Data Science Concepts

Lesson09–Time Series Concepts With ARIMA

Objective

After completing this lesson you will be able to:

- Describe the application areas of Time Series
- Explain Trend, Seasonality, Cyclic and Noise component in data
- Explain the various models in time series.

Forecasting demand of products and services for long range and short range planning.

Amazon sells more than 13 millions SKUS and has more than 2 million retailers selling them.

Demand of manpower in IT product and services to manage the business.

Netflix opened up a challenge to forecast movie ratings by customer. Can be further used for movie recommendation.

Time Series Data

Data on Response variable (Y) which is a random variable

Data points are collected at regular interval and arranged in chronological order

Univariate time series: data contains observation of a single variable

- warranty claims at time t

Multivariate time series: data contains observation of more than one variable

- warranty claims at time t
- sale of motor vehicles at time t
- Climatic conditions at time t

Time series components

- Trend Component (T_t)
  - Consistent upward or downward movement of data over a period of time
  - Trend can be identified by seasonality window or seasonality index

- Seasonality Component (S_t)
  - Seasonality, measured in terms of seasonality index, is fluctuations from the trend that occurs within a defined time period (seasons, quarters, months, days of the week, time interval within a day etc.)

- Cyclic Component (C_t)
  - Fluctuations around the trend line which happens due to macro economic changes (recession, unemployment etc.)
  - Fluctuations have repetitive pattern and time between repetition is more than a year but this time is random

- Irregular component (I_t)
  - White noise or random uncorrelated changes that follow a normal distribution with a mean of zero and constant variance

BSE SENSEX
INDEXBOM: SENSEX

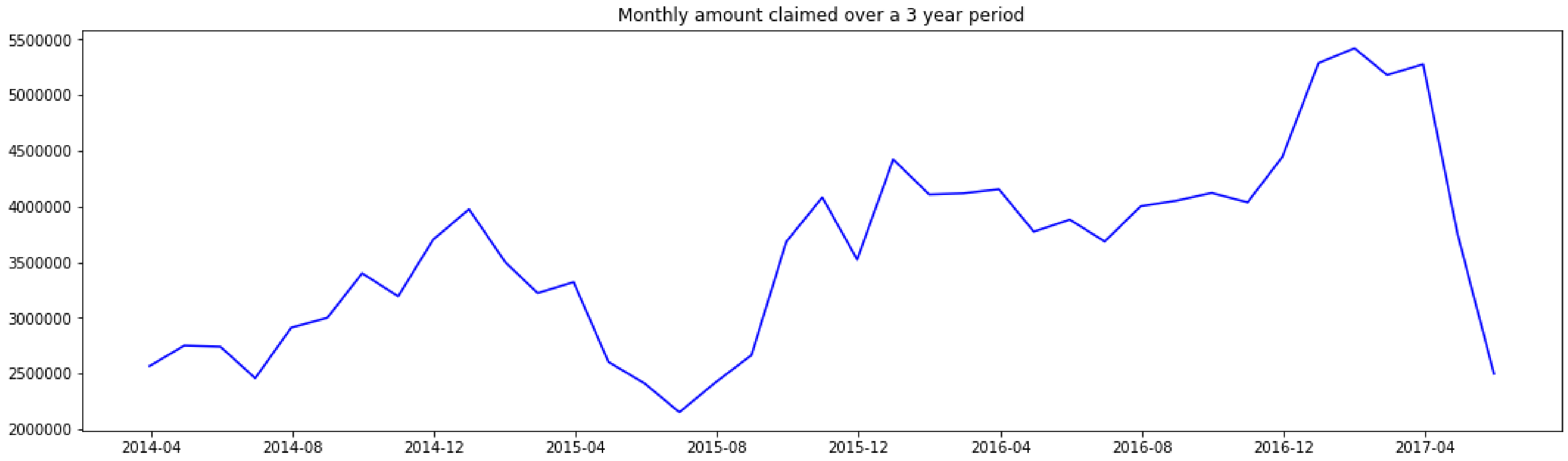
+ Follow

39,982.97 −100.57 (0.25%)

