

Introduction to Programming and Machine Learning in Python

Module 1

Lecture 1: Course Introduction

Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Let's Kahoot!
- ④ Overview to programming

Note on the 2-modules structure

2-modules structure: http://bit.ly/IProML2022_SNS

This course is the first module of a teaching unit of two modules:

- M1: Focuses on how to *program well*
- M2: Focuses on data analysis and machine learning

Students can attend single modules.

- M1 gives the necessary background for M2

These slides focus on M1

Course Responsibles

- Andrea Vandin
 - ★ andrea.vandin@santannapisa.it
 - ★ Tenure-track Assistant Professor in Computer Science at Institute of Economics & EMbeDS @ SSSA
 - ★ Adjunct Associate Professor at DTU Denmark
 - ★ Formerly:
 - ▶ Associate Professor in Computer Science at DTU Denmark
 - Most related teaching activity: responsible for 3 years of course 'Programming in C++ for non-computer scientists', 250 students
- Daniele Licari
 - ★ daniele.licari@santannapisa.it
 - ★ Data Engineer and Data Scientist at EMbeDS @ SSSA
 - ★ Expert of Python and machine learning
- Together we provided courses on programming and Machine Learning at SSSA and GSSI.

Course References & Material

- Webpage of the course:
 - ★ http://bit.ly/IProML2022_SNS
 - ▶ Slides and examples from the lectures, further materials and links
 - ▶ Coding assignments
- Suggested books:
 - ★ M. Lutz, Learning Python;
 - ★ W. McKinney, Python for Data Analysis.
- Well-done tutorial: <https://docs.python.org/3/tutorial/>
- Software
 - ★ Python: <https://www.python.org/>
 - ★ Python editor: JupyterLab <https://jupyter.org/>
 - ★ Setup your machine: http://bit.ly/IProML2022_SNS

Course Description

This course will

introduce the students to the fundamental principles of (object-oriented) structured programming with basic applications to data processing. Using Python as reference language, the course starts from basic notions of programming (variables, data types, collections, control & repetition structures, functions & modules), up to basic data processing functionalities (loading, manipulation, and visualization of CSV data).

Course Description

This course will

introduce the students to the fundamental principles of (object-oriented) structured programming with basic applications to data processing. Using Python as reference language, the course starts from basic notions of programming (variables, data types, collections, control & repetition structures, functions & modules), up to basic data processing functionalities (loading, manipulation, and visualization of CSV data).

A student who has met the objectives of the course will

acquire an understanding of the issues involved in computer programming, to be able to make informed decisions. The student will be able to write simple to medium python programs of various nature, including those for reading, manipulating and visualizing data.

Learning Objectives

A student who has met the objectives of the course will be able to:

- select and use the correct data types and collections for the problem at hand
- use and describe variables, operations, and control and repetition structures (if, loops)
- create and use functions and classes
- use libraries for I/O, data manipulation, and data visualization
- use principles of structured program design and methods
- discuss Python-related issues in a clear and concise way, possibly using on-line platforms

Self-Evaluation & Active Learning

These are *attività trasversali*

- There will not be an exam
- Attendance certificate (attestazione di presenza)
 - ★ Mandatory attendance $\geq 80\%$.
 - ▶ Monitored automatically and via rollcall

To get the best from this course, we will provide you

- Regular coding assignments
 - ★ Available in the wiki.
 - ▶ Automatic tests for your code and hints to fix bugs
 - ▶ (Soft) deadlines: before the following class
 - ▶ Feel free to contact us for support
 - ▶ Can be run remotely on Google Colab or locally on Jupyter
 - ★ A fundamental learning tool of this course

Tentative Lecture Plan

Module 1 – 16 hours

Class	Topic	Date	Time
1	Course Introduction Console I/O & Variables	Wed 20/04	17:00-19:00
2	Data types & Operations	Fri 22/04	17:00-19:00
3	Collections First plots <i>Practicum</i>	Wed 27/04	17:00-20:00
4	Control statements CSV manipulation on COVID19 data <i>Practicum</i>	Mon 02/05	17:00-20:00
5	Functions Application to epidemiological models	Fri 06/05	17:00-19:00
6	Creation of word clouds from online news Modules	Mon 09/05	17:00-19:00
7	Exceptions OOP	Mon 16/05	17:00-19:00

