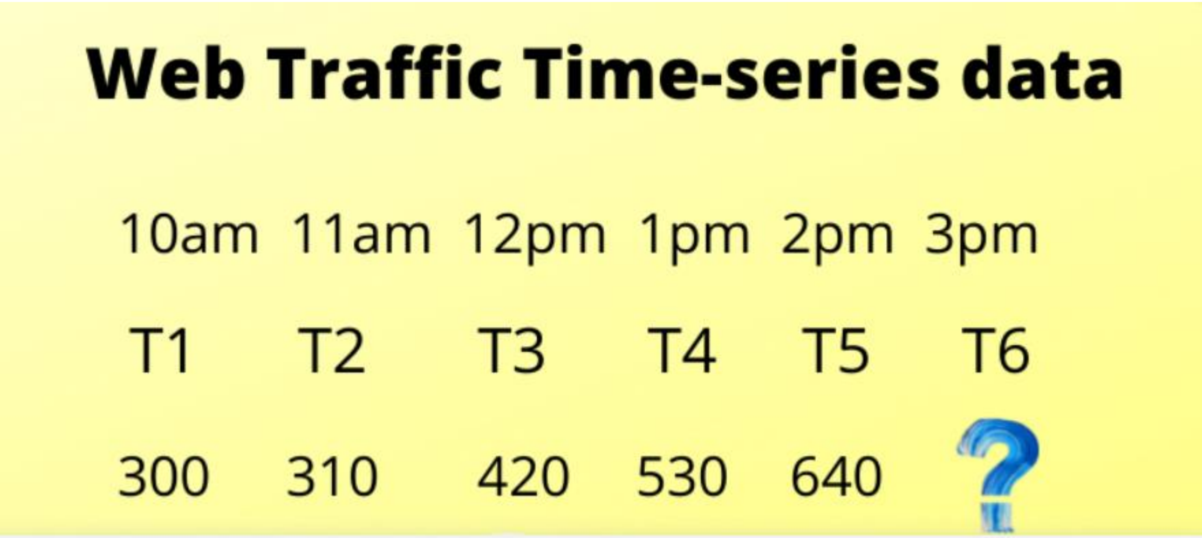
Time Series Forecasting Methods

Time series forecasting can broadly be categorized into the following categories:

- **Classical / Statistical Models** — Moving Averages, Exponential Smoothing, ARIMA, SARIMA, TBATS
- **Machine Learning** — Linear Regression, XGBoost, Random Forest, or any ML model with reduction methods
- **Deep Learning** — RNN, LSTM

Rolling statistics and stationarity in Time series



A stationary time series is a data that has a constant mean and constant variance. If I take a mean of T1 and T2 and compare it with the mean of T4 and T5 then is it the same, and if different, how much difference is there? So, constant mean means this difference should be less, and the same with variance.

If the time series is not stationary, we have to make it stationary and then proceed with Machine learning modelling. In order to do so, we must understand something called **Rolling statistics**.

MOVING AVERAGE

T1	T2	T3	T4	T5	T6
300	310	420	530	640	?
-----	305	365	475	585	

Rolling statistics help us in making time series stationary. So basically, rolling statistics calculates moving average. To calculate the moving average, we need to define the **window size** which is basically how much past values to be considered.

For example, if we take the window as 2 then to calculate a moving average in the above example then, at point T1 it will be blank, at point T2 it will be the mean of T1 and T2, at point T3 mean of T3 and T2, and so on. And after calculating all moving averages if you plot the line above actual values and calculated moving averages then you can see that the plot will be smooth.

