Course guide
295626 - 295MB122 - Data Analysis and Machine Learning

Last modified: 12/06/2025

Unit in charge: Barcelona East School of Engineering
Teaching unit: 749 - MAT - Department of Mathematics.

Degree: MASTER'S DEGREE IN ADVANCED BIOMEDICAL TECHNOLOGIES (Syllabus 2025). (Optional subject).

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

LECTURER

Coordinating lecturer: Pozo Montero, Francesc

Others: Pozo Montero, Francesc

PRIOR SKILLS

To follow this course, students should have prior knowledge of mathematics and statistics (linear algebra, probability, regression), some experience in Python programming (libraries such as NumPy, Pandas, Matplotlib, Scikit-learn), and familiarity with biomedical data analysis (medical images, physiological signals, clinical data). Basic knowledge of machine learning (classification, regression, PCA) and strong analytical skills for interpreting results are recommended.

LEARNING RESULTS

Knowledges:
K2. Recognise advanced data analysis and modelling structures.

Skills:
S6. Interpret biomedical data using data analysis, machine learning and deep learning techniques.

Competences:
C6. Integrate the values of sustainability and understand the complexity of systems, with the aim of undertaking or promoting actions that restore and maintain the health of ecosystems and improve justice, thereby generating visions of sustainable futures.
C4. Use information resources effectively, manage the acquisition, structure, analysis and visualisation of data and information in the area of specialisation and critically assess the results.
C5. Use scientific and technical information to respond to any demand for modification, innovation or improvement of devices, products and processes linked to biomedical engineering for new scientific or technological applications.

TEACHING METHODOLOGY

The course uses theoretical content exposition (AF.1) and the resolution of exercises, problems, and cases (AF.2) for 20%, laboratory or simulation sessions (AF.7) for 20%, and individual and cooperative work for 60%.

LEARNING OBJECTIVES OF THE SUBJECT

Upon completing the Data Analysis and Machine Learning course, students will have acquired the necessary knowledge and skills to:

1. Understand the fundamentals of machine learning applied to biomedicine, including types of biomedical data, analysis techniques, and major supervised and unsupervised learning algorithms.
2. Apply preprocessing and exploratory data analysis techniques to biomedical data, such as normalization, feature selection, and dimensionality reduction, using tools like Scikit-learn and PyTorch.
3. Develop and evaluate machine learning models for classification and regression in biomedical problems, implementing logistic regression, decision trees, SVM, and neural networks.
4. Implement deep learning models for biomedical applications, including convolutional neural networks (CNNs) for medical image processing and recurrent neural networks (RNNs) for data sequences.
5. Interpret the results obtained from artificial intelligence models and assess their reliability using appropriate metrics (accuracy, sensitivity, AUC-ROC, etc.), cross-validation, and overfitting detection techniques.
6. Apply interpretability and explainability methodologies to ensure the ethical and safe use of artificial intelligence in biomedicine, using tools such as SHAP and LIME.
7. Develop and deploy models in real biomedical environments, understanding the challenges and practical considerations of transitioning a model from the experimentation phase to clinical application implementation.
8. Integrate the acquired knowledge into a final project where students will design and implement a machine learning solution applied to a specific biomedical problem.

STUDY LOAD

Type	Hours	Percentage
Hours large group	21,0	14.00
Hours small group	21,0	14.00
Self study	108,0	72.00

Total learning time: 150 h

CONTENTS

Fundamentals of Programming and Machine Learning

Description:
- Introduction to Python and essential libraries (NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch).
- Basic concepts of machine learning: supervised vs. unsupervised learning, overfitting, cross-validation, evaluation metrics.

Related activities:
Practice: Exploratory analysis and preprocessing of biomedical data. A biomedical dataset (e.g., clinical records or physiological measurements) will be provided. Students will perform exploratory analysis, data cleaning, and normalization using Pandas, NumPy, and Scikit-learn.

Full-or-part-time: 6h
Theory classes: 3h
Laboratory classes: 3h

Regression and Classification Models analysis and preprocessing of biomedical data

Description:
- Linear and polynomial regression.
- Logistic regression and evaluation metrics in classification.
- Biomedical applications: prediction of clinical values, disease diagnosis.

Related activities:
Practice: Predicting clinical values with regression. Students will work with a patient dataset to predict a clinical variable (e.g., blood glucose levels). They will implement linear regression and Ridge/Lasso regression and compare performance using RMSE and R^2 metrics.

