Lab 3: Exercises with NumPy, Pandas, and Matplotlib[1] (Week 2.1)

1 – Objectives

With this work the student should be able to start making some programs in Pandas and Matplotlib, showing some results in graphical format.

2 – Outlier Removal

Use the file EURUSD_Daily_Ask_2018.12.31_2019.10.05v2.csv as your data file. Read the file and plot it using pandas and matplotlib. Make sure the program is able to open the file and plot it correctly.

Use the function "to_datetime" from pandas to transform the string containing the date to the internal format datetime. Also import "datetime" and create a new element using the following instruction:

```
startTime = datetime(year, month, day, 0, 0)
```

You can also use the function `now()` to get the current time.

Now create a function to detect and remove outliers. Start by plotting the data and creating two functions that do the following:

- Find a value;
- Detect the samples that are k*σ far from the average.

Now create three different functions to remove the outlier:

- Remove the line in the pandas dataset;
- Change the value to the previous point (in the data series);
- Change the value to the interpolation of the previous and the next points.

Finally plot the results without the outliers (and save them).

2 – Some statistical plots

Use the file DCOILBRENTEUv2.csv as your data file. Read the file and plot it using pandas and matplotlib. Make sure the program is able to open the file and plot it correctly.

Plot an histogram where in the x axis you plot the variation from the previous day (Close from the previous day to Close from the present day).

Plot another histogram with the variation in the day (the difference from High to Low in each day). Save both histograms to the previously created .pdf file

Use the file DCOILBRENTEUv2.csv and DCOILWTICOv2.csv as your data files. Brent is the price of oil in the UK in dollars, and WTI (West Texas Intermediate) is the price of oil in Texas, USA also in dollars. They have similar values, but some time differ a little because of transportation and refining issues. Read the files and plot them in a single figure using

[1] Rui Neves, João Paulo Carvalho

pandas and matplotlib. Make sure the program is able to open the file correctly and plot it correctly.

Now make a Scatter plot with the value oil of the different regions and see how they behave.