Data Analysis Challenge 🌶🌶🌶🌶🌶

Linear regression and mass data fitting

Python is well known to be used for massive data processing. In this challenge, we will read in, visualize and process a large quantity of measurement data in an automated fashion.

Download the sample data from `http://aypac.de/Dateien/fitting_data.zip`. Place the data in the same location as the notebook.

Try to solve each challenge without looking at the hints, but feel free to use whatever else information source. If you feel a bit stuck look at the hint. If you are still stuck, do not hesitate to reach out to one of the trainers. After finishing one of the challenges, you might find it useful to also reach out to one of the trainers to discuss it.

If you cannot finish it all, that does not matter. If you want to you can finish it at home. If you need help or have your answers checked, you can reach out to us at an time later.

Enjoy the challenges!

Challenge 0 (Warm-up)

Familiarize yourself with nested lists. Try to create a list of all members of your team. Each item is a list of your first and last name and your age. Then try to add the ages of everyone in that list.

Hint 0.1

Read the first section of lesson 9 on snakify.org.

Challenge 1

Find a way to read-in data from a csv file. Choose the file 'linear.csv' to test it (it should already be in the directory of the cloned notebook).

Hint 1.1

Use the package `csv`.

Challenge 2

Familiarize yourself with the selection of certain pieces of data (rows, columns, fields). Plot the data using the package `matplotlib`.

Challenge 3

Try to figure out how to fit a function to a data-set. Fit the data-set you have with a linear function: $f(x) = a \cdot x + b$. Don't worry if you have to use the hints below (but try without first) - this is a hard one!

Hint 3.1

You can use 'scipy.optimize.curve_fit'. Try to find a usage example and adapt it.

Challenge 4

Add the fitted function to the plot and save it into an svg and png file.

Challenge 5

Now copy your notebook and adjust it to read in one of the numbered csv files. Fit the function $g(x) = a \cdot (x + b)^2 + c$ to it. Plot the data and fit and save to a file.

Hint 5.1

The power function x^2 is `x**2` in Python!

Challenge 6 (Bonus)

Now automatically apply this step for all the numbered files in the data-set (make sure the file you save your plot into is using the same number as the csv file name). Then plot the parameter a as a function of the filename/-number.

Hint 6.1

You will need a command to list all files in the data folder.

Hint 6.2

Assume the filename is in a variable called `filename`. You can the get the number like this:

```
number = float(filename.split('.')[0])
```

Challenge 7 (Bonus)

Research the library `numpy` and translate all nested lists into numpy arrays.

Challenge 8 (Bonus)

Figure out how to obtain the residuals (figure of merit/"goodness" of the fit).