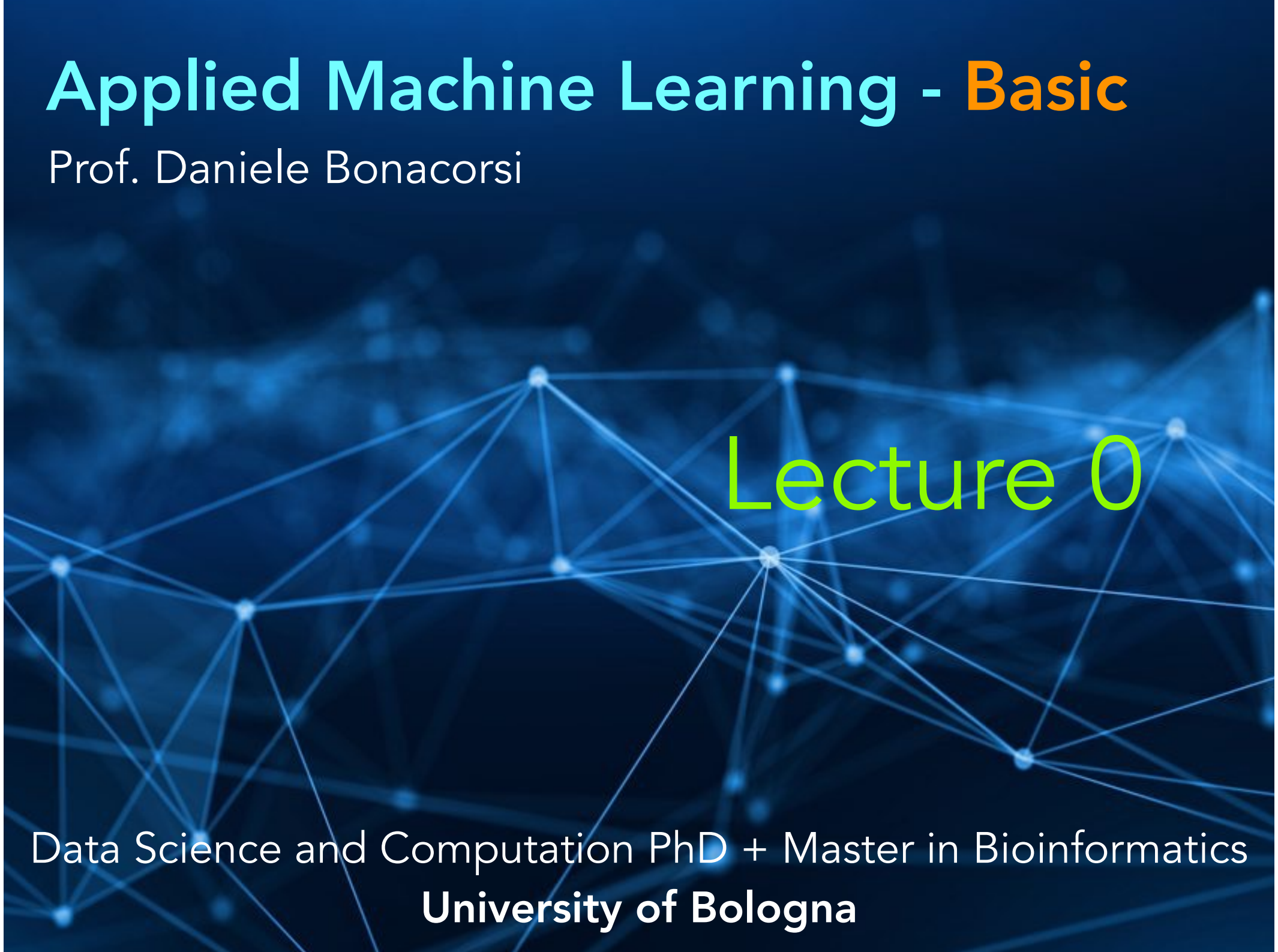
Applied Machine Learning - Basic

Prof. Daniele Bonacorsi



Lecture 0

