Mechatronics

Day 3 - 4

Supervised Learning

Regression

for Time-series data

Outline

- Time series forecasting
- ARIMA model

Time Series Forecasting

It is the process of using a statistical model to predict future values of a time series based on past results.

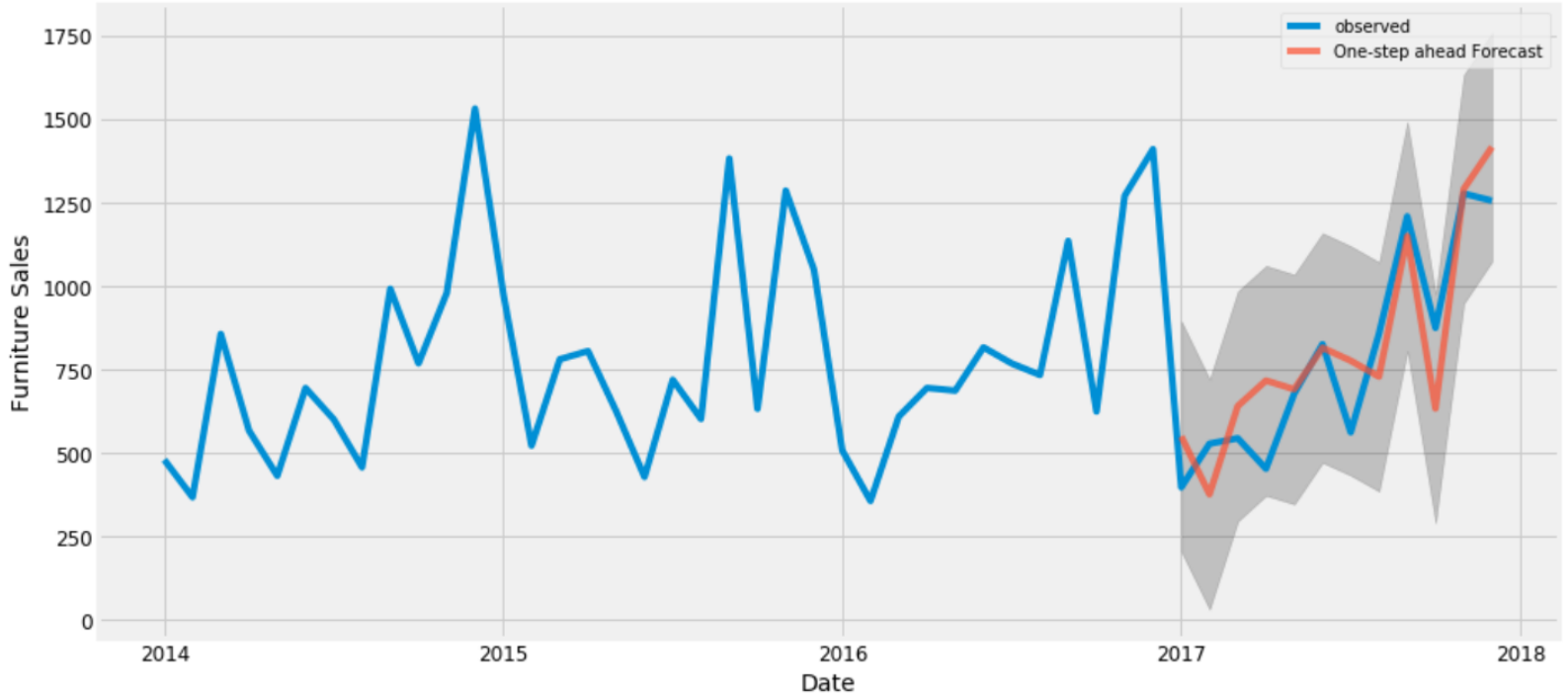
Some Use Cases

- To predict the number of incoming or churning customers.
- To explaining seasonal patterns in sales.
- To detect unusual events and estimate the magnitude of their effect.
- To Estimate the effect of a newly launched product on number of sold units.

Time series forecasting

Time series forecasting can broadly be categorized into the following categories:

- Machine Learning —ARIMA, SARIMA, TBATS, XGBoost, Random Forest etc.
- Deep Learning — RNN, LSTM



https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b

Components of a Time Series (1)

Trend

- It shows a general direction of the time series data over a long period of time.
- A trend can be increasing(upward), decreasing(downward), or horizontal(stationary).