Further info

- No previous experience on computer programming required
- Previous experience in writing small programs is advantageous
- We might adjust the course level according to your expertise and feedback
- You will never learn programming if you don't practice it!
 - ★ Therefore you should regularly do the assignments

Ideas for an Effective Course

Live Programming & Assignments/Examples

We have mostly classes of 2 hours (two of 3 hours):

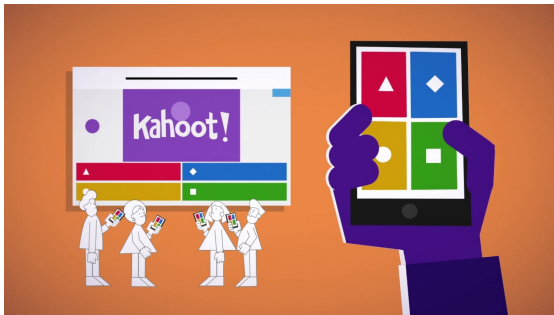
- First part (about 2 hours):
Intro to class's topics & Live programming
 - ★ No explicit slides!
 - ▶ Jupyter lab allows to mix slides-like material and executable code
 - ★ We present the new topics developing simple example programs
 - ▶ Please have your laptop ready! http://bit.ly/IProML2022_SNS
 - ▶ You find code in advance here http://bit.ly/IProML2022_SNS
- Second part: (about 1 hour):
 - ★ You consolidate your understanding with complex examples or assignments
 - ★ Complete the assignments before next class. Contact us if needed
 - ★ 'Home-assignments' for 2-hours classes, with our live support for the 3-hours classes

Ideas for an Effective Course

Gamified recap quizzes on Kahoot

We start each class with a Kahoot quiz to

- Recap the previous-class topics
 - Provide you an additional instrument to self-evaluate you learning process
 - Have fun and Stimulate healthy competition
- ★ You should regularly do the assignments if you want to win! :D



Ideas for an Effective Course

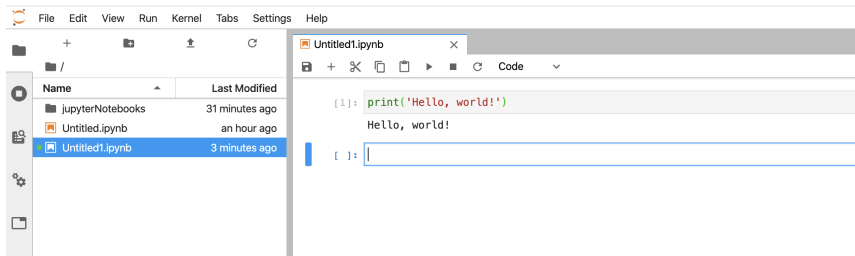
Gamified recap quizzes on Kahoot

A kahoot-related meme by your colleagues from previous year



Live Programming

Find the JupyterLab notebooks at http://bit.ly/IProML2022_SNS



Assignments on Colab

01ConsoleIOandVariables_Assignments

File Edit View Insert Runtime Tools Help

Table of contents

- RUN, BUT DO NOT MODIFY
- Assignment 01.01: Sum of three numbers
 - Write your solution here
 - Run the following cells to perform the provided tests
 - TEST 1
- Assignment 01.02: Hello, name1 name2
 - Write your solution here
 - Run the following cells to perform the provided tests
 - TEST Renzo Lucia
 - TEST Andrea Daniele
 - TEST Daniele Andrea
- Assignment 01.03: Hello, name1 and name2
 - Write your solution here
 - Run the following cells to perform the provided tests
 - TEST Renzo Lucia
 - TEST Andrea Daniele

Assignment 01.01: Sum of three numbers

Statement

Write a program that prints the Italian translation of 'Hello, world!'.

That is: 'Ciao, mondo!'

Example input

Example output

Ciao, mondo!

Theory

Well, there is not much theory to use here.

In the example, you can see how to print 'Hello, world!'.

