	170	174.2	178	180	186

2. Bivariate data

This type of data involves **two different variables**. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables. Example of bivariate data can be temperature and ice cream sales in summer season.

TEMPERATURE(IN CELSIUS)	ICE CREAM SALES
20	2000
25	2500
35	5000
43	7800

3. Multivariate data

When the data involves **three or more variables**, it is categorized under multivariate. Example of this type of data is suppose an advertiser wants to compare the popularity of four advertisements on a website, then their click rates could be measured for both men and women and relationships between variables can then be examined. It is similar to bivariate but contains more than one dependent variable. The ways to perform analysis on this data depends on the goals to be achieved. Some of the techniques are regression analysis, path analysis, factor analysis and multivariate analysis of variance (MANOVA).

A convolutional neural network (CNN) is mainly for image classification.

Difference between ANN, CNN and RNN

Artificial Neural Network (ANN):
Artificial Neural Network (ANN), is a group of multiple perceptrons or neurons at each layer. ANN is also known as a Feed-Forward Neural network because inputs are processed only in the forward direction.

Recurrent Neural Network (RNN):
Recurrent neural networks (RNN) are more complex. They save the output of processing nodes and feed the result back into the model (they did not pass the information in one direction only). This is how the model is said to learn to predict the outcome of a layer. Each node in the RNN model acts as a memory cell, continuing the computation and implementation of operations.

If the network's prediction is incorrect, then the system self-learns and continues working towards the correct prediction during backpropagation

A **Convolutional Neural Network (CNN)** is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data. When it comes to Machine Learning, Artificial Neural Networks perform really well. Neural Networks are used in various datasets like images, audio, and text. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use **Recurrent Neural Networks** more precisely an LSTM, similarly for image classification we use Convolution Neural networks.

In a regular Neural Network, there are three types of layers:

1. **Input Layers:** It's the layer in which we give input to our model. The number of neurons in this layer is equal to the total number of features in our data (number of pixels in the case of an image).
2. **Hidden Layer:** The input from the Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layer can have different numbers of neurons which are generally greater than the number of features.

The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by the addition of learnable biases followed by activation function which makes the network nonlinear.

3. **Output Layer:** The output from the hidden layer is then fed into a logistic function like sigmoid or softmax which converts the output of each class into the probability score of each class.

The data is fed into the model and output from each layer is obtained from the above step is called **feedforward**, we then calculate the error using an error function, some common error functions are cross-entropy, square loss error, etc. The error function measures how well the network is performing. After that, we backpropagate into the model by calculating the derivatives. This step is called **Backpropagation** which basically is used to minimize the loss.

Convolution Neural Network

Convolutional Neural Network (CNN) is the extended version of artificial neural networks (ANN) which is predominantly used to extract the feature from the grid-like matrix dataset. For example, visual datasets like images or videos where data patterns play an extensive role.