Seasonality

- It exhibits a trend that repeats with respect to timing, direction, and magnitude.
- Some examples include an increase in water consumption in summer due to hot weather conditions, or an increase in the number of airline passengers during holidays each year.

Cyclical Component

- These are the trends with no set repetition over a particular period of time.
- A cycle refers to the period of ups and downs, booms and slums of a time series, mostly observed in business cycles.
- These cycles do not exhibit a seasonal variation but generally occur over a time period of 3 to 12 years depending on the nature of the time series.

Components of a Time Series (2)

Irregular Variation

- These are the fluctuations in the time series data which become evident when trend and cyclical variations are removed.
- These variations are unpredictable, erratic, and may or may not be random.

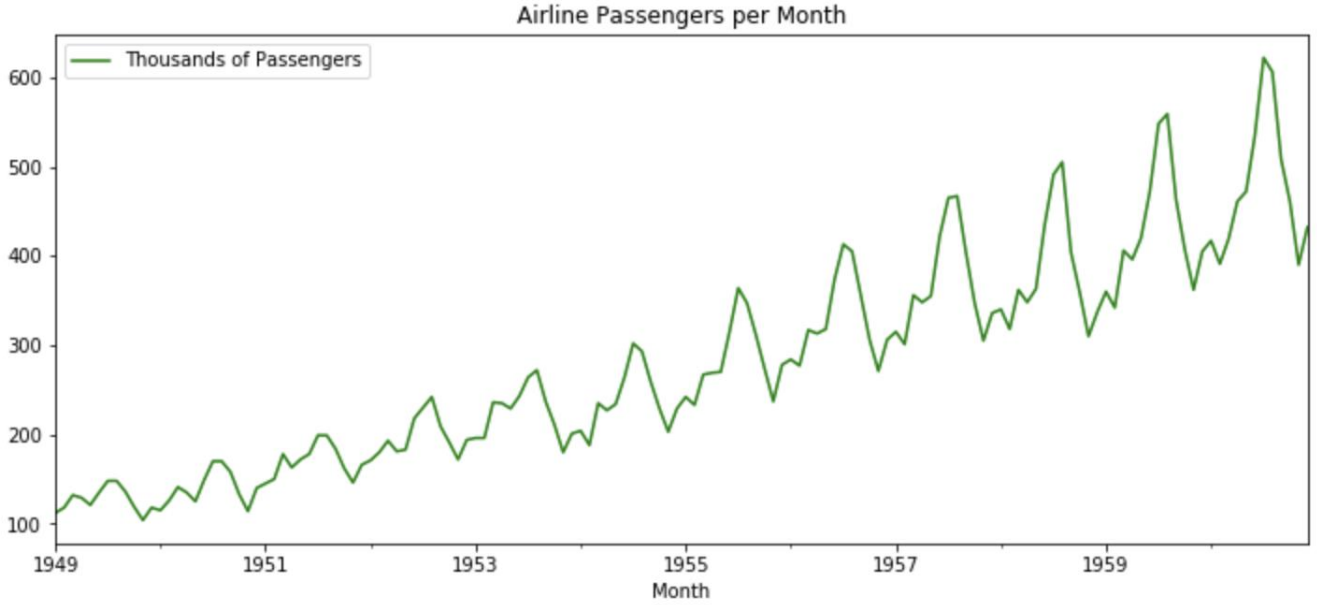
ETS Decomposition

- It is used to separate different components of a time series.
- The term ETS stands for Error, Trend, and Seasonality.

Example :

The example of a Time Series that illustrates the number of passengers of an airline per month from the year 1949 to 1960.

