COURSE SYLLABUS
2016-2017

GENERAL INFORMATION

Course information	
Name	Intelligent Data Analysis
Code	MRE-513
Degree	Official Master's Degree in Research in Engineering Systems Modeling (MRE)
Year	All
Semester	1 (Fall)
ECTS credits	3 ECTS
Type	Compulsory
Department	Industrial Organization
Area	Statistics and Operations Research
Coordinator	Eugenio Sánchez Úbeda

Instructor	
Name	Eugenio Sánchez Úbeda
Department	Industrial Organization
Area	Statistics and Operations Research
Office	SM26.D-304
e-mail	eugenio.sanchez@comillas.edu
Phone	91.542.28.00 – Ext. 2706
Office hours	Arrange an appointment through email

Instructor	
Name	Carlos Maté
Department	Industrial Organization
Area	Statistics and Operations Research
Office	AA23.D-404
e-mail	cmate@comillas.edu
Phone	91.542.28.00 – Ext. 2430
Office hours	Arrange an appointment through email

DETAILED INFORMATION

Contextualization of the course

Contribution to the professional profile of the degree

This course will contribute to the professional profile of the student by his/her training in research practices based on the use of advanced statistical techniques. The student will learn and practice during the course the use of multivariate analysis and data mining techniques in order to extract knowledge from raw data. In addition, the student will learn to formulate real problems as statistical learning problems for their solution using previous techniques. All these tools will improve the research potential of the student for facing problems where the analysis of data is a basic requirement.

Prerequisites

CONTENTS

Contents

Chapter 1: Introduction to data analysis

1.1 Types of problems.
1.2 General concepts.
1.3 Bayesian data analysis.

Chapter 2: Multivariate analysis techniques.

2.1 Methods for analyzing dependencies: analysis of variance.
2.2 Methods for analyzing dependencies: regression analysis.
2.3 Methods to reduce information: principal component analysis.
2.4 Methods to reduce information: factor analysis.
2.5 Methods for analyzing interdependencies: canonical correlation analysis

Chapter 3: Data mining techniques

3.1 Clustering techniques.
3.2 Classification and regression trees.
3.3 Neural Networks.

Chapter 4: Advanced techniques.

4.1 Structural equation models.
4.2 Symbolic data analysis.
4.3 Soft-computing techniques.
4.4 Orthogonal Projection based models.

Competences and Learning Outcomes

Competences

Basic and General Competences

CB2. Ability to apply and integrate their knowledge, their understanding of such, in addition to their scientific reasoning and problem-solving abilities in new and undefined environments, including multidisciplinary contexts involving highly specialized researchers and practitioners.

Specific Competences

CE1. Understanding and applying the different approaches currently used in data analysis, including both statistical and data mining techniques.

CE7. Knowing how to use the essential tools to cover a research topic.

Learning outcomes

By the end of the course students should be able to:

RA1. To know and understand the theoretical principles of the different multivariate analysis techniques.
RA2. To know and understand the theoretical principles of the different data mining techniques.
RA3. To select the most appropriate data analysis technique for a given problem.
RA4. To know how to apply multivariate analysis methods in real problems.
RA5. To know how to apply data mining techniques in real problems.

TEACHING METHODOLOGY

General methodological aspects	
The teaching method is focused on easing the learning of knowledge and increasing the student critical thinking on data analysis methods. Teaching objectives require the active participation of the student. In addition, the in-class activity should be complemented by the individual student work performed out of class. Both aspects are taken into account in the evaluation method.	
In-class activities	**Competences**
▪ **Lectures (28 hours)**: The lecturer will introduce the fundamental concepts of each chapter, along with some practical recommendations, and will go through worked examples to support the explanation. Active participation will be encouraged by raising open questions to foster discussion and by proposing short application exercises to be solved in class using a software package.	CB2, CE1, CE7
▪ **Tutoring** for groups or individual students will be organized upon request.	CB2, CE1, CE7
Out-of-class activities	**Competences**
▪ Personal study of the course material (30 hours).	CB2, CE1, CE7
▪ Final project. Each group of students proposes a real problem that allows the application of a minimum set of techniques or, alternatively, going deeper in a particular technique. The results will be included in a final report (30 hours).	CB2, CE1, CE7

