

Computing for Data Science

Fall term 3 ECTS

Mandatory Course

Prof. Joan Verdú

Prerequisites to Enroll

Although not mandatory, some knowledge of Python, Jupyter notebooks, and algebra is recommended.

Overview and Objectives

We will start each day with a short programming exercise that will be presented and discussed. This will be followed by lectures that will introduce a new topic, and there will be assignments each day to apply the theory we discussed in class. This is an essential time for you to get comfortable with programming before the theoretical onslaught of your first-term courses begins. I highly recommend you familiarize yourself as much as possible with R and Python before the course. This will allow me to give you more personalized guidance during the brushup.

Course Outline

Get to Know Your Computer

- Introduction to Unix
- Introduction to shell
- Installing packages, compiling executables, the path.
- Reproducible environments

Python

- Python basics: IDE's, editors, Jupyter
- Packages and folder structures
- Basic Syntax
- Easier to Ask for Forgiveness than Permission
- Object Oriented Programming in Python

R

- R basics: IDE's, editors, scripts, and interactive sessions.
- Packages
- Understanding and using Functions
- Working with empirical data: Tidyverse
- Functional Programming in R

Computing for Data Science

Fall term 3 ECTS

Mandatory Course

Prof. Joan Verdú

Programming Efficiently

- DRY
- Loose coupling and the principle of least knowledge
- Testability
- Object Oriented vs Functional Programming
- Working with a Team
- Git

Cloud computing

- AWS
- Google Collab
- Spark/pyspark

Required Activities

TBD

Evaluation

Attendance at classes, and submission of homework.

3 or 4 projects in total

- Will be given at the last class of each part
- Students will be given 2 weeks to submit their project.

The grade will be the average of these projects, plus eventually some extra tasks to be done in class.

Competences

- Construct a global vision of the situation of the problem based on knowledge of the synergies between advanced statistical methods, computing and business analysis to generate added value.
- Modeling and predicting high-dimensional data with advanced statistical methods in the field of data science in order to improve strategic decision making.
- Apply the knowledge of programming languages, computer programs and advanced services in the Cloud to solve the problems that are presented to the data scientist.
- Solve the real problems that arise in the fields of study through the accurate analysis of the data.

Computing for Data Science

Fall term 3 ECTS

Mandatory Course

Prof. Joan Verdú

- Visualize and interact with high-dimensional data in order to contextualize the information and facilitate subsequent decision-making.
- Communicate with conviction in English the results and implications of the required analytical study using a language related to the receiver.
- Work in a heterogeneous team of researchers in the field of the economic analyst using specific group techniques.
- Own and understand knowledge that provides a basis or opportunity to be original in the development and / or application of ideas, often in a research context.
- That students know how to apply the acquired knowledge and their ability to solve problems in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study.
- That the students be able to integrate knowledge and face the complexity of making judgments based on information that, being incomplete or limited, include reflections on the social and ethical responsibilities linked to the application of their knowledge and judgments.
- That the students know to communicate their conclusions and the knowledge and last reasons that sustain them to specialized and non-specialized publics in a clear and unambiguous way.
- That students have the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

Learning Outcomes

- Elaborate and estimate probabilistic prediction models based on certain data.
- Predict random processes.
- Apply supervised and semi-supervised learning algorithms.
- Apply search algorithms and estimation methodologies in networks through observation of data.
- Apply mathematical and computational analysis of social, business and economic networks knowing the theory and optimization algorithms.
- Predicting information needs based on decisions that must be made.
- Apply mathematical theory and statistics on data sets from disparate disciplines.

Computing for Data Science

Fall term 3 ECTS

Mandatory Course

Prof. Joan Verdú

Materials

TBD