https://www.kaggle.com/datasets/abhishekmamidi/air-passengers

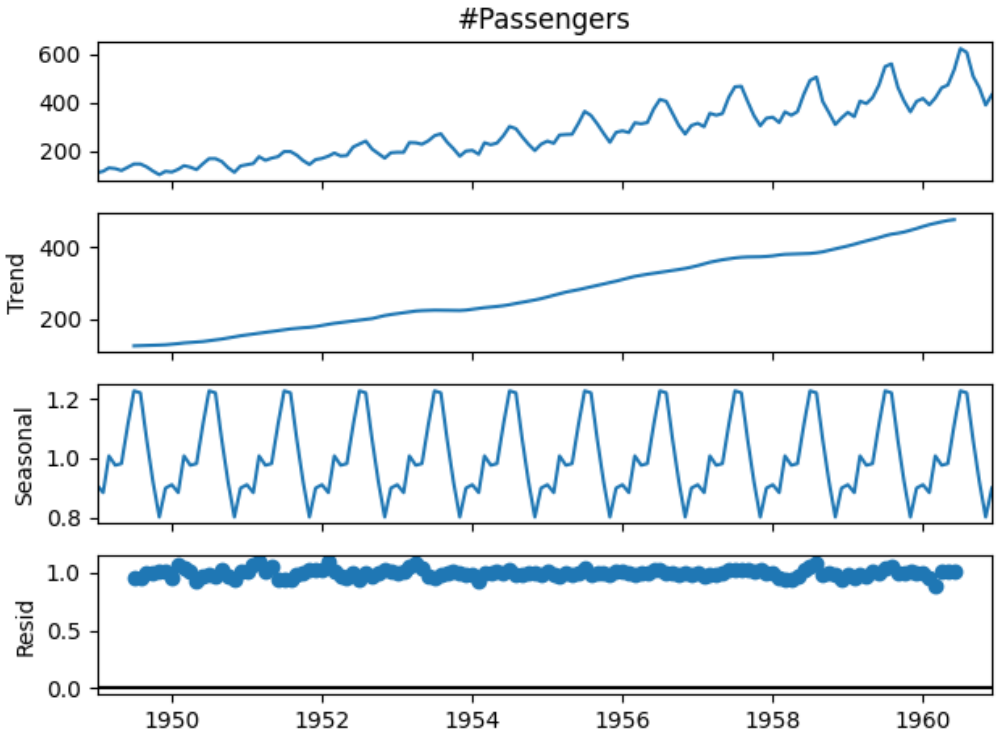


ETS Decomposition of Airline Passengers Dataset:

```
# Importing required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

# Read the AirPassengers dataset
airline = pd.read_csv('AirPassengers.csv',
                      index_col ='Month',
                      parse_dates = True)
# Print the first five rows of the dataset
airline.head()
# ETS Decomposition
result = seasonal_decompose(airline['#Passengers'],
                            model ='multiplicative')
# ETS plot
result.plot()
```

Month	#Passengers
1949-01-01	112
1949-02-01	118
1949-03-01	132
1949-04-01	129
1949-05-01	121



ARIMA Model for Time Series Forecasting

- ARIMA stands for autoregressive integrated moving average model and is specified by three order parameters: (p, d, q).

$$ARIMA(p, d, q)$$

$$y_t^{(d)} = c + \varepsilon_t + \underbrace{\phi_1 y_{t-1}^{(d)} + \phi_2 y_{t-2}^{(d)} + \ldots + \phi_p y_{t-p}^{(d)}}_{\text{Auto-Regressive}} + \underbrace{\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_p \varepsilon_{t-q}}_{\text{Moving Average}}$$

Integrated

ARIMA Model for Time Series Forecasting

AR(p) Autoregression

- A regression model that utilizes the dependent relationship between a current observation and observations over a previous period.
- An auto regressive (AR(p)) component refers to the use of past values in the regression equation for the time series.

I(d) Integration

- Uses differencing of observations (subtracting an observation from observation at the previous time step) in order to make the time series stationary.
- Differencing involves the subtraction of the current values of a series with its previous values d number of times.

MA(q) Moving Average

- A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.
- A moving average component depicts the error of the model as a combination of previous error terms.
- The order q represents the number of terms to be included in the model.

Types of ARIMA Model

ARIMA

- Non-seasonal Autoregressive Integrated Moving Averages

SARIMA

- Seasonal ARIMA

SARIMAX

- Seasonal ARIMA with exogenous variables

Fit ARIMA Model to AirPassengers dataset

```
# Split data into train / test sets
train = airline.iloc[:len(airline)-12]
# set one year(12 months) for testing
test = airline.iloc[len(airline)-12:]

# Fit a SARIMAX(0, 1, 1)x(2, 1, 1, 12) on the training set
from statsmodels.tsa.statespace.sarimax import SARIMAX

model = SARIMAX(train['#Passengers'],
                order = (0, 1, 1),
                seasonal_order =(2, 1, 1, 12))

result = model.fit()
result.summary()
```

