Time series

Introduction

A time series is a set of observations, xi each one being recorded at a specific time t. After being recorded, these data are rigorously studied to develop a model. This model will then be used to construct future values, in other words, to make a forecast. When looking at a time series, some questions must be asked: Does the time series have a trend or seasonality? Are their outliers?

Is there constant variance over time? Essential of Good time seriesThe data must be long enough. There must be equal time gap. There must be a normal period. Example1The following plot is a time series plot of the annual number of earthquakes in the world with seismic magnitude over 7. 0, for 99 consecutive years.

By a time series plot, we simply mean that the variable is plotted against time. Some features of the plot: There is no trend. The mean of the series is 20. 2. There is no seasonality as the data are annual data. There are no outliers. Example 2 This shows a time series of quarterly production of beer in Australia for 18 years. Some features are: There is an increasing trend.

There is seasonality. There are no outliers. The Components of Time SeriesThe components of time series are factors that can bring changes to the time series: Trend component, TtWhen there is an increase or a decrease over a long period of time in the data, then we say that there is a trend. Sometimes, a trend is said to be changing direction when it goes from an increasing trend to a decreasing one. It is the result of events such as price inflation, population growth or economic changes.

Seasonal component, StA seasonal pattern exists when the time series exhibits regular variations at specific time. It arises from influences such as natural conditions or social and cultural behaviors. For example, the sales of ice-cream are relatively high in summer. So, the salesman expects greater profit in summer than in winter.

Cyclic component, CtIf the time series shows an up and down movement around a given period of time, it is said to have a cyclical pattern. Irregular component, ItIrregular components consist of changes that are unlikely to be repeated in a time series. Examples are floods, fires, earthquakes or cyclones. Combining the time series componentsTime series is a combination of the components which were discussed above.

These components can be either combined additively or multiplicatively. Additive modelIt is linear, and the changes are made by the same amount over time. Yt = Tt + Ct + St + ItMultiplicative modelIt is non-linear such as quadratic or exponential, and the changes increase or decrease over time. Yt = Tt ×Ct × St × ItUsesTime series can be useful in the following fields:

StatisticsSignal processingEconometricsMathematical financeAstronomyEarthquake predictionsWeather forecastingImportance of Time series for businessesThere are many benefits of time series for business purposes: Helpful for study of past behaviorBusinessmen use time series to study the past behaviors and to see the trend of the sales or profit of their businesses.

Helpful in forecastingTime series is a great tool for forecasting. Businesses can make a time series of the past strategies of their competitors and make

an estimate of their future strategies. In this way, they make can built a better strategy and make more profits. Helpful in comparisonTime series can be used to calculate the trend of two or more branches of the same company and compare their performance. On their performances, rewards can be given.

However, time series can have some limitations for a business. Sales forecasting relies on the past results to predict future expectations. But, if a company is new, there is a limited amount of data to make predictions. Even so, past results do not always indicate what the future sales will be. To fully understand this topic, we will work out this example.

Example 2We will consider the actual arrival of passengers from an airport over the year 1949 to 1960. From these data, we will make a forecast. The first step is to plot the data and obtain descriptive measures such as trends or seasonal fluctuations. The second step is to check for the stationarity of the time series. StationarityA time series is said to be stationary if its mean and variance does not change over time. Obviously, not all the time series that we encounter are stationary.

It is important because, most of the models we work on, assumes that the time series is stationary. If the time series has the same behavior over time, there will be a high probability that it will follow the same trend in the future. How to check for stationarity? For the graph that was plotted, we can see that it has an increasing trend with some seasonal pattern.

But, it is not always evident to see whether a plot is increasing or has a seasonal trend. We can check for stationarity using the following: Plotting

rolling statisticsWe plot the moving average or variance and see whether it changes with time. But, as it is a visual technique, we will take more consideration for the next test.

Dickey-Fuller testIt is one of the statistical methods to check for stationarity. The null hypothesis is that the time series is non-stationary, and the alternative hypothesis is the converse. As shown below, the test consists of the test statistics and critical values at different significant levels. If the test statistics is less than the critical value, we reject the null hypothesis.

Results of Dickey-Fuller Test: Test Statistic 0. 815369p-value 0. 991880#Lags Used 13. 000000Number of Observations Used 130. 000000Critical Value (1%) -3. 481682Critical Value (5%) -2. 884042Critical Value (10%) -2. 578770According to the Dickey-Fuller test, the test statistics is less than the critical value.

