Course guide
295601 - AB - Biostatistical Learning

Last modified: 27/05/2024

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

Degree: BACHELOR'S DEGREE IN BIOMEDICAL ENGINEERING (Syllabus 2009). (Optional subject).

Academic year: 2024 **ECTS Credits:** 6.0 **Languages:** Spanish

LECTURER

Coordinating lecturer:
Alférez Baquero, Edwin Santiago
Mujica Delgado, Luis Eduardo
Ruiz Ordoñez, Magda Liliana

Others:
Alférez Baquero, Edwin Santiago
Ruiz Ordoñez, Magda Liliana
Mujica Delgado, Luis Eduardo

PRIOR SKILLS

During the undergraduate training stage you have seen some fundamentals of programming, as well as certain mathematical concepts. In this course you will be able to appreciate some utilities, specifically:

1. Python: The use of variables, data types, statements, expressions, operators and precedence.

2. Mathematical Concepts: Throughout the course you will see the application of some concepts of statistics, calculus and algebra.
- Statistics: Understanding basic statistical measures (mean, median, standard deviation), conditional probability and hypothesis testing.
- Calculus: Fundamental concepts of differentiation and integration, mainly as they relate to optimisation problems in data science.
- Algebra: Basic linear algebra, including matrix operations and vector spaces, as these are crucial for understanding data structures, algorithmic complexity and machine learning models.

While a firm grasp of these mathematical concepts will be helpful in understanding the course material, detailed performance of these specific skills will not be part of the course assessment.

REQUIREMENTS

This subject does not require any prerequisites. Yes, it is necessary to have a great interest in understanding and applying methods for the analysis of biomedical data. Enthusiasm and willingness to approach the concepts of new technologies to solve challenging problems are also essential.

TEACHING METHODOLOGY

This course adopts an integrated learning model, where theoretical concepts and practical application are seamlessly intertwined for four hours per week. Each session takes place inside a computer lab. After the introduction of each theoretical concept, its practical implementation will be immediately explored through Python programming, promoting an active learning environment (constructivist learning theory). At the beginning of each class, all Python materials will be provided. In this way, the focus is on learning and applying new concepts.

Additionally, the course leverages the flipped classroom model to some extent, allocating 60% of the learning path to self-study. Students are expected to engage in self-directed learning, encouraging independent research and consolidation of knowledge. This component is designed to cultivate a learner-centered environment, allowing learners to take responsibility for their learning and develop self-regulation skills, which are critical in the field of data management. The pedagogical strategy distributes the evaluation weights accordingly:

- Knowledge based on theory: 20%
- Computer laboratory tasks and projects: 20%
- Autonomous learning: 60%

As an assessment framework, three assignments will be developed aimed at fostering engagement and continuous application of concepts, ensuring a holistic learning experience that is both rigorous and contextually relevant.

LEARNING OBJECTIVES OF THE SUBJECT

The objectives are designed to foster a holistic learning experience that combines technical knowledge with essential soft skills such as critical thinking, collaboration and independent research:

- Gain a solid understanding of the fundamental principles of machine learning as they relate to the field of biomedical engineering.

- Learn to articulate and frame challenges in various biomedical contexts.

- Gain a comprehensive understanding of a spectrum of algorithms for data and image processing and analysis, including their potential benefits and limitations.

- Apply this new knowledge to address and solve biomedical problems of moderate complexity, promoting mastery of Python coding in data and image analysis contexts.

- Develop the capacity for critical evaluation of results by making well-informed methodological decisions.

- Encourage the development of autonomous learning skills, enabling students to navigate and master different fields independently.

- Cultivate a collaborative learning environment where the exchange of ideas and constructive group participation is valued and promoted.

- Instill a research-driven mindset, encouraging the student body to explore beyond the materials provided and synthesize information from diverse sources.

- Provide students with the ability to critically compare various techniques and recommend the most appropriate ones for certain biomedical issues.