6 Jun, 10:03 am IST · Disclaimer

1 day | 5 days | 1 month | 6 months | YTD | 1 year | **5 years** | Max

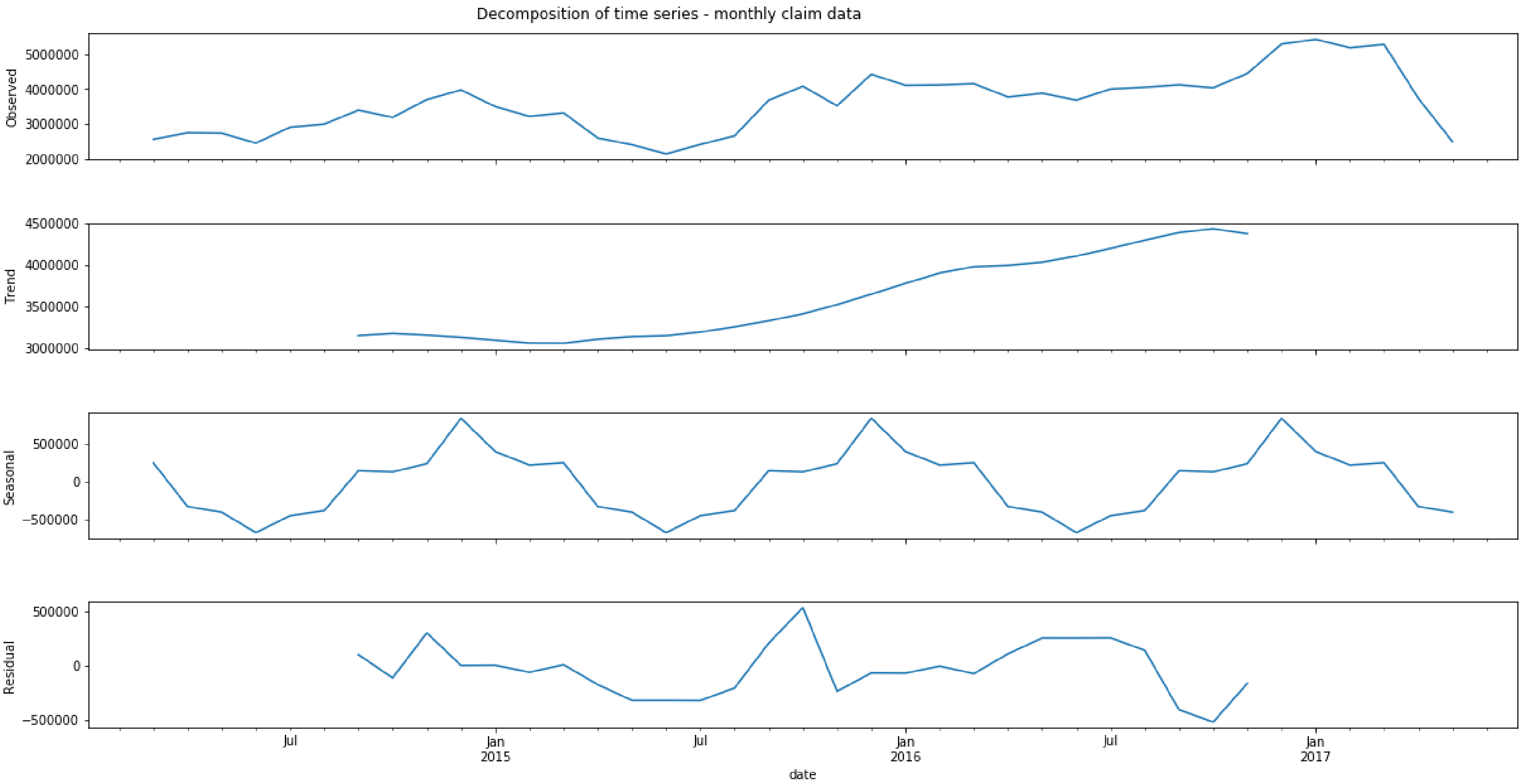
Open	40,136.43	Low	39,966.05
High	40,159.26		



Seasonality window

The idea of finding seasonality window is to find a window size of `s`, for which if the rolling average is calculated for each time point (-s/2<t<s/2), the zigzag motion in the time series data smoothens out.

The rolling average in the window `s` tries to smooth out noise and seasonality and what is observed is a pure trend.



The seasonality window in the claim data was found to be 13

Additive and Multiplicative model

Time series data can be modelled as addition or multiplication of the time series component.