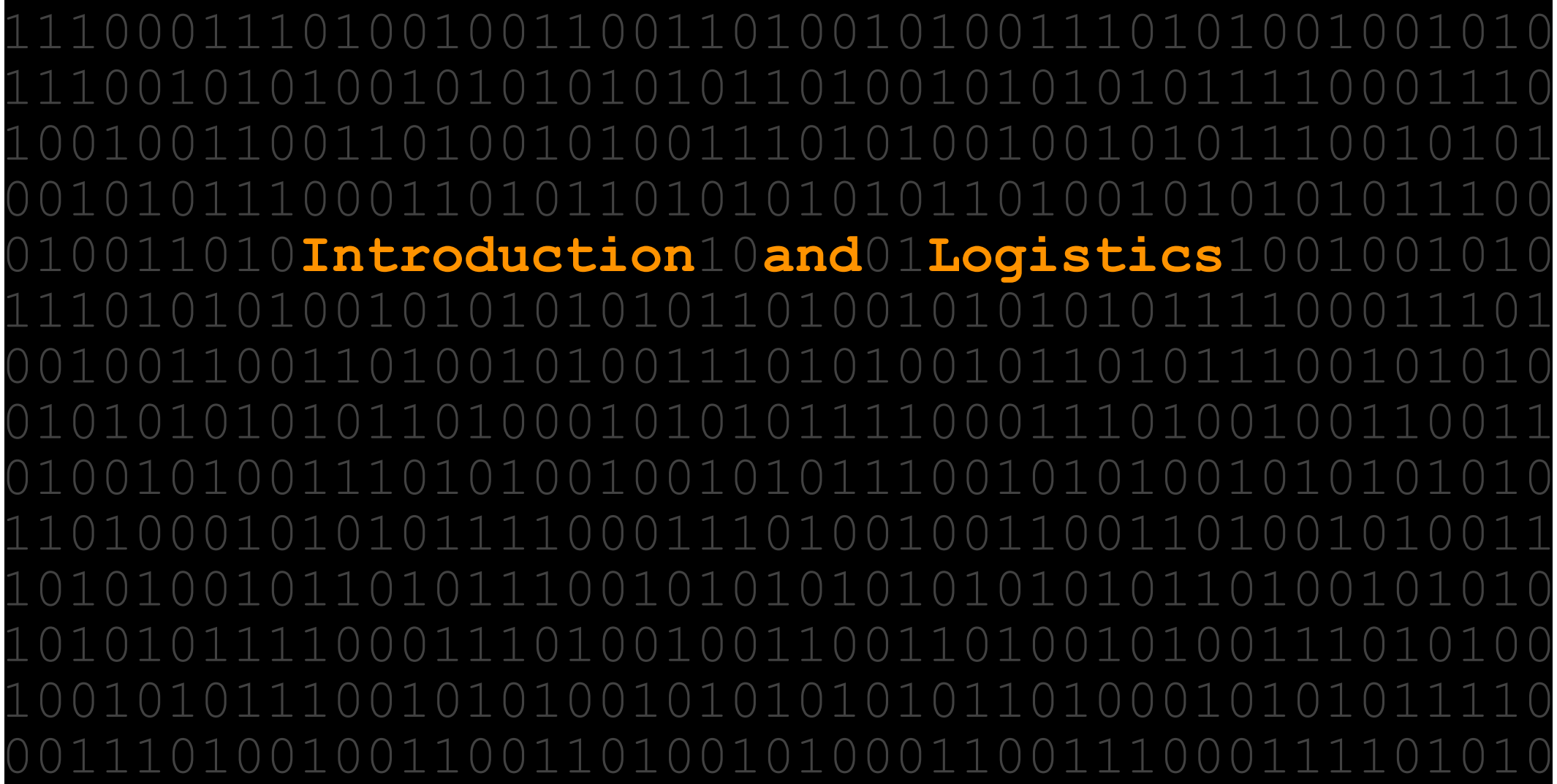
Data Science and Computation PhD + Master in Bioinformatics

University of Bologna

Lingua / Language ?