Dep. Variable:	y	**No. Observations:**	144
Model:	SARIMAX(0, 1, 1)x(2, 1, 1, 12)	**Log Likelihood**	-501.921
Date:	Fri, 31 Jan 2020	**AIC**	1015.842
Time:	18:07:17	**BIC**	1033.093
Sample:	0	**HQIC**	1022.852
	- 144		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	0.0004	0.033	0.011	0.991	-0.064	0.064
ma.L1	-0.4259	0.068	-6.223	0.000	-0.560	-0.292
ar.S.L12	0.6682	0.160	4.189	0.000	0.356	0.981
ar.S.L24	0.3302	0.096	3.447	0.001	0.142	0.518
ma.S.L12	-0.9734	1.188	-0.820	0.412	-3.301	1.354
sigma2	111.0288	108.931	1.019	0.308	-102.472	324.530

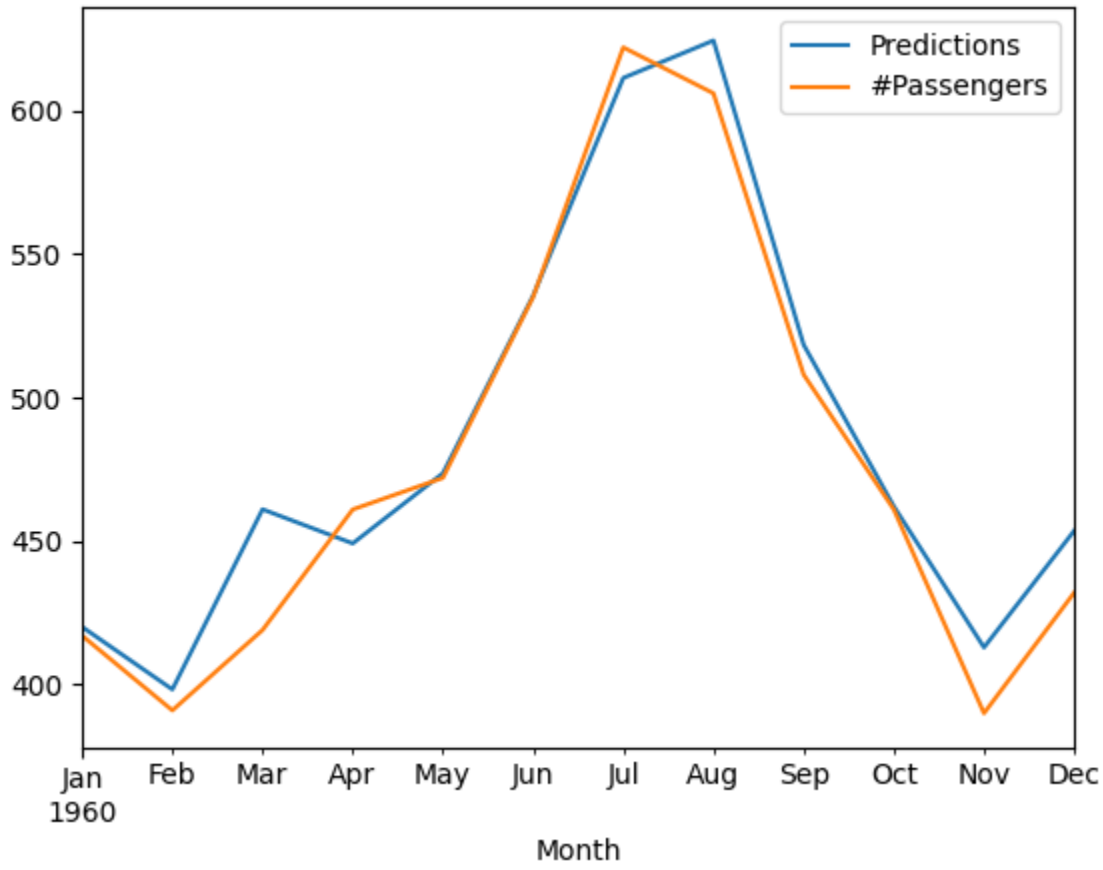
Ljung-Box (Q):	52.94	**Jarque-Bera (JB):**	7.41
Prob(Q):	0.08	**Prob(JB):**	0.02
Heteroskedasticity (H):	2.82	**Skew:**	0.10
Prob(H) (two-sided):	0.00	**Kurtosis:**	4.15

Predictions of ARIMA Model against the test set

```
start = len(train)
end = len(train) + len(test) - 1

# Predictions for one-year against the test set
predictions = result.predict(start, end,
                             typ =
'levels').rename("Predictions")

# plot predictions and actual values
predictions.plot(legend = True)
test['#Passengers'].plot(legend = True)
```



