Predictive Modeling

Academic Year: (2023 / 2024) **Review date: 27-04-2023**

Department assigned to the subject: Statistics Department

Coordinating teacher: GARCIA PORTUGUES, EDUARDO

Type: Compulsory ECTS Credits : 6.0

Year : 2 Semester : 2

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Calculus I and II
Linear Algebra
Programming
Probability and Data Analysis
Introduction to Statistical Modeling
Statistical Learning

OBJECTIVES

* General competences

- CG1: Adequate knowledge and skills to analyse and synthesise basic problems related to engineering and data science, solve them and communicate them efficiently.
- CG4: Ability to solve technological, computational, mathematical and statistical problems that may arise in engineering and data science.
- CG5: Ability to solve mathematically formulated problems applied to different subjects, using numerical algorithms and computational techniques.
- CG6: Synthesise the conclusions obtained from the analyses carried out and present them clearly and convincingly, both written and orally.

* Transversal competences

- CT1: Ability to communicate knowledge orally and in writing, before a specialised and non-specialised public.

* Specific competences

- CE1: Ability to solve mathematical problems that may arise in engineering and data science. Ability to apply knowledge about: algebra; geometry; differential and integral calculation; numerical methods; numerical algorithm; statistics and optimisation.
- CE2: Properly identify problems of a predictive nature corresponding to certain objectives and data and use the basic results of regression analysis as the basic basis of prediction methods.
- CE5: Understand and handle fundamental concepts of probability and statistics and be able to represent and manipulate data to extract meaningful information from them.
- CE7: Understand the basic concepts of programming and ability to carry out programs aimed at data analysis.

DESCRIPTION OF CONTENTS: PROGRAMME

This course is designed to give a panoramic view of several tools available for predictive modeling, at an introductory-intermediate level. This view covers in-depth the main concepts in linear models and gives an overview on their extensions. The focus is placed on providing the main insights on the statistical/mathematical foundations of the models and on showing the effective implementation of the methods through the use of the statistical software R.

1. Introduction
   1.1. Course overview
   1.2. Review on probability
   1.3. Random vectors
   1.4. Review on statistical inference
   1.5. What is predictive modeling?

2. Simple linear regression
  2.1. Model formulation and estimation
  2.2. Assumptions of the model
  2.3. Inference for model parameters
  2.4. Prediction
  2.5. ANOVA and model fit

3. Multiple linear regression
  3.1. Model formulation and estimation
  3.2. Assumptions of the models
  3.3. Inference for model parameters
  3.4. ANOVA and model fit
  3.5. Model selection
  3.6. Handling nonlinear relationships
  3.7. Use of qualitative predictors
  3.8. Model diagnostics and multicollinearity

4. Linear regression extensions
  4.1. Review on principal component analysis
  4.2. Principal components regression
  4.3. Partial least squares regression
  4.4. Regularized linear models
  4.5. Ridge and lasso regression

5. Logistic regression
  5.1. Model formulation and interpretation
  5.2. Maximum likelihood estimation
  5.3. Inference for model parameters
  5.4. Model selection and multicollinearity
  5.5. Regularized logistic models

The program is subject to minor modifications due to the course development and/or academic calendar.

LEARNING ACTIVITIES AND METHODOLOGY

The lessons primarily consist of theoretical expositions on the statistical methods of the course. These are complemented with illustrative examples. The laboratories are designed to carry out exercises and case studies that elaborate on the practical usage of the seen methods. The implementation of the methods is done with the statistical language R.

ASSESSMENT SYSTEM

The continuous evaluation is done by a mixture of:

(a) three theoretical-practical quizzes;
(b) active participation in lessons and voluntary exercises.

The Grade of the Continuous Evaluation (GCE) is

GCE = min(A + 0.10 * B, 10),

where:

- A (on the scale 0-10) is the weighted grade of the quizzes;
- B (on the scale 0-10) is the grade corresponding to (b).

If GCE < 5, the student has to attend the final exam to pass the course. For students who take the final exam, the Grade for the Ordinary Call (GOC) is

GOC = 0.4 * GCE + 0.6 * GFE,

where GFE is the Grade of the Final Exam. If GCE >= 5, the student can choose either (1) to take the final exam with the above weighting or (2) not to take the final exam, being in this case GOC = GCE.

The Grade for the Extraordinary Call (GEC) is determined by means of the Grade of the Extraordinary Exam (GEE) and the continuous evaluation:

GEC = max(0.4 * GCE + 0.6 * GEE, GEE).

Further details are provided in Aula Global. The evaluation is subject to modifications due to the course development and/or academic calendar.

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

BASIC BIBLIOGRAPHY

- James, G., Witten, D., Hastie, T., and Tibshirani, R. An Introduction to Statistical Learning, Springer-Verlag, 2013

- Papoulis, A. and Pillai, S. U. Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 2002

ADDITIONAL BIBLIOGRAPHY

- Hastie, T., Tibshirani, R., and Friedman, J. The Elements of Statistical Learning, Springer, 2013
- Kuhn, M. and Johnson, K. Applied Predictive Modeling, Springer, 2013
- Panaretos, V. M. Statistics for Mathematicians, Springer, 2016
- Peña, D. Regresión y Diseño de Experimentos, Alianza Editorial, 2002
- Seber, G. A. F. Linear Regression Analysis, John Wiley & Sons, 1977
- Wood, S. N. Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC, 2006