郭冠廷 Mimi
SID 107062274
2020/Oct/10

COVID-19-Forecast

KUO, KUAN-TING

This is a project that aims to predict the forthcoming patient cases from Russia, Greece, India, United_States_of_America and Turkey. The predicted figures of the forthcoming week, from Oct 9th to Oct 15th, are written into **result.csv**.

Model and Features

- AR model from statsmodels
- Set different window sizes for different countries

How to use it?

The predicted cases of these five countires, which are saved in the **result.csv**, can be regenerated simply by running the entire .ipynb file. This action would produce then produced a file named **result_regenerated.csv**.

Steps

The model training is entirely using **Google Colab**.

Installation - install statsmodels

```
pip install statsmodels --upgrade
```

Preprocess

- Load Data
- Mark all NA values with 0 + turn negative values into positive
- Split into different csv by countries

Create new csv file for saving result

```
result_df = pd.DataFrame({'':['10/9', '10/10','10/11', '10/12','10/13',
'10/14','10/15']})
for file in os.listdir(split_csv_folder):
```

郭冠廷 Mimi
SID 107062274
2020/Oct/10

```
  if (file == 'Russia.csv') or (file == 'Greece.csv') or (file == 'India.csv') or
(file == 'United_States_of_America.csv') or (file == 'Turkey.csv'):
    result_df[file[:-4]] = ''
result_df.to_csv(path + 'result.csv', index = False)
```

Autoregression

- Select different window sizes (1 to 50) for different countries to minimize the MAPE value
- Save the min MAPE values and window sizes for each countries in a dictionary

