



Introduction to Python

Madeira International Workshop in Machine Learning

Diogo Freitas

July 27, 2021



SEEHEALTH

Why Python for Data Science?

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

Motivation

- Python is a **general-purpose programming language**.
- **Simple and efficient**.
- Python and R are the most used languages for data science = **variety of libraries available + large online community**.
- Allows you to effectively and efficiently **analyze and visualize large amounts of data**.
- Allows working with **multiprocessing** (multithreading and multiprocessing).
- It also allows linking C/C++ and Fortran code.

What are the most common
libraries?

Libraries

- **NumPy**—allows you to work with multidimensional arrays, and has tools for numerical computation.
- **pandas**—allows you to create DataFrames for data manipulation and analysis.
- **Matplotlib**—allows you to get graphical representations of data through Python.
- **scikit-learn**—allows you to efficiently run the most popular machine-learning algorithms.

All these library are interoperable!

Libraries

- **NumPy**—allows you to work with multidimensional arrays, and has tools for numerical computation.
- **pandas**—allows you to create DataFrames for data manipulation and analysis.
- **Matplotlib**—allows you to get graphical representations of data through Python.
- **scikit-learn**—allows you to efficiently run the most popular machine-learning algorithms.

All these library are interoperable!

Libraries

- **NumPy**—allows you to work with multidimensional arrays, and has tools for numerical computation.
- **pandas**—allows you to create DataFrames for data manipulation and analysis.
- **Matplotlib**—allows you to get graphical representations of data through Python.
- **scikit-learn**—allows you to efficiently run the most popular machine-learning algorithms.

All these library are interoperable!

Libraries

- **NumPy**—allows you to work with multidimensional arrays, and has tools for numerical computation.
- **pandas**—allows you to create DataFrames for data manipulation and analysis.
- **Matplotlib**—allows you to get graphical representations of data through Python.
- **scikit-learn**—allows you to efficiently run the most popular machine-learning algorithms.

All these library are interoperable!

Libraries

- **NumPy**—allows you to work with multidimensional arrays, and has tools for numerical computation.
- **pandas**—allows you to create DataFrames for data manipulation and analysis.
- **Matplotlib**—allows you to get graphical representations of data through Python.
- **scikit-learn**—allows you to efficiently run the most popular machine-learning algorithms.

All these library are interoperable!