Week 14: Daily Morning Challenge

Day 1: Tuesday 07th April 2020

Question 1: Briefly describe the core components of a time series data

The trend is the long term pattern of a time series. A trend can be positive or negative depending on whether the time series exhibits an increasing long term pattern or a decreasing long term pattern. If a time series does not show an increasing or decreasing pattern then the series is stationary in the mean. Time series is the arrangement of data in accordance with their time of occurrence. It is the chronological arrangement of data. Here, time is just a way in which one can relate the entire phenomenon to suitable reference points. Time can be hours, days, months or years. A time series depicts the relationship between two variables. Time is one of those variables and the second is any quantitative variable. It is not necessary that the relationship always shows increment in the change of the variable with reference to time. The relation is not always decreasing too.

The components of a time series data are:

Trend:

The trend is the main component of a time series which results from long term effects of socio-economic and political factors. This trend may show the growth or decline in a time series over a long period. This is the type of tendency which continues to persist for a very long period. Prices and export and import data, for example, reflect obviously increasing tendencies over time.

Seasonality:

These are short term movements occurring in data due to seasonal factors. The short term is generally considered as a period in which changes occur in a time series with variations in weather or festivities. For example, it is commonly observed that the consumption of ice-cream during summer is generally high and hence an ice-cream dealer's sales would be higher in some months of the year while relatively lower during winter months. Employment, output, exports, etc., are subject to change due to variations in weather. Similarly, the sale of garments, umbrellas, greeting cards and fire-works are subject to large variations during festivals like Valentine's Day, Eid, Christmas, New Year's, etc. These types of variations in a time series are isolated only when the series is provided biannually, quarterly or monthly. Seasonality occurs when the time series exhibits regular fluctuations during the same month (or months) every year, or during the same quarter every year. For instance, retail sales peak during the month of December.

Cyclicity:
These are long term oscillations occurring in a time series. These oscillations are mostly observed in economics data and the periods of such oscillations are generally extended from five to twelve years or more. These oscillations are associated with the well-known business cycles. These cyclic movements can be studied provided a long series of measurements, free from irregular fluctuations, is available. Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern. The duration of a cycle depends on the type of business or industry being analysed.

Irregularity:

These are sudden changes occurring in a time series which are unlikely to be repeated. They are components of a time series which cannot be explained by trends, seasonal or cyclic movements. These variations are sometimes called residual or random components. These variations, though accidental in nature, can cause a continual change in the trends, seasonal and cyclical oscillations during the forthcoming period. Floods, fires, earthquakes, revolutions, epidemics, strikes etc., are the root causes of such irregularities.
This component is unpredictable. Every time series has some unpredictable component that makes it a random variable. In prediction, the objective is to "model" all the components to the point that the only component that remains unexplained is the random component.

Question 2: Briefly explain the following terms for smoothing a time series data

Moving average:

The moving average of a period (extent) m is a series of successive averages of m terms at a time. The data set used for calculating the average starts with first, second, third and etc. at a time and m data taken at a time.

In other words, the first average is the mean of the first m terms. The second average is the mean of the m terms starting from the second data up to (m + 1)th term. Similarly, the third average is the mean of the m terms from the third to (m + 2) th term and so on.

If the extent or the period, m is odd i.e., m is of the form (2k + 1), the moving average is placed against the mid-value of the time interval it covers, i.e., t = k + 1. On the other hand, if m is even i.e., m = 2k, it is placed between the two middle values of the time interval it covers, i.e., t = k and t = k + 1.

When the period of the moving average is even, then we need to synchronize the moving average with the original time period. It is done by centering the moving averages i.e., by taking the average of the two successive moving averages.

Drawbacks of Moving Average

The main problem is to determine the extent of the moving average which completely eliminates the oscillatory fluctuations.

This method assumes that the trend is linear but it is not always the case.

It does not provide the trend values for all the terms.

This method cannot be used for forecasting future trend which is the main objective of the time series analysis.

Exponential smoothing:

Exponential smoothing refers to the use of an exponentially weighted moving average (EWMA) to "smooth" a time series. Exponential Smoothing is a technique for smoothing univariate time-series by assigning exponentially decreasing weights to data over a time period.

Mathematically, the value of variable at time 't+1' given value at time t, y_(t+1|t) is defined as –

yt+1|t=αyt+α(1−α) yt−1+α(1−α) 2yt−2+...+y1yt+1|t=αyt+α(1−α) yt−1+α(1−α) 2yt−2+...+y1

where, 0≤α≤10≤α≤1 is the smoothing parameter, and

y1,....,yty1,....,yt are previous values of network traffic at times 1, 2, 3, ... ,t.

This is a simple method to model a time series with no clear trend or seasonality. But exponential smoothing can also be used for time series with trend and seasonality.