Additive Model	Multiplicative Model
$Y_t = T_t + S_t + C_t + I_t$	$Y_t = T_t * S_t * C_t * I_t$
• Assumes that, seasonal and cyclic component are independent of trend • Useful when seasonal component remain constant around the mean or does not vary with the mean of the series.	• Better fit to most cases • Useful when seasonal component is correlated with the mean of the series

Forecasting Accuracy

- Mean Absolute Error

$$MAE = \sum_{t=1}^{n} \frac{|Y_t - F_t|}{n}$$

- Mean Absolute Percentage Error

$$MAPE = \left(\frac{1}{n} * \sum_{t=1}^{n} \frac{|Y_t - F_t|}{Y_t}\right) * 100$$

- Mean Squared Error

$$MSE = \frac{1}{n} * \sum_{t=1}^{n} (Y_t - F_t)^2$$

- Mean Squared Error

$$RMSE = \sqrt{\frac{1}{n} * \sum_{t=1}^{n} (Y_t - F_t)^2}$$

MAPE is dimensionless and thus can be used to compare different models as well. RMSE is standard deviation of error. Lower RMSE implies better prediction.

Forecasting Techniques

Many forecasting techniques developed based on different logics.

- Moving average, exponential smoothening: Predict future value as a function of past observation.
- Regression based model (AR, MA, ARMA, ARIMA, ARIMAX): Predict future value as a function of past observation. However, are more advanced in terms of handling trend and seasonality effect.

Using complicated method does not guarantee better accuracy in forecasting.

Moving average

Forecast the future value using average or weighted average of past N observations.

$$F_{t+1} = \frac{1}{N} * \sum_{k=t+1-N}^{t} Y_k$$

In case of weighted moving average:

$$F_{t+1} = \sum_{k=t+1-N}^{t} W_k * Y_k$$

W_k is the weight given to past observation. Weight decreases as data becomes older.

$$\sum_{k=t+1-N}^{t} W_k = 1$$

Equation 1 is simple moving average as N past observations have been given equal weight. Assumes that the data does not have significant trend or seasonal component.

Single exponential smoothening

Forecast the future value using differential weights to past observations.

$$F_{t+1} = \alpha * Y_t + (1 - \alpha) * F_t$$

α is called the smoothening constant with a value between 0 to 1.

$$F_{t+1} = \alpha * Y_t + \alpha * (1 - \alpha) * Y_{t-1} + \alpha * (1 - \alpha)^2 * Y_{t-2} \ldots + \alpha * (1 - \alpha)^{t-1} * Y_1 + (1 - \alpha)^2 * F_1$$

Initial value of F_t is taken same as Y_t

Compared to SMA, SES uses the entire history data

Weights assigned declines exponentially with the most recent observation having the highest weight. Model does not perform better in presence of trend.

Double exponential smoothening – Holts Method

Forecast the future value using two equations. One for forecasting level (short term average) and the other for capturing trend

$$L_t = \alpha * Y_t + (1-\alpha) * F_t$$
$$T_t = \beta * (L_t - L_{t-1}) + (1-\beta) * T_{t-1}$$

α, β are called the level and trend smoothening constant with a value between 0 to 1.

$$F_{t+1} = L_t + T_t = \{\alpha * Y_t + (1-\alpha) * F_t\} + \{\beta * (L_t - L_{t-1}) + (1-\beta) * T_{t-1}\}$$

$$F_{t+n} = L_t + nT_t$$

- L_t is the level representing the smoothed value up to the last data point
- T_t is the slope of the line at period t. n is the number of time periods to forecast in future.
- Initial value of L_t is taken same as Y_t

Weights assigned declines exponentially with the most recent observation having the highest weight. Model does not perform better in presence of seasonal component.

Triple exponential smoothening – Holts Winter Method

Forecast the future value using three equations. One for forecasting level second for capturing trend and third for seasonality

$$L_t = \alpha * \frac{Y_t}{S_{t-c}} + (1-\alpha) * [L_{t-1} + T_{t-1}]$$

$$T_t = \beta * (L_t - L_{t-1}) + (1-\beta) * T_{t-1}$$

$$S_t = \gamma * (\frac{Y_t}{L_t}) + (1-\gamma) * S_{t-c}$$

- α, β, γ are called the level, trend and seasonality smoothening constant with a value between 0 to 1.
- c is the number of seasons. (monthly data c = 12, quarterly data c = 4, daily data c = 7 etc.)