Italiano / **English**

Lectures and material in English


Introduction and Logistics

Material (including these slides)

Access point to all you need → bit.ly/CourseAML2223Bas

(details in next slides)

People

This part (**Applied ML - Basic**):

Prof. **Daniele Bonacorsi** (daniele.bonacorsi@unibo.it)

- front lectures + hands-on

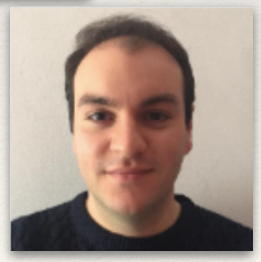
Dott. **Simone Rossi Tisbeni** (simone.rossitisbeni@unibo.it)

- PhD candidate in Data Science & Computation, tutor for this module

Next part (**Applied ML - Advanced**):

Prof. **Daniele Bonacorsi** (daniele.bonacorsi@unibo.it)

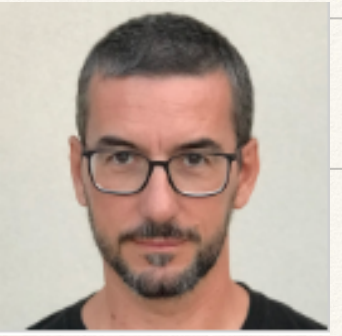
- front lectures + hands-on



Dott. **Luca Anzalone** (luca.anzalone2@unibo.it)

- PhD candidate in Data Science and Computation, tutor for this module

Who am I?



https://www.unibo.it/sitoweb/daniele.bonacorsi

I am a **physicist**. My field is "**experimental high-energy particle physics (HEP)**", in particular with particle accelerators.

Research:

- subnuclear physics in OPAL at LEP, in CMS at LHC
- over last >20 years: focus on Software/Computing for the CMS experiment
  - CERN: https://home.cern/; CMS: https://cms.cern/

Teaching:

- General Physics, Data Analysis, Physics Laboratory, Applied Machine Learning, Software&Computing (focus on Nuclear and Subnuclear Physics), Quantum Machine Learning
- "Scuole" (i.e. Faculties): Physics, Engineering, Natural Sciences
  - L, LM, PhD

What this course **is**, and what it **is not**

First of all, "expectations tuning".

Despite we will talk about…

- ML concepts, statistics, data science, computer science, software, hardware..

… this course:

- is <u>not</u> a "theoretical" ML / statistics course
- is <u>not</u> a "theoretical" {data, computer} science course

It is **a ~30 + ~50 hours journey through (selected) ML concepts and their application to a variety of use-cases through "intuitions" from theory + tutorial-like hands-on exercises**

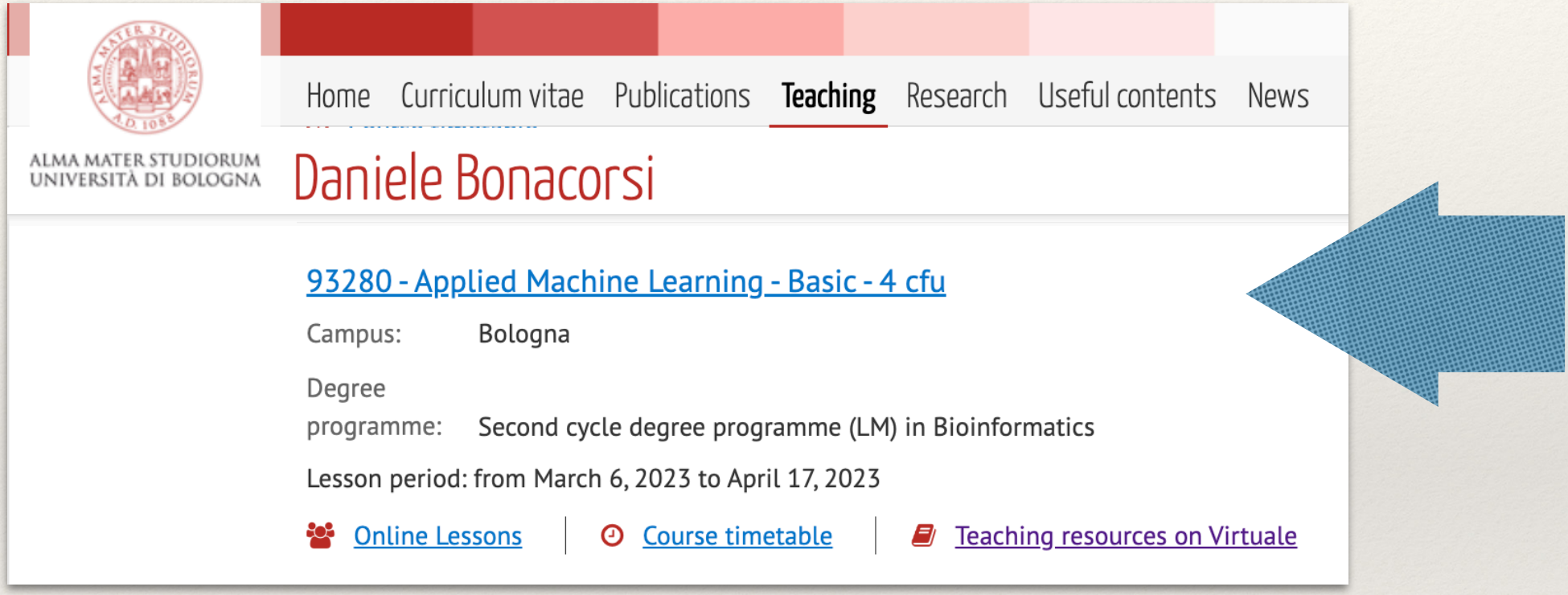
- allow me brevity, accelerations, simplifications, time-constraints driven choices