Therefore, the time series is not stationary. However, there are various methods to make a time series stationary. How to make a time series stationary? The assumption of stationarity is very important when modelling a time series, but most of the practical time series are not stationary. Eventually, we cannot make a time series one hundred percent stationary, most of the time, it will be with a confidence of 99%. Before going into detail, we will discuss on the reasons why the time series is not stationary.

There are two major reasons to that, trend and seasonality. Having discuss the reasons, we will now talk about the techniques to make the time series stationary: TransformationLog transformation is probably the most commonly used form of transformation.

DifferencingDifferencing is a widely used method to make the time series stationary. It is performed by subtracting the previousobservationfrom the current one. When making the forecast, the process of differencing must be inverted to convert the data back to its original scale. This can be done by adding the difference value to the previous value.

Using the Dickey-Fuller test we can see that the test statistic is -2. 717131 and that the critical values at 1%, 5% and 10% are -3. 482501, -2. 884398 and -2. 578960 respectivelyThe time series is stationary with 90% confidence. The second or third order differencing can be done to get better results. DecompositionIn decomposition, the time series is divided into several components mainly trend, cyclical, seasonal and irregular components.

The time series can sometimes be broken down into an additive or multiplicative model. We will assume a multiplicative model for our example. Since the trend and seasonality were separated from the residuals, we can check the stationarity of the residuals. Results of Dickey-Fuller Test is test statistic is -6. 332387e+00 and the critical values at 1%, 5% and 10% are -3. 485122e+00, -2. 885538e+00 and -2. 579569e+00 respectively.

We can conclude that the time series is stationary at 99% confidence. Now, we can go forward with the forecasting. Forecasting the time seriesWe will fit this time series using the ARIMA model, ARIMA is an acronym that stands for Autoregressive Integrated Moving Average. It is a linear equation similar to a linear regression.

The first goal is to find the values of the predictors (p, d, q), but before finding these values, two situations in stationarity must be discussed. A strictly stationary series without any dependence among the values. In this case, we can model the residual as white noise. The second case is a series with significant dependency among the values.

The predictors mainly depend on the parameters (p, d, q) of the ARIMA model: Number of AR(Auto-Regressive) terms (p)It is the number of lag observation that were included in the model. This term helps to incorporate the effect of the past values into the model. Number of MA (Moving Average) terms (q)It is the size of the moving average window, that is, this term sets the error of the model as a linear combination of the error values observed at previous time points in the past.

Number of differences(d)The number of times that the raw observations are differenced. In order to obtain the values of p and q, we will use the following two plots: Autocorrelation Function, ACFThis function will measure the correlation of the time series with its lagged version.

Partial Autocorrelation Function, PACFThis function measures the correlation between the time series with a lagged version of itself, controlling the values of the time series at all shorter lagsIn the ACF and PACF plots, the dotted lines are the confidence interval, these values are p and q. The value of p is obtained from the PACF plot and the value of q is obtained from the ACF plot. We can see that both p and q are 2.

Now, that we have obtained p and q, we will make three different ARIMA model: AR, MA and the combined model. The RSS of each of the model will

be given. AR modelMA modelCombined modelFrom the plots, it is clearly shown that the RSS of AR and MA are the same and that of the combined is much better. As the combined model give a better result, the following steps will take the values back to its original scale.

The predicted results are stored. The differencing is converted the log scale. This can be done by adding the differences consecutively to the base numbers. The exponent is taken and is compared to the original scale. Therefore, we have the final result.

References

1. Aarshay Jain(2016) A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python) [WWW] Available from https://www. analyticsvidhya. com/blog/2016/02/time-series-forecasting-codes-python/ [Accessed 14/04/18]Maxime Phillot (2017)

2. How do I interpret the results in an augmented Dickey-Fuller test? [WWW] Available from https://www. quora. com/How-do-I-interpret-the-results-in-an-augmented-Dickey-Fuller-test [Accessed 23/04/18]Jason Brownlee (2016)

3. What Is Time Series Forecasting? [WWW] Available from https://machinelearningmastery. com/time-series-forecasting/ [Accessed 23/04/18]Chris St. Jeor and Sean Ankenbruck (2018)

4. Time Series for dummies- The 3 step process [WWW] Available from https://www. kdnuggets. com/2018/03/time-series-dummies-3-step-process. html [Accessed 22/04/18]

5. Pennsylvania state university (n. d) Overview of Time Series Characteristics [WWW] Available from https://onlinecourses. science. psu. edu/stat510/node/47 [Accessed 22/04/18]