Data Mining - Python

Dr. Jean-Michel RICHER

2018

jean-michel.richer@univ-angers.fr

Outline

1. Introduction

What is python ?

Python is amongst the top 5 most popular and practical programming languages (see TIOBE Index).

- multipurpose (web, scientific computing, GUI, parallel, bioinformatics)
- high-level, object-oriented
- interactive, interpreted (can be compiled)
- extremely user-friendly
- large standard libraries to solve common tasks

History of python

- Dec' 1989: **Guido Van Rossum** (Dutch programmer) wrote Python as a hobby programming project
- the name comes from *The Monty Python's Flying Circus*
- January 1994: v1.0.0
- March 2017: v3.6.x

2. Installation

Installation under Windows

Under windows

- Go to `https://www.python.org`
- Download Python3.7.0a4-2018-09 Windows Executable Installer

Installation under Linux

Under Ubuntu 17.10

```
sudo apt-get install python
sudo apt-get install pip
python -m pip install package-name
```

Python on the web

Execute python code in a web browser

`http://quintagroup.com/cms/python/online-interpreter`

- `http://repl.it/languages`
- `http://labs.codecademy.com/`
- `http://pythonfiddle.com/`

3. The language

Features of python

Python special features

- you **don't need to declare** variable, just use them
- type of a variable depends on the assignment
- **indentations** of the program (tabulations) define the structure
- everything is an object

Comments

Two types of comments

- # to introduce comments on one line
- """ . . . """ on several lines

Variable declaration and assignment

- `a = 2`
- `b = 3.14`
- `c = "hello"`
- `a, b, c = 2, 3.14, "hello"`
- `x = y = z = "same"`

Type of an object

type() and isinstance()

- a = 5
- print(a, "is of type", type(a))

```
<type 'int'>
```

- a = 1+2j
- print(a, "is complex number?", isinstance(1+2j,complex))

```
True
```

Integers and floating point values

- integers can be of any length:

```
>>> a=11111111111111
>>> b=2222222222222
>>> c=a*b
>>> c
24691358024690864197530864 2L
```

- floating point values have 15 decimal places

Other python types

Other types

- list `l=[1,2,3]`
- tuple `t=(1, "john", 3500)`
- set `s={5,1,2,4}`
- dictionary `d = {1:'value','key':2}`

Lists operations

- `append(value)` - appends element to end of the list
- `count('x')` - counts the number of occurrences of 'x' in the list
- `index('x')` - returns the index of 'x' in the list
- `insert(y,'x')` - inserts 'x' at location 'y'
- `pop()` - returns last element then removes it from the list
- `remove('x')` - finds and removes first 'x' from list
- `reverse()` - reverses the elements in the list
- `sort()` - sorts the list alphabetically in ascending order, or numerical in ascending order
- `len(list)` - number of elements

Operations on tuple

```
t1 = (1, "john", 3500)
t2 = 2, "paul", 2700
print(t1)
print t2[1]
print(len(t1))
```

Operations on dictionary

```
d = { 1 : "red", 2 : "green", 3 : "blue" }
print(d[1]) # red
del d[1]
for key in d:
        print(key)
# for 2.x """
for key, value in d.iteritems():
        print(key,"=",value)
# for 3.x """
for key, value in d.items():
        print(key,"=",value)
```

Conversions between data types

`float()`, `int()`, `str()` and others

- `float(5)`, `float('2.5')`
- `int(10.6)`, `int('123')`
- `str(2)`, `str(3.14)`
- `set([1,2,3])`
- `tuple(5,6,7)`
- `list('hello')`*
- `dict([[1,2],[3,4]])`
- `dict([(3,26),(4,44)])`

if then else and others

- `if`, `elif`, `else`
- `for loop`
- `while`
- `break` and `continue`

if elif else

```
x = int(input("x ?"))
if x < 1:
        print("x < 1")
elif x < 4:
        print("2 < x < 4")
else:
        print("x >= 4")
```

for loop

for variable in range

```
for i in range(1,10):
        print(i)
```

Be careful it will print values from 1 to 9 !

while condition, eventually else

```
i, n, sum = 1, 10, 0

while i <= n:
        sum = sum + i
        i = i+1
else:
        print("sum=%s" %sum)
```

break and continue

break, continue

```
for car in "string":
        if car == "s" or car == "t":
                continue
        print(car)
        if car == "n":
                break
```

