Course guide
270029 - APA - Machine Learning

Last modified: 13/07/2023

Unit in charge:	Barcelona School of Informatics
Teaching unit:	723 - CS - Department of Computer Science.

Degree: BACHELOR'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2010). (Optional subject).

Academic year: 2023   **ECTS Credits:** 6.0   **Languages:** Spanish

LECTURER

Coordinating lecturer: JAVIER BÉJAR ALONSO

Others: Primer quadrimestre:
JAVIER BÉJAR ALONSO - 11, 12, 13
DANIEL HINJOS GARCÍA - 11, 12

PRIOR SKILLS

Elementary notions of probability and statistics.
Elementary linear algebra and real analysis
Good programming skills in a high-level language

REQUIREMENTS

- Prerequisite PE
- Corequisite PROP

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CCO2.1. To demonstrate knowledge about the fundamentals, paradigms and the own techniques of intelligent systems, and analyse, design and build computer systems, services and applications which use these techniques in any applicable field.
CCO2.2. Capacity to acquire, obtain, formalize and represent human knowledge in a computable way to solve problems through a computer system in any applicable field, in particular in the fields related to computation, perception and operation in intelligent environments.
CCO2.4. To demonstrate knowledge and develop techniques about computational learning; to design and implement applications and system that use them, including these ones dedicated to the automatic extraction of information and knowledge from large data volumes.

Generical:
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

TEACHING METHODOLOGY

The topics exposed in the main classes are very well motivated and motivating (why is it important to know), and supplemented with many examples. The theory lessons introduce all the knowledge, techniques, concepts and results necessary to achieve a well-grounded body of knowledge. These concepts are put into practice in the problem sessions, and in the laboratory.

Prior to each problem-solving session, the teacher proposes problems related to the current topic and the students have time to prepare them using their AA hours of personal work. In class, the students will be divided among several small groups depending on their number. The teacher provides guidance and eventually resolves questions, giving feedback and making the students progress in the resolution. The use of collaborative learning strategies is envisaged, in which one or more students take responsibility to lead the process. Students work again in these problems in their AA time and then deliver them. These deliveries are continued in time, have a uniform load and are evaluated. This strategy is also used to evaluate their skills for effective communication.

The theory lessons are weekly (two hours). The two hours of laboratory classes are twice a month. The problem sessions are weekly.

There is a deliverable practical work that solves a real problem (to be chosen by the student) that integrates all acquired knowledge and skills of the course. This practical work is also used to evaluate their skills for effective communication.

LEARNING OBJECTIVES OF THE SUBJECT

1.Formulate the problem of machine learning from data, and know the different machine learning tasks
2.Organize the flow for solving a machine learning problem, analyzing the possible options and choosing the most appropriate to the problem
3.Decide, defend and criticize a solution to a machine learning problem, arguing the strengths and weaknesses of the approach
4.To compare, judge and interpret a set of results after making a hypothesis about a machine learning problem
5.Understand and know how to apply least squares techniques for solving supervised learning problems
6.Understand and know how to apply techniques for single and multilayer neural networks for solving supervised learning problems
7.Understand and know how to apply support vector machines for solving supervised learning problems
8.Understand and formulate different theoretical tools for the analysis, study and description of machine learning systems
9.Understand and know how to apply the basic techniques for solving unsupervised learning problems
10.Understand and know how to apply basic techniques for solving reinforcement learning problems
11.Understand the most important modern machine learning and computational learning techniques

STUDY LOAD

Type	Hours	Percentage
Guided activities	6,0	4.00
Hours medium group	15,0	10.00
Hours small group	15,0	10.00
Self study	84,0	56.00
Hours large group	30,0	20.00

Total learning time: 150 h

CONTENTS

Introduction to Machine Learning

Description:
General information and basic concepts. Overview and approach to the problems tackled by machine learning techniques. Supervised (classification and regression), unsupervised (clustering) and reinforcement learning. Examples of modern applications.

Supervised Machine Learning Theory

Description:
The supervised Machine Learning problem setup. Classification and regression problems. Bias-variance tradeoff. Overfitting and underfitting. Generalization bounds. Empirical Risk minimization/Log Likelihood. Model selection and feature selection.

