VNIVERSITAT ĐE VALÈNCIA

COURSE DATA

Data Subject	
Code	44650
Name	Statistics and optimisation
Cycle	Master's degree
ECTS Credits	6.0
Academic year	2022 - 2023

Study (s)			
Degree	**Center**	**Acad. year**	**Period**
2221 - Master's Degree in Data Science	School of Engineering	1	First term

Subject-matter		
Degree	**Subject-matter**	**Character**
2221 - Master's Degree in Data Science	2 - Statistics and optimisation	Obligatory

Coordination	
Name	**Department**
MARTINEZ GAVARA, ANNA	130 - Statistics and Operational Research

SUMMARY

This course is divided into two main parts. The Statistics part covers the basic statistical concepts and techniques that are useful for data scientists. First the basic Probability Theory, then the main methods of parameter estimation and hypothesis testing, and finally the time series analysis, including the ARIMA models.

The Optimization part covers the basic concepts of unconstrained and constrained optimization and from them develops optimization algorithms (like algorithm EM), including local search and global search methods (genetic algorithms).

PREVIOUS KNOWLEDGE

VNIVERSITAT ĐĐVALÈNCIA

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

COMPETENCES (RD 1393/2007) // LEARNING OUTCOMES (RD 822/2021)

2221 - Master's Degree in Data Science

- Students should demonstrate self-directed learning skills for continued academic growth.

- Be able to assess the need to complete their technical, scientific, language, computer, literary, ethical, social and human education, and to organise their own learning with a high degree of autonomy.

- Ser capaces de acceder a herramientas de información (bibliográficas y de empleo) y utilizarlas apropiadamente.

- Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio, aplicando los conocimientos adquiridos en la identificación de salidas profesionales y yacimientos de empleo.

- Extraer conocimiento de conjuntos de datos en diferentes formatos.

- Entender la utilidad de la ciencia de datos y sus elementos asociados, así como su aplicación en la resolución de problemas, eligiendo las técnicas más adecuadas a cada problema, aplicando de forma correcta las técnicas de evaluación y, finalmente,
?interpretando los modelos y resultados.

LEARNING OUTCOMES (RD 1393/2007) // NO CONTENT (RD 822/2021)

To know the mathematical bases that give support to the different techniques of data analysis.
To give the students the basic knowledge for them to be able to develop new algorithms/techniques of data analysis.
To be able of processing data in which there is a time dependency (time series analysis)
To know how to use the tests of hypothesis in data analysis.
To apply local search and global search methods (including genetic algorithms).

DESCRIPTION OF CONTENTS

1. Data description and Time series.

Bayes Theorem and its implications in data analysis

---

2. Statistical Inference.

Probability. Random Sample. Likelihood function.
Point and interval estimation. Maximum Likelihood Method.
Hypothesis testing: Statistical hypothesis. Critical and acceptance region. Contrast statistic, decision rule, Type I and II errors. Significance level, power test function- p-value.

3. Statistical analysis of one or more populations.

One-sample statistical analysis. Tests for one proportion. Tests for one population mean.
Two-sample analysis and comparison (continuous data). Paired and independent two sample tests. Non parametrical alternatives.
Categorical data analysis. Contingency tables.

4. Unconstrained optimization

Basic concepts of unconstrained optimization. Local search methods. EM algorithm.

5. Constrained Optimization

Basic concepts of constrained optimization. Linear and integer programming. Local search methods

6. Global search methods. Metaheuristic algorithms.

Global search methods. Metaheuristic algorithms. Genetic algorithms.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theoretical and practical classes	60,00	100
Development of individual work	20,00	0
Study and independent work	12,00	0
Readings supplementary material	3,00	0
Preparation of evaluation activities	12,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	13,00	0
Resolution of case studies	10,00	0
TOTAL	150,00	

VNIVERSITAT ĠID VALÈNCIA

TEACHING METHODOLOGY

The course will combine the theoretical and the practical part, without separating sessions devoted to theory from those devoted to practice. The lessons will be taught in a computer equipped classroom.

In the theoretical part of the classes, the teacher will introduce the concepts and methods Statistics and Optimization, with examples and exercises to be solved by the students.

The practical sessions will be synchronized with the theory. In these sessions, the students will learn by solving problems, exercises and case studies, in order to acquire the skills of this course.

EVALUATION

The educational evaluation of knowledge and skills achieved by the students will be made continuously throughout the course, and will consist in the following blocks of evaluation:

1. Exercises and the class work submitted during the course and / or partial exams: 60% of the final grade.

2. Individual tests: 40% of the final grade.

The activities in part 1 are considered non-recoverable, that is, they cannot be evaluated by an exam. The marks will be kept for the whole academic year.

REFERENCES

Basic

- Barber, D., (2012). Bayesian Reasoning and Machine Learning. Ed. Cambridge University Press

- Peña, D. , (2001). Fundamentos de Estadística. Ed. Alianza Editorial.

- Shumway, R.H., Stoffer, D. S., (2011). Time Series Analysis and Its Applications, Ed. Springer.

- Walpole, R. E., Myers, R.H. y Myers, S.L. (1999). Probabilidad y Estadística para Ingenieros. Ed. Prentice Hall.

-   Lange, K. (2004) Optimization. Ed. Springer.

- Li, L. (2015) Selected Applications of Convex Optimization. Ed. Springer

- Nocedal, J; Wright, S. (2006) Numerical Optimization. Ed. Springer, 2nd edition

- Snyman, J.A. (2005) Practical Mathematical Optimization. Ed. Springer

- James G.; Witten D.; Hastie T.; Tibshirani R. (2017). An introduction to Statistical Learning with Applications in R. Springer.

- Bruce, P.; Bruce, A. (2017). Practical Statistics for Data Scientists. OReilly Media, Inc.

Additional

- John Braun Duncan, W., Murdoch, J. (2007). A first course in statistical programming with R. Cambridge University Press.

- Venables, V.N. (2013). An Introduction to R. http://cran.r-project.org.

- Bertsekas, D.P. (2016) Nonlinear Programming. Ed. Athena Scientific, 3rd Edition