

Academic Year: ( 2020 / 2021 )

Review date: 20-07-2020

Department assigned to the subject: Statistics Department

Coordinating teacher: DELGADO GOMEZ, DAVID

Type: Compulsory ECTS Credits : 6.0

Year : 4 Semester : 1

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

Basic multivariate analysis

**OBJECTIVES****SPECIFIC SKILLS and COMPETENCES**

1. To know and use advanced statistical techniques, with last generation software support.
2. To extract and analyze information from large data sets.

**TRANSVERSAL SKILLS and COMPETENCES**

1. Ability of information analysis and synthesis.
2. Modelization and resolution of practical problems in Data Mining.
3. Oral and written communication skills.

**DESCRIPTION OF CONTENTS: PROGRAMME**

1. R programing language
  - 1.1 Data types and importing data
  - 1.2 Loops and conditionals
  - 1.3 Functions
2. Exploratory Data Analysis
  - 2.2 Ggplot2 package
3. Supervised Classification
  - 3.1 K-nearest neighbors
  - 3.2 Decision Trees
  - 3.3 The Gaussian distribution and discriminant analysis
  - 3.4 Support Vector Machines
  - 3.5 Logistic Regression
4. Dimensionality Reduction and clustering techniques
  - 4.1 Principal Component Analysis
  - 4.2 K-means
  - 4.3 Hierarchical Clustering
5. How to write a report with R-Markdown

## LEARNING ACTIVITIES AND METHODOLOGY

The program consists of 14 theoretical classes with supporting material available in the global classroom and 14 sessions based on computer practice sessions. Every week students will have an optional collective tutorial where they can solve their doubts.

## ASSESSMENT SYSTEM

Continuous evaluation will account for 100% of the grade. Students who do not pass the course by continuous assessment may take a practical exam that will count 100% of the course.

Continuous assessment will consist of an exploratory analysis and a predictive analysis of a database. At the middle of the course students will hand in a report showing the most interesting results they have obtained in their exploratory analysis. This report will have a weight of 30% of the final grade of the continuous evaluation. Before the end of the course, students will complement their exploratory study with a predictive analysis in which they will predict the value of a variable of interest. This second assignment will be 30% of the continuous evaluation. Students will be required to present and defend this second assignment at the end of the course. The defense of their work will be the remaining 40% of the continuous evaluation.

<b>% end-of-term-examination:</b>	0
<b>% of continuous assessment (assignments, laboratory, practicals...):</b>	100

## BASIC BIBLIOGRAPHY

- Pathak, Manas A. Beginning data science with R, Springer, 2014