ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000660 - Intelligent data analysis

DEGREE PROGRAMME

10AP - Eit Digital Master's Programme In Data Science

ACADEMIC YEAR & SEMESTER

2018/19 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	103000660 - 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	10AP - Eit digital master's programme in data science
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2018-19

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

El plan de estudios Eit Digital Master's Programme In Data Science no tiene definidas asignaturas previas recomendadas para esta asignatura.

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

CB07 - 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

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.

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

RA68 - To perform a time series analysis using the proper statistical methodology.

RA67 - To build, interpret and conduct diagnostics analysis of regression models.

RA69 - To apply the expected utility paradigm to solve decision problems.

RA66 - To use multivariate data representation and dimensionality reduction techniques.

RA71 - To build, estimate and interpret probabilistic graphical models.

* 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. Descriptive statistics and statistical modelling.
    1.1. Aspects of multivariate data. Descriptive statistics. Introduction to R.
    1.2. Dimensionality reduction: Principal Component Analysis and biplots.
    1.3. Clustering methods.
    1.4. Regression models.
2. Time Series.
    2.1. Definitions, Applications and Techniques.
    2.2. Stationarity and Seasonality.
    2.3. Common approaches.

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

2.5. Forecasting.

3. Introduction to Decision Analysis.

3.1. Structure and representation of a decision problem.

3.2. Decision making under certainty and uncertainty.

3.3. Preferences and beliefs modelling.

3.4. Collective decision making.

4. Graphical Models for Decision Making.

4.1. Decision Trees and Influence Diagrams for optimal decisions.

4.2. Bayesian networks for diagnosis and prognosis.

4.3. Sensitivity Analysis for explanation of reasoning.

6. Schedule

6.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1	**Module 1: Lecture** Duration: 02:00 Lecture			**Individual Study** Other assessment Continuous assessment Duration: 02:00
2	**Module 1: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 1.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00
3		**Computer Lab. Module 1.** Duration: 02:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 1** Other assessment Continuous assessment Duration: 02:00
4		**Computer Lab. Module 1.** Duration: 02:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 1** Other assessment Continuous assessment Duration: 02:30
5		**Computer Lab. Module 1.** Duration: 02:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 2** Other assessment Continuous assessment Duration: 04:30 **Upload Homework 1** Group work Continuous assessment Duration: 00:00
6	**Module 1: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 1.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 2** Other assessment Continuous assessment Duration: 04:30

7	**Module 1: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 1.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 2** Other assessment Continuous assessment Duration: 04:00
8	**Module 1: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 1.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Upload Homework 2** Group work Continuous assessment Duration: 00:00
9	**Module 2: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 2.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 3** Other assessment Continuous assessment Duration: 04:30
10	**Module 2: Lecture** Duration: 01:00 Lecture	**Computer Lab. Module 2.** Duration: 01:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 3** Other assessment Continuous assessment Duration: 03:30
11	**Module 3: Lecture** Duration: 02:00 Lecture			**Individual Study** Other assessment Continuous assessment Duration: 02:00
12	**Module 3: Lecture** Duration: 02:00 Lecture			**Upload Homework 3** Group work Continuous assessment Duration: 00:00 **Individual Study** Other assessment Continuous assessment Duration: 02:00
13	**Module 4: Lecture.** Duration: 02:00 Lecture			**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 4** Other assessment Continuous assessment Duration: 04:30

14		**Computer Lab. Module 4.** Duration: 02:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 4** Other assessment Continuous assessment Duration: 04:30
15		**Computer Lab. Module 4.** Duration: 02:00 Additional activities		**Individual Study** Other assessment Continuous assessment Duration: 02:00 **Work on Homework 4** Other assessment Continuous assessment Duration: 04:30
16	**There will be 6 more hours of teaching classes throughout the semester. These hours will be properly announced to the students.** Duration: 06:00 Lecture			**Upload/Presentation Homework 4.** Group work Continuous assessment Duration: 03:00
17				**Final Exam** Group presentation Final examination Duration: 03:00

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

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 subject schedule is based on a previous theorical planning of the subject plan and might go to through experience some unexpected changes along throughout the academic year.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
1	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
2	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
3	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
3	Work on Homework 1	Other assessment	No Presential	02:00	%	0 / 10	
4	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
4	Work on Homework 1	Other assessment	No Presential	02:30	%	0 / 10	
5	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
5	Work on Homework 2	Other assessment	No Presential	04:30	%	0 / 10	
5	Upload Homework 1	Group work	No Presential	00:00	25%	3 / 10	CB07 CB10
6	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
6	Work on Homework 2	Other assessment	No Presential	04:30	%	0 / 10	
7	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
7	Work on Homework 2	Other assessment	No Presential	04:00	%	0 / 10	
8	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
8	Upload Homework 2	Group work	No Presential	00:00	25%	3 / 10	CB07 CB10 CG19

9	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
9	Work on Homework 3	Other assessment	No Presential	04:30	%	0 / 10	
10	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
10	Work on Homework 3	Other assessment	No Presential	03:30	%	0 / 10	
11	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
12	Upload Homework 3	Group work	No Presential	00:00	25%	3 / 10	CB07 CB10 CG19
12	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
13	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
13	Work on Homework 4	Other assessment	No Presential	04:30	%	0 / 10	
14	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
14	Work on Homework 4	Other assessment	No Presential	04:30	%	0 / 10	
15	Individual Study	Other assessment	No Presential	02:00	%	0 / 10	
15	Work on Homework 4	Other assessment	No Presential	04:30	%	0 / 10	
16	Upload/Presentation Homework 4.	Group work	Face-to-face	03:00	25%	3 / 10	CB07 CB10 CG19

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Final Exam	Group presentation	Face-to-face	03:00	100%	3 / 10	CB07 CB10 CG19

7.1.3. Referred (re-sit) examination

No se ha definido la evaluación extraordinaria.

7.2. Assessment criteria

The course grade will be based on the performance on four required homework sets (25% each one). These homework sets will consist on applying the methods and techniques studied in class to different data sets. Each homework 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 (or many) homework grade doesn't meet the requirments exposed above, the homework can be improved and presented to be graded again in the Final Exam. Once more, grade requirements apply.

If a homework (or homeworks) was not delivery when requested, it can be presented to be graded in the Final Exam. Again, grade requirements apply.

In case you don't want to follow the continuous evaluation process and you want to opt for the final exam in January's session, you have to warn the instructor(s) in advance (at the end of September, by email).

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).

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 and 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 techniques, 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.
Sharma, S (1996). Applied Multivariate Techniques. Wiley.	Bibliography	
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	
French, S. Decision Theory. Ellis Horwood Ltd.	Bibliography	
Koller,D., Friedman, N (2009) Probabilistic Graphical Models. Principles and techniques. MIT Press.	Bibliography	
Duda, R. Hart P.E., Stork D.G. Pattern Classification (2001). Wiley.	Bibliography	