Analysis of Time Series

In a normal machine learning dataset, the dataset is a collection of observations that are treated equally when future is being predicted. In time series the order of observations provides a source of additional information that should be analysed and used in the prediction process. Time series can have one or more variables that change over time. If there is only one variable varying over time, we call it **Univariate time series**. If there is more than one variable it is called **Multivariate time series.** Time series analysis extract meaningful statistics and other characteristics of the dataset in order to understand it.

Time series - Set of evenly spaced numerical data

- Obtained by observing response variable at regular time periods

Forecast based only on past values

- Assumes that factors influencing past, present, & future will continue

Time series forecasting involves taking models fit on historical data (the training set) and using them to predict future observations (the test set). At the first step past observations are collected and analysed to develop a suitable mathematical model which captures the underlying data generating process for the series. In the second step the future events are predicted using the model.

Types of Time Series

Deterministic time series

A deterministic time series is one which can be expressed explicitly by an analytic expression. It has no random or probabilistic aspects. In mathematical terms, it can be described exactly for all time in terms of a Taylor series expansion provided that all its derivatives are known at some arbitrary time. Its past and future are completely specified by the values of these derivatives at that time. If so, then we can always predict its future behaviour and state how it behaved in the past.

Non-deterministic time series

A non-deterministic time series is one which cannot be described by an analytic expression. It has some random aspect that prevents its behaviour from being described explicitly. A time series may be non-deterministic because:

1. All the information necessary to describe it explicitly is not available, although it might be in principle.

2. The nature of the generating process is inherently random.

Since non-deterministic time series have a random aspect, it follows probabilistic laws. Thus, the data is defined by statistical terms, i.e. by probability distributions and averages of various forms, such as means and variances.

Stationary time series

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc., do not depend upon time. A stationary series is relatively easy to predict: you simply forecast that its statistical properties will be the same in the future as they have been in the past. Thus, most statistical forecasting methods are based on the assumption that the time series are approximately stationary.

Non stationary time series

Non-stationary series is one whose statistical properties change over time. There are an infinite number of ways for a time series to be non-stationarity, such as changing variance, level shifts, and seasonality in the 6th moment, etc. Here are the most common non-stationarity patterns:

Trend component: The trend shows the general tendency of the data to increase or decrease during a long period of time. A trend is a smooth, general, long-term, average tendency. It is not always necessary that the increase or decrease is in the same direction throughout the given period of time. If a time series does not show an increasing or decreasing pattern then the series is stationary in the mean.
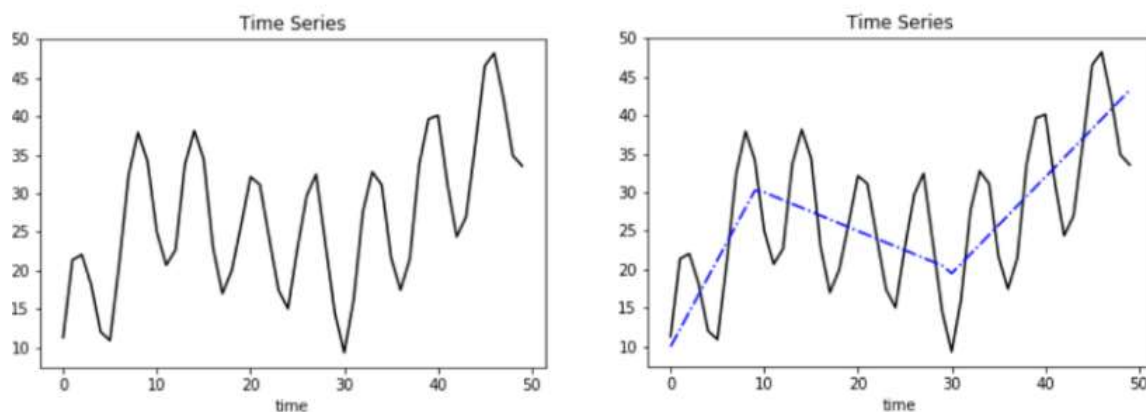
Trends for time series

Cyclical component: Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern. In cyclical pattern the up and down movements does not occur in a constant time intervals, they cannot be predicted.

Seasonal component: If the series peaks and trough occur in a regular intervals the pattern is called seasonal pattern (e.g. sales of ice cream).

Random component: the residual is what's left over when all the patterns have been removed. Residuals are random fluctuations. You can think of them as a noise component.
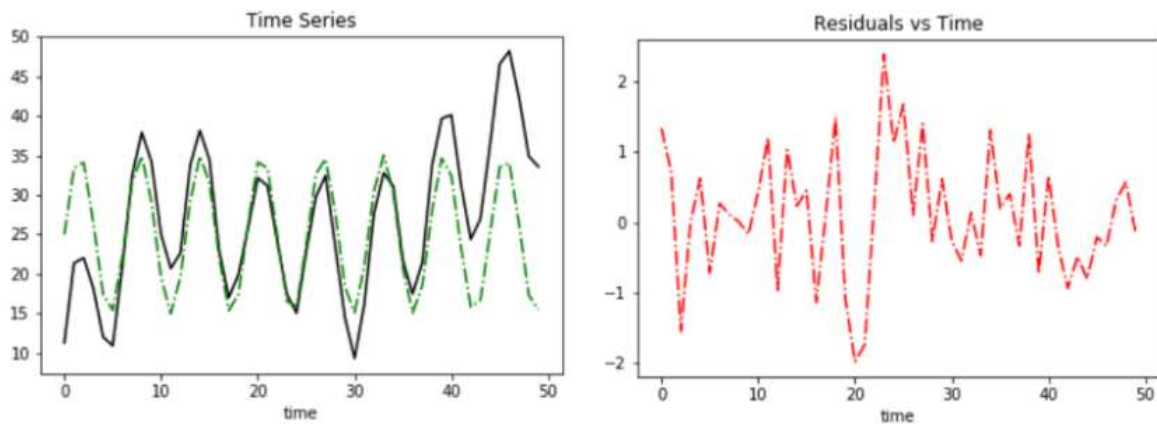
On the left Seasonal pattern, on the right the random residual