ASSESSMENT AND GRADING CRITERIA

Assessment activities	Grading criteria	Weight
Final exam	▪ Understanding of the theoretical concepts. ▪ Application of these concepts to problem-solving. ▪ Analysis and interpretation of results.	30%
Final project	▪ Understanding of the theoretical concepts. ▪ Application of theoretical concepts to problem-solving. ▪ Analysis and interpretation of the obtained results. ▪ Comparison of results using different techniques ▪ Writing presentation.	70%

GRADING AND COURSE RULES

Grading

Regular assessment

In order to pass the course, the mark obtained must be greater or equal to 5 out of 10 points.

The student can use until two periods of final evaluation during one academic year. The regular assessment will be carried out at the end of the course (end of the semester).

In case that the course was not passed, the student has another opportunity of final assessment at the end of the academic year. The dates for these evaluation periods will be announced in the web page.

The final grade of the course will be obtained by addition of the following criteria for any evaluation period:

- 30% final exam.
- 70% final project.

Retakes

See previous section

Course rules

- Class attendance is mandatory according to Article 93 of the General Regulations (Reglamento General) of Comillas Pontifical University and Article 6 of the Academic Rules (Normas Academicas) of the ICAI School of Engineering. Not complying with this requirement may have the following consequences:
  - Students who fail to attend more than 15% of the lectures may be denied the right to take the final exam during the regular assessment period.

Students who commit an irregularity in any graded activity will receive a mark of zero in the activity and disciplinary procedure will follow (cf. Article 168 of the General Regulations (Reglamento General) of Comillas Pontifical University).

WORK PLAN AND SCHEDULE[1]

In and out-of-class activities	Date/Periodicity	Deadline
Review and self-study of the concepts covered in the lectures	After each session	–
Final project proposal	7th-8th week	8th week
Final project	During the last third of the course	Last week
Final exam preparation	Last weeks	–
Final exam	Last week	–

STUDENT WORK-TIME SUMMARY		
IN-CLASS HOURS		
Lectures	Final exam	
28	2	
OUT-OF-CLASS HOURS		
Self-study	Final project	
30	30	
	ECTS credits:	3 (90 hours)

[1] A detailed work plan of the subject can be found in the course summary sheet (see following page). Nevertheless, this schedule is tentative and may vary to accommodate the rhythm of the class.

BIBLIOGRAPHY

Basic bibliography

- Notes prepared by the instructors (available in Moodle).

Complementary bibliography

MULTIVARIATE ANALYSIS AND BAYESIAN DATA ANALYSIS

- **Afifi, A. A.; Clark, V.A.; May, S. (2011).** *Practical Multivariate Analysis*. 5th Ed. Chapman & Hall, CRC Press
- **Cho, M., and Martinez, W. L. (2014).** *Statistics in MATLAB: A Primer*. Chapman & Hall/CRC Computer Science & Data Analysis.
- **Cuadras, C. M. (2014).** *Nuevos Métodos de Análisis Multivariante*. CMC Ediciones. Barcelona. (http://www.ub.edu/stat/personal/cuadras/metodos.pdf)
- **Everitt, B., and Hothorn, T. (2011).** *An Introduction to Applied Multivariate Analysis with R*. Springer.
- **Field, A. (2009).** *Discovering Statistics using SPSS*. 3rd Ed. Sage Publications Ltd.
- **Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2013).** *Bayesian Data Analysis (Third ed)*. Chapman & Hall: London.
- **Graffelman, J. (2011).** *Multivariate Analysis with Matlab and R.* Chapman & Hall, CRC Press.
- **Liu, J.S. (2001).** *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag: New York.
- **Hair, J. F.; Black, B.; Babin, B.; Anderson, R. E.; Tatham, R. L. (2009).** *Mulltivariate Data Analysis*. 7th Ed. Prentice-Hall.
- **Harlow, L. L. (2005).** *The Essence of Multivariate Thinking*. Lawrence Erlbaum Associates Pub. New Jersey.
- **Haslwanter, T. (2016).** *An Introduction to Statistics with Python.* With Applications in the Life Sciences. Springer.
- **Husson, F.; Le, S. and Pages, J. (2010).** *Exploratory Multivariate Analysis by Example Using R*. Chapman & Hall, CRC.
- **Maté, C. (1995).** *Curso General sobre Statgraphics. Procedimientos. Métodos Estadísticos. Aplicaciones. Ejercicios Resueltos*. Editorial Universidad Pontificia Comillas. Madrid.
- **Maté, C. (2014).** "*Big data. Un nuevo paradigma de análisis de datos*", ***Anales de Mecánica y Electricidad*.** vol. XCI, no. VI, pp. 10-16.
- **Maté, C. & Calderón, R. (2000).** *Exploring the characteristics of rotating electric machines with factor analysis*, **Journal of Applied Statistics**, Vol. 27, Nº 8, pp. 991-1006.
- **Peña, D. (2002).** *Análisis de Datos Multivariantes*. Editorial McGraw-Hill Interamericana de España.
- **Rencher, A. C. (2002).** *Methods of Multivariate Analysis*. Second edition. John Wiley & Sons. New York.
- **Visauta, B. (2007).** *Análisis Estadístico con SPSS 14 + CD. Estadística Básica. 2ª Edición.* McGraw-Hill. Madrid.

