Data Processing

Academic Year: (2020 / 2021) **Review date: 21-07-2020**

Department assigned to the subject: Signal and Communications Theory Department

Coordinating teacher: CID SUEIRO, JESUS

Type: Compulsory 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 underlying the general regression, classification and data analysis problems, and will become familiarized with the different approaches for dealing with them. Students will learn 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 following 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 classic methods for estimation and classifications, and skills for their correct application.
- ability to use machine learning tools: gaussian processes, support vector machines, non-parametric methods
- knowledge of other data analysis problems, like topic modeling or recommendations systems

DESCRIPTION OF CONTENTS: PROGRAMME

Unit 0: Introduction to data processing

Unit 1: Data preprocessing
- 1.1. Data normalization
- 1.2. Dimensionality reduction
- 1.3. Clustering

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. Other regression algorithms

Unit 3: Classification
- 3.1. Classification problema
- 3.2. Non-parametric methods: k-NN
- 3.3. Logistic regression
- 3.4. Neural Networks.
- 3.5. Other classification algorithms

Unit 4: Topic models
- 4.1. Text analysis

4.2. Latent Dirichlet Allocation

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 real data analysis problems, and will have to evaluate the performance of the implemented systems

ASSESSMENT SYSTEM

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

* Continuous assessment: 50%:
  - Classification or regression challenge: 25%
  - Data analysis project: 25%

* Final assessment: 50%:
  - Theory exam: 30 %
  - Laboratory test: 20 %.

The extraordinary call will consist of three parts: data analysis project, theory exam and lab exam. The student can preserve the score obtained during the ordinary call, for the extraordinary call. However, the attendance of any of these tests on the extraordinary call implies the automatic withdrawal of the corresponding score from the ordinary call

The challenge will not be celebrated during the extraordinary call. The student can keep the mark of the challenge from the ordinary call, or discard it. In the latter case, the final grade will be computed from the other components preserving the same relative weights.

% end-of-term-examination:	50
% of continuous assessment (assigments, laboratory, practicals…):	50

BASIC BIBLIOGRAPHY

- R. O. Duda, P. E. Hart, D. G. Stork Pattern Classification (2nd ed.), Wiley Interscience, 2001

ADDITIONAL BIBLIOGRAPHY

- A.C. Müller, S. Guido Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly Media, Inc, 2016
- C. M. Bishop Pattern Recognition and Machine Learning, Springer, 2006
- Trevor Hastie, Robert Tibshirani, Jerome H. Friedman The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer Series in Statistics, 2009