Introduction to ARIMA

Zeham Management Technologies BootCamp by SDAIA
August 20th, 2024

SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority

Introduction to ARIMA

Let’s start together…

Agenda

- AR model
- MA model
- ARIMA model
- ETS model

AR MODEL

AR Model

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) , widely used in various fields such as finance, economics, and climate science, AR models help in understanding and predicting temporal data by .modelling the current value of a time series as a function of its previous values

AR Model

Auto-Regressive Model

Autoregressive models are a type of time series model. The fundamental concept is that the current value of a time series can be represented as a linear combination of its past values, along with some .random noise

:Mathematically, an autoregressive model of order p, denoted as AR(p), is expressed as

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t$$

:Where

- X_t is the value at time t.
- c is a constant.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the model parameters.
- $X_{t-1}, X_{t-2}, \ldots, X_{t-p}$ are the lagged values.
- ε_t represents white noise (random error) at time t.

AR Model

Autocorrelation (ACF) in Autoregressive Models

Autocorrelation, often referred to as the Autocorrelation Function (ACF), is a crucial concept in time series analysis and autoregressive models. It measures the correlation between a time series and its lagged versions. In autoregressive models, autocorrelation indicates the degree to which the current .value of a time series is related to its past values at various time lags

AR Model

Autocorrelation (ACF) in Autoregressive Models

Let’s a breakdown of the concept of autocorrelation in autoregressive models:

- Autocorrelation involves calculating the correlation between a time series and a lagged version of itself. The "lag" represents the number of time units by which the series is shifted. For example, a lag of 1 compares the series with its previous time step, while a lag of 2 compares it with the time step before that, and so on. Lag values help calculate autocorrelation, which measures how each observation in a time series is related to previous observations.

- The autocorrelation at a particular lag provides insights into the temporal dependence of the data. If the autocorrelation is high at a certain lag, it indicates a strong relationship between the current value and the value at that lag. Conversely, if the autocorrelation is low or close to zero, it suggests a weak or no relationship.

AR Model

Autocorrelation (ACF) in Autoregressive Models

- To visualize autocorrelation, a common approach is to create an ACF plot. This plot displays the autocorrelation coefficients at different lags, with the horizontal axis representing the lag and the vertical axis representing the autocorrelation values. Significant peaks or patterns in the ACF plot can reveal the underlying temporal structure of the data.

- Autocorrelation plays a pivotal role in autoregressive models. In an autoregressive model of order p, the current value of the time series is expressed as a linear combination of its past p values, with coefficients determined through methods like least squares or maximum likelihood estimation. The selection of the lag order (p) in the AR model often relies on the analysis of the ACF plot.

- Autocorrelation can also be used to assess whether a time series is stationary. In a stationary time series, autocorrelation should gradually decrease as the lag increases. Deviations from this behaviour might indicate non-stationarity.

AR Model

Autocorrelation (ACF) in Autoregressive Models

:Types of Autoregressive Models

:AR(1) Model

- In the AR(1) model, the current value depends only on the previous value.
- It is expressed as $X_t = c + \phi_1 X_{t-1} + \varepsilon_t$

:AR(p) Model

- The general autoregressive model of order p includes p lagged values.
- It is expressed as shown in the introduction.

AR Model

Benefits and Drawbacks of Autoregressive Models

Benefits of Autoregressive Models:

- **Simplicity**: AR models are relatively simple to understand and implement. They rely on past values of the time series to predict future values, making them conceptually straightforward.
- **Interpretability**: The coefficients in an AR model have clear interpretations. They represent the strength and direction of the relationship between past and future values, making it easier to derive insights from the model.
- **Useful for Stationary Data**: AR models are well-suited for stationary time series data, which have stable statistical properties over time. This is a key assumption underlying AR models.
- **Efficiency**: AR models can be computationally efficient, especially for short time series or when you have a reasonable amount of data.
- **Modelling Temporal Patterns**: AR models are effective at capturing short-term temporal dependencies and patterns in the data, making them valuable for short-term forecasting.

AR Model

Benefits and Drawbacks of Autoregressive Models

Drawbacks of Autoregressive Models:

- **Stationarity Assumption**: AR models assume that the time series is stationary, meaning its statistical properties do not change over time. In practice, many real-world time series are non-stationary and require preprocessing steps like differencing.
- **Limited to Short-Term Dependencies**: AR models are not well-suited for capturing long-term dependencies in data. They are primarily designed for modelling short-term temporal patterns.
- **Lag Selection**: Choosing the appropriate lag order (p) in an AR model can be challenging. Selecting too few lags may lead to underfitting, while selecting too many may lead to overfitting. Techniques like ACF and PACF plots are used to determine the optimal lag order.

