Satyajit Pattnaik

Time Series Analysis

End to End Guide on Time Series Analysis
&
Forecasting

Introduction

Problem

Time series forecasting is an important area of machine learning that is often neglected.

It is important because there are so many prediction problems that involve a time component and these problems are often neglected because it is this time component that makes time series problems more difficult to handle.

Before getting started with Time Series Analysis, let's get our basics clear on **Anomaly Detection**

What is a Time Series

It is a series of observations taken at specified times basically at equal intervals. It is used to predict future values based on past observed values.

Components of Time Series

Satyajit Pattnaik

Trend	Seasonality	Irregularity	Cyclic
Increasing or Decreasing value in the series	A general systematic linear or (most often) non-linear component that changes over time and does repeat.	The data in the time series follows a temporal sequence, but the measurements might not happen at a regular time interval.	Pattern exists when data exhibit rises & falls that are not of fixed period.

Stationarity

What is Stationarity?

Satyajit Pattnaik

Points to Remember

In stationary TS, the mean, variance and covariance of the series should not be a function of time rather should be a constant.

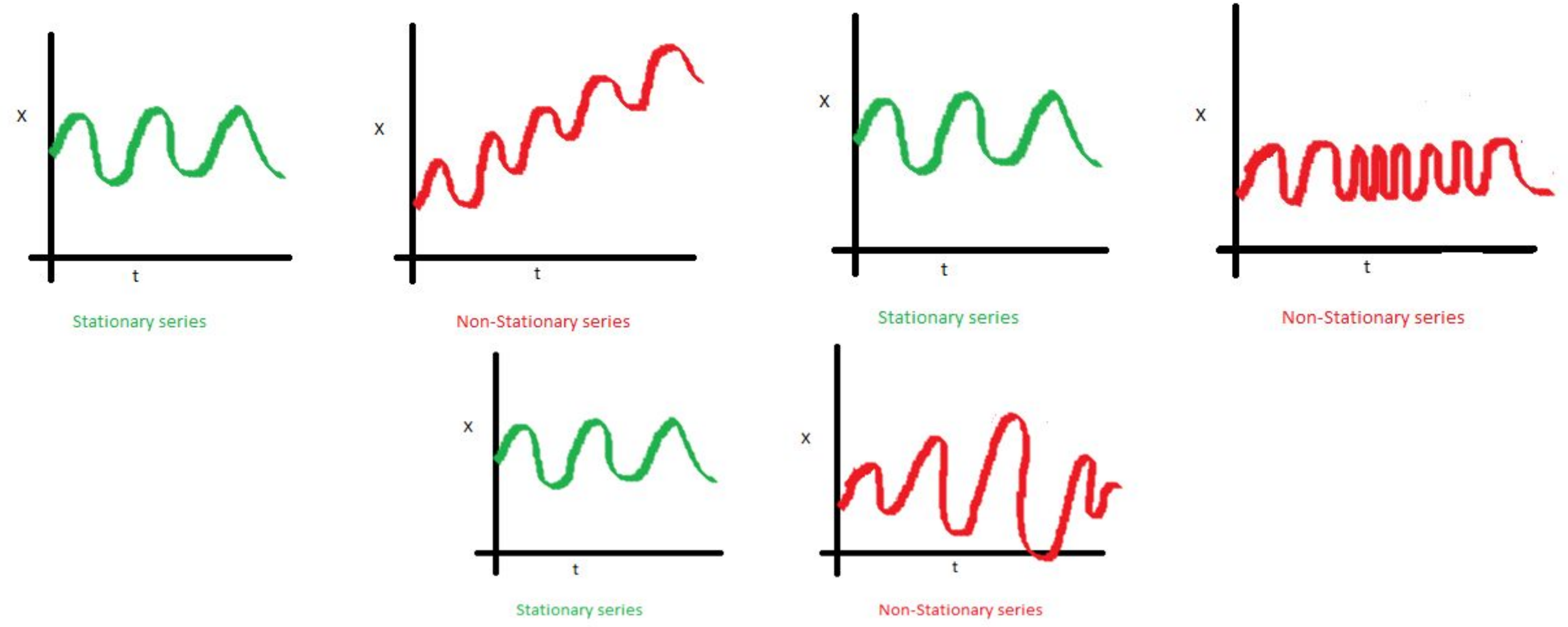
TS data can be stationary by removing their various components:

- Trend – Varying over time
- Seasonality – Variations at specific time
- Cyclic
- Irregularity

Satyajit Pattnaik

What is Stationarity?