Consider it a "portal" towards ***<u>your</u>*** travel into data science and ML

- you should expect to build some awareness + skills on toolkits, that enable you to embark into some "real world" ML projects during/after the end of the course

"Sito docente" → "Didattica"

ITA: https://www.unibo.it/sitoweb/daniele.bonacorsi/didattica
ENG: https://www.unibo.it/sitoweb/daniele.bonacorsi/teachings



You can refer to this area for Teams (comments laters on this), timetable, etc

- be careful of checking the _right_ entry in the page (as there are duplications..)

"Sito docente" → "Didattica"

https://www.unibo.it/en/teaching/course-unit-catalogue/course-unit/2022/455025

93279 - Applied Machine Learning

ACADEMIC YEAR 2022/2023

Learning outcomes

At the end of the course the student is able to handle different Machine Learning and Deep Learning models, to tune them to specific applications, and to design approaches that may scale with large amount of data. Moreover, the student has competences on how to exploit different hardware architectures for Machine Learning and Deep Learning solutions, both on-premise and via cloud. The student will be also introduced to most recent approaches and active areas of work in the Artificial Intelligence community worldwide.

Course contents

Advanced concepts of Applied Machine Learning.

Office hours

See the website of Daniele Bonacorsi

Course Unit Page

- Teacher
  Daniele Bonacorsi
- Learning modules
  Daniele Bonacorsi (Modulo 1)
  Daniele Bonacorsi (Modulo 2)
- Credits
  10
- SSD
  FIS/01
- Teaching Mode
  Traditional lectures (Modulo 1)
  Traditional lectures (Modulo 2)
- Language
  English
- Campus of Bologna
- Degree Programme
  Second cycle degree programme (LM) in Bioinformatics (cod. 8020)

Also valid for
Second cycle degree programme (LM) in Bioinformatics (cod. 8020)

Course Timetable from **Mar 06, 2023** to **Apr 17, 2023**

Course Timetable from **May 04, 2023** to **Jun 16, 2023**

Calendar [1/2]

AML - Basic: 6 March - 17 April

- 8 lectures → 8 * 4hrs = 32 hrs → **4 CFU**

AML - Advanced: 4 May - 16 June

- 12 lectures → 9 * 4hrs + 4 * 3 hrs = 48 hrs → **6 CFU**

Basic

mon Mar 6 14-18
fri Mar 10 14-18
mon Mar 13 14-18
fri Mar 17 14-18
mon Mar 27 14-18
Fri Mar 31 14-18
mon Apr 14 14-18
fri Apr 17 14-18

Advanced

thu May 4 14-18
fri May 5 14-18
thu May 11 14-18
fri May 12 14-17
thu May 18 14-18
fri May 19 14-17
thu May 25 14-18
fri May 26 14-17
thu Jun 1 14-18
thu Jun 8 14-18
thu Jun 9 14-18
Thu Jun 15 14-18
fri Jun 16 14-17

Dates are susceptible to change:
in case, you will be notified in advance.

Calendar [2/2]

Admittedly: tough.

- always 3 or 4 hours in a row
  - Note also the peculiarity of this room in the Dept closing at 18:45 roughly
  - Bas: some periods with no lectures
  - Adv: up to mid June..

I will do my best to mitigate everything for your protection.