Write your solution here

- Each lecture comes with a set of simple coding assignments
- ★ Links available in the wiki page for slides and further material

Colab

- Colab is a Google service similar to Google docs
 - ★ but for python notebooks.
 - ★ no installation required
- Each set of assignments is actually a python notebook
- We implemented in Colab autograding functionalities
 - ★ to test your solution
- If you prefer, you can also download them as jupyter notebooks

Colab: auto-testing

Write your solution here

- Do not change the first line (`def ...():`)
- Maintain the given indentation
- You can run some tests by yourself by uncommenting the last line

```
[3] def asgn01_01Hello_world():  
    # This program prints 'Hello, world!':  
    #print('Hello, world!')  
  
    # Can you change it so that it prints the same,  
    #but in Italian?  
  
    print('Ciao, mondo')  
  
#You can test independently your solution by executing the following line  
#asgn01_01Hello_world()
```

Run the following cells to perform the provided tests



TEST 1

```
Test []  
The program prints 1 lines as expected.
```

```
Line 0  
Test FAILED  
Expected: Ciao, mondo!  
Actual  : Ciao, mondo
```

Colab: auto-testing

▼ Write your solution here

- Do not change the first line (`def ...():`)
- Maintain the given indentation
- You can run some tests by yourself by uncommenting the last line

```
[ ] def asgn01_01Hello_world():  
    # This program prints 'Hello, world!':  
    #print('Hello, world!')  
  
    # Can you change it so that it prints the same,  
    #but in Italian?  
  
    print('Ciao, mondo!')  
  
#You can test independently your solution by executing the following line  
#asgn01_01Hello_world()
```

▼ Run the following cells to perform the provided tests



TEST 1

```
Test []  
The program prints 1 lines as expected.  
Line 0  
Expected and actual output match: Ciao, mondo!  
Test PASSED!
```

Outline

- 1 Course introduction
- 2 Sneak preview of Module 2**
- 3 Let's Kahoot!
- 4 Overview to programming

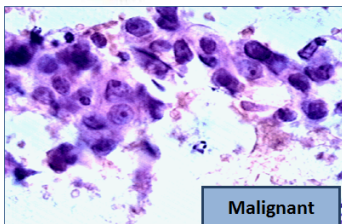
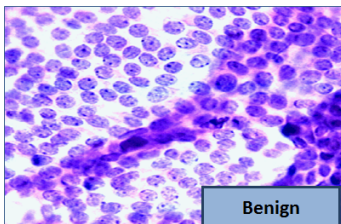
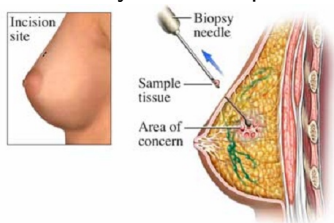
Tentative Lecture Plan Module 2

Module 2 – 14 hours

Class	Topic	Date	Time
1	Course & Project introduction Advanced Python IDEs (JuPyteR, Google Colab) Advanced libraries for data manipulation: NumPy	Fri 20/05	17:00-19:00
2	Advanced libraries for data manipulation: Pandas 1 Application to official Italian COVID'19 data Application to Yahoo! Finance stock prices	Mon 23/05	17:00-19:00
3	Advanced libraries for data manipulation: Pandas 2 Application to Yahoo! Finance stock prices Application to official Italian COVID'19 data First Exploratory Data Analysis tasks on the Titanic dataset	Fri 27/05	17:00-19:00
4	Data pre-processing Application to breast cancer diagnosis	Mon 30/05	17:00-19:00
5	Unsupervised ML Application to breast cancer diagnosis	Mon 06/06	17:00-19:00
6	Supervised ML Application to breast cancer diagnosis	Fri 10/06	17:00-19:00
7	KNIME a graphical language for complex data analysis OR Tutorial on Process-oriented Data Science: Process mining	Mon 13/06	17:00-19:00

Sneak preview of Module 2

Starting from the competences developed in the first module, we will study how to apply data analysis techniques from Machine learning



Can we classify them automatically?

