

Academic Year: (2018 / 2019)

Review date: 01-05-2018

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: CID SUEIRO, JESUS

Type: Compulsory ECTS Credits : 6.0

Year : 3 Semester : 2

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

It is recommended to have completed the subjects about mathematical foundations from the first year (Calculus I and II, Linear Algebra, Probability and Data Analysis), the subjects related to programming and algorithms (Programming and Data Structures and Algorithms), as well as subject Statistical Learning.

OBJECTIVES

CB1: That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply knowledge coming from the vanguard of their field of study.

CB2: That students know how to apply their knowledge to their work or vocation in a professional manner and possess the skills that are usually demonstrated through the elaboration and defense of arguments and the resolution of problems within their area of study

CB5: That the students have developed those learning skills necessary to undertake further studies with a high degree of autonomy.

CE1: Ability to solve mathematical problems that may arise in engineering and data science. Ability to apply knowledge about: algebra; geometry; differential and integral calculus; numerical methods; numerical algorithmics;

DESCRIPTION OF CONTENTS: PROGRAMME

This course is divided into 3 thematic blocks. The first concerns the problem of adapting and cleaning a database, a critical preprocessing step that is addressed prior to any machine learning application. The next two blocks address two industry-relevant applications where machine learning techniques have achieved a great success. The understanding of how the different machine learning techniques have to be adapted to solve specific problems of interest to industry and society will provide students with a practical and general vision of applied Machine Learning.

PART I: TECHNIQUES DATA CURATION AND CLEANING

1. Problem Introduction.
2. Organization and integration of databases from different sources.
3. Data cleaning: data characterization, detection and imputation of corrupt data. Outlier detection.

PART II: NATURAL LANGUAGE PROCESSING

4. Text processing with topic modeling
5. Sequential processing of text using neural networks. Text Vector representation and models for automatic translation.

PART III: RECOMMENDATION SYSTEMS

6. Content-based recommendation systems.
7. Matrix factorization decomposition. Collaborative filtering and recommendation systems.

LEARNING ACTIVITIES AND METHODOLOGY

AF1: THEORETICAL-PRACTICAL CLASSES. They will present the knowledge that students should acquire. They will receive the class notes and will have basic texts of reference to facilitate the follow-up of the classes and the development of the subsequent work. Exercises, practical problems on the part of the student will be solved and workshops and evaluation test will be held to acquire the necessary skills.

AF2: Updated to allegation

AF3: INDIVIDUAL OR GROUP WORK OF THE STUDENT.

AF9: FINAL EXAM. In which the knowledge, skills and abilities acquired throughout the course will be assessed globally.

MD1: CLASS THEORY. Exhibitions in the teacher's class with support of computer and audiovisual media, in which the main concepts of the subject are developed and the materials and bibliography are provided to complement the students' learning.

MD2: PRACTICES. Resolution of practical cases, problems, etc. raised by the teacher individually or in groups.

MD3: TUTORIALS. Individualized assistance (individual tutorials) or group (collective tutorials) to students by the teacher.

ASSESSMENT SYSTEM

SE1: FINAL EXAMINATION In which the knowledge, skills and abilities acquired throughout the course will be assessed globally.

SE2: CONTINUOUS EVALUATION. In it, work, presentations, debates, exhibitions in class, exercises, practices and work in the workshops throughout the course will be evaluated.

% end-of-term-examination:	60
-----------------------------------	----

% of continuous assessment (assignments, laboratory, practicals...):	40
---	----

BASIC BIBLIOGRAPHY

- C.C. Aggarwal Recommender Systems: The Textbook, Springer, 2016
- D. Juravsky, J.H. Martin Speech and Language Processing, Prentice Hall; 2nd edition, 2008
- J. Ham, M. Kamber Data Mining: Concepts and Techniques (3rd. ed), Morgan Kaufman, 2011

ADDITIONAL BIBLIOGRAPHY

- C. Manning, H. Schütze Foundations of Statistical Natural Language Processing, MIT Press, 1999
- K. Murphy Machine Learning: A probabilistic Perspective, The MIT Press, 2012
- M. W. Berry Survey of Text Mining Clustering, Classification, and Retrieval, Springer, 2004
- S. Bird, E. Klein, E. Loper Natural Language Processing with Python, O'Reilly, 2009