Lectures format

Format:

- Sort of "**hybrid**":
  - **lectures are in presence** (especially for Bioinformatics students),
    - I will make some use of Teams → more info live at the introductory lecture
    - If so, be aware that this modality requires a strong motivation and maturity!
- **Frontal** vs **hands-on**
  - **frontal lectures** to start..
  - .. then (later on) we will inject **hands-on sessions**
- every hourly slot will be in the format: **45 mins lecture +15 mins pause**
  - this is my choice to support your best fruition of the course - if you prefer it different, speak up!
- **ask questions!**

FAQs on lectures and material

Material available? Yes.

- you will have slides in advance (gdrive) or soon after each lecture
- PDF updates *after* each lecture are possible, so check back!

Mandatory **attendance**? No

- I may note down your presence, but it will not impact the final exam

Remote attendance? → see the "hybrid" comment

- For all students in Bioinformatics, lectures are in presence

Recordings? Not really.

- I may activate it, but for internal usage (e.g. tutors' training), not for students

Course material and organisation

Static and dynamic material

- **Slides** and **Jupyter/Colab** notebooks
- (brief comment at the lecture about copyrights and material usage)

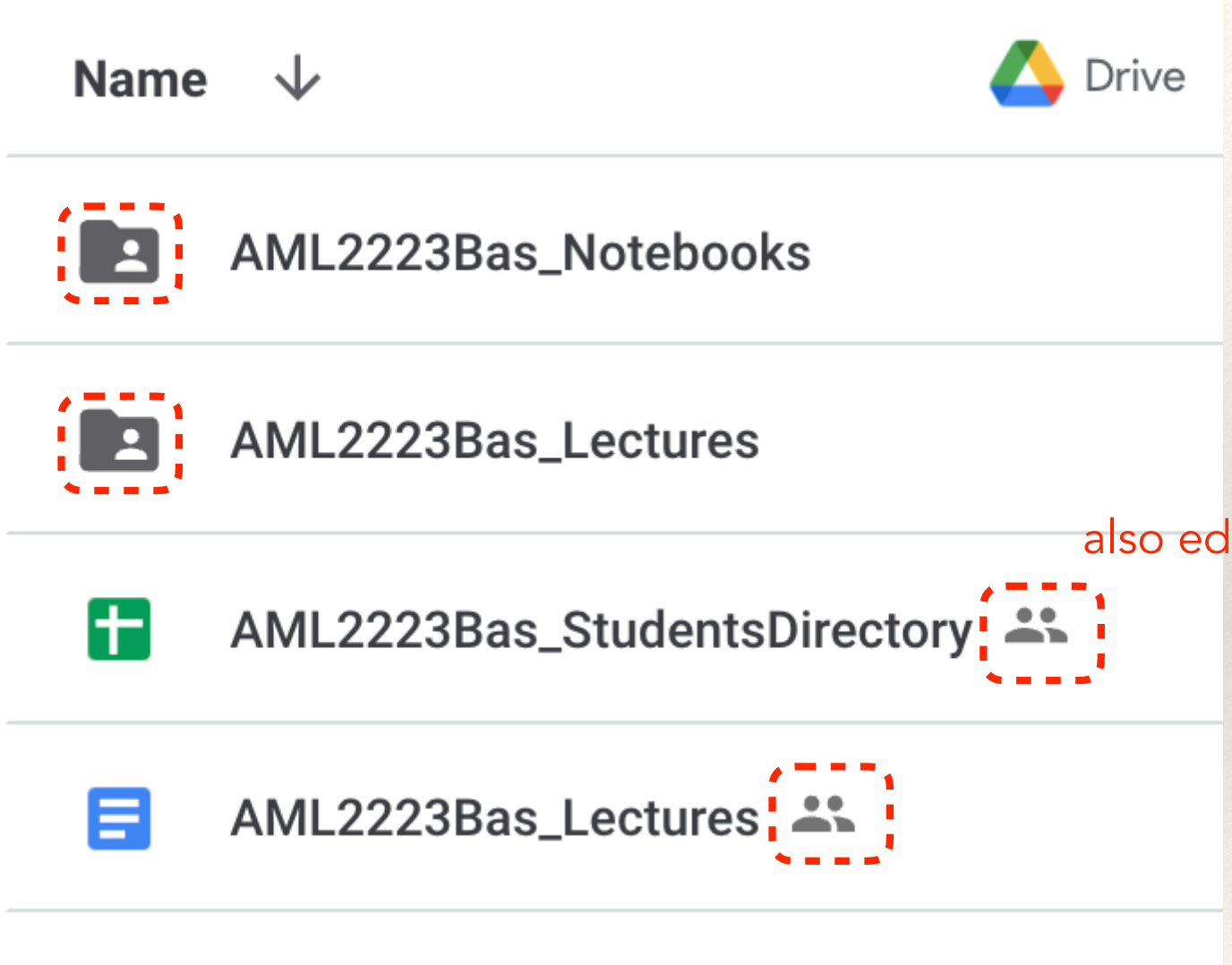
Material both on gdrive and github (perhaps IOL?)

