

Academic Year: (2024 / 2025)

Review date: 16-09-2024

Department assigned to the subject: Statistics Department

Coordinating teacher: BENITEZ PEÑA, SANDRA

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Basic knowledge of statistical software R or similar.

OBJECTIVES

The objective of this course is to train students in the use of advanced data analysis techniques applied to the connected industry. Data visualization tools will be used, and advanced machine learning models will be implemented. By the end of the course, students will be equipped to analyze data, identify patterns, and apply predictive solutions in industrial environments, contributing to process optimization and data-driven decision-making.

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction
 - 1.1 Basics of Multivariate Data Analysis
 - 1.2 Introduction to Statistical Learning
 - 1.3 Supervised vs. Unsupervised Learning
 - 1.4 Data Visualization Techniques
2. Supervised Learning: Regression
 - 2.1 Linear Regression
 - 2.2 Linear Model Selection and Regularization
 - 2.3 Cross-Validation on Regression problems
 - 2.4 Extensions
3. Supervised Learning: Classification
 - 3.1 Logistic Regression
 - 3.2 Bayes classifier
 - 3.3 Linear Discriminant Analysis
 - 3.4 k-Nearest Neighbor classifier
 - 3.5 Random Forests
 - 3.6 Support Vector Machines
 - 3.7 Cross-Validation on Classification problems
4. Unsupervised Learning and Dimensionality Reduction Techniques
 - 4.1 Clustering methods: k-means and hierarchical clustering
 - 4.2 Principal Component Analysis
 - 4.3 Multidimensional Scaling
 - 4.4 ISOMAP and Locally-Linear Embedding

LEARNING ACTIVITIES AND METHODOLOGY

LEARNING ACTIVITIES:

AF3: Theoretical-Practical classes.

AF6: Group work.

AF7: Individual student work.

AF8: Partial and final exams.

METHODOLOGY:

MD1: Theoretical lessons, with support material available on the Web, to present and develop the main concepts of the course. Teachers will provide students with supplementary material.

MD2: Critical reading of documents provided by the teachers: newspaper articles, reports, manuals and / or academic papers, either for later discussion in class, either to expand and consolidate the knowledge of the subject.

MD3: Resolution of practical cases, problems, etc. proposed by the teacher individually or in groups.

MD4: Preparation of projects individually or in group.

TUTORING SESSIONS:

- Weekly individual tutoring sessions
- Group tutorials might be possible

ASSESSMENT SYSTEM

% end-of-term-examination:	40
% of continuous assessment (assignments, laboratory, practicals...):	60

The assessment system (in the ordinary and extraordinary calls) is:

- In-class participation (SE1): 20%
- Individual or group work (SE2): 40%
- Final exam (SE3): 40%

If the student followed the continuous assessment process:

- The exam in the extraordinary call will have the same percentage value as in the ordinary call, and the final grade of the subject will take into account the grade of the continuous assessment and the grade obtained in the exam. However, the student will have the right to be graded in the extraordinary call taking into account only the grade obtained in the exam when it is more favourable.

For those students who do not follow the continuous assessment process:

- Students will be allowed to take an ordinary exam which will have a 60% of value in the final mark.
- The grade in the extraordinary call is exclusively obtained through an exam consisting of both theoretical questions and practical problems, with a 100% of value in the final mark.

BASIC BIBLIOGRAPHY

- G. James, D. Witten, T. Hastie and R. Tibshirani An Introduction to Statistical Learning, Springer, 2021
- H. Wickham ggplot2. Elegant Graphics for Data Analysis, Springer, 2016
- T. Hastie, R. Tibshirani and J. H. Friedman The Elements of Statistical Learning, Springer, 2017
- T. Hastie, R. Tibshirani and M. Wainwright Statistical Learning with Sparsity, CRC Press, 2015

ADDITIONAL BIBLIOGRAPHY

- Annansingh, F., Sesay, J. B. Data Analytics for Business: Foundations and Industry Applications, Taylor & Francis, 2022
- P. Kaliraj, T. Devi Big Data Applications in Industry 4.0., CRC Press, 2022