Python Notes¶

Introduction to Python for Econometrics, Statistics and Numerical Analysis: Third Edition

New material added to the third edition on January 3, 2018.

Download the Notes

Python is a widely used general purpose programming language, which happens to be well suited to econometrics, data analysis and other more general numeric problems. These notes provide an introduction to Python for a beginning programmer. They may also be useful for an experienced Python programmer interested in using NumPy, SciPy, matplotlib and pandas for numerical and statistical analysis (if this is the case, much of the beginning can be skipped).

Third edition, Update 1¶

- Verified that all code and examples work correctly against 2019 versions of modules. The notable packages and their versions are:
- Python 3.7 (Preferred version)
- NumPy: 1.16
- SciPy: 1.3
- pandas: 0.25
- matplotlib: 3.1
- Python 2.7 support has been officially dropped, although most examples continue to work with 2.7. **Do not Python 2.7 in 2019 for numerical code**.

Third edition update¶

- Rewritten installation section focused exclusively on using Continuum\'s Anaconda.
- Python 3.5 is the default version of Python instead of 2.7. Python 3.5 (or newer) is well supported by the Python packages required to analyze data and perform statistical analysis, and bring some new useful features, such as a new operator for matrix multiplication (@).
- Removed distinction between integers and longs in built-in data types chapter. This distinction is only relevant for Python 2.7.
- dot has been removed from most examples and replaced with @ to produce more readable code.
- Split Cython and Numba into separate chapters to highlight the improved capabilities of Numba.
- Verified all code working on current versions of core libraries using Python 3.5.
- pandas
    - Updated syntax of pandas functions such as resample.
    - Added pandas Categorical.
    - Expanded coverage of pandas groupby.
    - Expanded coverage of date and time data types and functions.
- New chapter introducing statsmodels, a package that facilitates statistical analysis of data. statsmodels includes regression analysis, Generalized Linear Models (GLM) and time-series analysis using ARIMA models.

Second edition update¶

- Improved Cython and Numba sections
- Added sections discussing interfacing with C code
- Added sections to the chapter on running code in Parallel covering IPython\'s cluster server and joblib
- Further improvements in the installation based on feedback from the Python Course
- Updated Anaconda to 1.9
- Added information about using Spyder as an initial IDE.
- Added packages for Spyder to the installation instructions.

NEW IN SECOND EDITION¶

- The preferred installation method is now Continuum Analytics\' Anaconda. Anaconda is a complete scientific stack and is available for all major platforms.
- New chapter on pandas. pandas provides a simple but powerful tool to manage data and perform basic analysis. It also greatly simplifies importing and exporting data