Day 58

DIY Solution

Q1. Explain the types of Time Series Methods Used for Forecasting.

Answer: Times series methods refer to different ways to measure timed data. Common types include Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving-Average.

The important thing is to select the appropriate forecasting method based on the characteristics of the time-series data.

Smoothing-Based Models: In time series forecasting, data smoothing is a statistical technique that removes outliers from a time series data set to make a pattern more visible. Inherent in collecting data taken over time is some form of random variation. Smoothing data removes or reduces random variation and shows underlying trends and cyclic components.

Moving-Average Model: In time series analysis, the MA model, also known as the MA process, is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

Together with the AR model (covered below), the moving-average model is a particular case and critical component of the more general ARMA and ARIMA models of time series, which have a more complicated stochastic structure.

Contrary to the AR model, the finite MA model is always stationary.

Exponential Smoothing Model: Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Exponential smoothing is a quickly learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Different types of exponential smoothing include single exponential smoothing, double exponential smoothing, and triple exponential smoothing (also known as the Holt-Winters method).

In single exponential smoothing, forecasts are given by:

$\hat{Y}_{(t+h|t)} = \alpha y_{(t)} + \alpha(1-\alpha)y_{(t-1)} + \alpha(1-\alpha)^2 y_{(t-2)} + \dots$

with $0 < \alpha < 1$.

Triple Exponential Smoothing or Holt-Winters is mathematically similar to Single Exponential Smoothing, except that the seasonality and trend are included in the forecast.

Q2. What is ACF?

Answer: Autocorrelation is the relationship between two values in a time series. Put another way, the time series data are correlated, hence the word. “Lags” are the term for these kinds of connections. When a characteristic is measured on a regular basis as daily, monthly, or yearly, time-series data is created.

The number of intervals between two measurements is known as the lag. For example, there is a one-second lag between current and past observations. The lag grows to two if you go back to another interval, and so on.

The observations at y_t and y_{t-k} are separated by k time units in mathematical terms K. Depending denotes the lag on the nature of the data. This lag can be measured in days, quarters, or years. When $k = 1$, you're evaluating observations next to each other.

There is a correlation with each latency. The autocorrelation function (ACF) evaluates the correlation between observations in a time series over a given range of lags. Corr(y_t, y_{t-k}), $k = 1, 2,$…. provides the ACF for the time series y. We generally use graphs to demonstrate this function.

The ACF can determine a time series' randomness and stationarity. You may also examine if there are any seasonal patterns or tendencies. In an ACF plot, each bar represents the size and direction of the connection. Bars that cross the red line are statistically significant.

Q3. What is PACF?

Answer: The partial autocorrelation function (PACF), like the ACF, indicates only the association between two data that the shorter lags between those observations do not explain. The partial autocorrelation for lag 3 is, for example, merely the correlation that lags 1 and 2 do not explain. In other words, the partial correlation for each lag is the striking correlation between the two observations after the intermediate correlations have been removed.

As previously stated, the autocorrelation function aids in determining the qualities of a time series. The PACF, on the other hand, is more beneficial during the definition phase for an autoregressive model. Partial autocorrelation plots can specify regression models with time series data and Auto-Regressive Integrated Moving Average models.

Q4. What us ARIMA and ARIMAX model?

Answer: It stands for "Auto-Regressive Integrated Moving Average", a set of models that defines a given time series based on its initial values, lags, and lagged forecast errors, so that equation is used to forecast forecasted values.

We have non-seasonal time series that manifests patterns and is not a stochastic white noise that can be molded with ARIMA models.

Three terms delineate an ARIMA model: *p*, *d*, *q*, where,

- *p* is a particular order of the AR term.
- *q* is a specific order of the MA term.
- *d* is the number of differences wanted to make the time series stationary.

If a time series has seasonal patterns, you require adding seasonal terms, and it converts to SARIMA, which stands for "Seasonal ARIMA".

The "Auto-Regressive" in ARIMA indicates a linear regression model that employs its lags as predictors. Linear regression models work best if the predictors are not correlated and remain independent of each other. We want to make them stationary, and the standard approach is to differentiate them. This means subtracting the initial value from the current value. Concerning how complex the series gets, more than one difference may be required.

Hence, the value of d is the merest number of differences necessitated to address the series stationary. If we already have a stationary time series, we proceed with d as zero.

"Auto-Regressive (AR)" term is indicated by "*p*". This relates to the number of lags of *Y* to be adopted as predictors. "Moving Average" (MA) term is associated with "*q*". This relates to the number of lagged prediction errors that should conform to the ARIMA

Model.

An Auto-Regressive (AR only) model has Y_t that depends exclusively on its lags. Such, Y_t is a function of the "lags of Y_t".

$$Y_t = \alpha + \epsilon_t + \emptyset_1 \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \cdots + \emptyset_q \epsilon_{t-q}$$

Furthermore, a Moving Average (MA only) model has Y_t that depends mainly on the lagged forecast errors.

The time series differencing in an ARIMA model is differenced at least once to make sure it is stationary, and we combine the AR and MA terms. Hence, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} \epsilon_t + \emptyset_1 \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \cdots + \emptyset_q \epsilon_{t-q}$$

We have continued operating by manually fitting various models and determining the best one. Therefore, we transpire to automate this process. It uses the data and fits several models in a different order before associating the characteristics. Nevertheless, the processing rate increases considerably when we seek to provide complicated models. This is how we move for Auto-ARIMA models.