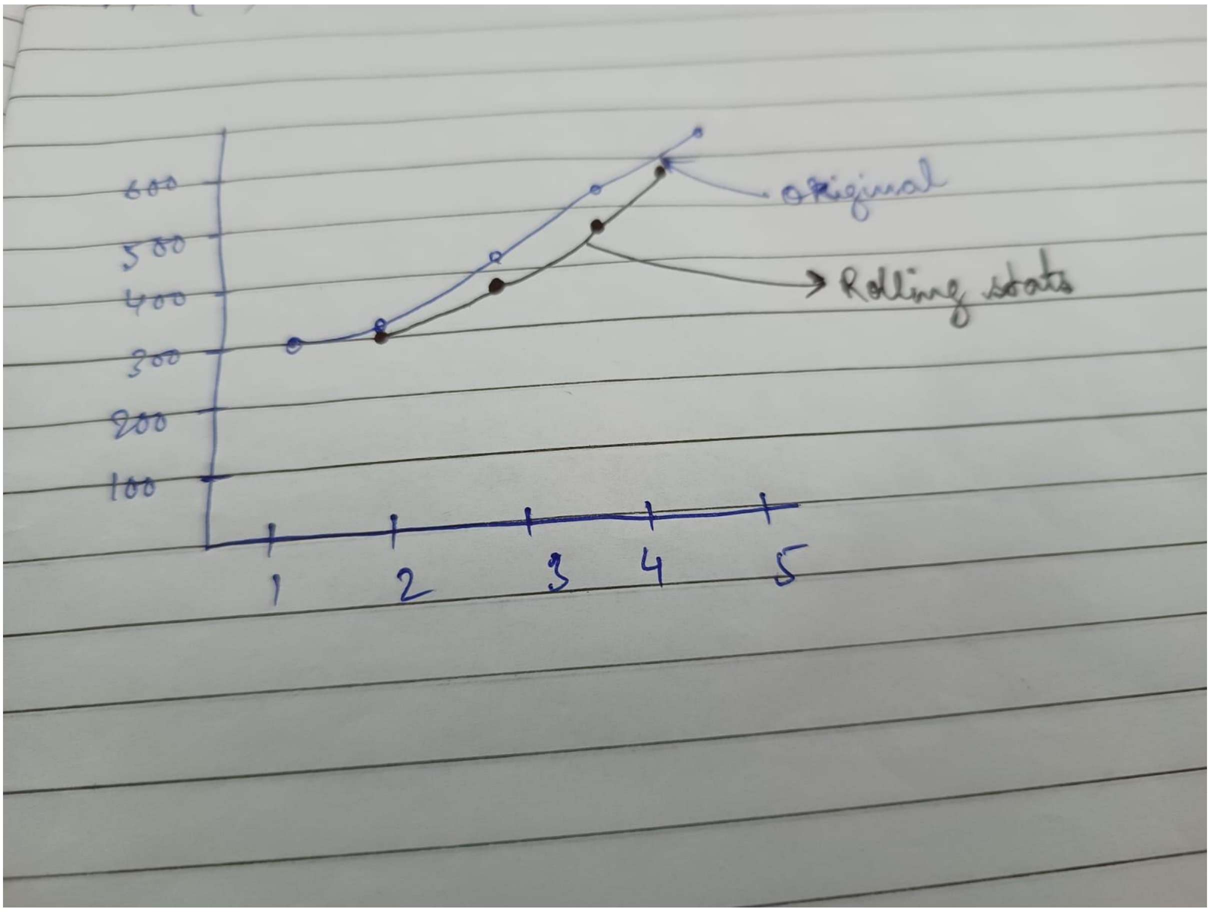
Applying MA(2)

MA(T2) = (300+310)/2 = 610/2 = 305

MA(T3) = (310+420)/2 = 730/2 = 365

And so on….

Following is the graph plotted after moving average technique.



EXPONENTIAL SMOOTHING

T1	T2	T3	T4	T5
10	20	30	40	50
10	20	40	80	160

Time series data can be modelled as addition or product of trend, seasonality, cyclical and irregularity components. So, we decompose time series data into multiplicative and additive and visualize the seasonal and trend components that they have extracted.

Additive time series is a combination(addition) of trend, seasonality, cyclical and Irregularity components.

$Y_t = T_t + S_t + C_t + I_t$

Multiplicative Time series is multiplication of the above terms.

$Y_t = T_t * S_t * C_t * I_t$

Exponential smoothing calculates the moving average by considering more past values and give them weightage as per their occurrence, as recent observation gets more weightage compared to past observation so that the prediction is accurate. hence the formula of exponential smoothing can be defined as.

$Y_t = \alpha * X_t + (1-\alpha) * Y_{t-1}$

where, α is the smoothing constant (hyperparameter that defines the weightage to give) and its value lies between 0 and 1. Y_t is the forecasted value at time t using actual value X_t at time t and forecasted value Y_{t-1} at time (t-1).

This is known as **simple exponential smoothing**.

T1	T2	T3	T4	T5
10	20	30	40	50

In the above time-series data,

$Y_5 = \alpha * 50 + (1-\alpha) * \ldots$ (observation at T4 and T3)

$= 0.8 * 50 + 0.2 * \ldots$

So, in this way, we are giving more weightage to the most recent observation and less weightage to the previous observation and also taking into account all the previous observations.

