VNIVERSITAT DE VALÈNCIA

Course Guide
44276 Exploratory data analysis

COURSE DATA

Data Subject	
Code	44276
Name	Exploratory data analysis
Cycle	Master's degree
ECTS Credits	3.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
2199 - M.D. in Electronic Engineering	School of Engineering	1	Second term
3131 - Electronic Engineering	Doctoral School	0	First term

Subject-matter

Degree	Subject-matter	Character
2199 - M.D. in Electronic Engineering	1 - Digital signal processing	Obligatory
3131 - Electronic Engineering	1 - Complementos de Formación	Optional

Coordination

Name	Department
MARTIN GUERRERO, JOSE DAVID	242 - Electronic Engineering
SERRANO LOPEZ, ANTONIO JOSE	242 - Electronic Engineering

SUMMARY

This course's main objective is to describe the most important techniques of Exploratory Data Analysis that can be drawn from knowledge of a problem through statistical analysis of the acquired data.

The course consists of five main topics. The first one deals with basic data analysis, such as probability, statistics, linear algebra and mathematical tools. The second one defines the characteristics of the type of problem to solve and the technique used for this (number of patterns, outliers, missing values, type of learning used, etc.). Later, in the third issue, the techniques are presented to perform early exploration of data to get a brief description of the type of information stored, emphasizing the study of clustering algorithms and extraction techniques and selection of variables (features). The last two topics describe the classification and regression models (linear and nonlinear, respectively) most widely used in Exploratory Data Analysis.

This is a mandatory course, which is taught in the first semester of the Master in Electronic Engineering. The total teaching load is 3 ECTS. The workload for the student is 75 hours over the semester, of which 30 are on-site and 45 are individual work.

Fort the lab sessions, Matlab will be used although the students can make use of other software solutions of their choice, such as Python or R.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

Es recomendable que el alumno tenga una base matemática mínima de Álgebra, Cálculo, Estadística y Probabilidad. Si no la posee, se le facilitará una serie de tutoriales para que se adapte al curso sin problemas.

OUTCOMES

2199 - M.D. in Electronic Engineering

- Students should apply acquired knowledge to solve problems in unfamiliar contexts within their field of study, including multidisciplinary scenarios.
- Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.
- Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.
- Students should demonstrate self-directed learning skills for continued academic growth.
- Take into account the economic and social context in engineering solutions, be aware of diversity and multiculturalism and ensure sustainability and respect for human rights and equality between men and women.
- Diseñar un sistema, componente o proceso que cumpla unas especificaciones desde diferentes puntos de vista: electrónico, económico, social, ético y medioambiental.
- Demostrar una comprensión sistemática de un campo de estudio y el dominio de las habilidades.
- Realizar un análisis crítico, evaluación y síntesis de ideas nuevas y complejas.
- Ser capaz de fomentar, en contextos académicos y profesionales, el avance tecnológico, social o cultural dentro de una sociedad basada en el conocimiento.

- Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la Ingeniería Electrónica y en particular los de tratamiento de la señal, sistemas digitales y de comunicaciones y electrónica industrial.
- Capacidad para el modelado matemático, cálculo y simulación en todos los ámbitos relacionados con la Ingeniería Electrónica y campos multidisciplinares afines. En especial los de tratamiento de la señal, sistemas digitales y de comunicaciones y electrónica industrial.
- Conocer las técnicas avanzadas de análisis de datos.
- Capacidad de analizar, especificar y diseñar sistemas de tratamiento digital de señales desde su concepción hasta su implementación en sistemas hardware de tiempo real..

LEARNING OUTCOMES

After taking this subject, students should have leant to tackle a problem of data analysis from the beginning, being able to filter the useful information from noise, and with capacity to find typical profiles that are observed in the data set.

They should also have sufficient evidence and the ability to establish the advantages and disadvantages of different alternatives in data processing. Students must be able to decide whether linear or nonlinear approaches are more appropriates for a given.