STUDY LOAD

Type	Hours	Percentage
Hours small group	30,0	20.00
Hours large group	30,0	20.00
Self study	90,0	60.00

Total learning time: 150 h

CONTENTS

1. Introduction

Description:
The sessions are designed to provide an overview of the historical development of industrial processes and lay the foundation for understanding the role of automated processing in contemporary data and image analysis within biomedical applications. The importance of data acquisition, process control and monitoring will be examined, along with various models used for data analysis and interpretation.

Specific objectives:
Become familiar with the conceptual foundations of machine learning and its application in biomedical engineering. This includes understanding the importance of data in driving process improvements, the basics of data acquisition and instrumentation, and an introduction to the types of algorithms you will encounter throughout the course.

Related activities:
Theoretical classes: Presentation of the course and Introduction.
Laboratory sessions: Introduction to Python, Matplotlib, NumPy, Pandas and data exploration.

Full-or-part-time: 10h
Theory classes: 2h
Laboratory classes: 2h
Self study : 6h

2. Supervised learning

Description:
Supervised learning is a type of machine learning in which an algorithm learns from labeled training data and uses it to make predictions or decisions without human intervention. The basic fundamentals of regression and classification will be presented. The main techniques used for the predictive model within the domain of data science will be studied.
Regression encompasses the statistical methodology for modeling the relationship between a quantitative dependent variable and one or more independent variables. Emphasis will be placed on understanding the theory, underlying assumptions, and practical implementation of regression models in the context of biomedical data analysis.
In the classification sessions, algorithms aimed at categorizing data into predetermined classes will be taught. A variety of classification techniques will be examined. Practical exercises will facilitate the understanding of the application of these methods in the classification of biomedical data sets.

Specific objectives:
Provide the necessary skills to build, interpret and evaluate predictive models using regression and classification algorithms. It will focus on applying these models to biomedical data sets, interpreting the results with a critical eye, and understanding the implications of the models in real-world biomedical scenarios. This knowledge will serve as a springboard to more advanced topics and decision making.

Related activities:
Theoretical classes and laboratory sessions: Linear Regression, Bayes-Naives, Logistic Regression, Discriminant Analysis, K Nearby Neighbors (KNN), Support Vector Machines (SVM). Indices for the evaluation of models and confusion matrix.
Homework 1.

Full-or-part-time: 40h
Theory classes: 8h
Laboratory classes: 8h
Self study : 24h

3. Resampling and ensemble learning

Description:
This part of the course will present advanced methodologies essential to improve predictive model performance. Resampling strategies, such as cross-validation, bootstrapping, and validation sets, are explored for their critical role in evaluating, selecting, and estimating model accuracy. Ensemble learning methods encompass integrative approaches such as bagging, boosting, and stacking, which combine multiple models to overcome the accuracy and stability of single predictive models.

Specific objectives:
Impart a comprehensive understanding of resampling techniques for robust validation and ensemble learning methods to increase model effectiveness. The application of these methodologies will be demonstrated using Python on biomedical data sets, providing you with the skills necessary to mitigate overfitting, variation and bias in model development. Mastering these techniques is crucial to developing advanced and generalizable models in data analysis.

Related activities:
Theoretical classes and laboratory sessions: Cross-validation, Bootstrapping, Ensemble Learning, bagging, boosting, stacking, decision trees.

Full-or-part-time: 20h
Theory classes: 4h
Laboratory classes: 4h
Self study : 12h

4. Unsupervised learning

Description:
These sessions will delve into the exploration of data structures without pre-labeled results. Special attention will be paid to clustering techniques, such as k-means and hierarchical clustering, which group data points based on similarity metrics. The usefulness of dimensionality reduction methods, such as principal component analysis (PCA), will be examined to simplify complex data sets and thus reveal intrinsic patterns.

Specific objectives:
Effectively apply unsupervised learning techniques to discern underlying structures in biomedical data. This includes identifying natural groupings within data sets, reducing the complexity of high-dimensional data for better visualization and interpretation, and preparing the data for further analysis. Through this, competence will be acquired to reveal patterns and trends.

Related activities:
Theoretical classes and laboratory sessions: k-means, Density-Based Spatial Clustering of Applications with noise (DBSCAN), hierarchical clusterin, Principal Component Analysis (PCA).

Homework 2.