$$F_{t+1} = [L_t + T_t] * S_{t+1-c}$$

Holts Winter Method – Initial value calculation

- Level

$$L_t = Y_t$$

or

$$L_t = \frac{1}{c} * (Y_1 + Y_2 + Y_3 \dots + Y_c)$$

- Trend

$$T_t = \frac{1}{c} * (\frac{Y_t - Y_{t-c}}{12} + \frac{Y_{t-1} - Y_{t-c-1}}{12} + \cdots + \frac{Y_{t-c+1} - Y_{t-2c+1}}{12})$$

- Season: Initial seasonality can be calculated using method of simple averages. Seasonality index can be interpreted as percentage change from the trend line.
  - Apart from simple averages, several other techniques also exists

This technique is very sensitive to initial values of level, trend and seasonal index

Regression based Models

When the data consists of $Y_t \; and \; X_t$, regression based models may be better suited than the previous methods. Y_t may have seasonal variation.

$$F_{d,t+1} = \beta_0 + \beta_1 * x_{1,t} + \beta_2 * x_{2,t} + \cdots + \beta_n * x_{n,t} + \varepsilon$$

- Where $F_{d,t+1}$ represents forecasted demand on de-seasonalized data.

Steps to de-seasonalize data:

1. Estimate seasonality index (S_t) using moving average technique.
2. De-seasonalize data using multiplicative model: $Y_{d,t} = Y_t / S_t$
3. Develop forecasting model on de-seasonalized data ($F_{d,t+1}$)
4. The actual forecast for t+1 is $F_{t+1} = F_{d,t+1} * S_t$

Regression models with variable being regressed on itself at different time periods.

Fundamental assumption: Time series data (Y_t) is stationary.

- The mean value of Y_t at different time period is stationary.
- The variance of Y_t at different time periods are constant.
- The co-variance of Y_t and Y_{t-k} for different lags depends only on k and not on time t (ACF, PACF).

AR Model

- AR(1) is auto-regressive model with a lag of 1

$$Y_{t+1} = \beta * Y_t + \varepsilon_{t+1}$$

Can be written as

$$Y_{t+1} - \mu = \beta * (Y_t - \mu) + \varepsilon_{t+1}$$

- ε_{t+1} is a sequence of uncorrelated error with zero mean and constant standard deviation
- $Y_{t+1} - \mu$ can be interpreted as deviation from the mean value μ. This is know as mean centered series. Recursive expansion of the equation

$$Y_{t+1} - \mu = \beta^t * (Y_0 - \mu) + \sum_{k=1}^{t-1} \beta^{t-k} * \varepsilon_k + \varepsilon_{t+1}$$

β is estimated using OLS technique:

- $|\beta| > 1, not\ useful\ for\ practical\ purpose.$
- $|\beta| = 1, implies\ that\ future\ value\ depends\ on\ entire\ past\ (non - stationary).$
- $|\beta| < 1, for\ all\ practical\ purpose.$

AR model with p lags, AR(p) process is

$$Y_{t+1} = \beta_0 + \beta_1 * Y_t + \beta_2 * Y_{t-1} + \cdots + \beta_p * Y_{t-p+1} + \varepsilon_{t+1}$$

How many lags should be used?

Auto-correlation is the correlation between Y_t measured at different time periods.

$$Y_t \; with \; Y_{t-1} \quad Y_t \; with \; Y_{t-2} \ldots \quad Y_t \; with \; Y_{t-k}$$

Auto-correlation can be thought of as a measure which depicts “how far back in time the current observation remembers the past event”

ACF and PACF

Auto-correlation with k-lags (between $Y_t\ with\ Y_{t-k}$)

$$\rho_k = \frac{\left(\sum_{t=k+1}^{n}(Y_{t-k} - \bar{Y}) * (Y_t - \overline{Y})\right)}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}$$

n is the number of observations in the sample. A plot of correlation for different values of k is called as auto-correlation function (ACF).

Partial auto-correlation function (PACF) is the plot of correlation between $Y_t\ with\ Y_{t-k}$, when the influence of all the intermediate values ($Y_{t-1}, Y_{t-2,} \dots, Y_{t-k+1}$) is removed from $Y_t\ and\ Y_{t-k}$

ACF and PACF

- Hypothesis to check whether the ACF and PACF is different than zero.