DATA MINING:

- T. J. Hastie and R. J. Tibshirani and J. Friedman (2009), The Elements of Statistical Learning, Springer-Verlag (also on-line).
- T. J. Hastie and R. J. Tibshirani (1990), Generalized Additive Models, Chapman and Hall, London.
- V. Cherkassky and F. Mulier (1998), Learning from data: concepts, theory, and methods, John Wiley and Sons, New York.
- C. M. Bishop (1995), Neural Networks for Pattern Recognition, Oxford University Press , New York.
- K. M. Bossley (1997), Neurofuzzy Modelling Approaches in System Identification, Ph.D. thesis, University of Southampton, Faculty of Engineering and Applied Science. See http://www.isis.ecs.-soton.ac.uk/publications.
- S. Haykin (1994), Neural Networks. A comprehensive Foundation, Macmillan College Publishing Company Inc, IEEE Press.
- E. F. Sánchez-Úbeda (1999), Models for data analysis: contributions to automatic learning, Ph.D. thesis, Universidad Pontificia Comillas, no. 255/1999.

	In-class activities			Out-of-class activities			Learning outcomes
Week	Time [h]	Lecture & Problem-solving	Assessment	Time [h]	Self-study	Other activities	Code
1	2	Course presentation (1h) Introduction to data analysis (1h)		0			RA1, RA2, RA3, RA4, RA5
2	2	Basics of data mining (2h)		2	Review and self-study (2h)		RA2
3	2	Clustering techniques (2h)		2	Review and self-study (2h)		RA2
4	2	Clustering techniques (1h), Classification and regression trees (1h)		2	Review and self-study (2h)		RA2
5	2	Classification and regression trees (2h)		2	Review and self-study (2h)		RA2
6	2	Neural networks (2h)		2	Review and self-study (2h)		RA2
7	2	Neural networks (1h), Advanced techniques in data mining (1h)		3	Review and self-study (2h)	Final project proposal (1h)	RA2
8	2	Analysis of variance (2h)		3	Review and self-study (2h)	Final project proposal (1h)	RA1
9	2	Regression analysis (2h)		2	Review and self-study (2h)		RA1
10	2	Principal component analysis (2h)		7	Review and self-study (2h)	Final project development (5h)	RA1
11	2	Factor analysis (2h)		7	Review and self-study (2h)	Final project development (5h)	RA1
12	2	Canonical correlation analysis (2h)		7	Review and self-study (2h)	Final project development (5h)	RA1
13	2	Bayesian data analysis (2h)		7	Final exam preparation (2h)	Final project development (5h)	RA1
14	2	Advanced techniques in multivariate analysis (2h)		10	Final exam preparation (2h)	Final project development (8h)	RA1
15	2		Final exam (2h)	4	Final exam preparation (4h)	Final project submission	RA1, RA2, RA3, RA4, RA5