- Gdrive: **bit.ly/CourseAML2223Bas**
- Github: — as soon as we start using notebooks —

Access to course material:

- during the course: open to everyone having the link
- after the course: might restrict access to material from old courses (maintain it by providing your gmail in the Students Directory)

More on material location(s)



Students' directory

Students directory

Course	AML Basic
AY	2022-23
Teacher	Prof. Daniele Bonacorsi

"nb matricola"	Family name	Name	enrolled in:	Preferred email for communication	gmail o altro	mobile (PLEASE NOT FI
-	BONACORSI	DANIELE	-	daniele.bonacorsi@unibo.it	-	
						no

Details at the lecture.

Lectures

Drive

Applied ML (Basic)

AY 2022-23 - D.Bonacorsi

Material from lectures

Selected material from the lectures is listed below.

This material is intended to be used solely by students attending the AML course. Any sharing MUST be discussed with the teacher in advance. Thanks for the collaboration.

- **Mar 6** (4 hrs)
  - PDFs: Lecture0, Lecture1 (in progress)

Details at the lecture.

C.Y.U. quizzes

C.Y.U. = check your understanding

Throughout the course:

- very small number of multiple-choice (easy) questions
- answer online (QR code) during the lecture
- we just comment the results and continue..

Goal: as from the name, just to C.Y.U.

- they will NOT be used to evaluate your proficiency in the course! (anonymous)
- so, just answer at the best of your understanding

The exam

Exam (for LM students in Bioinformatics) [1/2]

The exam is composed of:

- **written exam(s)** (multiple choice) → **15/30** of the score
  - Almaesami for written exams dates and sign-up
- an **end-2-end ML project** (code+documentation) → **15/30** of the score
  - proposed by the student or selected in a list proposed by the teacher
  - In the latter case, I give you problem+dataset+objective. You always give me approach+code+documentation.
  - max 2 submission for the same project - if still unhappy with the score, you need to change project

"Can I work in a team?"

- In principle yes, but → constraints discussed at the introductory lecture

Exam (for LM students in Bioinformatics) [2/2]

Note: AML course composed of Basic (4 CFU) + Advanced (6 CFU):

- which score goes for "verbalizzazione"? It depends on your "Study plan"
  - written exams → 2 separate ones: one for Basic, one for Advanced
  - project → if you planned for 10 CFU, you can work on a unique project

To deliver a project, **send it by mail to teacher + tutors**

- You get an acknowledgement, then you enter the correction queue
- In case of URGENT correction needs, submit in due time and notify the teacher in advance!

Exam (for DSC/Physics/other PhD researchers)

[*DSC = Data Science and Computation*]

The exam is composed of:

- an **end-2-end ML project** (code+documentation) → no score, just a **YES/NO**
  - Same rules as before, for 4 or 6 or 10 credits, plus…
  - … some PhD researchers might want to discuss with the teacher their own specific case, depending on the learning plan their PhD programme foresees/requires (**CONTACT ME**)

No Almaesami. You submit the project when you are ready

- Be careful of your own (PhD-specific) deadlines

Exam (for other PhD students, and "guests")

No exam.

Enjoy the course and I hope you find it useful.

Jargon and icons in the slides/notebooks

Many acronyms throughout this course, e.g. obvious ones:

- (AI = Artificial Intelligence)
- ML = Machine Learning
- AML = Applied Machine Learning
- DL = Deep Learning
- algo = algorithm
- …

As we will introduce concepts, we will create new acronyms.

There is a video in this slide.

[<tag>] This is a reference. Find it in the references gdoc.

My contacts

daniele.bonacorsi@unibo.it

NOTE: to avoid unreceived/unseen mails, please ALWAYS write to teacher(s)+tutor

daniele.bonacorsi

Good for quick exchange of (not urgent) information

@DBonacorsi

phone -> see gdoc

The tutor will take care of