VNIVERSITAT DE VALÈNCIA

Course Guide
46574 Machine learning (I)

COURSE DATA

Data Subject	
Code	46574
Name	Machine learning (I)
Cycle	Master's degree
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
2262 - M.U. en Ciencia de Datos	School of Engineering	1	First term

Subject-matter

Degree	Subject-matter	Character
2262 - M.U. en Ciencia de Datos	6 - Machine learning (I)	Obligatory

Coordination

Name	Department
FERRI RABASA, FRANCESC JOSEP	240 - Computer Science
GOMEZ CHOVA, LUIS	242 - Electronic Engineering
NAVARRO QUILES, ANA	130 - Statistics and Operational Research

SUMMARY

This course is based on learning linear models and their extensions, the main neural architectures and unsupervised learning algorithms. To know what a support vector machine is and its difference regarding MLPs. To learn the concepts of entropy and information gain. To know decision trees and their construction. To know the concept of clustering and basic algorithms including adaptive and self-organizing systems. To know the advantages of classifier/model combination. To know the extensions of classic machine learning approaches.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

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

Other requirements

OUTCOMES

2262 - M.U. en Ciencia de Datos

- Students should demonstrate self-directed learning skills for continued academic growth.
- 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.
- Be able to defend criteria with rigor and arguments and to present them properly and accurately.
- Ability to access and manage information in different formats for subsequent analysis in order to obtain knowledge from data.
- Capacidad para trabajar en equipo para llegar a soluciones de problemas interdisciplinarios usando técnicas de análisis de datos.
- 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.
- Extraer conocimiento de conjuntos de datos en diferentes formatos.
- Entender la utilidad de la ciencia de datos y sus elementos asociados, así como su aplicación en la resolución de problemas, eligiendo las técnicas más adecuadas a cada problema, aplicando de forma correcta las técnicas de evaluación y, finalmente,
  interpretando los modelos y resultados.
- Ability to solve classification, modelling, segmentation and prediction problems from a set of data.
- Modelar la dependencia entre una variable respuesta y varias variables explicativas, en conjuntos de datos complejos, mediante técnicas de aprendizaje máquina, interpretando los resultados obtenidos.

LEARNING OUTCOMES

- Knowing how to construct and interpret simple and multiple linear regression models.
- Diagnosis and validation of regression models.
- Comparison of regression models. Selection of variables.
- To learn the main neural constructions, multilayer perceptron (MLP) and self organizing maps (SOM)
- To know what a support vector machine is and its difference regarding MLPs.

- To learn the concepts of entropy and information gain.
- To know decision trees and their construction. To know the concept of clustering and basic algorithms. To know the advantages of classifier/model combination.

DESCRIPTION OF CONTENTS

1. Supervised Learning

Regression, correlation and causality; Multivariate normal distribution; Simple regression and multiple regression; Diagnosis and validation of multiple regression models; Hypothesis estimation and testing; ANOVA table; Prediction; Comparison of regression models; Variable selection. Multilayer perceptron (learning algorithms, input preprocessing, architecture selection, Bayesian approach); support vector machines (kernel design, multiple kernel learning, multiclass, one-class), decision trees (pruning, rule extraction).

2. Ensemble models

Bagging, boosting, random forest, extremely randomized trees

3. Unsupervised Learning

k-means (EM algorithm), hierarchical clustering, selecting the number of clusters, adaptive and density-based methods. Semisupervised learning.

WORKLOAD

ACTIVITY	Hours	% To be attended
Laboratory practices	29,00	100
Theory classes	27,00	100
Theoretical and practical classes	4,00	100
Development of individual work	20,00	0
Study and independent work	12,00	0
Readings supplementary material	3,00	0
Preparation of evaluation activities	12,00	0
Preparing lectures	20,00	0
Preparation of practical classes and problem	13,00	0
Resolution of case studies	10,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

Theoretical activities. Exposition of the contents with student participation at solving particular problems. Individual quiz solving.

Practical activities. Learning by problem solving, exercises and case studies to acquire competences about different aspects of the subject.

Lab/computer work. Learning through individual or small group activities carried out in the lab.

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: 40% of the final grade.

2. Final exam: 60% of the final grade.

Grades earned in paragraph 1 shall be kept in the two examinations of the academic year in which they were made, since their evaluation is only possible in the teaching period.

REFERENCES

Basic

- Richard O. Duda (2016) Pattern Classification, Third Edition, John Wiley & Sons Inc.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2011) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer (Series in Statistics).
- Christopher Bishop (2010) Pattern Recognition and Machine Learning, First Edition, Springer (Information Science and Statistics).
- Ethem Alpaydin (2014) Introduction to Machine Learning, Third Edition, The Mit Press (Adaptive Computation and Machine Learning Series).
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)
- Peter K. Dunn, Gordon K. Smyth (2018) Generalized Linear Models with Examples in R Springer (Springer Texts in Statistics).

Additional

- Sebastian Raschka (2015) Python Machine Learning, Packt Publishing