The image below has the left hand graph satisfying the condition whereas the graph in red has a time dependent mean



How to make TS stationary?

Satyajit Pattnaik

Points to Remember

Two common techniques:

- Differencing
- Transforming - Log, Double Log, etc.

Testing TS Stationarity

Satyajit Pattnaik

Checks for Stationarity

There are many methods to check whether a time series is stationary or non-stationary.

1. **Look at Plots:** You can review a time series plot of your data and visually check if there are any obvious trends or seasonality.
2. **Summary Statistics:** You can review the summary statistics for your data for seasons or random partitions and check for obvious or significant differences.
3. **Statistical Tests:** You can use statistical tests to check if the expectations of stationarity are met or have been violated.

Testing TS Stationarity .. contd

Let’s talk about Summary Statistics

You can split your time series into two (or more) partitions and compare the mean and variance of each group. If they differ and the difference is statistically significant, the time series is likely non-stationary.

Testing TS Stationarity .. contd

Satyajit Pattnaik

Let’s talk about ADF Test

Statistical tests make strong assumptions about your data. They can only be used to inform the degree to which a null hypothesis can be rejected or fail to be reject. The result must be interpreted for a given problem to be meaningful.

Nevertheless, they can provide a quick check and confirmatory evidence that your time series is stationary or non-stationary.

ADF Test is otherwise known as **unit root test.**

Testing TS Stationarity .. contd

Satyajit Pattnaik

ADF Hypothesis

The null hypothesis of the test is that the time series can be represented by a unit root, that it is not stationary (has some time-dependent structure). The alternate hypothesis (rejecting the null hypothesis) is that the time series is stationary.

- **Null Hypothesis (H0)**: If failed to be rejected, it suggests the time series has a unit root, meaning it is non-stationary. It has some time dependent structure.
- **Alternate Hypothesis (H1)**: The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary. It does not have time-dependent structure.

We interpret this result using the p-value from the test. A p-value below a threshold (such as 5% or 1%) suggests we reject the null hypothesis (stationary), otherwise a p-value above the threshold suggests we fail to reject the null hypothesis (non-stationary).

- **p-value > 0.05**: Fail to reject the null hypothesis (H0), the data has a unit root and is non-stationary.
- **p-value <= 0.05:** Reject the null hypothesis (H0), the data does not have a unit root and is stationary.

Algorithms

Satyajit Pattnaik

- ARIMA (AR, MA, ARMA, ARIMA)
- Facebook Prophet
- LSTMs
- Holt’s Winter Exponential Smoothing
- GARCH
- SARIMA/SARIMAX
- VAR
- VARMA etc. etc.

Quick Link:
https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

ARIMA

Satyajit Pattnaik

AR	MA	ARMA	ARIMA
Autoregressive	**Moving Average**	**Autoregressive Moving Average (No differencing)**	**Autoregressive Integrated Moving Average**

ARIMA

Satyajit Pattnaik

Identifying the right parameters

There are different techniques to find the right parameters for ARIMA(p,d,q)

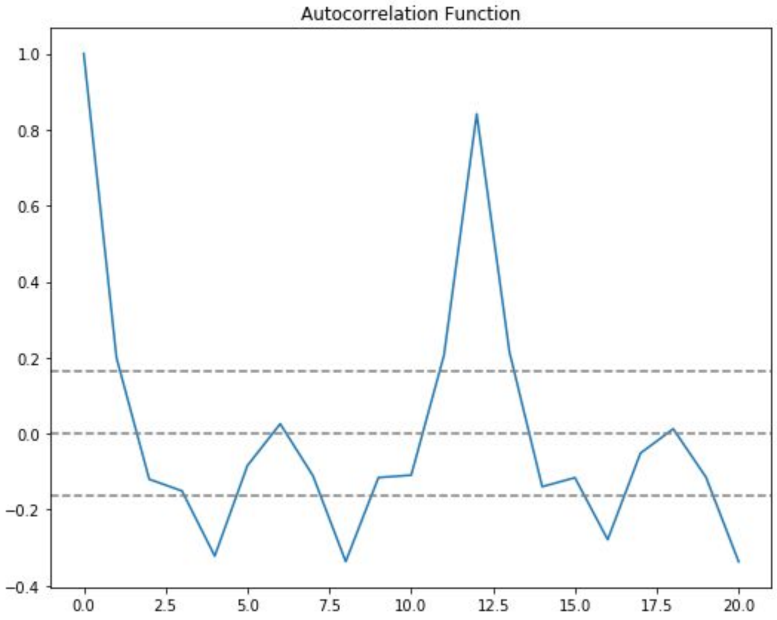
1. ACF/PACF Plots
2. Grid Search
3. Auto_Arima

