ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000607 - Intelligent Data Analysis

DEGREE PROGRAMME

10AN - Master Universitario En Ingenieria Informatica

ACADEMIC YEAR & SEMESTER

2022/23 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	103000607 - Intelligent Data Analysis
No of credits	4.5 ECTS
Type	Compulsory
Academic year ot the programme	First year
Semester of tuition	Semester 1
Tuition period	September-January
Tuition languages	English
Degree programme	10AN - Master Universitario en Ingenieria Informatica
Centre	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Academic year	2022-23

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Jacinto Gonzalez Pachon	2105	jacinto.gonzalez.pachon@upm.es	Sin horario. Check office hours in September
Arminda Moreno Diaz (Subject coordinator)	2112	arminda.moreno@upm.es	Sin horario. Check office hours in September

Juan Antonio Fdez Del Pozo De Salamanca	2105	juan.fdezpozo.salamanca@upm.es	Sin horario. Check office hours in September
Antonio Jimenez Martin	2110	antonio.jimenez@upm.es	Sin horario. Check office hours in September

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- Basic knowledge of statistics and inference will be helpful.

4. Skills and learning outcomes *

4.1. Skills to be learned

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CG19 - Capacidad para el modelado matemático, cálculo y simulación en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación, desarrollo e innovación en todos los ámbitos relacionados con la Ingeniería en Informática

4.2. Learning outcomes

RA44 - Formular, analizar y validar modelos de regresión, análisis discriminante, clasificación y "clustering".

RA43 - Conocer y aplicar técnicas de reducción de la dimensionalidad en un conjunto de datos multivariantes.

RA42 - Conocer y aplicar las principales técnicas de análisis de datos multivariantes.

RA46 - Ser capaz de estructurar problemas de toma de decisiones bajo el paradigma bayesiano.

RA45 - Aplicar la metodología apropiada para el ajuste de series temporales.

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

The course is intended to be a non-exhaustive survey of techinques to convert multivariate data into useful information so that good decisions can be made. The perspective is twofold, theorical and applied, covering topics such as: exploratory data analysis, statistical summaries and graphical representations, dimensionality reduction, regression techniques, time series analysis, decision theory and probabilistic graphical models. There will be an emphasis on hands-on application of the theory and methods throughout, with extensive use of R.

5.2. Syllabus

1. Exploratory data analysis and descriptive statistics.
    1.1. Aspects of multivariate data. Exploratory Data Analysis.
    1.2. Aspects of multivariate data. Descriptive statistics. Introduction to R.
    1.3. Dimensionality reduction: Principal Component Analysis and biplots.
2. Statistical Modelling
    2.1. Simple Linear Regression Model
    2.2. Multiple Linear Regression Model
    2.3. The General Linear Model
3. Time Series.

3.1. Definitions, Applications and Techniques.

3.2. Stationarity and Seasonality.

3.3. Common approaches.

3.4. Box-Jenkins model identification, estimation and validation.

3.5. Forecasting.

4. Introduction to Decision Analysis.

4.1. Structure and representation of a decision problem.

4.2. Decision making under certainty and uncertainty.

4.3. Preferences and beliefs modelling.

4.4. Collective decision making.

6. Schedule

6.1. Subject schedule*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	**Module 1: Lecture** Duration: 02:00 Lecture			
2	**Module 1: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 1.** Duration: 01:00 Additional activities		
3		**Computer Lab. Module 1.** Duration: 02:00 Additional activities		
4		**Computer Lab. Module 1.** Duration: 02:00 Additional activities		
5	**Module 2: Lecture** Duration: 02:00 Lecture			**Upload Assignment 1** Group work Continuous assessment Not Presential Duration: 00:00
6	**Module 2: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 2.** Duration: 01:00 Additional activities		
7	**Module 2: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 2.** Duration: 01:00 Additional activities		
8	**Module 2: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 2.** Duration: 01:00 Additional activities		**Upload Assignment 2** Group work Continuous assessment Not Presential Duration: 00:00
9	**Module 3: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 3.** Duration: 01:00 Additional activities		
10	**Module 3: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 3.** Duration: 01:00 Additional activities		
11	**Module 3: Lecture** Duration: 02:00 Lecture			
12	**Module 3: Lecture** Duration: 02:00 Lecture			**Upload Assignment 3** Group work Continuous assessment Not Presential Duration: 00:00

13	**Module 4: Lecture.** Duration: 02:00 Lecture			
14	**Module 4: Lecture** Duration: 02:00 Lecture			
15	**Module 4: Lecture** Duration: 02:00 Lecture			
16				**Upload/Presentation Assignment 4.** Individual work Continuous assessment Presential Duration: 03:00
17				**Global exam: upload all 4 assignments** Other assessment Final examination Not Presential Duration: 03:00

