Programming for Data Science (with Python)

Le Trong Ngoc – http://fit.iuh.edu.vn/giangvien@letrongngoc

Contents

- Introduction to Data Science and Data Analysis
- **Introduction to Python for Data Science**
- Data Visualization with Python
- Statistical Thinking in Python
- Applied Machine Learning in Python

Introduction to Python for Data Science

- Set up the Lab Environment
- Python basics
- List – A Data Structure
- **Functions and Packages**
- Numpy
- Plotting with Matplotlib
- Control Flow and Pandas

Functions and Packages

- Functions
  - Nothing new
  - Piece of reusable code
  - Solves particular task
  - Call function instead of writing code yourself

Functions and Packages

- Functions – Example
  - In [**1**]: fam =[1.68, 1.72, 1.55, 1.71]
  - In [**2**]: fam
  - Out[**2**]: [1.68, 1.72, 1.55, 1.71]
  - In [**3**]: max(fam)
  - Out[**3**]: 1.72
  - In [**4**]: tallest=max(fam)
  - In [**5**]: tallest
  - Out[**5**]: 1.72

Functions and Packages

- Function – Example
  - In [**6**]: round(1.68,1)
  - Out[**6**]: 1.7
  - In [**7**]: round(1.68)
  - Out[**7**]: 2
  - In [**8**]: help(round)

Help on built-in function round in module builtins:

round(...)

round(number[, ndigits]) -> number

Round a number to a given precision in decimal digits (default 0 digits).This returns an int when called with one argument, otherwise thesame type as the number. ndigits may be negative.

Functions and Packages

- Methods
  - Everything = Object
  - Object have methods associated, depending on type.
  - Methods : functions that belong to objects
  - Methods : call functions on objects

Functions and Packages

- Methods – Example
  - List methods
    - In [**1**]: fam=["ngoc", 1.69, "nhi", 1.55, "duc",1.74, "nhan",1.4]
    - In [**2**]: fam.index("duc")
    - Out[**2**]: 4
    - In [**3**]: fam.count(1.74)
    - Out[**3**]: 1

Functions and Packages

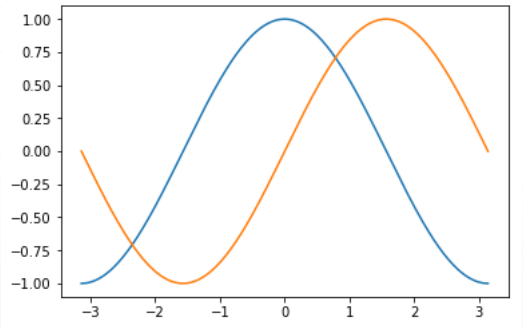
- Methods – Example
  - str methods
    - In [**6**]: name="ngoc"
    - In [**7**]: name.capitalize()
    - Out[**7**]: 'Ngoc'
    - In [**8**]: name.replace("c","n")
    - Out[**8**]: 'ngon'

Functions and Packages

- Packages
  - Director of python scripts
  - Each script =module
  - Specify functions, methods, types
  - Thousands of packages available
    - Numpy
    - Matplotlib
    - Scikit- Learn

Functions and Packages

- Packages – Example
  - import numpy as np
  - import matplotlib.pyplot as plt
  - X = np.linspace(-np.pi, np.pi, 256, endpoint=True)
  - C, S = np.cos(X), np.sin(X)
  - plt.plot(X, C)
  - plt.plot(X, S)
  - plt.show()



Numpy

- Numeric Python
- Alternative to Python List: Numpy Array
- Calculations over entire arrays
- Easy and Fast
- Installation
  - In the terminal: pip3 install numpy

Numpy

- Comparison
  - In [**9**]: height = [1.73, 1.68, 1.71, 1.89, 1.79]
  - In [**10**]: weight = [65.4, 59.2, 63.6, 88.4, 68.7]
  - In [**11**]: weight / height ** 2

    TypeError: unsupported operand type(s) for ** or pow(): 'list' and 'int'
  - In [**12**]: np_height = np.array(height)
  - In [**13**]: np_weight = np.array(weight)
  - In [**14**]: np_weight / np_height ** 2
  - Out[**14**]: array([21.85171573, 20.97505669, 21.75028214, 24.7473475 , 21.44127836])

Numpy

- Remark
  - In [**15**]: np.array([1.0, "is", True])
  - Out[**15**]: array(['1.0', 'is', 'True'], dtype='<U32')
  - In [**16**]: python_list = [1, 2, 3]
  - In [**17**]: numpy_array = np.array([1, 2, 3])
  - In [**18**]: python_list + python_list
  - Out[**18**]: [1, 2, 3, 1, 2, 3]
  - In [**19**]: numpy_array + numpy_array
  - Out[**19**]: array([2, 4, 6])