Foundations of Data Analysis - WS24
Instructions for Python environments (Lab Assignment)

Date: 13.11.2024

Intro

This is a guide (explaining a few ways) to recreate the Python environment for the lab assignment either manually or using a virtual environment (venv) or through installing a Python distribution called Anaconda/Miniconda (conda). This guide shall be additional help for you to set everything up correctly and I do not guarantee correctness nor claim that it is exhaustive. It is your responsibility to produce code that works in the provided environment. If you cannot find help within this guide, please do your own google search, use chat-gpt or ask your colleagues for help. If you have specific questions, feel free to email me (Christoph Luther) at `christoph.luther@univie.ac.at` or post on the discussion forum on Moodle.

Remark

I tested the solutions on Mac and Linux but not Windows. If you experience any problems, you can try any Python 3.10.XX version and follow the instructions from the last section of this document (i.e., manually recreating the environment). But please indicate so in your submission.

Python virtual environment

Virtualenv is a module that comes with every Python version > 3.4. You can use it to create a (isolated) Python virtual environment. When you do so, make sure to use version 3.10.14. You can look up the following two pages, that I find helpful, for that:

- https://docs.python.org/3/library/venv.html
- https://python.land/virtual-environments/virtualenv

Then activate the environment and execute **pip install -r requirements.txt**, where requirements.txt is the file we provide and has to be in your current directory (else change the path).

Anaconda/Miniconda

You can install Anaconda directly from `https://www.anaconda.com/download`.

For a minimal version that uses less space, you can install Miniconda, `https://docs.anaconda.com/free/miniconda/index.html` (see quick command line install at bottom of page).

Once installed, you should be able to use the command `conda` in your terminal.

Setting up a new environment (with specified Python version):

- `conda create --name <name> python=3.10.14`
- `conda activate <name>`

Inside the environment, you can install libraries using:

- `conda install <library>`, e.g. `conda install pip`
- then you can also use `pip install <library>`

More helpful commands around conda environments: https://docs.conda.io/projects/conda/en/4.6.0/_downloads/52a95608c49671267e40c689e0bc00ca/conda-cheatsheet.pdf

Jupyter notebooks

We use the computing platform Jupyter (Jupyter notebooks) to write code. To make sure that you use the correct Python version as kernel in your Jupyter notebook, it is easiest to start the notebook inside your environment. To do so execute the following command after you have activated your environment:

- `jupyter notebook`

Inside your Jupyter notebook, you can then simply choose the default (Python 3 (ipykernel)) as your kernel. Alternatives to this way are using an IDE that supports the notebook format (`.ipynb`)[1] or setting your conda environment as kernel for your Jupyter notebook and start the notebook outside of your conda environment (see this post: https://stackoverflow.com/questions/58068818/how-to-use-jupyter-notebooks-in-a-conda-environment).

Manually recreating the environment

In case, none of the above solutions work for you, you can resort to manually setting up everything.

NOTE: If you do this, please indicate so in the first cell of your submission.

1. Install Python version 3.10.14 or setup virtual environment using version 3.10.14 or a conda environment using python=3.10.14 (or any 3.10.XX)
2. If you did not use a conda environment, the package manager `pip` is included in your python version; if you use a conda environment, first run `conda install pip`
3. Install the necessary libraries according to `pip install <library>` and replace `<library>` with [2]:
   - `jupyter`
   - `pandas`
   - `matplotlib`
   - `numpy`
   - `seaborn`
   - `scikit-learn`

[1] e.g., vsc or pycharm

[2] See also the import statements in the template files.