upGrad

#LifeKoKaroLift

Post-Graduate Diploma in ML/AI

upGrad

Course : Machine Learning

Lecture On : TimeSeries

Instructor : Arihant Jain

Today’s Agenda

1. Introduction to time series and its components
2. Classical decomposition of a time series
3. Stationarity time series - ARMA Modelling
4. Time Series Differencing and ARIMA Modelling
5. Time Series smoothing

What is Time Series and Time Series Data ?

upGrad

A **time series** is a collection of observations of well-defined data items obtained through repeated measurements over time. The data points are indexed in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, the daily closing value of the Dow Jones Industrial Average, etc.

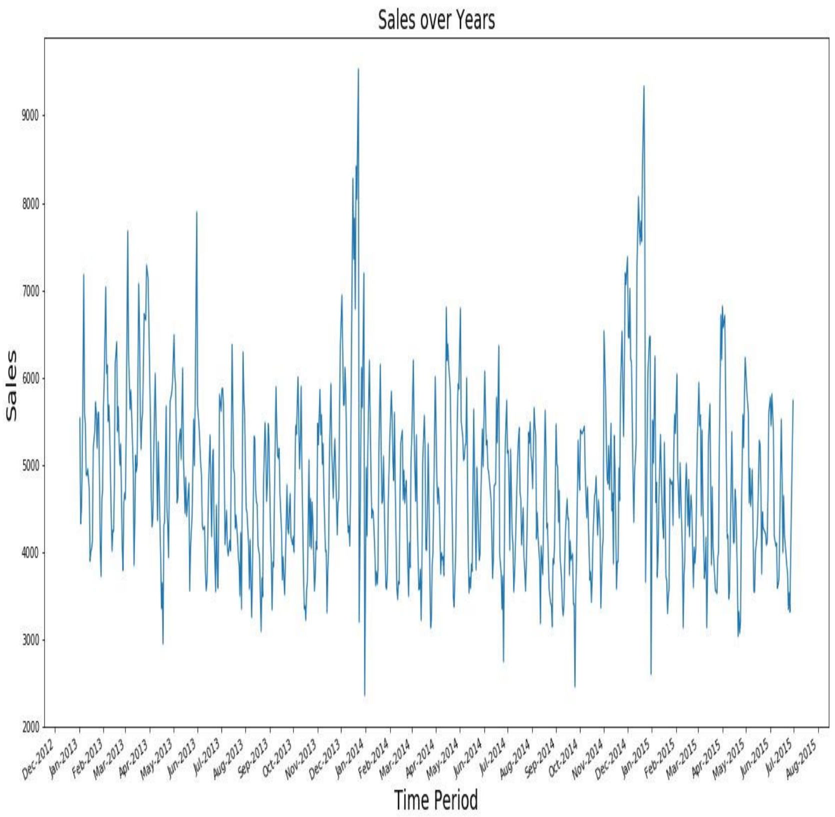
Time series data is data that is collected at different points in time. This is opposed to cross-sectional data which observes individuals, companies, etc. at a single point in time. Because data points in time series are collected at adjacent time periods there is potential for correlation between observations. This is one of the features that distinguishes time series data from cross-sectional data

Time Series Analysis

upGrad

Time series analysis is a <u>statistical technique</u> that deals with time series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered in two types:

- Time series data: A set of observations on the values that a variable takes at different times
- Cross-sectional data: Data of one or more variables, collected at the same point in time.



Regression Vs Timeseries

upGrad

Regression	Time Series
1. In Regression it does not matter if we reshuffle the data.	1. Time Series consists time from before to the latest and they cannot be shuffled.
2. In Regression data points are independent.	2. In Time Series there is strong correlation between successive values.
3. Regression model predicts only on the basis of the values given.	3. Time Series not only depends on the values given, but also on depends on the sequence in which the values are given.

Question : Which of the following is an example of time series problem?

1. Estimating number of rooms booking in next 12 months.
2. Estimating the total sales in next 3 years of financial company.
3. Estimating the number of calls for the next one week.

