Predictive modeling

Academic Year: (2020 / 2021) **Review date: 21/07/2020 20:08:08**

Department assigned to the subject: Statistics Department

Coordinating teacher: GARCIA PORTUGUES, EDUARDO

Type: Compulsory ECTS Credits : 3.0

Year : 1 Semester : 1

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Mathematics for data analysis
Statistics for data analysis

OBJECTIVES

* Basic competences

- CB6: Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context.
- CB9: That students know how to communicate their conclusions and the knowledge and ultimate reasons that sustain them to specialised and non-specialised audiences in a clear and unambiguous way.
- CB10: That the students have the learning skills that allow them to continue studying in a way that will be largely self-directed or autonomous.

* General competences

- CG1: Ability to apply the techniques of analysis and representation of information, in order to adapt it to real problems.
- CG4: Synthesise the conclusions obtained from data analyses and present them clearly and convincingly in a bilingual environment (Spanish and English) both written and orally.
- CG5: Be able to generate new ideas (creativity) and anticipate new situations, in the contexts of data analysis and decision making.
- CG6: Use social skills for teamwork and to relate to others autonomously.
- CG7: Apply advanced techniques of analysis and representation of information, in order to adapt it to real problems.

* Specific competences

- CE1: Apply in the development of methods of analysis of real problems, advanced knowledge of statistical inference.
- CE2: Use free software such as R and Python for the implementation of statistical analysis.
- CE5: Apply the advanced statistical foundations for the development and analysis of real problems, which involve the prediction of a variable response.
- CE6: Apply nonparametric models for the interpretation and prediction of random phenomena.
- CE10: Apply statistical modeling in the treatment of relevant problems in the scientific field.

DESCRIPTION OF CONTENTS: PROGRAMME

This course is designed to give a panoramic view of several tools available for predictive modeling, at an intermediate-advanced level. This view covers in-depth the main concepts in linear models and generalized linear models (with their shrinkage versions), and more superficially the model-free approach based on nonparametric regression. 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 statistical software. This is achieved by a mixture of theory and

reproducible code.

1. Introduction
   1.1 Course overview
   1.2 What is predictive modeling?
   1.3 General notation and background

2. Linear models I: multiple linear model
   2.1 Model formulation and least squares
   2.2 Assumptions of the model
   2.3 Inference for model parameters
   2.4 Prediction
   2.5 ANOVA
   2.6 Model fit

3. Linear models II: model selection, extensions, and diagnostics
   3.1 Model selection
   3.2 Use of qualitative predictors
   3.3 Nonlinear relationships
   3.4 Model diagnostics
   3.5 Dimension reduction techniques

4. Linear models III: shrinkage and big data
   4.1 Shrinkage
   4.2 Big data considerations

5. Generalized linear models
   5.1 Model formulation and estimation
   5.2 Inference for model parameters
   5.3 Prediction
   5.4 Deviance
   5.5 Model selection
   5.6 Model diagnostics
   5.7 Shrinkage
   5.8 Big data considerations

6. Nonparametric regression
   6.1 Nonparametric density estimation
   6.2 Kernel regression estimation
   6.3 Kernel regression with mixed multivariate data
   6.4 Prediction and confidence intervals
   6.5 Local likelihood

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

LEARNING ACTIVITIES AND METHODOLOGY

The lessons consist on a mixture of theory (methods description) and practice (implementation and practical usage of methods). The implementation of the methods is done with the statistical language R, so good coding abilities on it re fundamental to understand the implementations. Students are expected to bring their own laptops to experience with the code during some parts of the lessons.

ASSESSMENT SYSTEM

% end-of-term-examination/test:	0
% of continuous assessment (assigments, laboratory, practicals…):	99

The evaluation in the ordinary call is done entirely by continuous evaluation. This is done by a mixture of:

A) a set of midterm exams;
B) participation in lessons and voluntary activities.

The continuous evaluation grade (in the scale 0-10) is

min(A + 0.10 * B, 10)

% end-of-term-examination/test:	0
% of continuous assessment (assigments, laboratory, practicals…):	99

where

- A (in the scale 0-10) is the weighted grade of the midterm exams.
- B (in the scale 0-10) is the degree of participation in the lessons and voluntary activities.

The grade in the extraordinary call is established by an exam.

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

BASIC BIBLIOGRAPHY

- James, G., Witten, D., Hastiee, T. and Tibshirani, R. An Introduction to Statistical Learning with Applications in R, Springer-Verlag, 2013

ADDITIONAL BIBLIOGRAPHY

- Kuhn, M. and Johnson, K. Applied Predictive Modeling, Springer, 2013

- Li, Q. and Racine, J. S. Nonparametric Econometrics, Princeton University Press, 2007

- Peña, D. Regresión y Diseño de Experimentos, Alianza Editorial, 2002

- Wand, M. P. and Jones, M. C. Kernel Smoothing, Chapman & Hall, 1995

- Wasserman, L. All of Nonparametric Statistics, Springer-Verlag, 2006

- Wasserman, L. All of Statistics, Springer-Verlag, 2004

- Wood, S. N. Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC, 2006