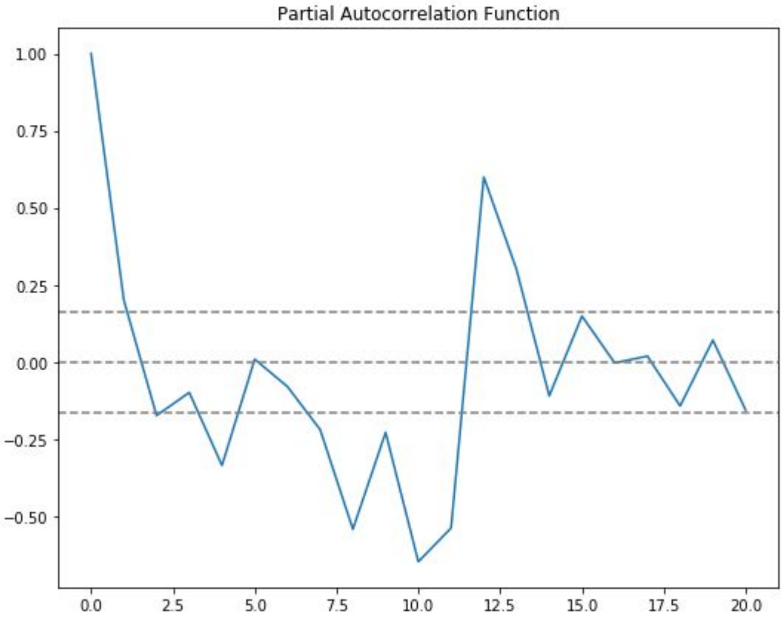
Let’s learn about these techniques in the next slides.

Auto Correlation

Partial Auto Correlation

Satyajit Pattnaik





1. **p** – The lag value where the **PACF** chart crosses the upper confidence interval for the first time. If you notice closely, in this case p=2.
2. **q** – The lag value where the **ACF** chart crosses the upper confidence interval for the first time. If you notice closely, in this case q=2.

Grid Search

ACF/PACF plots are some traditional methods of obtaining p & q values, and are sometimes misleading, hence we need to perform a hyper parameter optimization step in Time Series Analysis to get the optimum p,d & q values

Auto Arima

Grid Search techniques are manual ways, the same task can be achieved in few lines of coding and with a better efficiency using Auto Arima

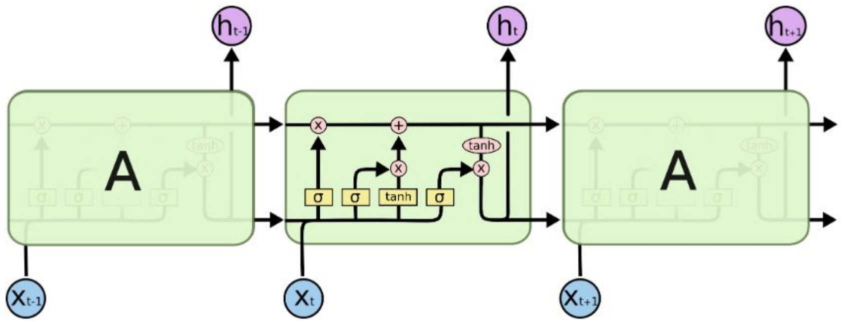
Facebook Prophet

Quick Link: https://facebook.github.io/prophet/docs/quick_start.html#python-api

Features:

1. Very fast
2. An additive regression model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects
3. Robust to missing data & shifts in trend, and handles outliers automatically.
4. Easy procedure to tweak & adjust forecast while adding domain knowledge or business insights.

LSTMs



Quick Link:
https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Features:

LSTM cell in place of standard neural network layers:

1. Input gate
2. Forget gate
3. Output gate

Thanks

Satyajit Pattnaik