

Academic Year: ( 2024 / 2025 )

Review date: 14-03-2025

Department assigned to the subject: Statistics Department

Coordinating teacher: AUSIN OLIVERA, MARIA CONCEPCION

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

**REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)**

Regression methods

Multivariate Analysis

It is not recommended to take this course if you have already taken the course on Statistical Learning.

**LEARNING OUTCOMES****LEARNING OUTCOMES**

2.To be able, by means of arguments or procedures developed and supported by themselves, to apply their knowledge,

understanding and problem-solving skills in complex or professional and specialised fields of work that require the use of creative and innovative ideas.

3.Have the ability to collect and interpret data and information on which to base their conclusions, including, when necessary and relevant, reflection on social, scientific or ethical issues in their field of study

4.To be able to cope with complex situations or those requiring the development of new solutions in the academic, work or professional sphere within their field of study

**BASIC COMPETENCES**

2.Students are able to apply their knowledge to their work or vocation in a professional manner and possess the competences usually demonstrated through the development and defence of arguments and problem solving within their field of study.

3.Students have the ability to gather and interpret relevant data (usually within their field of study) in order to make judgements which include reflection on relevant social, scientific or ethical issues.

5.Students will have developed the learning skills necessary to undertake further study with a high degree of autonomy.

**GENERAL COMPETENCES**

4.Identify or create the appropriate model for the specific problem that arises in each business activity (finance, marketing, production planning and control, etc.). Computationally and analytically manipulate the models established, taking advantage of the power of statistical and optimisation methods, etc., and analyse the results obtained.

5.Communicate the results, the conclusions of the models and the proposed solutions in a way that is intelligible to the rest of the company, so that they are accepted and implemented by decision-makers.

**SPECIFIC COMPETENCES**

1. Describe, synthesise and graphically represent a set of data.

2.Modelling and analysing both static and dynamic data using statistical techniques.

3.Mathematically model and solve problems using stochastic models and Operations Research techniques.

4.Demonstrate basic knowledge of programming. Be able to use and develop statistical packages.

13. Knowing how to access, analyse and classify high-volume data of a highly heterogeneous nature (Big Data). Use analysis tools for big data in typical scenarios in finance, marketing and business management applications.

16. Plan integrated offline and online strategies, through the use of communication in social networks, display advertising, affiliation marketing, email, remarketing, gamification, big data.

#### TRANSVERSAL COMPETENCES

2. Being able to establish good interpersonal communication and to work in multidisciplinary and international teams.

3. Being able to organise and plan their work, making the right decisions based on the information available, gathering and interpreting relevant data to make judgements and critical thinking within their area of study.

#### OBJECTIVES

1. Understand the importance of transforming large volumes of data into relevant information for decision making and business development in organizations, companies and individuals.

2. Learn the basic techniques of preprocessing and visualization of data. Gain knowledge on methods to work with missing and atypical data. Acquire the ability to use of dimension reduction techniques.

3. Gain knowledge on the main methods of supervised learning in regression and their usefulness in prediction problems. Distinguish between linear and non-linear models and understand the importance of model selection methods.

4. Become familiar with the usual supervised learning procedures for classification. Understand the most common classifiers and their limitations. Gain knowledge in advanced methods for classification and their benefits in business.

5. Be able to identify the appropriate Big Data techniques in real business problems: customer classification, scoring, risk management, fraud detection, bankruptcy prediction, etc.

#### DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction.

2. Data collection, sampling and preprocessing.

2.1. Types of data.

2.2. Sampling.

2.3. Data visualization tools.

2.4. Missing values.

2.5. Outlier detection and treatment.

2.6. Data transformations.

2.7. Dimension reduction.

2.8. Application: Risk management in the stock market.

3. Supervised learning: regression.

3.1. Linear and polynomial regression.

3.2. Cross-validation.

3.3. Model selection and regularization methods (ridge and lasso).

3.4. Nonlinear models, splines and generalized additive models.

3.5. Application: credit-scoring prediction.

4. Supervised learning: classification.

4.1. Bayes classifiers

4.2. Logistic regression.

4.3. K-nearest neighbors.

4.4. Random forest.

4.5. Support-vector machines.

4.6. Boosting.

4.7. Application: Credit risk.

4.8. Application: Fraud detection.

4.9. Application: Bankruptcy prediction

## LEARNING ACTIVITIES AND METHODOLOGY

Theory (2 ECTS). Lectures with available material posted in internet. Problems (4 ECTS) Problem Solving classes. Computational exercises at computer room. Work assignments in groups. Weekly office hours to assist students on an individual and group basis.

## ASSESSMENT SYSTEM

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

Final exam (60%). Presentations in class, explanatory videos and project (40%)

## BASIC BIBLIOGRAPHY

- Bradley Efron, Trevor Hastie. Computer Age Statistical Inference: Algorithms, Evidence and Data Science., Cambridge University Press, 2016
- E. Alpaydin Introduction to Machine Learning, MIT Press., 2010
- James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer, 2013.
- T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer. , 2009
- Trevor Hastie, Robert Tibshirani, Martin Wainwright Statistical Learning with Sparsity: the Lasso and Generalizations, Chapman & Hall, 2015