Full-or-part-time: 20h
Theory classes: 4h
Laboratory classes: 4h
Self study : 12h

5. Introduction to deep learning

Description:
This section of the course will address deep learning, placing special emphasis on neural networks and their various architectures. Fundamental concepts such as convolutional neural networks (CNN) and recurrent neural networks (RNN) will be introduced, as well as essential techniques such as parameter tuning and regularization. The practical use of deep networks in biomedical engineering will be explored, particularly in image and document classification and time series prediction. Students will learn how to implement and train deep learning models using Python and popular deep learning libraries.

Specific objectives:
Impart a solid understanding of the structure and functioning of artificial neural networks, allowing the practical implementation of these models to solve real-world biomedical engineering problems. Participants will gain the skills to design and train neural network architectures, understand their advantages and limitations, and leverage this knowledge to develop effective solutions to complex biomedical data analysis challenges. This foundation aims to pave the way for advanced studies in machine learning and innovative applications within the biomedical field.

Related activities:
Theoretical classes and laboratory sessions: Structure of neural networks (neurons, layers and activation functions). Algorithms with Feedforward and backpropagation networks. Different types of neural network architectures such as Multilayer Perceptrons (MLP), convolutional neural networks (CNN) and recurrent neural networks (RNN). Strategies to prevent overfitting, such as regularization and dropout methods.

Full-or-part-time: 40h
Theory classes: 8h
Laboratory classes: 24h
Self study : 8h

6. Introduction to Transformers

Description:
This final module of the course introduces the Transformers deep neural network architecture, starting with the attention module and studying the fundamentals of the original Transformer encoder-decoder architecture. The basic architectures derived from the encoder such as BERT, and the decoder part such as GPT, will be explored. Also, it will introduce a modification of the Transformer to apply it in biomedical images, called Vision Transformers (ViT). You will explore these advanced architectures and understand the fundamental principles that drive their performance and the innovative ways they are applied in solving complex problems. The module will also provide practical exposure to implement these algorithms using modern deep learning frameworks and demonstrate their application.

Specific objectives:
Provide a deep understanding of advanced deep learning architectures and their applications, providing the skills necessary not only to understand the theoretical foundations of these sophisticated models. Mastering these advanced techniques is crucial to significantly contributing to the field of biomedical data analytics, driving innovation and pushing the boundaries of what is possible with AI in healthcare and medicine.

Related activities:
Theoretical classes and laboratory practices to understand and apply the Transformers architecture, starting with the study of the attention mechanism and how it is used in the encoder-decoder structure. Architectures derived from Transformer such as BERT for natural language processing and GPT for text generation will be analyzed and put into practice, as well as the adaptation of Transformers to the field of computer vision with ViTs. We will learn about the importance of attention and how it has revolutionized sequential data analysis. Students will have the opportunity to train Transformer models on image classification and natural language understanding tasks, using current deep learning frameworks. Students will be encouraged to experiment with these advanced models to develop innovative solutions to biomedical data analysis problems.

Homework 3

Full-or-part-time: 20h
Theory classes: 4h
Laboratory classes: 4h
Self study : 12h

GRADING SYSTEM

The evaluation will be carried out continuously during the development of each module through practical work.

EXAMINATION RULES.

.

BIBLIOGRAPHY

Basic:
- James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013. ISBN 9781461471370.
- Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006. ISBN 9780387310732.
- Géron, Aurélien. Hands-on machine learning with scikit-learn & tensorflow : concepts, tools, and techniques to build intelligent systems [on line]. Sebastopol, CA: O'Reilly Media, Inc, 2017 [Consultation: 15/04/2020]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=4822582. ISBN 9781491962268.
- Raschka, Sebastian. Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow [on line]. 2nd ed. Birmingham, UK: Packt Publishing Ltd, 2017 [Consultation: 14/04/2020]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=5050960. ISBN 9781787126022|.

Complementary:
- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction [on line]. 2nd ed. New York, NY: Springer Series in Statistics, 2001 [Consultation: 27/08/2018]. Available on: http://dx.doi.org/10.1007/978-0-387-84858-7. ISBN 9780387848587.

RESOURCES

Other resources:
Material available in ATENEA from those responsible for the course.