A) Only 3
B) 1 and 2
C) 2 and 3
D) 1 and 3
E) 1,2 and 3

Application of Time Series

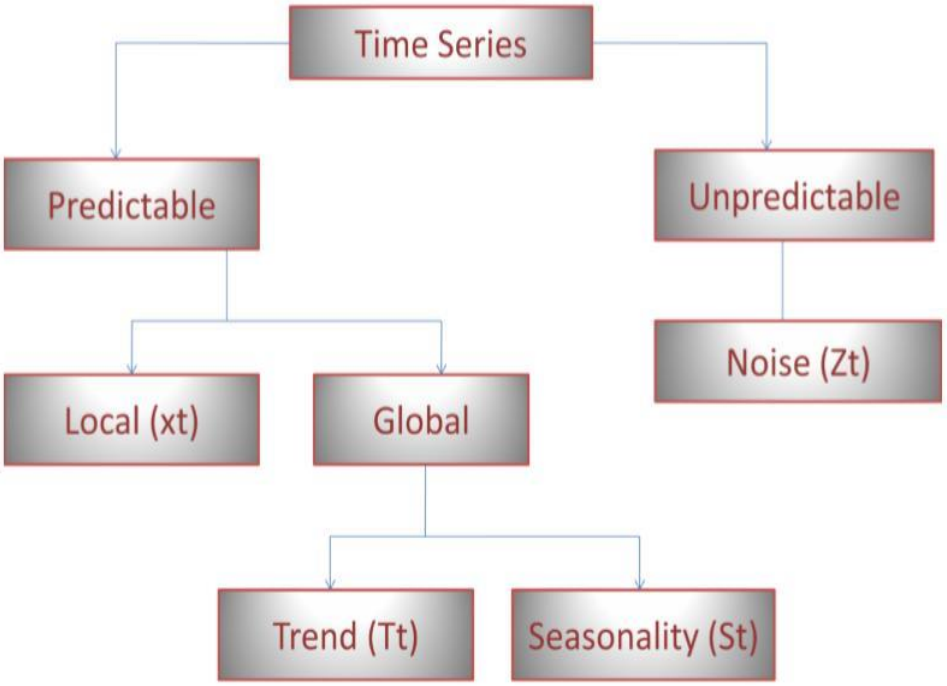
upGrad

Time Series

- Election Forecast
- Supply Chain Forecasting
- Advertising spend
- Web traffic
- Government Predictions
- Budget Forecast
- Turnover/Sales Prediction
- Bookings forecasting
- Science & Engineering
- Econometrics GDP, Inflation
- Earthquake prediction
- signal processing
- Health management
- mathematical finance
- Economic Growth
- Control engineer
- Weather Forecast
- Medicine
- Astronomy
- Healthcare Diagnosis
- Biomedical Monitoring
- Trading Market

Components of Time Series

upGrad



Source – dataanalyticsedge.com

Local Predictable part (Xt)

upGrad

Local predictable part consists of auto regressive behavior. It shows how time series is influenced by its immediate past.

$$X_t = a + bX_{t-1} + cX_{t-2}$$

If we know few time stamps immediately preceding time, we can predict the time series values.

Example: Give yesterday’s price we can predict today’s price but not price after a year

Global Predictable part

upGrad

In Global predictable part the present time series does not depend on the immediate past.

Example: Temperature in January does not depend on the December or November temperature.

This Global predictable part consists of two parts Trend and Seasonality.

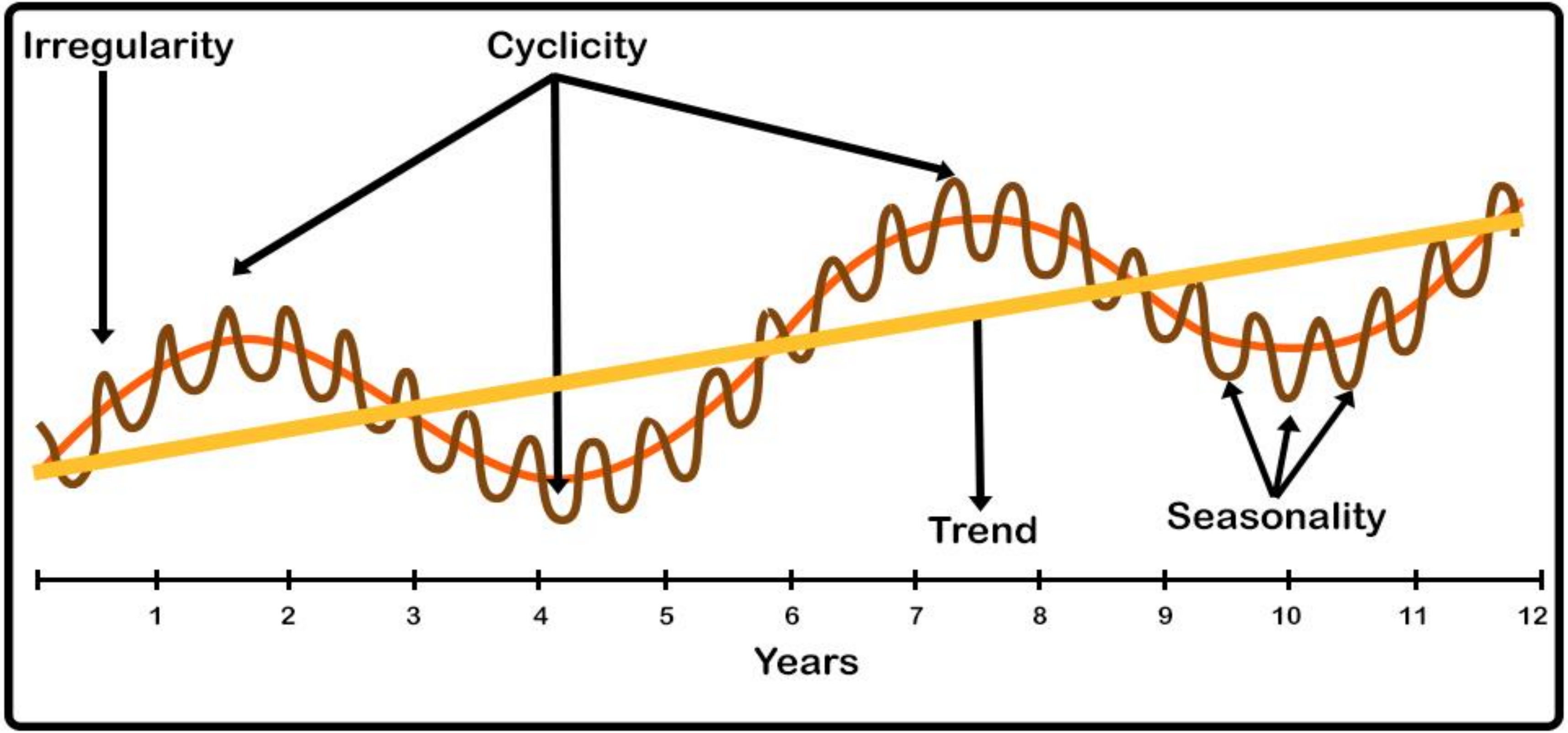
Trend: Trend refers to any pattern that talks about the overall increase or decrease in the values

Seasonality: Seasonality refers to a repeating pattern of values seen in the data

Time Series With Trend and Seasonality

Imp Components of Time Series

upGrad



Classical decomposition of a time series

upGrad

Additive & Multiplicative Model

let’s understand what the analysis features of a Time Series are –

- **Level**: is the average value in the series

- **Trend**: is the increasing or decreasing value in the series

- **Seasonality**: is the repeating the short-term cycle in the series

- **Noise**: is the random variation in the series.

Additive & Multiplicative Models

upGrad

The Additive Model

It is a model of data in which the effects of the individual factors are differentiated and added to model the data. An additive model is such a time series in which the magnitude of the seasonal fluctuations does not vary with the level of time series.

y(t) = Level + Trend + Seasonality + Noise

In the additive model, the behavior is linear where changes over time are consistently made by the same amount, like a linear trend. In this situation, the linear seasonality has the same amplitude and frequency.

Consider the scenario of sales in an umbrella manufacturing company. In every rainy season (June — August) of all years, we observed an increase of 20,000 umbrellas in sales. Then it can be termed as an additive model.

Additive & Multiplicative Models

upGrad

The Multiplicative Model

In this situation, trend and seasonal components are multiplied and then added to the error component. The multiplicative model is such a time series in which seasonal fluctuations increase or decrease proportionally with increases and decreases in the level of the series. It is not linear, can be exponential or quadratic and represented by a curved line as below:

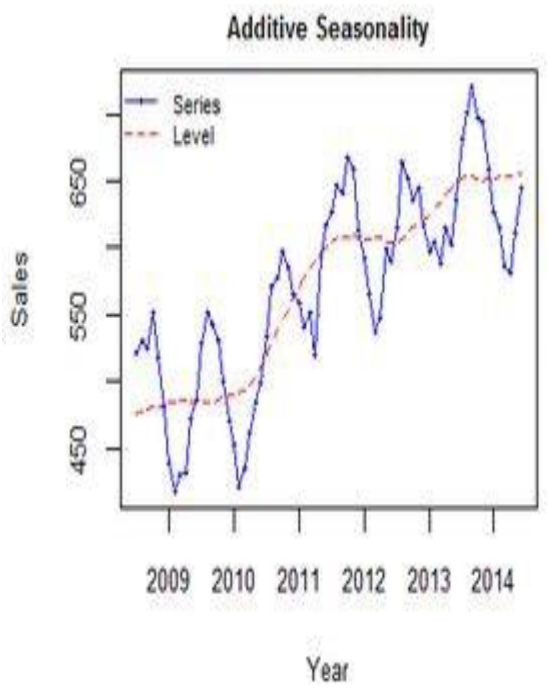
y(t) = Level * Trend * Seasonality * Noise

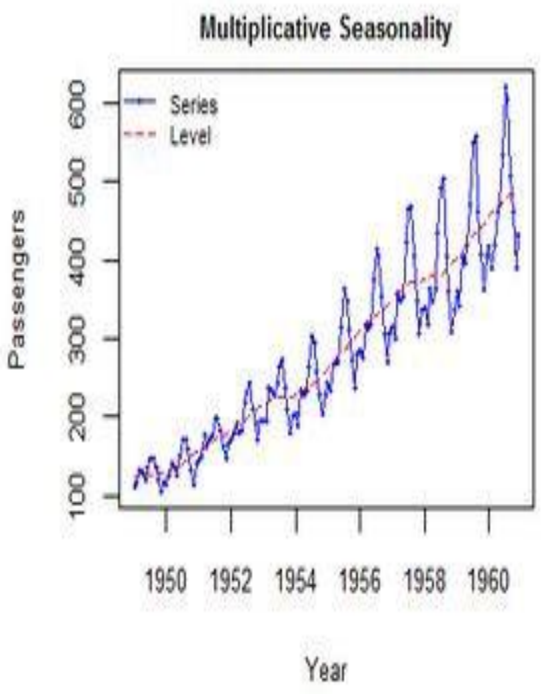
Different from the additive model, the multiplicative model has an increasing or decreasing amplitude and/or frequency over time.

Consider the same scenario of sales in an umbrella manufacturing company as above. In every rainy season (June — August) of all years, we observed an increase of 10% umbrellas in sales. Then it can be termed as a multiplicative model.

Additive vs Multiplicative

upGrad





Source: https://goo.gl/szKmgo

Defining Stationarity in a time series

upGrad

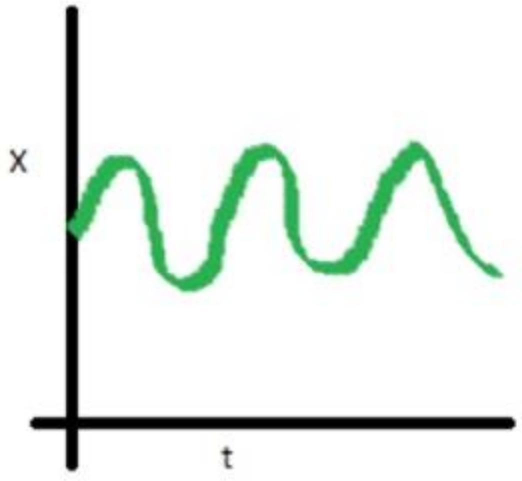
In the most intuitive sense, stationarity means that the statistical properties of a process generating a time series do not change over time. **A stationary series is one in which the properties – mean, variance and covariance, do not vary with time.**

It does not mean that the series does not change over time, just that the *way* it changes does not itself change over time.

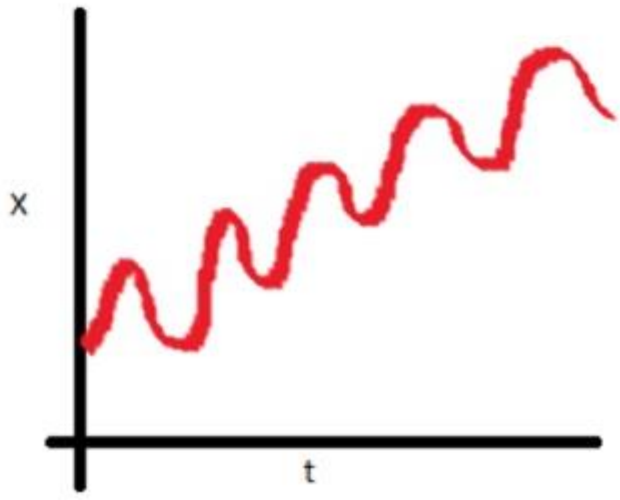
linear function - The value of a linear function changes as x grows, but the *way* it changes remains constant — it has a constant slope; one value that captures that rate of change.

Stationary and Non-Stationary Series

upGrad



Stationary series



Non-Stationary series

Poll

upGrad

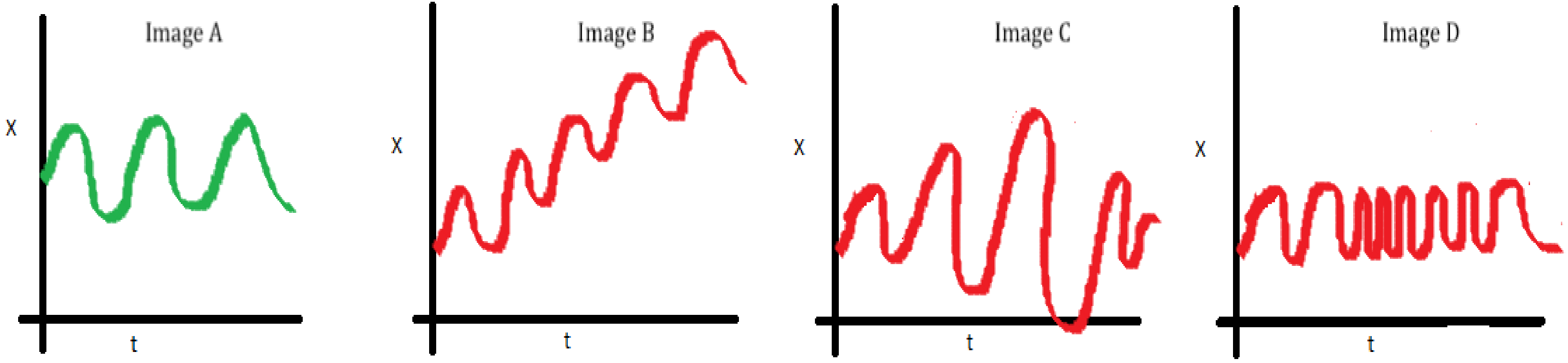
Stationarity is a desirable property for a time series process.

A) TRUE
B) FALSE

upGrad

The Principles of Stationarity

Time-Dependent Mean | Time-Dependent Variance | Time-Dependent Covariance



Stationary series | Non-Stationary series | Non-Stationary series | Non-Stationary series

Different Methods to Check Stationarity

statistical tests like the unit root stationary tests. Unit root indicates that the statistical properties of a given series are not constant with time, which is the condition for stationary time series.

Suppose we have a time series :
$y_t = a*y_{t-1} + \varepsilon_t$

where y_t is the value at the time instant t and ε_t is the error term.

In order to calculate y_t we need the value of y_{t-1}, which is :
$y_{t-1} = a*y_{t-2} + \varepsilon_{t-1}$

If we do that for all observations, the value of y_t will come out to be:
$y_t = a^n*y_{t-n} + \Sigma\varepsilon_{t-i}*a^i$

If the value of a is 1 (unit) in the above equation, then the predictions will be equal to the y_{t-n} and sum of all errors from t-n to t, which means that the variance will increase with time. This is knowing as unit root in a time series. We know that for a stationary time series, the variance must not be a function of time. The unit root tests check the presence of unit root in the series by checking if value of a=1.

Methods to Check Stationarity -
1. ADF Test
2. KPSS Test

ADF Test

ADF (Augmented Dickey Fuller) Test

The Dickey Fuller test is one of the most popular statistical tests. It can be used to determine the presence of unit root in the series, and hence help us understand if the series is stationary or not.

The null and alternate hypothesis of this test are:

Null Hypothesis: The series has a unit root (value of a =1)

Alternate Hypothesis: The series has no unit root

If we fail to reject the null hypothesis, we can say that the series is non-stationary

Test for stationarity: If the test statistic is less than the critical value, we can reject the null hypothesis (aka the series is stationary). When the test statistic is greater than the critical value, we fail to reject the null hypothesis (which means the series is not stationary)

upGrad

Thank You !