COM6018 Data Science with Python

Lab 3: Introducing NumPy

Jon Barker

In this lab

- Reading datasets from CSV files
- Smoothing data
- Finding local maxima and minima in data
- Using NumPy
- Comparing NumPy with Python list processing

The Task

- You will be analysing the atmospheric gas concentration data.
- The idea it to try and better understand the oscillations in the data.
- We will be using NumPy to help us with the analysis.

The Data

We are starting with the `co2e.csv` that we constructed in the previous lab.

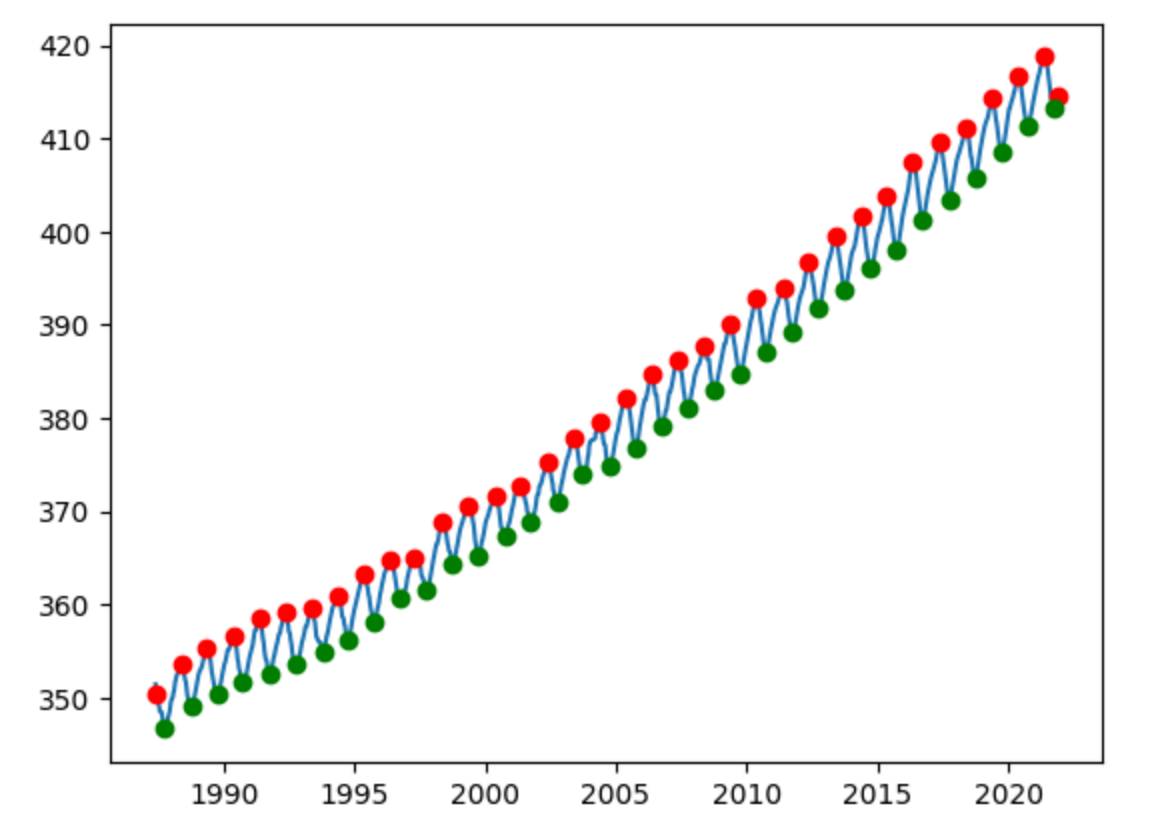
This is stored in the file `data/co2e.csv` and looks like

```
year,month,day,co2_concentration,ch4_concentration,co2e,decimal_year
1987,4,3,350.84,1.70002,393.34049999999996,1987.2554794520547
1987,4,4,350.28,1.69427,392.63674999999995,1987.2582191780823
1987,4,5,350.4,1.69821,392.85524999999996,1987.2609589041097
1987,4,9,349.47,1.67721,391.40025,1987.271917808219
1987,4,10,349.81,1.67019,391.56475,1987.2746575342467
1987,4,11,350.01,1.68991,392.25775,1987.277397260274
1987,4,13,350.13,1.68973,392.37325,1987.2828767123287
...
```

The Aim

We want to locate the local maxima and minima in the `co2e` values.

We will produce a figure like the one below.



The Aim

Once we have located the local maxima (peaks) and minima (dips) we will answer the following questions:

- What is the average difference between the annual peak and dip values?
- What is the average time (in days) between a peak and the next dip?
- What is the average time (in days) between a dip and the next peak?
- On which day of the year (on average) do the peaks and dips occur?

The Aim

The lab has two stages:

- Processing the data using pure Python
- Processing the data using NumPy

The purpose is to compare the two approaches and to see the advantages of using NumPy.

Obtaining the Jupyter Notebook

If you have cloned the module's GitHub repository then you should see,

```
materials/labs/
├── 030_introducing_numpy.ipynb
├── ... etc
├── data
│   ├── co2e.csv
│   ├── ... etc
```

(*Remember to run* `git pull` *to get the latest version of the repository.*)

The lab is `030_introducing_numpy.ipynb` and it will need the file `data/co2e.csv`.

Or you can download the notebook and data via links on Blackboard.