Course guide
295102 - 295II012 - Data Analysis and Pattern Recognition

Last modified: 19/01/2026

Unit in charge: Barcelona East School of Engineering
Teaching unit: 723 - CS - Department of Computer Science.
749 - MAT - Department of Mathematics.
707 - ESAII - Department of Automatic Control.

Degree: MASTER'S DEGREE IN CHEMICAL ENGINEERING (Syllabus 2019). (Compulsory subject).
MASTER'S DEGREE IN INTERDISCIPLINARY AND INNOVATIVE ENGINEERING (Syllabus 2019). (Compulsory subject).
ERASMUS MUNDUS MASTER IN SUSTAINABLE SYSTEMS ENGINEERING (EMSSE) (Syllabus 2024). (Optional subject).

Academic year: 2025 **ECTS Credits:** 6.0 **Languages:** English

LECTURER

Coordinating lecturer: CHRISTIAN MATA MIQUEL

Others: Primer quadrimestre:
CHRISTIAN MATA MIQUEL - Grup: T11, Grup: T12
FRANCESC POZO MONTERO - Grup: T11, Grup: T12
KEVIN IVAN BARRERA LLANGA - Grup: T11, Grup: T12

PRIOR SKILLS

Skills in algebraic computation.
Capacity for abstraction.
Self-sufficiency in learning.

REQUIREMENTS

Programming at intermediate/advanced level (Python)
Proficiency in interactive programming environments (Google Colab, Jupyter Notebook...)

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEMUEQ-01. To apply knowledge of mathematics, physics, chemistry, biology and other natural sciences, obtained through study, experience and practice, with critical reasoning, to establish economically viable solutions to technical problems
CEMUEQ-03. Conceptualize engineering models, apply innovative methods in the resolution of problems and adequate computer applications, for the design, simulation, optimization and control of processes and systems
CEMUEII-02. Apply techniques of pattern recognition, artificial intelligence and statistical data analysis that allow decisions to be made objectively, quantitatively and reproducibly in problems of a multidisciplinary nature.

Generical:
CGMUEQ-04. To carry out the appropriate research, undertake the design and manage the development of engineering solutions, in new or little known environments, relating creativity, originality, innovation and technology transfer
CGMUEQ-05. Know how to establish mathematical models and develop them through appropriate information technology, as a scientific and technological base for the design of new products, processes, systems and services, and for the optimization of others already developed
CGMUEII-01. Participate in technological innovation projects in multidisciplinary problems, applying mathematical, analytical, scientific, instrumental, technological and management knowledge.

Transversal:
05 TEQ. TEAMWORK. Being able to work as a team player, either as a member or as a leader. Contributing to projects pragmatically and responsibly, by reaching commitments in accordance to the resources that are available.
06 URI. EFFECTIVE USE OF INFORMATION RESOURCES. Managing the acquisition, structure, analysis and display of information from the own field of specialization. Taking a critical stance with regard to the results obtained.
03 TLG. THIRD LANGUAGE. Learning a third language, preferably English, to a degree of oral and written fluency that fits in with the future needs of the graduates of each course.

LEARNING RESULTS

Skills:
S04. Understand advanced digital technologies so that they can be applied critically in diverse contexts such as data analysis, multiscale modelling, technoeconomic analysis and environmental systems analysis.
S05. Apply pattern recognition, artificial intelligence and statistical data analysis techniques to make objective, quantitative and reproducible decisions in multidisciplinary problems.

Competences:
C03. Manage the acquisition, organisation, analysis and presentation of data and information in the field of complex systems engineering and critically assess the results obtained.
C05. Propose advanced scientific and technological solutions to complex industrial challenges in areas such as intelligent production, robotic systems, logistics, fault detection and predictive maintenance.
C02. Work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.

TEACHING METHODOLOGY

The teaching methodology will be based on weekly theory and laboratory sessions. In the in-person theory learning sessions, faculty will introduce the concepts, methods, and results of the subject through theoretical explanations and illustrative examples. Part of the theory will be based on problem-solving sessions, where the professor will guide students in completing exercises and problems related to the subject matter.

In the laboratory sessions, students will put into practice the concepts, methods, and results of the subject with the professor's assistance, working directly on real biomedical images. Students will independently study to assimilate concepts and solve proposed exercises, and work on a group application case.

Finally, throughout the course, various seminars will be held by leading experts and professionals. These seminars will complement the learning of theoretical and practical concepts of the subject, providing students with a direct view of the real-world applicability of the acquired knowledge and demonstrating the concrete impact this sector has on current real-world challenges.

LEARNING OBJECTIVES OF THE SUBJECT

The main objective of the course is to introduce students to advanced data analysis techniques, performance evaluation and result interpretation, pattern recognition, and provide a solid introduction to machine learning and deep learning in the field of Artificial Intelligence. A thorough study will be conducted of their essential characteristics —supervised and unsupervised algorithms, neural networks, optimization and validation techniques— as well as their technical limitations and extensive interdisciplinary field of application. Special emphasis will be placed on practical applicability to biomedical, industrial, and computer engineering through real-world use cases such as medical image segmentation, prediction, classification, and feature extraction in complex engineering problems.

STUDY LOAD

Type	Hours	Percentage
Hours small group	27,0	18.00
Hours large group	27,0	18.00
Self study	96,0	64.00

Total learning time: 150 h

CONTENTS

Exploratory data analysis

Description:
Exploratory Data Analysis (EDA) is a fundamental stage in any data science or machine learning project. It involves the process of examining datasets to summarize their main characteristics, uncover patterns, detect anomalies, test hypotheses, and check assumptions through both statistical methods and data visualization techniques. The goal of EDA is to understand the structure of the data before applying more complex models, ensuring that the following analytical steps are based on well-grounded insights.

