Python data science

Nicolas Rousset

Monday, 24th of October

The most popular language according to tiobe index !

How is it used ?

- Machine learning / AI
  - tensorflow
  - pytorch
- Data analysis / visualisation
- Web (not a main actor)
- Scripting

Python characteristics

- Created in 1991 by Guido van Rossum
- Interpreted langage using compiled library
- Dynamic typing
- Opensource
- multi-paradigm

Strength of python against other tools

Compared to excel, SAS, BI, matlab and other tools commonly used for data analysis and reporting, python has some strong points :

- free (little bit overrated strong point)
- industrial level programming language
- generic

On the other hand, it will be often beaten by other tools on their specialized features

Interpreted language

The program is not directly executed by the operating system, but by a interpreter

```
python hello_world.py
```

And not :

```
hello_world.py
```

Python is extraordinary permissive

You can modify everything !

```
import math
math.pi = "Hello"
```

Python will always try to do what you tell it to do, not what you want it do.

This machine has no brain, use your own

When going bigger you have to use TDD

TDD => Test Driven Development

Testing is a very important of software development in python (and other interpreted language)

Python is multi-paradigm

OOP (Object Oriented Programming) is not the most common programming paradigm in python.

Object are not the easiest to test, and OOP goes hand to hand with compilation

OOP in python in a nutshell $=>$ why don't use C#, java or C++ ?

For data analysis purpose, the simplest way is to use a distribution, like anaconda :

https://www.anaconda.com/products/distribution

It will provide python and all the most commonly used scientific library

After installing anaconda, the way to launch the jupyter notebook server is :

```
jupyter notebook
```

It will open a browser with the local files

Python module requirements

The tool used to manage package in python is pip :

```
pip install tensorflow
pip install -r requirements.txt
```

When you have multiple modules dependancies, you should put them in a **requirements.txt** file, that looks like this :

```
setuptools==45
asgiref==3.2.10
attrs==19.3.0
autobahn==20.12.3
Automat==0.8.0
bokeh==2.0.2
```