Sneak preview of Module 2

We will go through a classic pipeline for these data analysis tasks

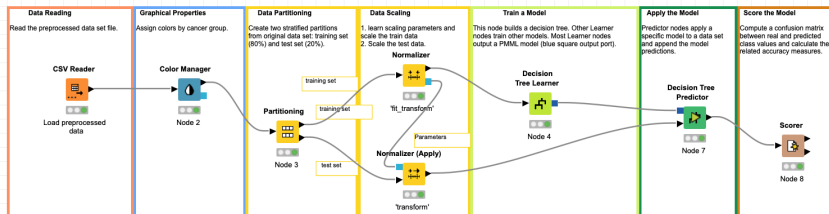
- with emphasis on data pre-processing.

We will use two alternative approaches

- Python: **main focus**
 - ★ NumPy, Pandas, Scikit-learn, Seaborn
- Knime: a graphical workflow language

Model Training for Classification

This workflow demonstrates how a simple classifier is built and applied to new data.
Task: Classification of Benign and Malignant Breast Cancer



Sneak preview of Module 2

Alternatively to Knime, we might consider further advanced research-oriented topics of data-driven analysis like Process Mining

1. **Lucy** takes your order

2. **Lucy** notes down your address

3. **Lucy** notes down your preferred payment method

4. **Luigi** prepares your burger

5. **Lucy** grabs your can of soda

6. **Luigi** puts your burger in a box

7. **Lucy** wraps your order

8. **Mike** delivers your order

1. **Randy** takes your order

2. **Randy** notes down your preferred payment method

3. **Randy** notes down your address

4. **Luigi** prepares your burger

5. **Luigi** puts your burger in a box

6. **Randy** wraps your order

7. **John** delivers your order

What is the process underlying my data?¹



¹Example from <https://pm4py.fit.fraunhofer.de/>

Sneak preview of Module 2

Self-Evaluation & Active Learning

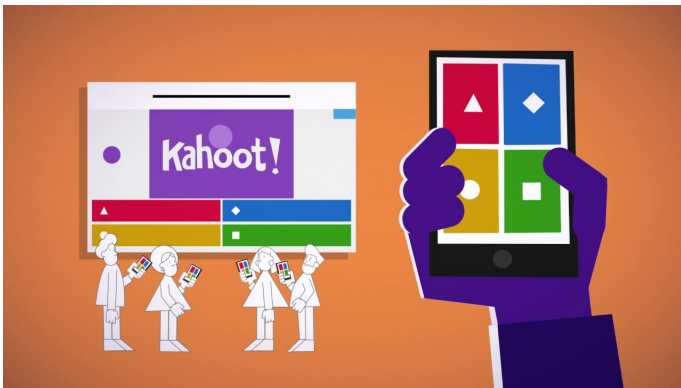
You will have the opportunity to do the same on data of interest or on data on titanic sinking

- Would you have survived the sinking of the titanic?
- An important learning tool for module 2

Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Let's Kahoot!**
- ④ Overview to programming

Let's play a game on Kahoot!



- Using your smartphone or a second monitor
- Visit www.kahoot.it
- Type the code we will give you during the class

Outline

- ① Course introduction
- ② Sneak preview of Module 2
- ③ Let's Kahoot!
- ④ Overview to programming

What is a program?

- A sequence of code instructions to control a machine
 - ★ Input/output
 - ★ Mathematical operations
 - ★ Conditional and repetitive executions
- A recipe to instruct a machine to execute instructions.
 - ★ We can't use a *natural language*.
 - ★ We need a **programming language**

Programming languages

The 7 Most In-Demand Programming Languages of 2019

March 15, 2019

Aspiring developers need to know what languages to learn; they need to select the right education and work on a skill set that will impress future employers and land their dream job. So what are the top programming languages? And what is the best one to learn? We've compiled a list for you that highlights the most in-demand programming languages based off current job postings on the market.

Here are the Top 7 programming languages with most job posting on [Indeed](https://www.indeed.com) as of January 2019:

- Java – 65,986 jobs
- Python – 61,818 jobs
- Javascript – 38,018 jobs
- C++ – 36,798 jobs
- C# – 27,521 jobs
- PHP – 16,890 jobs
- PERL – 13, 727 jobs

1	Java		11	MATLAB	
2	C		12	R	
3	Python		13	Perl	
4	C++		14	Assembly Language	
5	Visual Basic .NET		15	Swift	
6	Javascript		16	Go	
7	C#		17	Delphi/Object Pascal	
8	PHP		18	Ruby	
9	SQL		19	PL/SQL	
10	Objective-C		20	Visual Basic	

<http://www.codingdojo.com/blog/the-7-most-in-demand-programming-languages-of-2019>

Programming languages

<https://www.tiobe.com/tiobe-index/>

The index can be used to check whether your programming skills are still up to date or to make a strategic decision about what programming language should be adopted when starting to build a new software system. The definition of the TIOBE index can be found [here](#).