- H_0: $\rho_k = 0$ *and* H_a: $\rho_k \neq 0$, *where* ρ_k *is auto corelation of order* k
- H_0: $\rho_{pk} = 0$ *and* H_a: $\rho_{pk} \neq 0$, *where* ρ_k *is partial auto corelation of order* k
- *Reject* H_0 *when* $|\rho_k| > \frac{1.96}{\sqrt{n}}$ *and* $|\rho_{pk}| > \frac{1.96}{\sqrt{n}}$

- Thumb rule to select AR(p) :

- The number of lags is p when $|\rho_{pk}| > \frac{1.96}{\sqrt{n}}$ for the first p values and cuts-off to zero.
- The ACF decreases exponentially

After making the model, residual should be checked to see if it follows white noise. This can be done using ACF and PACF plot again. All the correlation values should be within the critical values.

MA Model

$Intution$ behind MA process is that the error at the current period ε_t and at the next period ε_{t+1} drives the next value of the time series

- Past residuals are used to forecast the future values. MA(1) model with a lag of 1

$$Y_{t+1} = \mu + \ \alpha * \varepsilon_t + \varepsilon_{t+1}$$

- ε_{t+1} is a sequence of uncorrelated error with zero mean and constant standard deviation

MA model with q lags, MA(q) process is

$$Y_{t+1} = \mu + \alpha_1 * \varepsilon_t + \alpha_2 * \varepsilon_{t-1} + \cdots + \alpha_q * \varepsilon_{t-q+1} + \varepsilon_{t+1}$$

- Thumb rule to select MA(q) :

- The number of lags is q when $|\rho_k| > \frac{1.96}{\sqrt{n}}$ for the first q values and cuts-off to zero.
- The PACF decreases exponentially

ARMA

$$Y_{t+1} = \beta_0 + \beta_1 * Y_t + \beta_2 * Y_{t-1} + \cdots + \beta_p * Y_{t-p+1} + \mu + \alpha_1 * \varepsilon_t + \alpha_2 * \varepsilon_{t-1} + \cdots + \alpha_q * \varepsilon_{t-q+1} + \varepsilon_{t+1}$$

Thumb rule to select ARMA(p, q) :

- The number of lags is q when $|\rho_k| > \frac{1.96}{\sqrt{n}}$ for the first q values and cuts-off to zero.
- The number of lags is p when $|\rho_{pk}| > \frac{1.96}{\sqrt{n}}$ for the first p values and cuts-off to zero

ARMA can be only used when the time series is stationary

ARIMA Process

- ARIMA can be used when the time series is non-stationary. Also know as Box and Jenkins Methodology.

Three components:

- AR component with p lags
- Integration component (d)
- MA component with q lags.

Objective of integration component is to convert non-stationary series into stationary series.

When series in non-stationary:

- ACF will not cut-off to zero quickly but will show a very slow decline (Visual inspection)
- Dickey fuller test or augmented dickey fuller (when error does not follow white noise) test can be used to test non-stationarity.

$$H_0: \beta = 1 \ (Time\ series\ is\ non - stationary)$$

DF test statistics is

$$DF\ Test\ statistics = \frac{\psi}{se(\psi)}$$

Transforming non-stationary series into stationary:

- Differencing
    - Identity order of differencing (d)
    - First order differencing (d=1) is difference between consecutive values of the time series ($Y_t\ and\ Y_{t-1}$)

$$\nabla Y = Y_t - Y_{t-1}$$

    - Second order differencing is difference of the first difference.

$$\nabla(\nabla Y) = Y_t - 2Y_{t-1} + Y_{t-2}$$

In most case, first or second order differencing will convert a non-stationary process to a stationary process.

ARIMA Model Building Steps

1. Plot ACF and PACF and perform Dickey Fuller Test

2. Is process stationary
   - Yes, set d = 0
   - No, identify the order of differencing. Perform a Dickey fuller test again.

3. Identify the value of (p and q). Perform a grid search to identify the value, if needed.

4. ARIMA model is ARIMA(p,d,q)

5. Analyze if error follows white noise. ACF and PACF plot of error should cut to zero.
   - Ljung-Box test, if the visual inspection is not conclusive.
     - $H_0: The\ auto-correlation\ for\ the\ errors\ are\ equal\ to\ zero$
     - $H_a: The\ auto-correlation\ for\ the\ errors\ are\ different\ from\ zero$

End of Lesson09–Time Series Concepts With ARIMA