Visualitation and dimensionality reduction

Description:
The curse of dimensionality. Methods for variable selection and data transformation. Linear methods for dimensionality reduction (Principal Components Analysis). Nonlinear dimensionality reduction methods (LLE, t-SNE)

Supervised machine learning (I): Linear Methods

Description:
Linear learning methods for regression and classification problems. Linear regression by least squares and regularization. Probabilistic linear models and discriminated by classification: Discriminant analysis (LDA/QDA), Bayesian models, logistic regression. Linear models interpretation.

Supervised Machine Learning (II): Métodes no lineals

Description:
Nonlinear models for regression and classification. Nonparametric models (k-nearest neighbors). Neural networks and multi-layer perceptron, deep learning. Vector support machines and kernels. Decision trees and combination of classifiers. Non linear models interpretation.

Unsupervised machine learning

Description:
Definition and approaches for unsupervised machine learning. Clustering algorithms: EM algorithm and k-means algorithm.

Reinforcement learning

Description:
Description of reinforcement learning. Markov processes. Bellman equations. Values and Temporal Difference methods. Q-learning and the Sarsa algorithm. Applications.

ACTIVITIES

Development of item 1 of the course

Description:
The student sees an overview and basic concepts of machine learning as well as modern examples of application.

Specific objectives:
1, 11

Full-or-part-time: 10h
Theory classes: 2h
Laboratory classes: 2h
Self study: 6h

Development of item 2 of the course

Description:
The teacher explains the theory of unsupervised machine learning, focusing on clustering algorithms.

Specific objectives:
1, 2, 9

Full-or-part-time: 8h
Theory classes: 2h
Laboratory classes: 2h
Self study: 4h

Resolution of the problems of item 2 of the course

Description:
The teacher poses up to 3 problems within the current topic and the students prepare them at home. In the class, the teacher solves difficulties, gives guidance towards full resolution of the problem and eventually answers general questions about the topic. The students work again on these problems and deliver them.

Specific objectives:
1, 3, 8

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 10h
Guided activities: 2h
Self study: 8h

Development of item 3 of the course

Description:
The teacher exposes the problem of supervised Machine Learning. Explains the differences between regression and classification problems, and the notions of bias-variance tradeoff, over and underfitting as well as other theoretical tools for models selection.

Specific objectives:
1, 8

Full-or-part-time: 12h
Theory classes: 6h
Laboratory classes: 2h
Self study: 4h

Resolution of the problems of item 3 of the course

Description:
The teacher poses up to 3 problems within the current topic and the
students prepare them at home. In the class, the teacher solves
difficulties, gives guidance towards full resolution of the problem
and eventually answers general questions about the topic. The students
work again on these problems and deliver them.

Specific objectives:
3, 4, 5, 6

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 13h
Guided activities: 3h
Self study: 10h

Development of item 4 of the course

Description:
The teacher explains the basics of algorithms for separating hyperplanes: the Perceptron algorithm and maximum margin separation. The kernel functions and support vector machines for classification are introduced. Finally two neural networks for classification are covered: the multilayer perceptron and the radial basis function network.

Specific objectives:
1, 6, 7, 11

Full-or-part-time: 11h
Theory classes: 5h
Laboratory classes: 2h
Self study: 4h

Resolution of the problems of item 4 of the course

Description:
The teacher poses up to 3 problems within the current topic and the
students prepare them at home. In the class, the teacher solves
difficulties, gives guidance towards full resolution of the problem
and eventually answers general questions about the topic. The students
work again on these problems and deliver them.

Specific objectives:
2, 3, 4, 6

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas
orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 13h
Guided activities: 3h
Self study: 10h

Development of item 5 of the course

Description:
The teacher introduces the methods for regression problems: basically least squares (analytical and iterative methods). Error
functions for the regression case are introduced. Multilayer neural networks are adapted to the regression case. The support
vector machine for regression is introduced.

Specific objectives:
1, 5, 6, 7

Full-or-part-time: 11h
Theory classes: 5h
Laboratory classes: 2h
Self study: 4h

Resolution of the problems of item 5 of the course

Description:
The teacher poses up to 3 problems within the current topic and the
students prepare them at home. In the class, the teacher solves
difficulties, gives guidance towards full resolution of the problem
and eventually answers general questions about the topic. The students
work again on these problems and deliver them.

