Charanraj Shetty

Summary

The website provides an overview of time series forecasting models AR

Time Series Models

AR, MA, ARMA, ARIMA

Image by Gerd Altmann from Pixabay

AR, MA, ARMA, and ARIMA models are used to forecast the observation at (t+1) based on the historical data of previous time spots recorded for the same observation. However, it is necessary to make sure that the time series is stationary over the historical data of observation overtime period. If the time series is not stationary then we could apply the differencing factor on the records and see if the graph of the time series is a stationary overtime period.

ACF (Auto Correlation Function)

observations irrespective of its effect on the future or present time period. It calculates the correlation between the t and (t-k) time period. It includes all the lags or intervals between t and (t-k) time periods. Correlation is always calculated using the Pearson Correlation formula.

PACF(Partial Correlation Function)

The PACF determines the partial correlation between time period t and t-k. It doesn't take into consideration all the time lags between t and t-k. For e.g. let's assume that today's stock price may be dependent on 3 days prior stock price but it might not take into consideration yesterday's stock price closure. Hence we consider only the time lags having a direct impact on future time period by neglecting the insignificant time lags in between the two-time slots t and t-k.

How to differentiate when to use ACF and PACF?

Let's take an example of sweets sale and income generated in a village over a year. Under the assumption that every 2 months there is a festival in the village, we take out the historical data of sweets sale and income generated for 12 months. If we plot the time as month then we can observe that when it comes to calculating the sweets sale we are interested in only alternate months as the sale of sweets increases every two months. But if we are to consider the income generated next month then we have to take into consideration all the 12 months of last year.

month.

AR (Auto-Regressive) Model

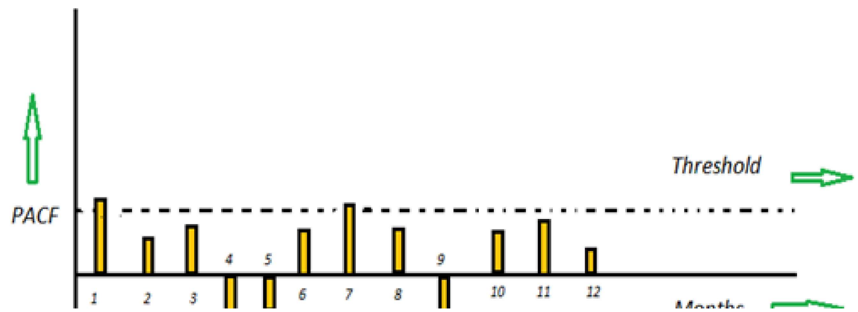

Image by Author

The time period at t is impacted by the observation at various slots t-1, t-2, t-3, ….., t-k. The impact of previous time spots is decided by the coefficient factor at that particular period of time. The price of a share of any particular company X may depend on all the previous share prices in the time series. This kind of model calculates the regression of past time series and calculates model.

$$Yt = \beta_1 * y_{-1} + \beta_2 * y_{t}\text{-}2 + \beta_3 * y_{t}\text{-}3 + \ldots\ldots\ldots\ldots + \beta_k * y_{t}\text{-}_{k}$$

Consider an example of a milk distribution company that produces milk every month in the country. We want to calculate the amount of milk to be produced current month considering the milk generated in the last year. We begin by calculating the PACF values of all the 12 lags with respect to the current month. If the value of the PACF of any particular month is more than a significant value only those values will be considered for the model analysis.

For e.g in the above figure the values 1,2, 3 up to 12 displays the direct effect(PACF) of the milk production in the current month w.r.t the given the lag t. If we consider two significant values above the threshold then the model will be termed as AR(2).