We need to capture trend and seasonality components so there is **double exponential smoothing** which is used to capture the trend components. Double exponential smoothing is used in time-series forecasting when the data has a linear trend but no seasonal pattern. Only a little bit of modification in the above equation is there. The method supports trends that change in additive ways (smoothing with linear trend) and trends that change in multiplicative ways (smoothing with exponential trend).

$Y_t = \alpha * X_t + (1-\alpha)\ (y_{t-1} + b_{t-1})$ #trend component

where, $b_t = \text{beta} * (Y_t – Y_{t-1}) + (1-\text{beta}) * b_{t-1}$

[Where b_t=best estimate of the trend at time t

and β = trend smoothing factor; $0 < \beta <1$]

Triple exponential smoothing is more reliable for parabolic trends or data that shows trends and seasonality. So, it is the variation of exponential smoothing that's most advanced and is used for time series forecasting when the data has linear trends and seasonal patterns.

When to use Different types of exponential smoothening and why?

1. **Simple Exponential Smoothing (**SES) is usually used to make short term forecasts. It is more effective than SMA because it gives a higher weight to more recent data points vs equal weightage given by **simple moving average** (SMA).
2. SES cannot do longer term forecasts reliably primarily because this method does not consider any trend in the data. Hence, we need to look at extensions of the SES model such as **double exponential smoothening** and **triple exponential smoothening.**

What is meant by exponential smoothing?

Exponential smoothing is a method for forecasting univariate time series data. It is based on the principle that a prediction is a weighted linear sum of past observations or lags. The Exponential Smoothing time series method works by assigning exponentially decreasing weights for past observations. The technique is so called because the weight assigned to each demand observation exponentially decreases.

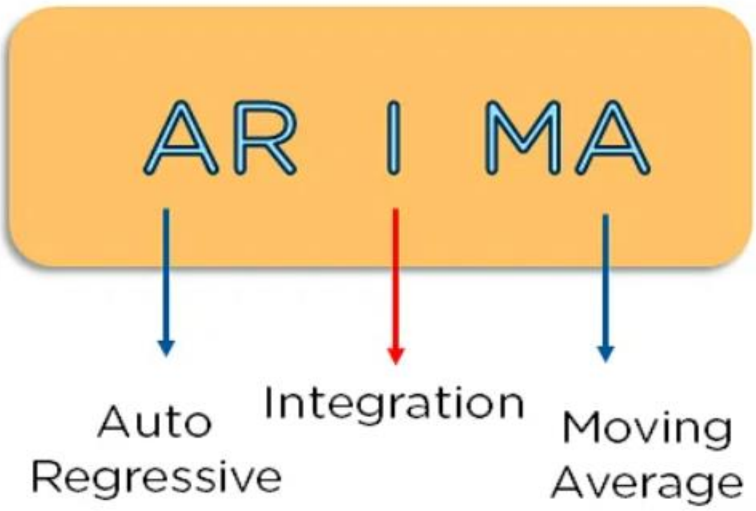
How is exponential smoothing used in forecasting?

Exponential smoothing is a widely preferred forecasting method for smoothing univariate time series data using the exponential window function. The method works by assigning exponentially decreasing weights for past observations. Larger weights are assigned to more recent observations, while exponentially decreasing weights are assigned as the observations get more and more distant.

Exponential smoothing assumes that the future will be somewhat the same as the recent past and, therefore, provides forecasts of time-series data based on prior assumptions by the user, such as seasonality or systematic trends. We can use it most effectively to make short-term forecasts when the time series parameters vary slowly over time.