AR Model

Benefits and Drawbacks of Autoregressive Models

Drawbacks of Autoregressive Models:

- **Sensitivity to Noise:** AR models can be sensitive to random noise in the data, which can lead to overfitting, especially when dealing with noisy or irregular time series.
- **Limited Forecast Horizon:** AR models are generally not suitable for long-term forecasting as they are designed for capturing short-term dependencies. For long-term forecasting, other models like ARIMA, SARIMA, or machine learning models may be more appropriate.
- **Data Quality Dependence:** The effectiveness of AR models is highly dependent on data quality. Outliers, missing values, or data irregularities can significantly affect the model's performance.

MA MODEL

MA Model
Moving Average

Moving Average (MA) Models are a type of time series analysis model commonly used in econometrics to forecast trends and understand patterns in time series data

In moving average (MA) models, the current value of the time series is determined by a linear combination of the past random error terms (white noise) of the time series. In time series analysis, the order of the moving average model is denoted by the letter "q," which indicates how many past error terms influence :the present value. Therefore, a moving average model of order q can be represented as

$$\mathbf{X_t} = c + \epsilon_t + \theta_1.\epsilon_{t-1} + \theta_2.\epsilon_{t-2} + \ldots + \theta_q.\epsilon_{t-q}$$

:where
.is the value of the time series at time t
$\mathrm{X_t}$ is the mean of the series.
c is the white noise error term at time t.
ϵ_t θ_2, ..., θq are the coefficients of the model.

MA MODEL

Concept Related to Moving Average:

:Now let's discuss some concepts that can help us better understand the moving average model

- Stationarity: Stationarity is a key principle in time series data, indicating that the statistical properties of the data do not change over time. This means that the mean remains constant or fluctuates around a certain value, the standard deviation remains nearly constant, and there is no seasonality or periodic behaviour in the data. Stationarity can be assessed visually or through the Augmented Dickey-Fuller (ADF) Test. It is considered one of the most important aspects of time series data for accurate modelling, as many time series models require the data to be stationary.

- Differencing: Differencing is a crucial step in time series analysis. If the original time series data is not stationary and exhibits significant trends, differencing should be considered to achieve stationarity. Accurate time series analysis requires stationary data. In regular differencing, the current data point is subtracted by the previous data point, denoted as $\Delta y_t = y_t - y_{t-1}$ .This method removes trends from the data, making it more suitable for modelling.

MA MODEL

Concept Related to Moving Average:

:Now let's discuss some concepts that can help us better understand the moving average model

- White Noise: White noise refers to an error term with a mean of zero, a constant standard deviation, and no correlation between data points. In time series modelling, white noise serves as a benchmark in the forecasting process. If the forecast error is not white noise, further modifications to the model are necessary. However, if the forecast errors reach a state of white noise, no further improvements are needed. The values in a white noise series are random and unpredictable, making it impossible to model or forecast a time series that is purely white noise.

- **ACF Plot**: Autocorrelation Function plot or the ACF plot is the plot of correlation between the time series and its lagged version. It shows how similar the time series is with it’s different **lagged** values. Here the lag term is a fixed time displacement, in the ACF plot the x-axis is the lagged time series and the y-axis is the correlation which ranges from -1 to 1.

MA MODEL

Benefits and Drawbacks of Moving Average Models

Benefits of Moving Average Models:

- **Simplicity**: MA models are straightforward and easy to understand. They model the time series as a linear combination of past error terms.
- **Stationarity** : MA models are inherently stationary, making them suitable for time series that do not exhibit trends or seasonality.
- **Effective for Short-Term Forecasting**: MA models can effectively capture the short-term dependencies and patterns in the data, making them useful for short-term forecasting.
- **Handling of Noise**: By modelling the time series as a function of past error terms, MA models can handle and smooth out random noise in the data, providing clearer insights into the underlying patterns.
- **Good for Series with Strong Autocorrelation** :MA models are particularly effective when the time series exhibits strong autocorrelation at lagged periods.
- **Flexibility**: MA models can be combined with other models (like AR models) to form ARMA or ARIMA models, providing more flexibility in modelling various types of time series data.

