

COM6018 Data Science with Python

Week 9: Curve Fitting with scikit-learn

Jon Barker

Copyright © 2023–2025 Jon Barker, University of Sheffield. All rights reserved.

In this lab

We will use scikit-learn to fit the atmospheric CO₂ concentration data.

We will then use our model to predict future atmospheric CO₂ concentration.

On what date will it reach 450 ppm?

The Task

The stages of the task are as follows:

- Loading the CO₂ data.
- Fitting a polynomial curve to describe the long-term growth trend.
- Fitting a periodic function to describe the seasonal variation.
- Tuning the model hyperparameters (order of polynomial, etc.).
- Evaluating the model.
- Using the model to make a prediction.

Obtaining the Jupyter Notebook

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

```
materials/labs/
├── 080_curve_fitting_with_scikit_learn.ipynb
└── data
    └── co2.csv
        ... etc.
```

The lab is `080_curve_fitting_with_scikit_learn.ipynb` and it will need the data file `data/co2.csv`.

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

Getting Help

- If you are stuck, just raise a hand to ask for help.
- Feel free to discuss the lab with your neighbours.
- Re-read the scikit-learn tutorial notes:
 - In the Git repo at
`materials/tutorials/080_Curve_Fitting_with_Scikit_Learn.ipynb`
 - or online at <https://uos-com-6018.github.io/COM6018>
- Use the scikit-learn API documentation for reference: <https://scikit-learn.org/stable/modules/classes.html>