Advanced Statistical Data Analysis

Academic Year: (2017 / 2018)   Review date: 30-03-2017

Department assigned to the subject: Statistics Department

Coordinating teacher: CABRAS , STEFANO

Type: Compulsory ECTS Credits : 6.0

Year : 4 Semester : 1

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.   Statistical problems for multivariate data: interpretation versus prediction.

2.   Interpretation by means of visual representation methods and Cluster analysis
2.1   Multidimensional Scaling.
2.2   Biplots.
2.3   Perceptual Mappings.
2.4.   Cluster Analysis. Hierarchical Methods, k-means and mixture models.
2.4.1   Bottom up hierarchical clustering algorithms.
2.4.2   k-means and related algorithms.

3.   Prediction of a random outcome: Parametric and non-parametric regression methods.
3.1   Linear and quadratic discriminant analysis.
3.2   Regression for quantitative and categorical data.
3.3   Regression trees and Random Forests

4. Text Mining.
   4.1 Main concepts.
   4.2 Word clouds.
   4.3 Term by document matrix.

LEARNING ACTIVITIES AND METHODOLOGY

14 Theoretical support materials available on the Web, and 14 sessions based on problem-solving sessions and practical computing. No group tutorials except during the last week.

ASSESSMENT SYSTEM

60% of the final grade will be achieved by a final examination for assessing the knowledge acquired. A minimum of 4 points (out of 10) is required in the final exam. The remaining 40% is obtained by two midterm exams (15%+20%) and the compulsory tasks assigned in the computational labs (5%). Theoretical questions as well as queries on computational laboratories can be asked in the exams.

% end-of-term-examination:	60
% of continuous assessment (assigments, laboratory, practicals…):	40

BASIC BIBLIOGRAPHY

- E. Alpaydin Introduction to Machine Learning, 2nd Edition, MIT Press, 2010

- T. Hastie, R. Tibshirani, J. Friedman Elements of Statistical Learning, 2d Ed, Springer, 2009