Evaluate the model using MSE and RMSE

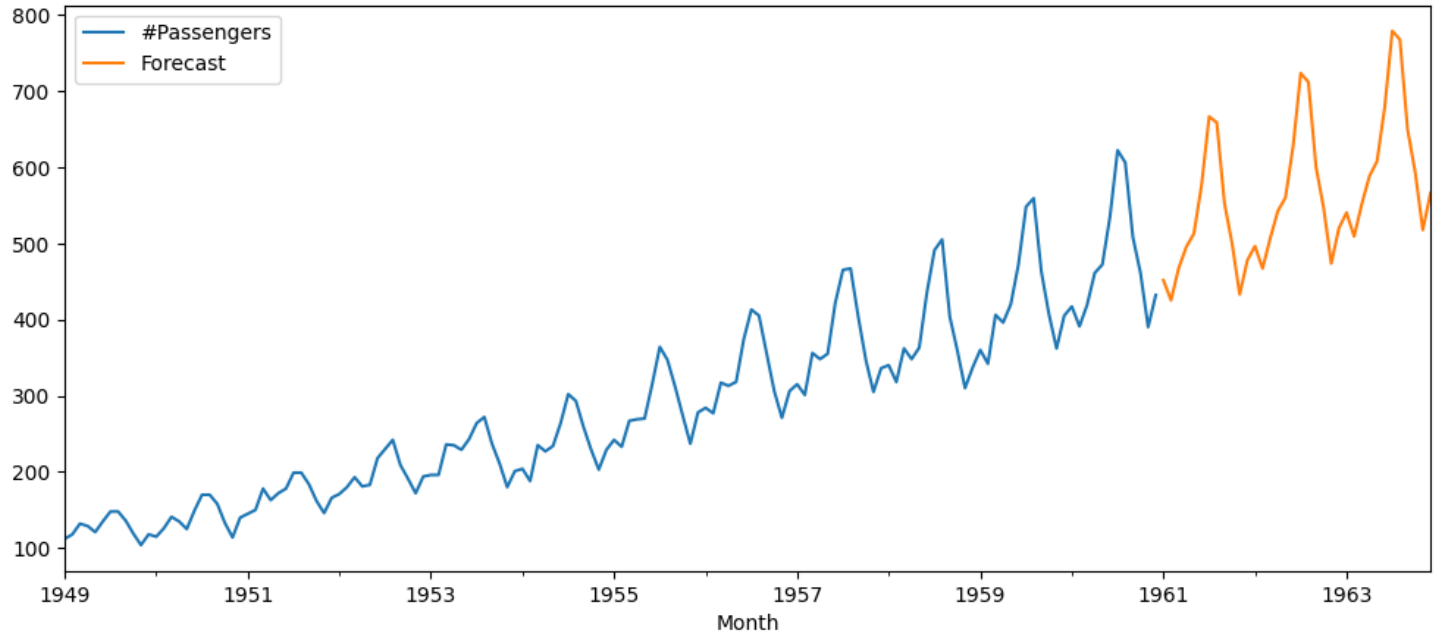
```
# Load specific evaluation tools
from sklearn.metrics import mean_squared_error

# Calculate mean squared error
mse = mean_squared_error(test["#Passengers"], predictions)
print("MSE=",mse)
# Calculate root mean squared error
rmse = np.sqrt(mse)
print("RMSE=",rmse)
```

MSE= 293.8842792561804
RMSE= 17.143053381943965

Forecast using ARIMA Model

```
# Train the model on the full dataset
model = model = SARIMAX(airline['#Passengers'],  order = (0, 1, 1),  seasonal_order =(2, 1, 1, 12))
result = model.fit()
# Forecast for the next 3 years
forecast = result.predict(start = len(airline),
                          end = (len(airline)-1) + 3 * 12,
                          typ = 'levels').rename('Forecast')
# Plot the forecast values
airline['#Passengers'].plot(figsize = (12, 5), legend = True)
forecast.plot(legend = True)
```



Read more

Time Series Forecasting Tutorial

https://www.datacamp.com/tutorial/tutorial-time-series-forecasting

ARIMA for Time Series Forecasting: A Complete Guide

https://www.datacamp.com/tutorial/arima

Gold Price Prediction Using ARIMA Model

https://www.kaggle.com/code/fahadrehman07/gold-price-prediction-using-arima-model/notebook


SUT
SURANAREE
UNIVERSITY OF
TECHNOLOGY
INSTITUTE OF
ENGINEERING
Mecha tronics
Modern Automotive