Full-or-part-time: 6h
Theory classes: 3h
Laboratory classes: 3h

Dimensionality Reduction and Clustering

Description:
- PCA and advanced dimensionality reduction techniques.
- Clustering: K-means, DBSCAN, and hierarchical clustering.
- Interpretability and applications in biomedical data.

Related activities:
Practice: Visualization of biomedical data in 2D using PCA and t-SNE. A dataset with many variables (e.g., images transformed into vectors or genomic data) will be used to reduce dimensionality with PCA and t-SNE and visualize latent groups.

Full-or-part-time: 6h
Practical classes: 3h
Laboratory classes: 3h

Advanced Models: Decision Trees and Ensemble Methods

Description:
- Decision trees, Random Forest, and Gradient Boosting.
- Feature selection and extraction.
- Applications in biomarkers and assisted diagnosis.

Related activities:
Problem-Solving: Model comparison for clinical anomaly detection. Various models (Decision Tree, Random Forest, Gradient Boosting) will be provided, and students will analyze which one performs best in terms of accuracy, recall, and feature importance in an anomaly detection case using ECGs.

Full-or-part-time: 6h
Theory classes: 3h
Laboratory classes: 3h

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Neural Networks and Deep Learning

Description:
- Introduction to deep neural networks.
- Fully Connected Neural Networks.
- Regularization and optimization in neural networks.

Related activities:
Practice: Building a neural network for medical diagnosis. PyTorch or TensorFlow will be used to construct an MLP for medical data classification (e.g., diabetes detection with the Pima Indians Diabetes dataset). Students will experiment with optimization, dropout, and L2 regularization.

Full-or-part-time: 9h
Theory classes: 4h 30m
Laboratory classes: 4h 30m

Convolutional Networks and Applications in Biomedical Imaging

Description:
- Fundamentals of CNNs: convolutions, pooling, and classic architectures.
- Biomedical applications: medical image classification, segmentation, automated diagnosis.
- Model evaluation and ethical considerations.

Related activities:
Practice: Medical image classification using CNNs. A dataset of medical images (e.g., chest X-rays for pneumonia detection) will be used. Students will implement a basic CNN with Keras/PyTorch, analyze confusion matrices, interpretability (Grad-CAM), and ethical biases.

Full-or-part-time: 9h
Theory classes: 4h 30m
Laboratory classes: 4h 30m

GRADING SYSTEM

First midterm exam = 35%
Second midterm exam = 45%
Projects, assignments, directed activities = 20%

The evaluation will be conducted through faculty assessment. Students can pass the course through continuous assessment, consisting of two midterm exams (the first midterm midway through the course and the second during the official exam period) and the completion of projects, assignments, and directed activities. Each exam consists of a theoretical block (questions) and a practical block (problems), with a total duration of two hours.

There will be no re-assessment exam for this course.

BIBLIOGRAPHY

Basic:
- Deprez, Maria; Robinson, Emma C. Machine learning for biomedical applications : with Scikit-learn and PyTorch. London: Elsevier, 2024. ISBN 9780128229040.

Complementary:
- Kose, Utku; Deperlioglu, Omer; Hemanth, Jude. Deep learning for biomedical applications [on line]. Washington: Taylor & Francis Group, 2021 [Consultation: 19/09/2025]. Available on: https://www-taylorfrancis-com.recursos.biblioteca.upc.edu/books/edit/10.1201/9780367855611/deep-learning-biomedical-application

s-utku-kose-omer-deperlioglu-jude-hemanth. ISBN 9780367855611.
- Machine learning and artificial intelligence in healthcare systems: tools and techniques. Boca Raton, FL: CRC Press, 2023. ISBN 9781003265436.
- Handbook of deep learning in biomedical engineering : techniques and applications. Academic Press, 2021. ISBN 0128230479.
- Machine learning for healthcare applications. Hoboken, New Jersey: Wiley, 2021. ISBN 9781119792598.
- Evolution of machine learning and internet of things applications in biomedical engineering. CRC Press, 2025. ISBN 9781003476207.
- Machine learning for healthcare systems : foundations and applications. Denmark: River Publishers, 2023. ISBN 9788770228114.
- Machine learning in healthcare : fundamentals and recent applications. Milton: CRC Press, 2024. ISBN 9781000540406.

RESOURCES

Hyperlink:
- Machine Learning (MIT OCW). Machine Learning (MIT OCW)

Other resources:
https://ocw.mit.edu/courses/6-867-machine-learning-fall-2006/