Specific objectives:
The specific objectives for this module will focus on learning:

- Data visualization (histograms, box-plot, qq-plot, multi-dimensional scatter plots, etc).
- Data clustering (k-means, agglomerative, Gaussian Mixture Models)
- Dimensionality reduction and Principal Component Analysis
- Data representation and feature extraction
- Data metrics, distances, norms, etc.

Related activities:
Laboratory session 1: Data visualization and cluster analysis
Laboratory session 2: Principal Component Analysis and dimensionality reduction

Full-or-part-time: 10h
Theory classes: 6h
Laboratory classes: 4h

Supervised machine learninig

Description:
Supervised Machine Learning is a technique where a model learns to predict outcomes from labeled data by finding relationships between inputs and outputs. It involves key methods such as Support Vector Machines (SVM), which find optimal class boundaries, and Decision Trees, which use hierarchical rules for classification. Model performance is evaluated using metrics like accuracy, precision, and F1-score, along with cross-validation to ensure good generalization.

Specific objectives:
The specific objectives for this module will focus on learning:
• Introduction to Classification/Regression problems
• Distance-based methods: kNN & Centroids algorithm
• Probabilistic methods: Naïve Bayes & LDA
• Rule-based methods: Decision Trees & AdaBoost
• Hyperplane-based methods: kernels & SVM

Related activities:
Laboratory session 3: Supervised classification I
Laboratory session 4: Supervised classification II

Full-or-part-time: 12h
Theory classes: 8h
Laboratory classes: 4h

Performance evaluation

Description:
Performance Evaluation is a key stage in the data analysis and machine learning pipeline. It involves assessing how well a model or statistical method performs when making predictions, estimating parameters, or generalizing to new data. The main objective is to ensure that the conclusions drawn from data are both statistically valid and reproducible. Throughout this part of the course, students develop a deep understanding of hypothesis testing, statistical inference, and model validation techniques that are essential for rigorous data analysis.

Specific objectives:
The specific objectives for this module will focus on learning:

• Type I and type II errors
• Univariate and multivariate hypothesis testing approaches
• Statistical inference and parameter estimation (Maximum-likelihood, Bayesian, bootstrapping)
• Validation procedures: cross-validation; leave-one-out, etc.

Related activities:
Laboratory session 5: Hypothesis testing
Laboratory session 6: Model validation procedures

Full-or-part-time: 12h
Theory classes: 8h
Laboratory classes: 4h

Neural networks and deep learning

Description:
Neural Networks and Deep Learning are computational models designed to learn complex patterns from large datasets through layers of interconnected nodes. They are widely used also in data processing due to their ability to automatically extract high-level features. Common classification techniques include convolutional and fully connected networks that categorize data into predefined classes. Various segmentation and classification techniques will also be explored.

Specific objectives:
The specific objectives for this module will focus on learning:

- Introduction to artificial neural networks (ANNs) and deep learning
- Feed-forward ANNs for classification and regression
- Training ANNs: backpropagation algorithm, optimization stages, advanced strategies (network complexity, early stopping, dropout, weight regularization)
- Specialized architectures: recurrent neural networks, autoencoders, generative adversarial networks, convolutional neural networks

Related activities:
Laboratory session 7: Artificial Neural Networks
Laboratory session 8: Deep Learning

Full-or-part-time: 10h
Theory classes: 6h
Laboratory classes: 4h

Advanced topics and applications

Description:
Seminars by experts, applied projects with real biomedical data, critical analysis of recent publications in high-impact journals, discussion of current topics such as generative AI in medical diagnosis, practical workshops, clinical case studies, and debates on emerging trends in medical image processing.

Specific objectives:
Educational supplement to achieve the knowledge explained during previous sessions and real-world applicability in the biomedical field focused on the use of artificial intelligence.

Related activities:
Final project

Full-or-part-time: 8h
Theory classes: 4h
Laboratory classes: 4h

GRADING SYSTEM

Partial exam 30%
Final exam 30%
Projects and exercises 40%

"Those students who meet the requirements set by the EEBE in their Assessment and Permanence Regulations will be able to access the re-assessment test (https://eebe.upc.edu/ca/estudis/estudis-de-master/documents-masters/assessment-and-academic-progress-regulations-for-bachelors-and-masters-degrees-at-the-eebe.pdf)"

EXAMINATION RULES.

The "midterm" exam is theoretical, so no laptops or tablets are required—everything will be done on paper. It consists of two parts: the first with theoretical questions (no notes allowed), and the second focused on practical exercises (using only a basic, non-programmable calculator and any useful handwritten notes on paper). Questions are based exclusively on the theoretical material available on ATENEA.

The final exam will not include the concepts from the midterm exam; it is completely independent. The exam will consist of theoretical questions and concept activities; no notes, laptops, or tablets will be allowed.

BIBLIOGRAPHY

Basic:

- Bishop, Christopher M. Pattern recognition and machine learning. New York: Springer, cop. 2006. ISBN 9780387310732.
- Duda, Richard O; Hart, Peter E; Stork, David G. Pattern classification. 2nd ed. New York [etc.]: John Wiley & Sons, cop. 2001. ISBN 0471056693.

RESOURCES

Other resources:

Python programming: https://www.python.org/ />Numpy mathematical libraries: http://www.numpy.org/ />Machine learning library: https://scikit-learn.org/stable/ />Graphical representation: https://matplotlib.org/ />Data repository: https://archive.ics.uci.edu/ml/index