Q5. Top five applications of Time Series.

Answer:

1. Time Series in Financial and Business Domain

Most financial, investment, and business decisions are considered based on future changes and demand forecasts in the economic domain.

Time series analysis and forecasting are essential processes for explaining financial markets' dynamic and influential behavior. By examining economic data, an expert can predict required forecasts for critical financial applications in several areas such as risk evolution, option pricing & trading, portfolio construction, etc.

For example, time series analysis has become the intrinsic part of financial analysis. It can predict interest rates, foreign currency risk, volatility in stock markets, etc. Policymakers and business experts use financial forecasting to make decisions about production, purchases, market sustainability, allocation of resources, etc.

In investment, this analysis is employed to track the price fluctuations and price of a security over time. For instance, the price of a security can be recorded.

- For the short term, such as the observation per hour for a business day, and
- For a long time, such as observation at the month end for five years.

Time series analysis is instrumental to observe how a given asset, security, or economic variable behaves/changes over time. For example, it can be deployed to evaluate how the underlying changes associated with some data observation behave after shifting to other data observations in the same period.

2. Time Series in Medical Domain

Medicine has evolved as a data-driven field and continues to contribute to human knowledge with enormous developments in time series analysis.

Consider the case of combining time series with a medical method of case-based reasoning and data mining. These synergies are essential as the pre-processing for feature mining from time-series data can help study patients' progress over time.

In the medical domain, it is essential to examine the transformation of behavior over time to derive inferences depending on the absolute values in the time series. For example, diagnosing heart rate variability in occurrence with respiration based on the sensor readings is the characteristic illustration of connecting time series with case-based monitoring.

However, time series in the context of the epidemiology domain have emerged very recently and incrementally as time series analysis approaches demand recordkeeping systems such that records should be connected over time and collected precisely at regular intervals.

As soon as the government has placed sufficient scientific instruments on accumulating good and lengthy temporal data, healthcare applications using time series analysis have resulted in massive prognostication for the industry and individuals’ health diagnoses.

Medical Instruments

Time series analysis has made its way into medicine with the advent of medical devices such as

- Electrocardiograms, invented in 1901, diagnose cardiac conditions by recording the electrical pulses passing through the heart.
- Electroencephalogram, developed in 1924, measures electrical activity/impulses in the brain.

These inventions made more opportunities for medical practitioners to deploy time series for medical diagnosis.

With the advent of wearable sensors and intelligent electronic healthcare devices, now persons can take regular measurements automatically with minimal inputs, resulting in a good collection of longitudinal medical data for both sick and healthy individuals consistently.

3. Time Series in Astronomy

One of the contemporary and modern applications where time series play a significant role are different areas of astronomy and astrophysics, Being specific in its domain, astronomy hugely relies on plotting objects, trajectories, and accurate measurements. Due to the same, astronomical experts are proficient in time series, calibrating instruments, and studying things of their interest.

Time series data had an intrinsic impact on knowing and measuring anything about the universe. It has a long history in the astronomy domain. For example, sunspot time series were recorded in China in 800 BC, making sunspot data collection a well-recorded natural phenomenon.

Similarly, in past centuries, time series analysis was used

- To discover variable stars that are used to surmise stellar distances, and
- To observe transitory events such as supernovae to understand the mechanism of changing the universe with time.

Such tools result from constant monitoring of live streaming of time series data depending upon the wavelengths and intensities of light that allow astronomers to catch events.

In the last few decades, data-driven astronomy introduced novel research areas such as astroinformatics and astrostatistics; these paradigms involve significant disciplines such as statistics, data mining, machine learning, and computational intelligence. And here, the role of time series analysis would be detecting and classifying astronomical objects swiftly along with the characterization of novel phenomena independently.

4. Time Series in Forecasting Weather

Anciently, the Greek philosopher Aristotle researched weather phenomena to identify the causes and effects of weather changes. Later, scientists started accumulating weather-related data using the instrument “barometer” to compute the state of atmospheric conditions; they recorded weather-related data hourly or daily and kept them in different locations.

With the time, customized weather forecasts began to be printed in newspapers, and later, with the advancement in technology, currently forecasts are beyond the general weather conditions.

To conduct atmospheric measurements with computational methods for fast compilations, many governments have established thousands of weather forecasting stations worldwide.

These stations are equipped with highly functional devices. They are interconnected to accumulate weather data at different geographical locations and forecast weather conditions at every bit of time as per requirements.

5. Time Series in Business Development

Time series forecasting helps businesses to make informed business decisions; as the process analyzes past data patterns, it can help forecast future possibilities and events in the following ways.

Reliability: When the data incorporates a broad spectrum of time intervals in massive observations for a more extended period, time series forecasting is highly reliable. It provides elucidated information by exploiting data observations at various time intervals.

Growth: To evaluate the overall financial performance and growth as well as endogenous, the time series is the most suitable asset. Endogenous growth is the progress within organizations' internal human capital resulting in economic growth. For example, studying the impact of any policy variables can be manifested by applying time series forecasting.

Trend estimation: Time series methods can be conducted to discover trends; for example, these methods inspect data observations to identify when measurements reflect a decrease or increase in sales of a particular product.

Seasonal patterns: Recorded data points variances could unveil seasonal patterns & fluctuations that act as a base for data forecasting. The obtained information is significant for markets whose products fluctuate seasonally and assist organizations in planning product development and delivery requirements.