MA MODEL

Benefits and Drawbacks of Moving Average Models

Drawbacks of Moving Average Models:

- **Lag Dependency**: MA models depend heavily on the past error terms, which can lead to the propagation of errors over time. This can make the model less reliable if the errors are not truly random.
- **Parameter Estimation**: Estimating the parameters of an MA model can be more complex compared to AR models, especially for higher-order MA processes. This often requires iterative methods like Maximum Likelihood Estimation.
- **Choice of Order (q)**: Determining the appropriate order of the MA model (the value of q) can be challenging and may require extensive diagnostic checks and model selection criteria.
- **Limited Long-Term Forecasting**: MA models are more suited for short-term forecasting and may not perform well for long-term forecasts due to their reliance on recent error terms.
- **Not Suitable for Trend or Seasonal Data**: MA models assume stationarity and may not be appropriate for time series data that exhibit trends or seasonal patterns unless combined with other methods to address these components.
- **Complexity with Higher Order Models**: Higher-order MA models (larger values of q) can become increasingly complex to interpret and estimate, leading to potential overfitting issues.

MA MODEL
Practical Application

MA models are used in various fields for time series analysis and forecasting. They are often applied to financial data, economic indicators, and other areas where understanding and predicting temporal dependencies is essential.

- **Finance**: Modelling and forecasting stock prices, returns, and volatility.
- **Economics**: Analysing and predicting economic indicators like GDP growth, inflation rates, etc.
- **Engineering**: Filtering and smoothing signals in control systems.

MA MODEL

The Moving Average (MA) process is a powerful tool in time series analysis that models the dependency between a time series value and past white noise error terms. The MA(q) model, characterized by its order q, is used to capture the temporal structure and make forecasts based on past errors.

Understanding and applying MA models require careful identification, estimation, and validation to ensure accurate and reliable results.

Let's Practice

TOTURIAL

ARIMA

ARIMA MODEL

Before diving into ARIMA, it's essential to understand the following concepts:

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

Auto-Correlation Function (ACF) :

The ACF measures the similarity between a value in a time series and its previous values. In other words, it quantifies the degree of correlation between a time series and its lagged versions at different intervals.

The Python Statsmodels library provides tools to calculate autocorrelation, helping to identify patterns in the data and understand how past observations influence the current values.

ARIMA MODEL

Auto-Correlation Function (ACF) :

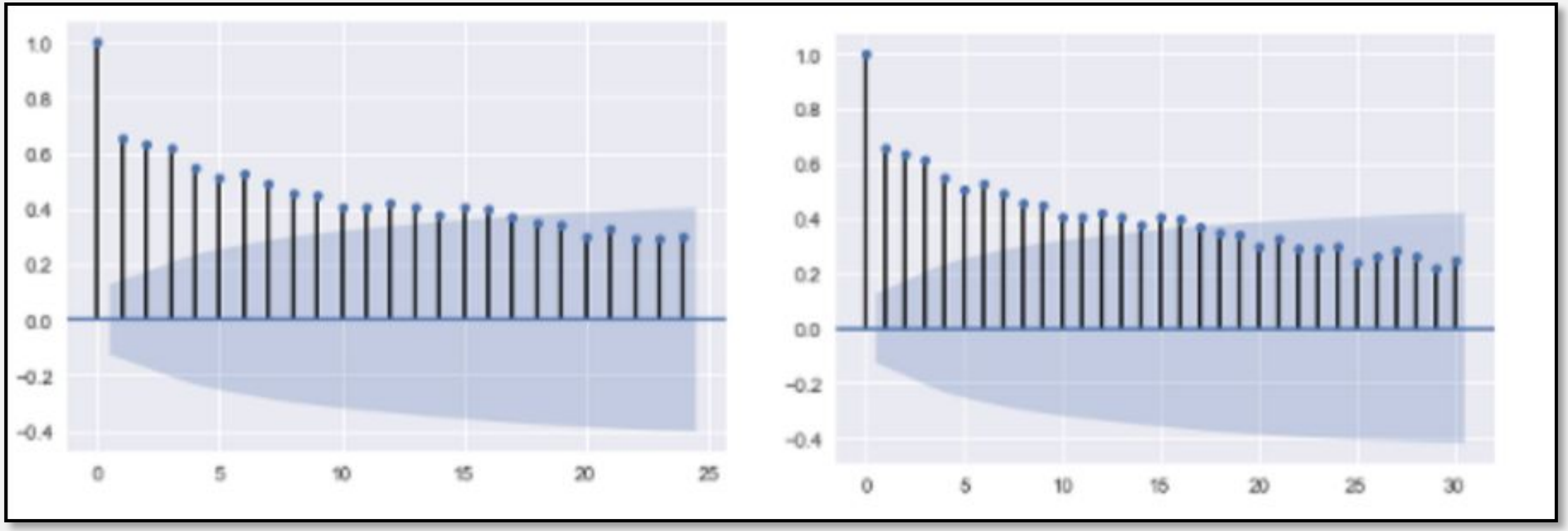
The ACF measures the similarity between a value in a time series and its previous values. In other words, it quantifies the degree of correlation between a time series and its lagged versions at different intervals. The Python Statsmodels library provides tools to calculate autocorrelation, helping to identify patterns in the data and understand how past observations influence the current values.