MA (Moving Average) Model

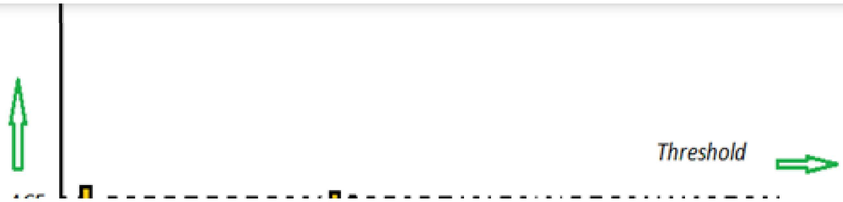

Image by Author

The time period at t is impacted by the unexpected external factors at various slots t-1, t-2, t-3, ….., t-k. These unexpected impacts are known as Errors or Residuals. The impact of previous time spots is decided by the coefficient factor α at that particular period of time. The price of a share of any particular company X may depend on some company merger that happened overnight or maybe the company resulted in shutdown due to bankruptcy. This kind of model calculates the residuals or errors of past time series and calculates the present or future values in the series in know as Moving Average (MA) model.

$$\mathbf{Yt = \alpha_1 * \varepsilon_{t-1} + \alpha_2 * \varepsilon_{t-2} + \alpha_3 * \varepsilon_{t-3} + \ldots\ldots\ldots\ldots + \alpha_k * \varepsilon_{t-k}}$$

that your mom asks you to bring pastries to the party. Every year you miss judging the no of invites to the party and end upbringing more or less no of cakes as per requirement. The difference in the actual and expected results in the error. So you want to avoid the error for this year hence we apply the moving average model on the time series and calculate the no of pastries needed this year based on past collective errors. Next, calculate the ACF values of all the lags in the time series. If the value of the ACF of any particular month is more than a significant value only those values will be considered for the model analysis.

For e.g in the above figure the values 1,2, 3 up to 12 displays the total error(ACF) of count in pastries current month w.r.t the given the lag t by considering all the in-between lags between time t and current month. If we consider two significant values above the threshold then the model will be termed as MA(2).

ARMA (Auto Regressive Moving Average) Model

Image by Author

This is a model that is combined from the AR and MA models. In this model, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series. Here β represents the coefficients of the AR model and α represents the coefficients of the MA model.

$$\mathbf{Yt = \beta_1 * y_{t-1} + \alpha_1 * \varepsilon_{t-1} + \beta_2 * y_{t-2} + \alpha_2 * \varepsilon_{t-2} + \beta_3 * y_{t-3} + \alpha_3 * \varepsilon_{t-3} + \ldots\ldots\ldots\ldots + \beta_k * y_{t-k} + \alpha_k * \varepsilon_{t-k}}$$

Consider the above graphs where the MA and AR values are plotted with their respective significant values. Let's assume that we consider only 1 significant value from the AR model and likewise 1 significant value from the MA model. So the ARMA model will be obtained from the combined values of the other two models will be of the order of ARMA(1,1).

Image by Author

We know that in order to apply the various models we must in the beginning convert the series into Stationary Time Series. In order to achieve the same, we apply the differencing or Integrated method where we subtract the t-1 value from t values of time series. After applying the first differencing if we are still unable to get the Stationary time series then we again apply the second-order differencing.

The ARIMA model is quite similar to the ARMA model other than the fact that it includes one more factor known as Integrated(I) i.e. differencing which stands for I in the ARIMA model. So in short ARIMA model is a in order to forecast future values.

Consider the above graphs where the MA and AR values are plotted with their respective significant values. Let's assume that we consider only 1 significant value from the AR model and likewise 1 significant value from the MA model. Also, the graph was initially non-stationary and we had to perform differencing operation once in order to convert into a stationary set. Hence the ARIMA model which will be obtained from the combined values of the other two models along with the Integral operator can be displayed as ARIMA(1,1,1).

Conclusion :

All these models give us an insight or at least close enough prediction about any particular time series. Also, it depends on the users that which model perfectly suffices their needs. If the chances of error rate are less in any one model compared to other models then it's preferred that we choose the one which gives us the closest estimation.

Hope this article helps you to understand things better !!

Recommended from ReadMedium