DESCRIPTION OF CONTENTS

1. Basic foundations for exploratory data analysis

UNIT 1.
1.1 Probability (axioms and conditional joint prob.)
1.2 Bayes Theorem. Applications
1.3 Random Variable. Typical distributions; moments
1.4 Statistics. Hypothesis testing
1.5 Review of Linear Algebra
1.6 Optimization of functions

PRACTICE: Application of statistics and hypothesis for the data analysis.

2. Introduction to data analysis

UNIT 2.
2.1 Variables: types. Patterns
2.3 Types of problems to solve
2.4 Learning. Types

3. Descriptive exploratory analysis

UNIT 3.
3.1 Acquisition and data cleansing
3.2 Characterization of the variables (statistics and graphs)
3.3 Data Transformations
3.4 Selection and Feature Extraction
3.5 Algorithms for clustering (HCM, FCM and SOM)

PRACTICE: Data Preprocessing
Consolidation, standardization, removal of outliers, missing data processing, obtaining typical profiles with clustering methods.

4. Linear models

UNIT 4.
4.1 Description of linear models
4.2 Obtaining parameters: normal equations. regularization
4.3 Model validation

PRACTICE: Applying linear models to data processing
Linear least squares regression for modeling functions. Regularization. Logistic regression applied to classification problems. Regression with robust cost functions.

5. Non-linear models

ITEM 5.
5.1 Trees (regression and classification)
5.2 Artificial Neural Networks
5.3 Support Vector Machines
5.4 Extraction Rule

PRACTICE: Application of nonlinear models to data processing (I)
Application of nonlinear models viewed in theory; sets similar to the previous practical session will be used to compare their performance.

PRACTICE: Application of nonlinear models to data processing (II)
Since several nonlinear models are shown, there is a second session to analyze those models not studied in the preious one.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	15,00	100
Laboratory practices	15,00	100
Study and independent work	10,00	0
Readings supplementary material	5,00	0
Preparation of evaluation activities	5,00	0
Preparing lectures	5,00	0
Preparation of practical classes and problem	10,00	0
Resolution of case studies	10,00	0
TOTAL	**75,00**	

TEACHING METHODOLOGY

The teaching methods employed in the development of the course are:

a) Theoretical activities.

Expository development of matter with the student's participation in the resolution of specific issues.

b) Practical activities.

Solving practical problems

c) Student's personal work.

Description: Out-of-classroom development of problems as well as the preparation of classes and exams (study). This task will be performed individually and try to promote self-employment.

We will use e-learning platforms (Aula Virtual) to support communication with students; Aula Virtual will be used to access the course materials used in class, as well as solving problems and exercises.

EVALUATION

The evaluation of the course will be carried out by means of one or more than one tests in the format of exam and/or individual assignment and/or collaborative assignment.

REFERENCES

Basic

- Advances in knowledge discovery and data mining. Edited by Usama M. Fayyad [et al.]. MIT Press, 1996.
- Data mining for scientific and engineering applications. Edited by Robert L. Grossman [et al.]. Kluwer, 2001.
- Análisis de datos experimentales. Emilio Soria, José D. Martín, Antonio J. Serrano, Daniel Aguado. Universidad Politécnica de Valencia, 2007.
- Machine Learning. Ethem Alpaydin, MIT Press, 2009.

Additional

- Neural Networks and Learning Machines. Simon Haykin. Pearson Education, 2009.
- Time Series Prediction: Forecasting the Future and Understanding the Past: Proceedings of the NATO by Andreas Weigend y Neil Gershenfeld (Editores). Addison-Wesley, 1993.
- Pattern Classification (2nd Edition). Richard O. Duda, Peter E. d G. Stork. Wiley-Interscience, 2n edition, November 2000.
- Kernel Methods for Pattern Analysis. John Shawe-Taylor and Nello Cristianini. Cambridge University Press, New York, NY, USA, 2004.
- The elements of statistical learning: data mining, inference, and prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer, 2001.