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
5	Upload Assignment 1	Group work	No Presential	00:00	25%	3 / 10	CG19 CB7
8	Upload Assignment 2	Group work	No Presential	00:00	25%	3 / 10	CB10 CG19 CB7
12	Upload Assignment 3	Group work	No Presential	00:00	25%	3 / 10	CG19 CB7
16	Upload/Presentation Assignment 4.	Individual work	Face-to-face	03:00	25%	3 / 10	CB10

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Global exam: upload all 4 assignments	Other assessment	No Presential	03:00	100%	3 / 10	CB10 CG19 CB7

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Extraordinary exam: upload all 4 assignments	Other assessment	Face-to-face	00:00	100%	3 / 10	CG19 CB7 CB10

7.2. Assessment criteria

Progressive evaluation:

The course grade will be based on the performance on four assignments (25% each one). These assignments will consist on applying the methods and techniques studied in class to different data sets. Each assignment should obtain a grade equal or greater than 3 (0-10 scale) for the final grade to be computed. This final grade must be equal or greater than 5 to pass the course. Eventually, students may be asked to present orally the conclusions of their work.

If any assignments' grade doesn't meet the requirments exposed above, the homework can be improved and presented to be graded again in the January exams session. Again, grade requirements apply to compute the course overall grade.

January Exam:

If an assignment was not deliver when requested, it can be presented to be graded in the January exams session. If this is the case, the work submitted will only be corrected once and there will be no further opportunity to improve it. Again, grade requirements apply.

Extraordinary exam:

In case you are opting for the extraordinary exam in July's session, you have to warn the instructor(s) in advance (at least two weeks before the exam, by email, but the sooner the better). In this case, all 4 assignments are to be delivered. Grade requirements apply to compute the overall grade.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Moodle	Web resource	Main communication channel with students. Repository: slides, scripts, data sets, other resources.
Computer Lab	Equipment	Computer Room for hands-on sessions.
Johnson, R.A., Whichern, D.W. (2007) Applied Multivariate Statistical Analysis. Pearson Education.	Bibliography	Friendly exposition of the most important multivariate techiques, including clustering. They also introduce other Artificial Intelligence techniques like neural networks.
Rencher, A.C. Methods of Multivariate Analysis.	Bibliography	Clear exposition of Multivariate Analysis Techniques, from a statistical point of view. Many examples.
Everitt, B.S. and Dunn G. (1997) Applied Multivariate Data Analysis. Arnold.	Bibliography	Excellent exposition of multivariate techniques. They make the Generalised Linear Model easily understandable.
Hair, J.F., Black, W.C., Babin, B.J., Anderson R.E. Multivariate Data Analysis.	Bibliography	A Global Perspective on multivariate Techniques. Very detailed examples. In almost every topic, a "Rules of Thumb" section summarizes the relevant facts.
Multivariate Analysis of Ecological Data. Greenacre, Primicerio. Fundación BBVA.	Web resource	http://www.fbbva.es/TLFU/tlfu/esp/publicaciones/libros/fichalibro/index.jsp?codigo=769 Excellent compilation and description of multivariate techniques applied to Ecological Data. Clear, educational, didactic explanations. Lots of examples

Biplots in Practice. Michael Greenacre. Fundación BBVA.	Web resource	http://www.multivariatestatistics.org/biplots.html Excellent Monograph on Biplots.
Rawlings, J.O., Pantula, S.G., Dickey, D.A. Applied Regression Analysis.	Bibliography	Almost everything about Regression Models.
Chatfield, C. (2003) The Analysis of Time Series: An Introduction. Chapman and Hall.	Bibliography	
Forecasting: Principles and Practice. R. J. Hyndman and G. Athanasopoulos. Monash University, Australia.	Web resource	A comprehensive introduction to time series analysis and other forecasting methods. Available online at: https://otexts.com/fpp2/
Statistical forecasting: notes on regression and time series analysis. Robert Nau. Fuqua School of Business. Duke University.	Web resource	Available on line at: https://people.duke.edu/~rnau/411home.htm
French, S. Decision Theory. Ellis Horwood Ltd.	Bibliography	

9. Other information

9.1. Other information about the subject

Regarding the Sustainable Developing Goals, broadly speaking, proper data analysis leads to an understanding of processes and the generation of information available to everyone. This information can help clarify and make more transparent the way in wich institutions, for instance, try to meet people's needs. This information icreases transparency, accountability and citizen participation in institutions and facilitates information-sharing between different actors. In this particular case, proper analysis of data might help in achieving SDG 16: Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.