Specific objectives:
3, 4, 7

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas
orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 13h
Guided activities: 3h
Self study: 10h

Development of item 6 of the course

Description:
The teacher introduces the basic techniques for ensemble learning: bagging, boosting and ECOC, and explains them in light of the bias / variance tradeoff.

Specific objectives:
2, 8, 11

Full-or-part-time: 10h
Theory classes: 4h
Laboratory classes: 2h
Self study: 4h

Resolution of the problems of item 6 of the course

Description:
The teacher poses up to 3 problems within the current topic and the
students prepare them at home. In the class, the teacher solves
difficulties, gives guidance towards full resolution of the problem
and eventually answers general questions about the topic. The students
work again on these problems and deliver them.

Specific objectives:
2, 3, 4, 9, 10

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 10h
Guided activities: 2h
Self study: 8h

Development of item 7 of the subject

Description:
The teacher explains the basics of reinforcement learning and its applications and briefly introduces transductive learning.

Specific objectives:
10, 11

Full-or-part-time: 10h
Theory classes: 3h
Laboratory classes: 3h
Self study: 4h

Resolution of the problems of item 7 of the subject

Description:
The teacher poses up to 3 problems within the current topic and the
students prepare them at home. In the class, the teacher solves
difficulties, gives guidance towards full resolution of the problem
and eventually answers general questions about the topic. The students
work again on these problems and deliver them.

Specific objectives:
3, 4, 11

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas
orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

Full-or-part-time: 10h
Guided activities: 2h
Self study: 8h

Final exam

Specific objectives:
5, 6, 7, 8, 9, 10

Full-or-part-time: 9h
Guided activities: 3h
Self study: 6h

Delivery of the practical work

Specific objectives:
2, 3, 4, 5, 6, 7, 8, 9, 10

Related competencies :
G4. EFFECTIVE ORAL AND WRITTEN communication: To communicate with other people knowledge, procedures, results and ideas
orally and in a written way. To participate in discussions about topics related to the activity of a technical informatics engineer.

GRADING SYSTEM

The course can be passed with continuous assessment, as follows:

NProbs = Average mark of problems completed during the course
NPract = Mark for the practical work
NPart = Mark used to evaluate skills for effective communication

NF1 = 50% NProbs + 40% NPract + 10% NPart

The course can also be passed with a final exam, as follows:

NExF = Mark obtained in a final exam (during the exams period)

NF2 = 40% NExF + 20% NProbs + 30% NPract + 10% NPart

In any case, the final mark is the highest of the two:

FINAL MARK = max (NF1, NF2)

BIBLIOGRAPHY

Basic:
- Bishop, C.M. Pattern recognition and machine learning. New York: Springer, 2006. ISBN 0387310738.
- Deisenroth, M.P.; Faisal, A.A.; Ong, Ch. S. Mathematics for machine learning. Cambridge, United Kingdom: Cambridge University Press, 2020. ISBN 9781108470049.
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction [on line]. 2nd ed. Springer, 2009 [Consultation: 18/07/2023]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7. ISBN 9780387848570.
- Molnar, C.. Interpretable machine learning: a guide for making black box models explainable. 2nd ed. Munich, Germany: Christoph Molnar, 2022. ISBN 9798411463330.
- Alpaydin, E. Introduction to machine learning. 4th ed. Cambridge, Massachusetts; London: The MIT Press, 2020. ISBN 9780262043793.

Complementary:
- Duda, R.O.; Hart, P.E.; Stork, D.G. Pattern classification [on line]. 2nd ed. John Wiley & Sons, 2001 [Consultation: 18/07/2023]. Available on: https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=699526. ISBN 0471056693.
- Bishop, C.M. Neural networks for pattern recognition. Clarendon Press, 1995. ISBN 0198538642.
- Géron, A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems [on line]. 3rd ed. Sebastopol, CA: O'Reilly, 2023 [Consultation: 18/07/2023]. Available on: https://ebookcentral-proquest-com.recursos.biblioteca.upc.edu/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=30168989. ISBN 9781098125974.
- Murphy, Kevin P. Probabilistic machine learning : an introduction. Cambridge, Mass: The MIT Press, [2022]. ISBN 9780262046824.

RESOURCES

Hyperlink:
- https://scikit-learn.org/stable/index.html- https://www.cs.columbia.edu/~amueller/comsw4995s20- https://www.python.org/