Aniket Anand PES1201700185

Kundan Kumar PES1201700109

ASSIGNMENT 6

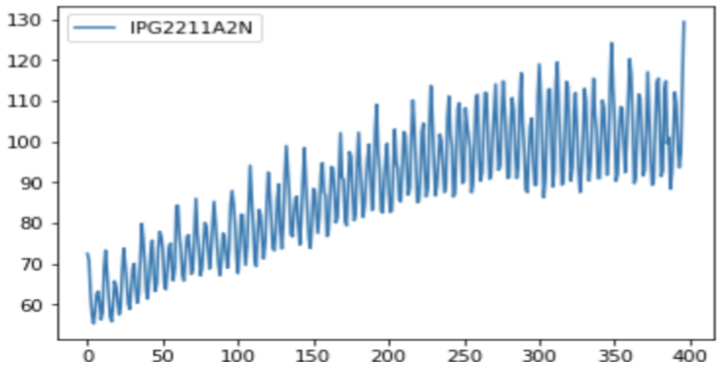
Problem statement:

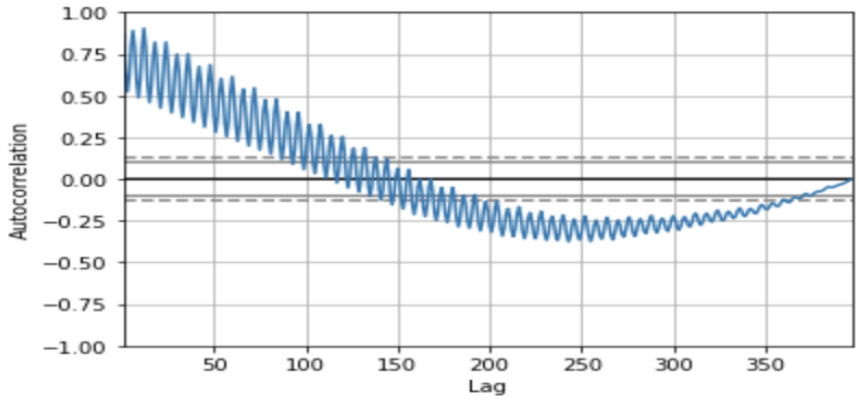
On a chosen time-series dataset, analyse the autoregressive moving average (ARMA) and from it, the autoregressive integrated moving average (ARIMA) and their respective Mean Squared Errors (MSE).

1)ARIMA

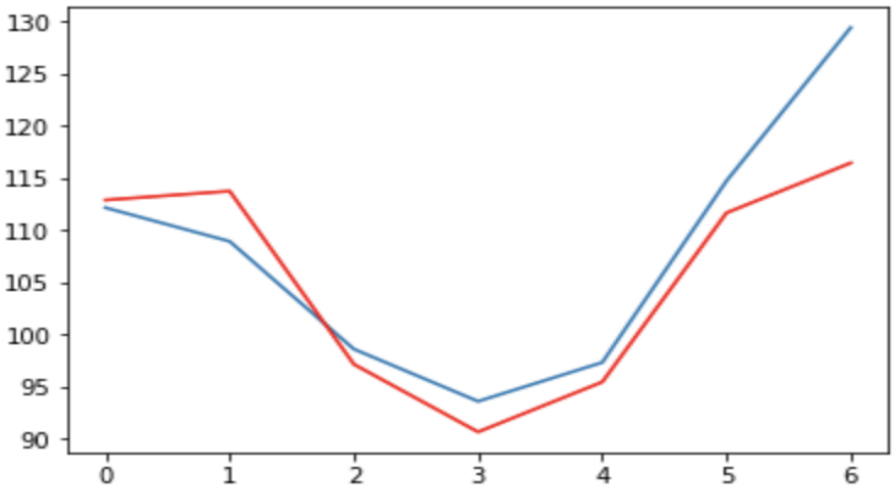
```
from pandas import read_csv
from pandas import datetime
from pandas import DataFrame
from matplotlib import pyplot
from pandas import DataFrame
from pandas import concat
from pandas.plotting import autocorrelation_plot
from sklearn.metrics import mean_squared_error
from statsmodels.tsa.ar_model import AR


series = read_csv('Electric_Production.csv')
X = list(series['IPG2211A2N'])
print(series.head())
series.plot()
pyplot.show()


autocorrelation_plot(X)
pyplot.show()


train, test = X[1:len(X)-7], X[len(X)-7:]
# train autoregression
model = AR(train)
model_fit = model.fit()
window = model_fit.k_ar
coef = model_fit.params
# walk forward over time steps in test
history = train[len(train)-window:]
history = [history[i] for i in range(len(history))]
predictions = list()
for t in range(len(test)):
    length = len(history)
    lag = [history[i] for i in range(length-window,length)]
    yhat = coef[0]
    for d in range(window):
        yhat += coef[d+1] * lag[window-d-1]
    obs = test[t]
    predictions.append(yhat)
    history.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
error = mean_squared_error(test, predictions)
print('Test MSE: %.3f' % error)
# plot
pyplot.plot(test)
pyplot.plot(predictions, color='red')
pyplot.show()
```

	DATE	IPG2211A2N
0	1/1/1985	72.5052
1	2/1/1985	70.6720
2	3/1/1985	62.4502
3	4/1/1985	57.4714
4	5/1/1985	55.3151






```
predicted=112.890539, expected=112.153800
predicted=113.744904, expected=108.931200
predicted=97.161840, expected=98.615400
predicted=90.680653, expected=93.613700
predicted=95.465050, expected=97.335900
predicted=111.663226, expected=114.721200
predicted=116.438962, expected=129.404800
Test MSE: 30.771
```



2)ARIMA

In [14]:
```
from pandas import datetime
from matplotlib import pyplot
import pandas as pd
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_squared_error


series = pd.read_csv('Electric_Production.csv')
X = list(series['IPG2211A2N'])

size = int(len(X) * 0.66)
train, test = X[0:size], X[size:len(X)]
history = [x for x in train]
predictions = list()
for t in range(len(test)):
    model = ARIMA(history, order=(5,1,0))
    model_fit = model.fit(disp=0)
    output = model_fit.forecast()
    yhat = output[0]
    predictions.append(yhat)
    obs = test[t]
    history.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
error = mean_squared_error(test, predictions)
print('Test MSE: %.3f' % error)
# plot
pyplot.plot(test)
pyplot.plot(predictions, color='green')
pyplot.show()
```

```
predicted=92.957877, expected=92.356600
predicted=102.045371, expected=103.066000
predicted=111.623817, expected=112.057600
predicted=109.152008, expected=111.839900
predicted=102.331260, expected=99.192500
predicted=90.888934, expected=90.817700
predicted=92.979633, expected=92.058700
predicted=102.055367, expected=100.967600
predicted=110.018704, expected=107.568600
predicted=107.106122, expected=114.103600
predicted=106.837345, expected=101.531600
predicted=92.125922, expected=93.006800
predicted=92.883488, expected=93.912600
predicted=101.107785, expected=106.752800
predicted=115.587192, expected=114.833100
predicted=112.864392, expected=108.235300
predicted=98.709460, expected=100.438600
predicted=94.557913, expected=90.994400
predicted=94.448943, expected=91.234800
predicted=102.201825, expected=103.958100
predicted=95.478835, expected=97.335900
predicted=105.678418, expected=114.721200
predicted=119.382867, expected=129.404800
Test MSE: 19.297
```