

## SUBJECT TEACHING GUIDE

M1966 - Data Mining

University Master's Degree in Data Science

Academic year 2019-2020

1. IDENTIFYING DATA					
Degree	University Master's Degree in Data Science			Type and Year	Compulsory. Year 1
Faculty	Faculty of Sciences				
Discipline	METHODS IN SCIENCE DATA				
Course unit title and code	M1966 - Data Mining				
Number of ECTS credits allocated	6	Term	Semester based (1)		
Web					
Language of instruction	Spanish	English Friendly	Yes	Mode of delivery	Face-to-face

Department	DPTO. MATEMATICA APLICADA Y CIENCIAS DE LA COMPUTACION				
Name of lecturer	SIXTO HERRERA GARCIA				
E-mail	sixto.herrera@unican.es				
Office	E.T.S. de Ingenieros de Caminos, Canales y Puertos. Planta: + 1. DESPACHO PROFESORES (1042)				
Other lecturers	STEVEN JOHAN MARIA VAN VAERENBERGH JOAQUIN BEDIA JIMENEZ RODRIGO GARCIA MANZANAS MAIALEN ITURBIDE MARTINEZ DE ALBENIZ				

### 3.1 LEARNING OUTCOMES

- To understand and correctly apply the methods commonly used to statistically describe the datasets .
- To understand and apply the linear regression model .
- To learn the regularization concept .
- To recognize the data mining problems (association, segmentation, classification and prediction) and know what method should be apply in a particular case .
- To understand the different methodologies of machine learning (case-based, supervised and non supervised).
- To understand and know how to use the typical data mining techniques .
- To know and understand the generalization concept and the over-fitting problem .
- To properly validate the learned models and identify the better model in a specific case .

### 4. OBJECTIVES

To gain knowledge on the statistical methods for data description , presentation and analysis. To learn how to apply the machine learning techniques to big data problems.

### 6. COURSE ORGANIZATION

CONTENTS	
1	Introduction and historical perspective Canonical problems and Learning Paradigms Non Supervised Learning with Association Rules Supervised Learning with K-Nearest Neighbors Cross-validation, learning and over-fitting
2	Regularization and features selection Dimension reduction with linear and non-linear techniques
3	Classification and regression trees Ensemble methods: Bagging and Random Forests
4	Non-supervised learning. Segmentation.

### 7. ASSESSMENT METHODS AND CRITERIA

Description	Type	Final Eval.	Reassessn	%
Evaluation of the reports of practices.	Activity evaluation with Virtual Media	Yes	Yes	60,00
Exam covering the concepts explained in the subject.	Activity evaluation with Virtual Media	Yes	Yes	40,00
<b>TOTAL</b>				<b>100,00</b>
Observations				
To recover the course, each practice failed should be repeated.				
Observations for part-time students				
Same evaluation will be applied for both full and partial time students				

## 8. BIBLIOGRAPHY AND TEACHING MATERIALS

### BASIC

An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Editorial Springer. Available at <http://www-bcf.usc.edu/~gareth/ISL/>

Dimension Reduction: A Guided Tour, Christopher J. C. Burges, Foundations and Trends® in Machine Learning: Vol. 2: No. 4, pp 275-365. <http://dx.doi.org/10.1561/22000000002>

Machine learning: a probabilistic perspective, Kevin P. Murphy. ISBN 978-0-262-01802-9, MIT Press Cambridge, Massachusetts, London, England