will print r, i, n

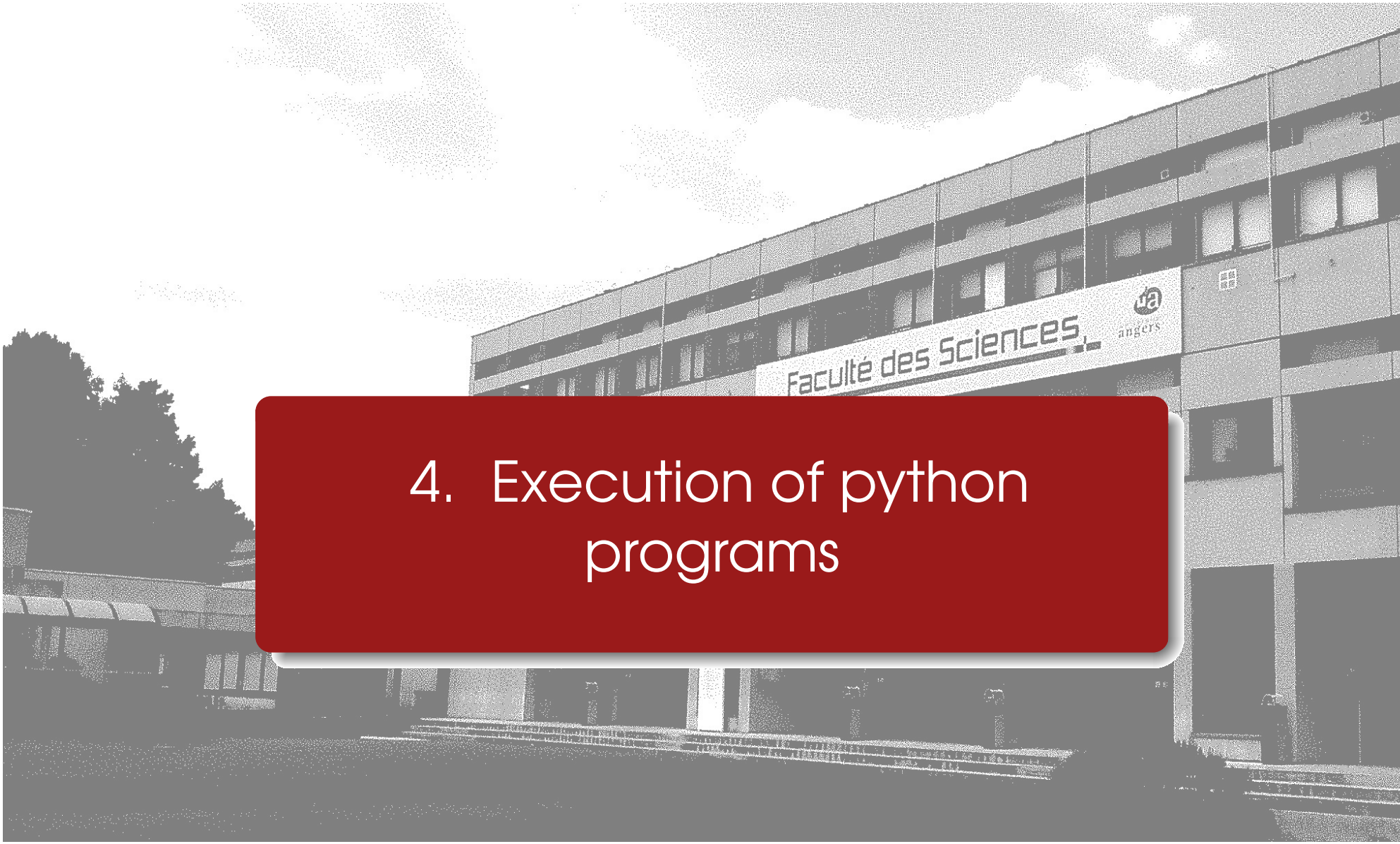
Functions

```
def sum(a,b):
        """This function sums two
        values passed as parameters"""
        return a + b

print("sum(1,2)=%s" %sum(1,2))
print("sum('the',' cat')=%s" %sum("the"," cat"))

def sum2(*args):
        a, b = args
        return a + b

print("sum2(1,2)=%s" %sum2(1,2))
print("sum2('the',' cat')=%s" %sum2("the"," cat"))
```



4. Execution of python programs

Modules in python

Modules

- modules in Python are simply Python files with the .py extension
- implement a set of functions
- use `import module-name`
- and `help(module-name)` to get help

Packages

- Packages are namespaces which contain multiple packages and modules
- they are simply directories
- each package must contain a special file called `__init__.py`
- use `import package-name`
- or `from package-name import module`

Note: to reinstall a package use `sudo pip install -upgrade -force-reinstall package-name`

Run a python program

Execution

You have two possibilities under Linux

- use `python program.py`
- put as first line of your program #!/usr/bin/python and make it executable

5. the SciPy package

What is SciPy ?

SciPy

SciPy is a Python-based ecosystem of open-source software for mathematics, science, and engineering:

- **NumPy** numerical array package
- **SciPy library** fundamental library for scientific computing
- **Matplotlib** comprehensive 2D Plotting
- **Sympy** symbolic mathematics
- **Pandas** data structures and analysis

For more information go to `https://scipy.org/`

What is numpy ?

Numpy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

For more information go to `http://www.numpy.org/`

Arrays of numpy

ndarray

NumPy's multidimensional array class is called ndarray (numpy.array)

- not the same as the Standard Python Library class array.array (1 dimension)

For more information go to `http://www.numpy.org/` and follow the **NumPy Tutorial**

How to use a numpy array

Functions

```
import numpy as np

t1 = np.array([(1.5,2,3), (4,5,6)])
t2 = np.array( [ [1,2], [3,4] ], dtype=complex )
t3 = np.zeros( (3,4) )
t4 = np.ones( (2,3,4), dtype=np.int16 )
t5 = np.empty( (2,3) )
t6 = np.arange( 1, 25, 2 ).reshape(4,3)
```

5. End

FACULTÉ
DES SCIENCES
Unité de formation
et de recherche
DÉPARTEMENT
INFORMATIQUE

UA - Angers
2 Boulevard Lavoisier 49045 Angers
Cedex 01
Tel: (+33) (0)2-41-73-50-72

université
angers
FACULTÉ
DES SCIENCES
Unité de formation
et de recherche
DÉPARTEMENT
INFORMATIQUE