VNIVERSITAT DE VALÈNCIA

Course Guide
36404 Data mining and machine learning

COURSE DATA

Data Subject	
Code	36404
Name	Data mining and machine learning
Cycle	Grade
ECTS Credits	6.0
Academic year	2019 - 2020

Study (s)

Degree	Center	Acad. year	Period
1400 - Degree in Computer Engineering	School of Engineering	4	First term
1403 - Degree in Telematics Engineering	School of Engineering	4	First term
1407 - Degree in Multimedia Engineering	School of Engineering	4	First term

Subject-matter

Degree	Subject-matter	Character
1400 - Degree in Computer Engineering	16 - Optional subject	Optional
1403 - Degree in Telematics Engineering	19 - Optional subjects	Optional
1407 - Degree in Multimedia Engineering	19 - Optatividad	Optional

Coordination

Name	Department
FERRI RABASA, FRANCESC JOSEP	240 - Computer Science

SUMMARY

Data mining and machine learning fundamentals are introduced from a computer science point of view. In particular, multimodal data processing is introduced along with associated learning algorithms including, parametric and nonparametric statistical, neural and metaheuristic methods. Several applications are considered to illustrate the course contents, e.g. content-based image retrieval, emotion detection from writing or web-browsing patterns, identity recognition, etc.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

None

OUTCOMES

1400 - Degree in Computer Engineering

- C1 - Ability to know the fundamentals, paradigms and techniques in the field of intelligent systems, and to analyse, design and build computer systems, services and applications that use these techniques in any field of application.
- C2 - Ability to acquire, obtain, formalise and represent human knowledge in a computable form for solving problems through a computer system in any field, particularly in those related to aspects of computing, perception and action in intelligent environments.
- C3 - Ability to recognise and develop computational learning techniques and to design and implement applications and systems that use them, including those for the automatic retrieval of information and knowledge from large volumes of data.

1405 - Grado en Ingenieria Multimedia

- G1 - Be able to relate and structure information from different sources and to integrate ideas and knowledge. (RD1393/2007)
- MM28 - Be able to solve problems with initiative, decision-making and creativity and to communicate and transmit the knowledge, abilities and skills of a multimedia engineer.

LEARNING OUTCOMES

To reason with data models knowing about their main properties.

To deal with basic operations and use corresponding tools for data processing with different purposes.

To solve simple problems about prediction from a computational point of view.

To know about bounds and computational burden of some of the most important problems in Machine learning and data mining.

To implement practical solutions for problems involving reasoning from data.

DESCRIPTION OF CONTENTS

1. Machine learning foundations and statistical background

Introduction to machine learning. Data representation, preprocessing and visualization

2. Parametric/nonparametric learners and distance-based methods

Bayes rule. Errors. Discriminant functions. Distance-based and neighbor-based techniques

3. Linear machines and extensions: kernels, layers and depth

Perceptrons. Adaline and extensions. Support vector machines. Introduction to kernels.

4. Unsupervised methods, estimates and clustering.

Clusters and quantization. Hierarchical methods. Prototype-based methods. Parametric and non parametric estimation. Semisupervised methods.

5. Extension and applications

Content-based multimedia information retrieval. Behavior pattern discovery. Optimal representations. Automatic identity recognition.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	30,00	100
Laboratory practices	20,00	100
Classroom practices	10,00	100
Development of group work	5,00	0
Development of individual work	10,00	0
Study and independent work	25,00	0
Readings supplementary material	5,00	0
Preparation of evaluation activities	10,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	10,00	0
Resolution of case studies	5,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

-Theory and problem teaching with student participation

-Discussion sessions and problem solving

-Lab sessions

-Quiz solving both in class and remotely

-Monograph writing and bibliographic search, both individually and group wise

EVALUATION

Weighted average of the following items

(Weights for the 2nd round in brackets):

Assistance and participation: 10% (5%)

Partial tests: 15% (7.5%)

Labs: 25% (12.5%)

Final test: 50% (75%)

All individual marks must be superior to 4 out of 10 in order to compute the final mark.

In any case, the evaluation of this subject will be done in compliance with the University Regulations in this regard, approved by the Governing Council on 30th May 2017 (ACGUV 108/2017)

REFERENCES

Basic

- C.M. Bishop. Pattern recognition and machine learning, 2006
- D.J. Hand, H. Mannila, P. Smith. Principles of data mining, 2001
- R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2n ed, 2001
- R. Garreta, G. Moncecchi. Scikit-learn. Machine learning in Python, 2013

Additional

- E. Alpaydin, Introduction to machine learning, 2010
- W. McKinney. Python for data analysis, 2013
- D.G. Stork, E. Yom-Tov, Pattern Classification. Computer manual in Matlab, 2004
- S. Theodoridis, K. Koutroumbas, Pattern Recognition, 3r ed, 2006

ADDENDUM COVID-19

This addendum will only be activated if the health situation requires so and with the prior agreement of the Governing Council

English version is not available