Machine Learning

Academic Year: (2019 / 2020) **Review date: 30-04-2019**

Department assigned to the subject:

Coordinating teacher: CID SUEIRO, JESUS

Type: Electives ECTS Credits : 6.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

This is a first term course, so no other courses of the Master programme are key for this course. However, it is highly desirable that students are familiarized with basic concepts from statistics.

OBJECTIVES

After this course students will understand the principles of estimation, decision, and grouping problems, and will become familiarized with the different approaches for dealing with them. Students will understand that, for the correct understanding of these problems, it is necessary to master three basic probability theory elements: 1) the likelihood, 2) the difference between a priori and a posteriori uncertainty, and 3) Bayes' Theorem.

From a practical point of view, students will be presented different approaches for learning from data to solve these problems: non-parametric techniques, methods based on empirical risk minimization, or those that follow Bayesian principles.

More specifically, the followint list summarizes the main objectives of this course, enumerated as competences to be acquired by the students:
- knowledge of the theoretic principles underlying several of the most important techniques for learning from data.
- ability to apply such techniques on real problems, and to extract results and conclusions.
- understanding of clasic methods for estimation and clasifications, and skills for their correct application.
- ability to use machine learning tools: neural networks, support vector machines, etc.
- knowledge of last generation approaches, including those based on machine ensembles

DESCRIPTION OF CONTENTS: PROGRAMME

Unit 0: Introduction to data processing

Unit 1: Bayes' Estimation and Decision Theory
- 1.1. General overview of the estimation and decision problems
- 1.2. Bayes' Theorem
- 1.3. Bayes Estimation Theory. MSE, MAD, and MAP estimators
- 1.4. ML Estimation
- 1.5. Optimum Bayes' classifier for the binary and multiclass cases
- 1.6. Characterization of binary classifiers

Unit 2: Regression
- 2.1. The regression problem
- 2.2. Non-parametric regression: k-NN
- 2.3. Linear and polynomial least squares regression
- 2.4. Bayesian regression
- 2.5. Gaussian processes

Unit 3: Classification
- 3.1. Classification problema
- 3.2. Non-parametric methods: k-NN
- 3.3. Logistic regression
- 3.4. Support vector machines

Unit 4: Data clustering

4.1. k-means clustering
4.2. Spectral clustering

LEARNING ACTIVITIES AND METHODOLOGY

LECTURES AND PRACTICAL SESSIONS
Theory sessions consist of lectures in which the basic concepts of the course will be introduced, illustrating them with a large number of examples. Exercises and problems similar to those to be proposed in the exam will also be solved along the course.

LAB SESSIONS
Sessions in which students will apply the concepts presented in the course with the help of a computer. Students will deal with estimation and classification problems with real data, and will have to evaluate the performance of the implemented systems

RESEARCH PROJECT AND DISCUSSIONS
Students will be given a list of topics related to the research areas of the course, so that they can prepare a project on one of them. The work will be presented to the class on specific sessions.

ASSESSMENT SYSTEM

During the ordinary period, students will be graded according to:

* Continuous assessment (75%):
  - Intermediate assessment (Unit 1): 20%
  - Work project about a scientific article:
    * Content: 25%
    * Oral presentation: 15%
  - Lab Sessions: 15%

* Final exam (Unit 2-4): 25%

The extraordinary call will consist of three parts: data analysis project, theory exam (Units 1-4) and lab exam. The student can keep any scores obtained during the ordinary call, for the extraordinary call. However, the attendance to any of these tests on the extraordinary call implies the automatic withdrawal of the corresponding score from the ordinary call.

The work presentation will not be repeated for the extraordinary call. Students can keep their score from the ordinary call, or discard it. In the latter case, the final grade will be computed as follows:

- Final project (30 %)
- Lab exam (20 %)
- Theory assessment (50 %)

% end-of-term-examination:	25
% of continuous assessment (assigments, laboratory, practicals…):	75

BASIC BIBLIOGRAPHY

- C. E. Rasmussen Gaussian Processes for Machine Learning, MIT Press, 2006
- R. O. Duda, P. E. Hart, D. G. Stork Pattern Classification (2nd ed.), Wiley Interscience, 2001

ADDITIONAL BIBLIOGRAPHY

- C. M. Bishop Pattern Recognition and Machine Learning, Springer, 2006