

Information Theory for Data Science

Master of Science in Data Science and Engineering
Academic Year 2024/2025

Introduction and rules

Vers. 1.0 24/09/2024



**Politecnico
di Torino**

Prof. Roberto Garelo (roberto.garelo@polito.it)

Information Theory for Data Science

- Master of Science in Data Science and Engineering
- First year, first semester
- 8 credits
- Third Edition

Course Description 1/2

- **Information Theory** studies how to quantify, represent, store, compress, transmit, protect information.
- It is a **fundamental tool for a data scientist** and has many important applications in Classifiers, Machine Learning, Artificial Intelligence, Data compression, Data transmission and Cryptography.
- The course aims to provide the **basis of Information Theory** and show some of the most important **applications for Data Science**, including Classifiers, Data Compression and Cryptography.

Course Description 2/2

- The course has a **learning-by-doing approach** where for each topic an **assignment** is proposed that must be solved by writing **Matlab** (or Python) programs and writing a **report**.
- About half of the hours are devoted to **tutoring**, where the teachers are available to help students with the assignments.

Pre-requirements

- Basic notions of probability theory
- Very basic programming skills.

Course Structure

- The course is organized into **three sections**.

Course Topics

1. Introduction to Information Theory and application to Classifiers (Garello, 25 hours)
2. Application of Information Theory to Data Compression (Taricco, 30 hours)
3. Information Theory and Cryptography. Algorithms for Cryptography (Garello, 25 hours)

Exam

➤ Exam 1:

- 3 Assignments (max = 32, 70%)
- Written exam A (4/20 questions, max = 30, 30%)

➤ Exam 2:

- Written exam B (4/50 questions, max = 25)
- optional oral exam (entire program, may increase or decrease the grade)

Exam 1: Assignments

- 3 assignments (one for each section)
- Groups of 1 to 6 students
- You must deliver:
 1. A presentation (slides, A1 and A3) or a report (A2) containing all the results, the figures and the required answers.
 2. All the written Matlab programs.
- Grade: max = 30
- Deadlines:
 - 2 weeks: +2 → max = 32
 - 3 weeks: +1 → max = 31
 - 4 weeks: +0 → max = 30
 - **Later: assignment is not accepted**
- Assignment grade = weighted average (number of hours/80) of the 3 grades (max = 32)
- 70% of exam 1 grade

Exam 1: Written exam A

- 4 questions selected from a list of approximately 20 (time available = 1.5 hours, closed-book exam, maximum grade = 30)
- 30% of exam 1 grade

Exam 1: Final grade

- 70% average assignment grade, max = 32
- 30% written exam (A), max = 30
- Max grade = 30L (for students who get at least 30.5)

Exam 2: Written exam B

- 4 questions selected from a list of approximately 50 (time available = 1.5 hours, closed-book exam, maximum grade = 25)

Exam 2: optional oral exam

- Some questions on the entire program
- May increase or decrease the grade
- If an assignment had been submitted and received a grade of $X/30$, the written exam grade is increased by $X/30$

Exam 2: final grade

- Written exam (B), max = 25
- Optional oral exam
- Points for assignments
- Max grade = 30L

Groups 1/2

- You can work in groups of 1 to 6 students
- Ideal size = 4 students. Smaller groups are discouraged, as for some assignments it is suggested to divide the workload among multiple students. Additionally, one of the course objectives is to get students used to teamwork.
- Students who are having trouble finding others to form groups can post a message on the course forum on the portal

Groups 2/2

- The group composition must be finalized within the first two weeks and **cannot be changed** afterward.
- Once the group is formed, you must elect a **representative**, which will be the student whose last name comes first alphabetically.
- This student must send an email to roberto.garello@polito.it and giorgio.taricco@polito.it, CC'ing all other members, with the group composition. Additionally, this student will be responsible for uploading the assignment solutions to the portal and sending the PDF via email.

Language

- Matlab is preferred.
- Python is accepted only as a backup option and only in the format “Jupyter Notebook Format”.

