COURSE DATA

DATA SUBJECT

Code: 46570
Name: Statistics and optimisation
Cycle: Master's Degree
ECTS Credits: 6
Academic year: 2025-26

STUDY (S)

Degree	Center	Acad. year	Period
2262 - Master's Degree in Data Science	Escola Tècnica Superior d'Enginyeria	1	First quarter

SUBJECT-MATTER

Degree	Subject-matter	Character
2262 - Master's Degree in Data Science	Statistics and optimisation	COMPULSORY

COORDINATION

MARTINEZ GAVARA ANNA

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

RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

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

OTHER REQUIREMENTS

COMPETENCES / LEARNING OUTCOMES

-

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.

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

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

PRESENCIAL ACTIVITIES

Activity	Hours
Theory	38,00
Theoretical and practical classes	4,00
Laboratory	18,00
Total hours	**60,00**

NON PRESENCIAL ACTIVITIES

Activity	Hours
Attendance at other activities	0,00
Individual or group project	20,00
Independent study and work	15,00
Preparation of lessons	33,00
Preparation for assessment activities	12,00
Resolution of case studies	10,00
Total hours	**90,00**

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

- 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.
- 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