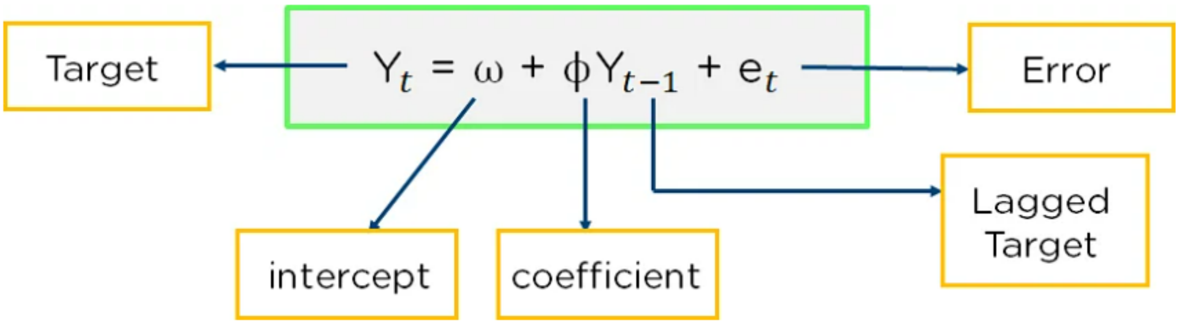
ARIMA MODEL

ARIMA is one of the most popular classical methods for time series forecasting. It stands for autoregressive integrated moving average and is a type of model that forecasts given time series based on its own past values, that is, its own lags and the lagged forecast errors. It can be used for any non-seasonal series of events. e.g., sales data from a clothing store would be a time-series because it was collected over a period of time. One of the key characteristics is the data is collected over a series of constant regular intervals. A modified version can be created to model predictions over multiple seasons. For a period of multiple seasons, the data must be corrected to account for differences between the seasons.



ARIMA(p,d,q) where p,d,q are non-negative integers. It consists of following three components:

- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values. Auto-Regressive models predict future behaviour using past behaviour where there is some correlation between past and future data.



Auto-Regressive Model

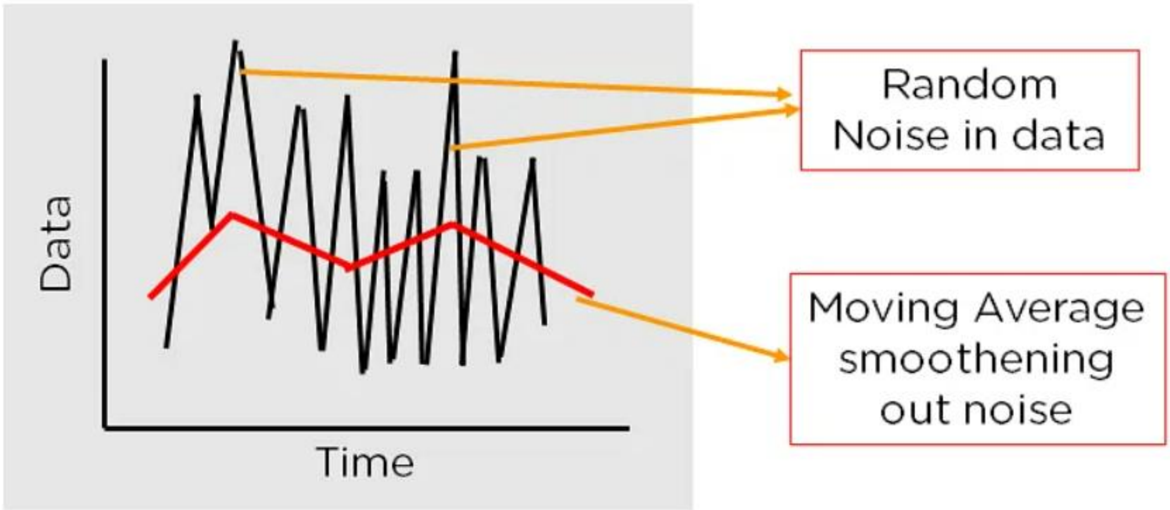
p is the order (number of time lags) of the auto-regressive model.

- Integrated (I): represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values). This means, it is the difference between present and previous observations.

d is the degree of differencing (the number of times the data have had past values subtracted)

- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations. Moving Average is a statistical method that takes the updated average of values to help cut down on noise. It takes the average over a specific interval of time.

q is the order of the moving-average model.



Stationarity using Moving Average

Each of the above acts as a parameter for the ARIMA model. Following are the parameters -

- Previous lagged values for each time point. Derived from the Auto-Regressive Model.
- Previous lagged values for the error term. Derived from the Moving Average.
- Number of times data is differenced to make it stationary. It is the number of times it performs integration.

The "AR" part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior observed) values. The "MA" part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The "I" (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Limitations of Classical models: (Exponential Smoothing models, ARIMA — based, models)

Classical forecasting models have several limitations:

1. Missing values are not supported

2. Assumption of linearity in the relationship. This problem is partly overcome by transforming the data using transformations such as logs, etc

3. These models work on uni-variate data. Most of the models in time series forecasting don't support multiple variables to be taken as inputs

Due to the above limitations there is a strong case for using deep learning techniques in forecasting.

Extras

AR models – Often we forecast a series based solely on the past values in the series called lags. A model that depends only on one lag in the past is called AR model of the order of 1 or AR(1) model.

A statistical model is autoregressive if it predicts future values based on past values.

ACF definition – A function which gives us values of auto-correlation of any series with its lagged values.

PACF definition – An indirect function to find Auto correlation after removing the relationship explained by previous lags.