Results

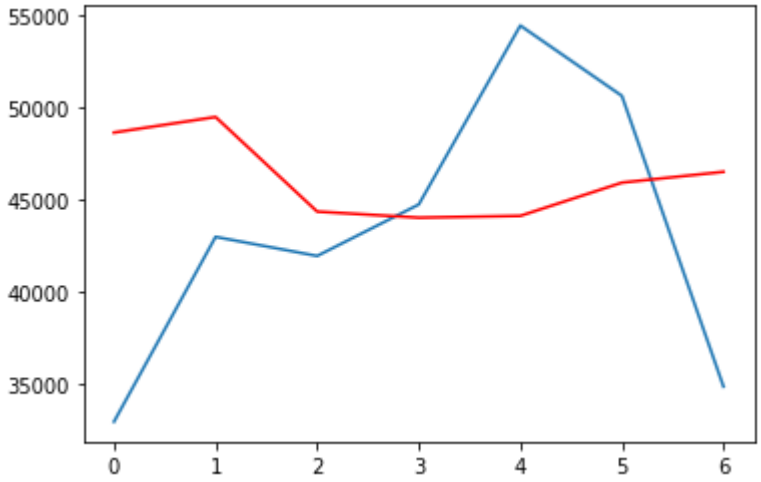

Figure 1: United_States_of_America

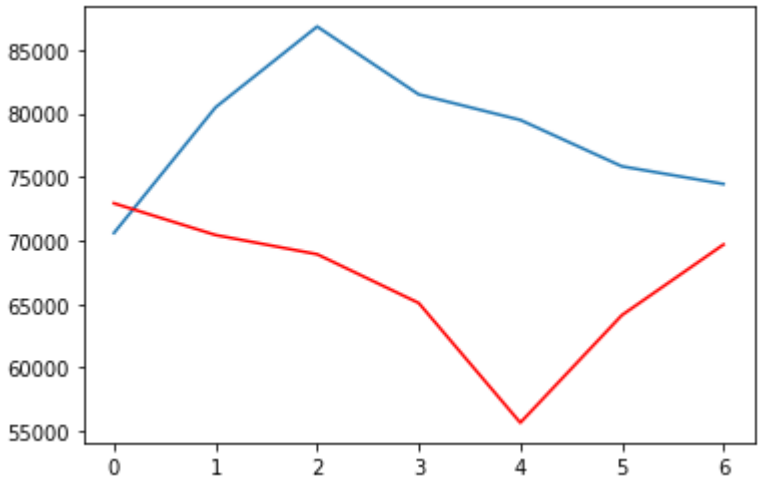

Figure 2: India

郭冠廷 Mimi
SID 107062274
2020/Oct/10

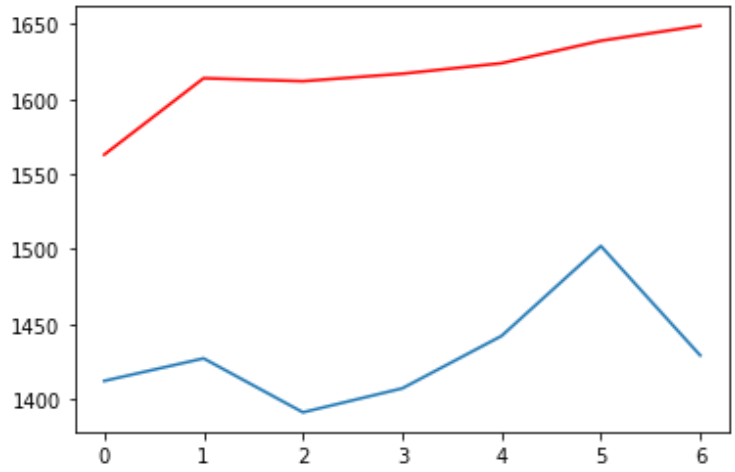

Figure 3: Turkey

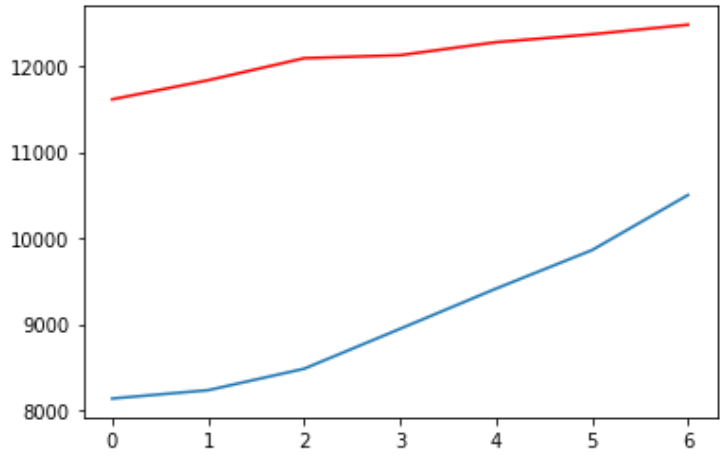

Figure 4: Russia

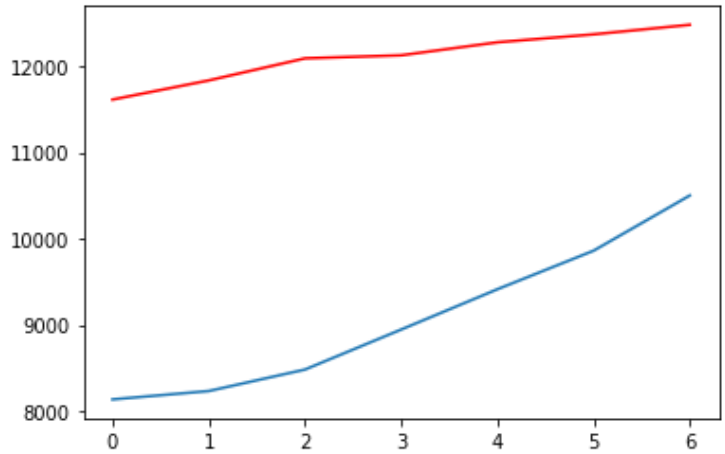

Figure 5: Greece

郭冠廷 Mimi
SID 107062274
2020/Oct/10

Problem faced

- Bad MAPE results with linear regression --> switched to autoregression
- Still bad MAPE results with autoregression with a fixed window size --> try to give each country a customized window size

Reference

AR model

- https://pythondata.com/forecasting-time-series-autoregression/
- https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/

Dataset

Collected from https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide on **Oct 8th** (included)

Github Link

https://github.com/ting1123/COVID-19-Forecast