Partial Auto-Correlation Function (PACF) :

Similar to the Auto-Correlation Function but is a bit more complex to grasp. PACF represents the correlation of a time series with its own lagged values while isolating the direct effects by removing the influence of any intermediate lags. This helps in showing only the direct relationship between the time series and its previous values over specific time intervals.

ARIMA MODEL

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



ARIMA Model for Time Series Forecasting

AR(p) Autoregression : This is a regression model that captures the dependent relationship between a current observation and observations from previous time periods. In an autoregressive (AR(p)) model, past values are used in the regression equation to predict future values of the time series. Here, \(p \) denotes the number of lagged observations included in the model.

ARIMA Model for Time Series Forecasting

I(d) Integration : Integration involves differencing observations to make the time series stationary, which means the statistical properties of the series like mean and variance do not change over time. Differencing is performed by subtracting the current value of the series from its previous value, and this process can be repeated d times to achieve stationarity.

ARIMA Model for Time Series Forecasting

MA(q) Moving Average : This model leverages the dependency between an observation and the residual errors from a moving average model applied to lagged observations. The moving average (MA(q)) component treats the error of the model as a combination of previous error terms. The order \(q \) specifies the number of past error terms included in the model.

ARIMA MODEL

ARMA is effective for predicting stationary time series, while ARIMA was developed to handle both stationary and non-stationary data.

- AR (AutoRegressive) : Uses past values to predict future values.
- MA (Moving Average) : Utilizes past error terms to forecast future values.

The "I" (Integrated) component of ARIMA applies differencing to the data, making it stationary.

Together, AR + I + MA = ARIMA.

ARIMA Components:

- p (AutoRegressive order) : The number of lag observations included in the model.
- d (Degree of Differencing) : The number of times the raw observations are differenced to make the data stationary.
- q (Order of Moving Average) : The size of the moving average window

Types of ARIMA Model

- ARIMA:Non-seasonal Autoregressive Integrated Moving Averages
- SARIMA:Seasonal ARIMA
- SARIMAX:Seasonal ARIMA with exogenous variables

ARIMA MODEL

Implementation Steps for ARIMA :

1. Plot the Time Series Data : Start by visualizing the time series to understand its behavior.
2. Apply Differencing : Make the series stationary on the mean by removing the trend.
3. Log Transformation : Apply a log transform to stabilize the variance and make the data stationary.
4. Difference the Log-Transformed Data : Ensure the data is stationary in both mean and variance.

ARIMA MODEL

Implementation Steps for ARIMA :

5. Plot ACF and PACF : Analyze the AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) to identify potential AR and MA models.
6. Identify the Best-Fit ARIMA Model : Discover the ARIMA model that best fits the data.
7. Forecast with ARIMA : Use the best-fit ARIMA model to predict future values.
8. Residual Analysis : Plot ACF and PACF for the residuals of the ARIMA model to confirm that no significant information remains unmodeled.

Pyramid Auto-ARIMA

The ‘auto_arima’ function from the ‘pmdarima’ library helps us to identify the most optimal parameters for an ARIMA model and returns a fitted ARIMA model.

Let's practice

Tutorial :

5- Time Series Forecasting/3- Time Series Forecasting(ARIMA)/LAB/ARIMA_Tutorial.ipynb

Datasets :

5- Time Series Forecasting/3- Time Series Forecasting(ARIMA)/LAB/Datasets/airline-passengers.csv

Let's practice

Exercise :

5- Time Series Forecasting/3- Time Series Forecasting(ARIMA)/
LAB/ARIMA_Exercise.ipynb

Datasets :

5- Time Series Forecasting/3- Time Series Forecasting(ARIMA)/
LAB/Datasets/traffic.csv

Thank you!

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority