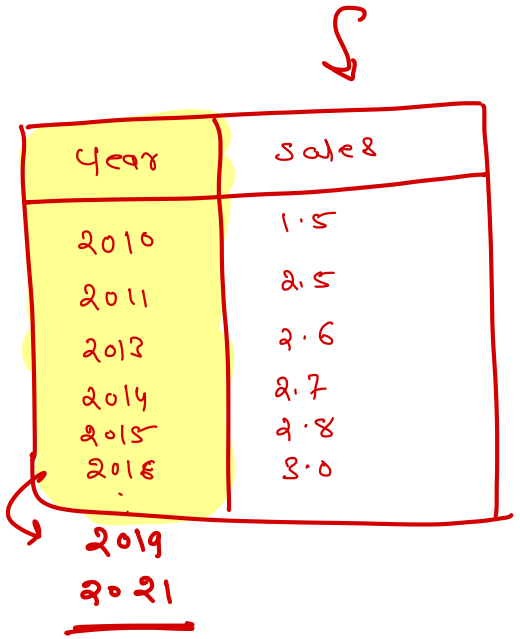
Machine Learning

Time Series Analysis

What is time series?

- An ordered sequence of values of a variable at equally spaced time intervals
- In other words, a time series is a series of data points indexed (or listed or graphed) in time order
- Such data may be collected at regular time intervals, such as monthly, weekly, quarterly or annually
- Time series are used in statistics, economics, mathematical finance, weather forecasting, earthquake predictions and many more other applications
- E.g.
  - Sales of a company over last decades
  - List of temperatures for few weeks
  - Marks of students for few years



Univariate time series

- The term "univariate time series" refers to a time series that consists of single (scalar) observations recorded sequentially over equal time increments
- E.g.
  - Monthly CO2 concentration
  - Revenue taken yearly

What is time series analysis?

- Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data
- Time series forecasting is the use of a model to predict future values based on previously observed values
- There are two main goals of time series analysis
  - identifying the nature of the phenomenon represented by the sequence of observations
  - forecasting (predicting future values of the time series variable)
- Both of these goals require that the pattern of observed time series data is identified and more or less formally described
- Once the pattern is established, we can interpret and integrate it with other data (i.e., use it in our theory of the investigated phenomenon, e.g., seasonal commodity prices)
- Regardless of the depth of our understanding and the validity of our interpretation (theory) of the phenomenon, we can extrapolate the identified pattern to predict future events.

Components of time series

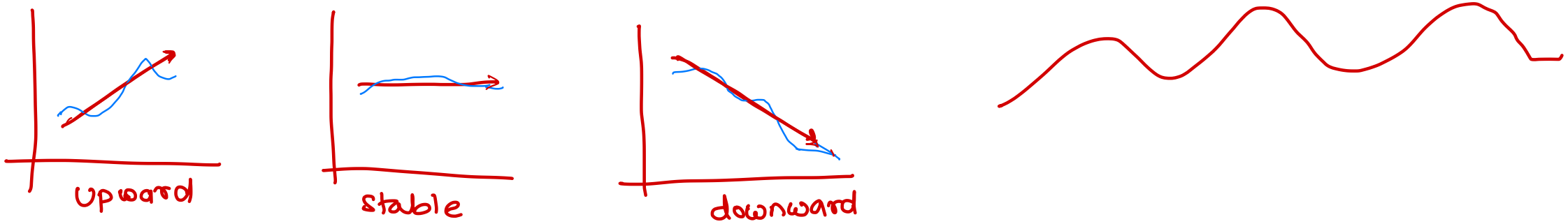
- Trend
- Seasonal Variations (Seasonality)
- Random or Irregular movements
- Cyclic Variations

Trend

- 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
- It is observable that the tendencies may increase, decrease or are stable in different sections of time
- But the overall trend must be upward, downward or stable
- The population, agricultural production, items manufactured, number of births and deaths, number of industry or any factory, number of schools or colleges are some of its example showing some kind of tendencies of movement

Trend

- **Linear and Non-Linear Trend**
  - If we plot the time series values on a graph in accordance with time t
  - The pattern of the data clustering shows the type of trend
  - If the set of data cluster more or less round a straight line, then the trend is linear otherwise it is non-linear (Curvilinear)



Seasonal Variations

- These are the rhythmic forces which operate in a regular and periodic manner over a span of less than a year
- They have the same or almost the same pattern during a period of 12 months
- This variation will be present in a time series if the data are recorded hourly, daily, weekly, quarterly, or monthly
- These variations come into play either because of the natural forces or man-made conventions
- The various seasons or climatic conditions play an important role in seasonal variations. Such as production of crops depends on seasons, the sale of umbrella and raincoats in the rainy season, and the sale of electric fans and A.C. shoots up in summer seasons.
- The effect of man-made conventions such as some festivals, customs, habits, fashions, and some occasions like marriage is easily noticeable. They recur themselves year after year. An upswing in a season should not be taken as an indicator of better business conditions.

Cyclic Variations

- The variations in a time series which operate themselves over a span of more than one year are the cyclic variations
- This oscillatory movement has a period of oscillation of more than a year
- One complete period is a cycle. This cyclic movement is sometimes called the 'Business Cycle'.
- It is a four-phase cycle comprising of the phases of prosperity, recession, depression, and recovery
- The cyclic variation may be regular are not periodic
- The upswings and the downswings in business depend upon the joint nature of the economic forces and the interaction between them.

Random or Irregular Movements

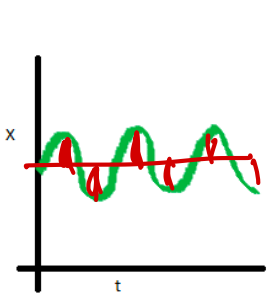
- There is another factor which causes the variation in the variable under study
- They are not regular variations and are purely random or irregular
- These fluctuations are unforeseen, uncontrollable, unpredictable, and are erratic
- These forces are earthquakes, wars, flood, famines, and any other disasters

Stationary vs non-stationary series

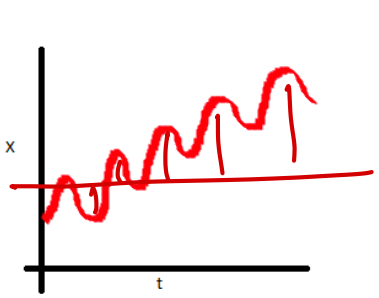
- A series is said to be strictly stationary if the marginal distribution of Y at time t is same as at any other point in time

  p(Yt) = p(Yt+k) and p(Yt, Yt+k) does not depend on t

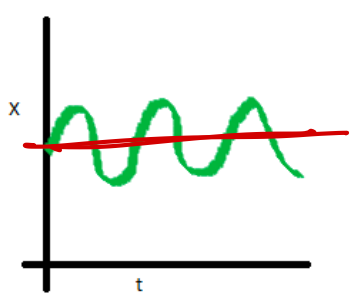
- This implies that mean, variance and covariance of the series Yt are time invariant
- There are three basic criterion for a series to be classified as stationary series
  - The mean of the series should not be a function of time rather should be a constant
  - The variance of the series should not a be a function of time. This property is known as homoscedasticity.
  - The covariance of the i th term and the (i + m) th term should not be a function of time



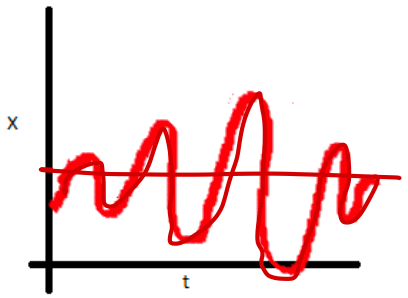
Stationary series



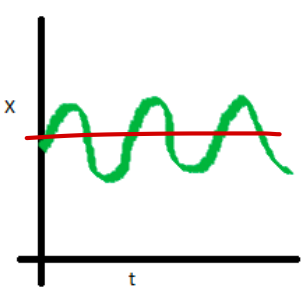
Non-Stationary series



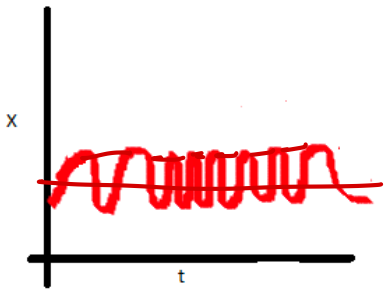
Stationary series



Non-Stationary series



Stationary series



Non-Stationary series

Time series forecasting models

- Naive Approach
- Simple average
- Moving Average
- Exponential Smoothing
- Autoregressive Movie Average
- Autoregressive Integral Moving Average

Year	Profit	MA(2)	MA(3)
2016	1	-	-
2017	1	-	-
2018	3	1	-
2019	2	2	1.6
2020	2	2.5	3
2021	?	2	2.3

Naive approach ⇒ (2)

Simple average / simple mean ⇒ $\frac{9}{5} = 1.8$

MA(2) ⇒ (2) MA(3) ⇒ (2.3)

x	y
1	1
2	4
3	9
4	16
5	(25)

$y = x^2$

Moving Average (MA)

- Moving average smoothing is a naive and effective technique in time series forecasting.
- It can be used for data preparation, feature engineering, and even directly for making predictions
- A common representation of MA model where it depends on q of its past values is called as MA(q) model and is represented as

$$Yt = B0 + Et + \ldots$$

Moving Average - Drawbacks

- 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

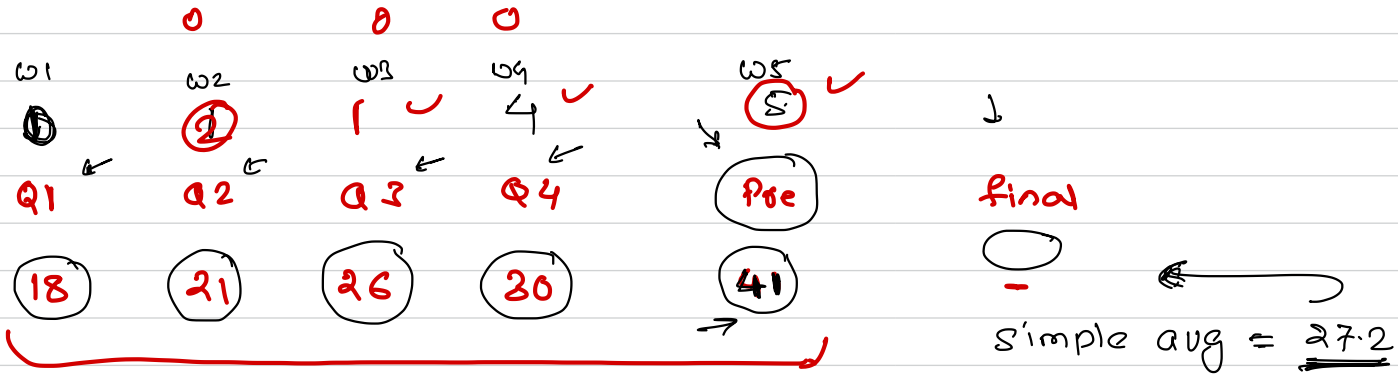
Exponential Smoothing

- Exponential smoothing uses a similar logic to moving average, but this time, a different *decreasing weight* is assigned to each observations
- In other words, *less importance* is given to observations as we move further from the present
- Mathematically it can be represented as:

$$y = \alpha x_t + (1 - \alpha) y_{t-1}, t > 0$$

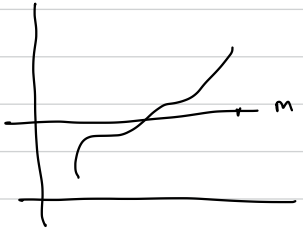
2019 2018

- Methods for exponential smoothing
  - Holt's Trend Method
  - Holt-Winters method



$$\text{Weighted average} = \frac{18\times w1 + 21\times w2 + 26\times w3 + 30\times w4 + 41\times w5}{\text{sum of weights}}$$

$$= \frac{398}{12} = 33.16$$



$$\text{simple average} = \frac{18 + 21 + 26 + 30 + 41}{5} = 27.2$$

Auto Regressive Moving Average (ARMA)

- Forecast a series without trend and seasonality
- It is the model where current value of variable (Y), depends upon only the values that the variable took in previous periods plus an error term

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t$$

Auto Regressive Integrated Moving Average (ARIMA)

- Also known as Box-Jenkins methodology which has three steps
  - Identification
  - Estimation
  - Diagnostic Checking
- Identification
  - Autocorrelation function (ACF)
    - Refers to the way the observations in a time series are related to each other and is measured by a simple correlation between current observation (Yt) and the observation p periods from the current one Yt-p

$$\rho_k = \text{Corr}(Y_t, Y_{t-p}) = \frac{\mathbf{Cov}\,(Yt, Yt-p)}{\sqrt{var(\mathbf{Yt})}\sqrt{var(\mathbf{Yt-p})}} = \frac{\gamma p}{\gamma 0}$$

  - Partial Autocorrelation Function (PACF)
    - Used to measure the degree of association between Yt and Yt-p when the effects at other time lags 1, 2, 3 .. P-1 are removed

Uses of time series analysis

- The most important use of studying time series is that it helps us to predict the future behaviour of the variable based on past experience
- It is helpful for business planning as it helps in comparing the actual current performance with the expected one
- From time series, we get to study the past behaviour of the phenomenon or the variable under consideration
- We can compare the changes in the values of different variables at different times or places, etc