

Curso Académico: (2018 / 2019)

Fecha de revisión: 24-04-2018

Departamento asignado a la asignatura: Departamento de Estadística

Coordinador/a: GARCIA PORTUGUES, EDUARDO

Tipo: Obligatoria Créditos ECTS : 6.0

Curso : 2 Cuatrimestre : 2

REQUISITOS (ASIGNATURAS O MATERIAS CUYO CONOCIMIENTO SE PRESUPONE)

Cálculo I y II
Álgebra Lineal
Programación
Probabilidad y Análisis de Datos
Introducción a la Modelización Estadística
Aprendizaje Estadístico

OBJETIVOS

* 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.

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

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 (simple and multiple) linear models, gives an overview on their extensions, and treats more superficially regression trees. 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 What is predictive modeling?
- 1.3 Review on statistical inference
- 1.4 Review on probability

- 1.5 Software
- 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 Use of qualitative predictors
 - 3.7 Model diagnostics and multicollinearity
- 4. Linear regression extensions
 - 4.1 Dimension reduction techniques
 - 4.2 Regularization
 - 4.3 Handling nonlinear relationships
 - 4.4 Regression splines
 - 4.5 Local linear regression
 - 4.6 Logistic regression
- 5. Regression trees
 - 5.1 Decision trees
 - 5.2 Bagging
 - 5.3 Random forest
 - 5.4 Boosting

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

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

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.

SISTEMA DE EVALUACIÓN

Continuous evaluation is done by means of:

- A) two midterms;
- B) one group project with presentation;
- C) participation in lessons.

The grade (in the scale 0-10) in the ordinary call is

$$\min(0.65 * A + 0.30 * B + 0.05 * C, 10)$$

where

- A (in the scale 0-10) is the weighted grade of two midterms.
- B (in the scale 0-12) is the group project and presentation grade. The grade is the same for all the members in the group.
- C (in the scale 0-10) is the degree of participation in the lessons.

The grade in the extraordinary call is established by an exam that will consist of theoretical questions and practical problems.

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

Peso porcentual del Examen Final:

0

Peso porcentual del resto de la evaluación:

100

BIBLIOGRAFÍA BÁSICA

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

BIBLIOGRAFÍA COMPLEMENTARIA

- Kuhn, M. and Johnson, K. Applied Predictive Modeling, Springer, 2013
- Peña, D. Regresión y Diseño de Experimentos, Alianza Editorial, 2002
- Wood, S. N. Generalized Additive Models: An Introduction with R, Chapman & Hall/CRC, 2006