Feb 2021	Feb 2020	Change	Programming Language	Ratings	Change
1	2	▲	C	16.34%	-0.43%
2	1	▼	Java	11.29%	-6.07%
3	3		Python	10.86%	+1.52%
4	4		C++	6.88%	+0.71%
5	5		C#	4.44%	-1.48%

Programming languages

<https://www.tiobe.com/tiobe-index/>

The index can be used to check whether your programming skills are still up to date or to make a strategic decision about what programming language should be adopted when starting to build a new software system. The definition of the TIOBE index can be found [here](#).

Feb 2021	Feb 2020	Change	Programming Language	Ratings	Change
1	2	▲	C	16.34%	-0.43%
2	1	▼	Java	11.29%	-6.07%
3	3		Python	10.86%	+1.52%
4	4		C++	6.88%	+0.71%
5	5		C#	4.44%	-1.48%

May 2021	May 2020	Change	Programming Language	Ratings	Change
1	1		C	13.38%	-3.68%
2	3	▲	Python	11.87%	+2.75%
3	2	▼	Java	11.74%	-4.54%
4	4		C++	7.81%	+1.69%
5	5		C#	4.41%	+0.12%

The Python Programming language



- High-level: almost human readable. Abstracts from hardware
- Beginner-friendly:
 - ★ streamlined syntax
 - ★ it is easy to write your *first programs*
- Free, open-source and multi-platform
- Developed since the 90s, therefore it has
 - ★ A wide community, and its popularity keeps increasing
 - ★ Many predefined software modules

Python programs

- A sequence of python instructions to control a machine
- Python supports the most common programming styles
 - ★ Imperative: Statements are executed in sequence changing the state of the program (the variables)
 - ★ Procedural: The program is structured in reusable units named functions
 - ★ Object-oriented: The program is structured as a collection of interacting objects that send messages to each other.
 - ★ Functional: Statements are not written/executed as an ordered sequence of instructions. A computation is treated as the evaluation of a mathematical function.

Variables

Basic abstraction to represent units of data

A variable has a name and a value

- Names can contain any letter, number, or the underscore _

False	await	else	import	pass
None	break	except	in	raise
True	class	finally	is	return
and	continue	for	lambda	try
as	def	from	nonlocal	while
assert	del	global	not	with
async	elif	if	or	yield

Note:

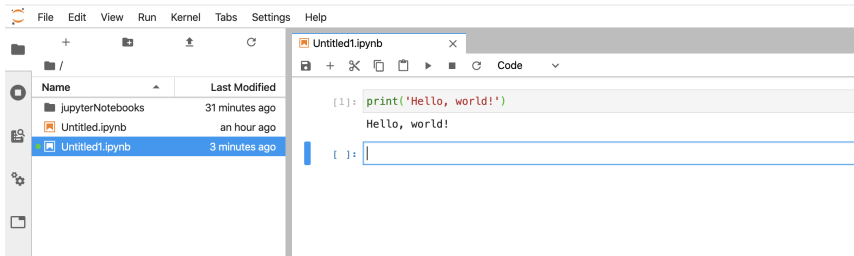
- ★ Cannot start with numbers
- ★ Cannot be a keyword
- ★ Names are case-sensitive

- We assign/update values to variables using assignment statements

```
month_number=3
month_name="April"
print("The number of",month_name,"is",month_number)
month_number=4
print("The number of",month_name,"is",month_number)
```

Live Programming

Find the JupyterLab notebooks at http://bit.ly/IProML2022_SNS



Configure your machine

If you have not done it yet

Follow the instructions in

http://bit.ly/IProML2022_SNS

“But it works . . .”



“Can You Learn To Ski Without Lessons?”



<https://www.skibro.com/blog/en/can-you-learn-to-ski-without-lessons/>

Most of the times you get to the valley.
The problem is how you get there . . .