Lab of Applied Computational Intelligence

IST

2024/2025

Some Examples in Pandas and Matplotlib

Guide 3

17 September 2024

(Week 2)

1 – Objetives

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 correctly and plot it.

Use the function "to_datetime" from pandas to create a new column with the information about the time in the file that is read as a string to the internal format datetime. Also import the library "datetime" and use it to create a new element in the format datetime. You can use the following function to create a new datetime where you explicitly give the values startTime = datetime(year, month, day, 0, 0) or use the function now() to get the current time.

Now make a function to detect and remove the outlier. Plot the data. Create two different functions to detect the outlier:

1) Find the outlier value (just by looking at the figure).
2) Detect the samples that are k*σ far from the average.

Now make three different functions to remove the outlier.

3) Remove the line in the pandas dataset.
4) Change the value with the previous one.
5) Change the value with the interpolation of the previous and the next one.

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

2 –Histogram

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

Plot a 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).

Now use your creativity to plot some interesting statistic for this data file.

3 –Scatter Plot

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 then in a single figure using 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.