How to deliver your assignments 1/2

- An assignment is made by some exercises.

- For each assignment you must prepare:
 1. A separate Matlab program file for each exercise (not a single file for the entire assignment).
 2. A presentation (slides) or a report containing all the answers, the figures and the comments for all the exercises of that assignment. Write your name and student number on the first page of the report. Presentations/Reports must be delivered in pdf.
 3. A zip archive containing the pdf report and the program files
 4. File names: using 'SURNAME' as the last name of the representative student, name the files as SURNAME_A1.pdf and SURNAME_A1.zip

How to deliver your assignments 2/2

To deliver your assignment, the representative student must:

- Upload the zip archive on the portal (elaborati section).
- Send an email to roberto.garello@polito.it (for topics 1 and 3) or giorgio.taricco@polito.it (for topic 2) with the pdf report only (**no zip, no programs**) in attachment. Include all other members of the group in cc.

Written exams: instructions

1. **Registration for the exam is mandatory.**
2. Students taking **Exam Type B** must also send an email (at least one week before the exam) to roberto.garello@polito.it.
3. Bring only plain white sheets (such as those used in printers).
4. Bring your student ID card.
6. No cheat sheets.
7. Remember to write your name and student code on all used sheets and number them.
8. Solve each question on separate sheets.
9. At the end of the exam, you will be required to scan each used sheet and upload the PDF file to the "elaborati" section of the portal. We recommend testing this process with your phone the day before.
10. Upload two separate PDFs: one for sections 1 & 3 (named **exam0802_yoursurname_garello.pdf**) and one for section 2 (named **exam0802_yoursurname_taricco.pdf**).

Very Important 1/3

- **The presentation (or report) and the programs of an assignment can be delivered only once per academic year (you cannot change them).**

Very Important 2/3: Ethic Code

- You must solve the assignments individually or in groups of 2/3/4/5/6.
- You are not allowed to share your programs with other students/groups and/or to use part of the programs written by other students.
- We will compare the programs you deliver looking for similarities with state-of-the-art anti-plagiarism programs (check against previous years, too).
- If we conclude that an assignment is not original, trespassers of the ethic code (both the author of the original code and who partially copied the program) will be referred to the disciplinary committee.
- The committee may decide to prevent the students to take the exam during the current semester or the current academic year.

Very Important 3/3

An important rule that also applies to those who have passed the exam.

- The publication of the solution of the assignments on any websites or repositories is prohibited.
- In fact, at this university, the publication of individual exam solutions is prohibited, and the solutions of the assignments are an integral part of the exam.

Course's organization

- Schedule:
 - Tuesday 16.00/19.00 room 7t
 - Wednesday 10.00/13.00 room 11i
 - (Rooms may change)

Material

- Notes written by the teachers, uploaded on Dropbox
- Whenever possible (unless there are portal issues, errors, or other problems), the recording of the lesson will be made available. (Tutoring is never registered.)

Tutoring

- The teacher will answer your questions on the assignments.
- Offered on the Zoom Virtual Classroom, too, where the students will be admitted on an individual basis.

Teachers

- Roberto Garelo: responsible of the course, teacher for the topics of assignments 1 and 3
- Giorgio Taricco: teacher and tutor for the topics of assignment 2

Roberto Garelo

roberto.garelo@polito.it

Affiliation:

- Department of Electronics and Telecommunications
- Associate Professor in Communications Engineering

Research:

- Error Correcting Codes, Space Communication Systems

Teaching 2024/2025

- Information Theory for Data Science
- Communication Systems
- Applied Signal Processing Laboratory
- Space technologies for application and services

Giorgio Taricco

giorgio.taricco@polito.it

Affiliation:

- Department of Electronics and Telecommunications
- Full Professor in Communications Engineering

Research:

- Information Theory, MIMO wireless systems, space communications

Teaching 2024/2025

- Information Theory for Data Science
- Advanced wireless communications and coding