

Course guide

270217 - AA1 - Machine Learning 1

Last modified: 31/01/2025

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

Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Compulsory subject).

Academic year: 2024 **ECTS Credits:** 6.0 **Languages:** Catalan

LECTURER

Coordinating lecturer: MARTA ARIAS VICENTE - JOSE VIDAL MANZANO

Others: Segon quadrimestre:
ALEXIS MOLINA MARTINEZ DE LOS REYES - 11, 12
JOSE VIDAL MANZANO - 11, 12

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.
CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

Generical:

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

Transversal:

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.
CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.
CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

TEACHING METHODOLOGY

The theory classes introduce all the knowledge, techniques, concepts and results necessary to reach a well-founded and insightful level of maturity. These concepts are put into practice in the laboratory classes. In these labs, Python code is provided that allows solving certain aspects of a data analysis problem with the techniques corresponding to the current topic of study. This laboratory also serves as a guide for the corresponding part of the term project, which must be developed by the students throughout the course. Some laboratory hours may be used to solve problems (without a computer) in the theory classroom.

There is a graded practical project which works out a real problem to be chosen by the student and which collects and integrates the knowledge and skills of the entire course. The generic competence of effective written communication is also evaluated by means of this practical work.

LEARNING OBJECTIVES OF THE SUBJECT

1. Formulate the problem of automatic learning from data, and get to know the types of tasks that can be given.
2. Organize the resolution flow of a machine learning problem, analyzing the possible options and choosing the most suitable for the problem.
3. Decide, defend and criticize a solution to a machine learning problem, arguing the strong and weak points of the approach.
4. Know and know how to apply linear techniques to solve supervised learning problems.
5. Know and know how to apply mono and multilayer neural network techniques to solve supervised learning problems.
6. Know and know how to apply support vector machines to the resolution of supervised learning problems.
7. Know and know how to apply the basic techniques for the resolution of unsupervised learning problems, with emphasis on data clustering tools.
8. Know and know how to apply the basic techniques for solving reinforcement learning problems.
9. Know and know how to apply ensemble techniques to solve supervised learning problems.

STUDY LOAD

Type	Hours	Percentage
Hours small group	30,0	20.00
Hours large group	30,0	20.00
Self study	90,0	60.00

Total learning time: 150 h

CONTENTS

Introduction to Machine Learning

Description:

General information and basic concepts. Description and approach of problems attacked by automatic learning. Supervised learning (regression and classification), non-supervised (clustering) and semi-supervised (reinforcement and transductive). Modern examples of application.

Unsupervised machine learning: clustering

Description:

Definition and approach of unsupervised machine learning. Introduction to clustering. Probabilistic algorithms: k-means and Expectation-Maximization (E-M).

Supervised machine learning (I): linear regression methods

Description:

Maximum likelihood for regression. Errors for regression. Least squares: analytical (pseudo-inverse and SVD) and iterative (gradient descent) methods. Notion of regularization. L1 and L2 regularized regression: algorithms ridge regression, LASSO and Elastic Net.

Supervised machine learning (II): linear methods for classification

Description:

Maximum likelihood for classification. Error functions for classification. Bayesian Generative Classifiers: LDA/QDA/RDA, Naïve Bayes and k-nearest neighbours.

Hierarchical methods: decision trees

Description:

General construction of decision trees. Split criteria: gain in entropy and Gini. Regularization in decision trees. CART trees for regression and classification.

Ensemble methods

Description:

Introduction to ensemble methods. Bagging and Random Forests. Boosting. Adaboost and variants.

Kernel based learning methods

Description:

Introduction to learning with kernel functions. Regularized kernelized linear regression. Basic kernel functions. Complexity and generalization: Vapnik-Chervonenkis dimension. Support Vector Machine.

ACTIVITIES

Development of topic 1

Specific objectives:

1

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

Full-or-part-time: 5h 18m

Self study: 3h 18m

Theory classes: 2h

Development of topic 2

Specific objectives:

1, 3, 7

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

Full-or-part-time: 11h 36m

Self study: 6h 36m

Theory classes: 3h

Laboratory classes: 2h

Development of topic 3

Specific objectives:

1, 4

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

Full-or-part-time: 20h

Self study: 10h

Theory classes: 8h

Laboratory classes: 2h

Development of topic 4

Specific objectives:

1, 2, 4

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 16h 18m

Self study: 8h 18m

Theory classes: 6h

Laboratory classes: 2h

Development of topic 5+6

Specific objectives:

1, 2, 5

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 18h 36m

Self study: 11h 36m

Theory classes: 5h

Laboratory classes: 2h

Development of topic 7

Specific objectives:

1, 2, 5

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 13h

Self study: 5h

Theory classes: 6h

Laboratory classes: 2h

Control session for the practical work

Specific objectives:

1, 2, 3, 4, 5, 6, 7, 8, 9

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 2h

Guided activities: 2h

Delivery of the practical work

Specific objectives:

1, 2, 3, 4, 5, 6, 7, 8, 9

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 3h

Guided activities: 3h

Seguiment i tutories de la pràctica

Description:

Seguiment i tutories de la pràctica

Specific objectives:

2, 3, 4, 6, 7, 9

Related competencies :

CE1. Skillfully use mathematical concepts and methods that underlie the problems of science and data engineering.

CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

CE9. Ability to choose and employ a variety of automatic learning techniques and build systems that use them for decision making, even autonomously.

CE8. Ability to choose and employ techniques of statistical modeling and data analysis, evaluating the quality of the models, validating and interpreting them.

CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.

CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

CT3. Efficient oral and written communication. Communicate in an oral and written way with other people about the results of learning, thinking and decision making; Participate in debates on topics of the specialty itself.

CT4. Teamwork. Be able to work as a member of an interdisciplinary team, either as a member or conducting management tasks, with the aim of contributing to develop projects with pragmatism and a sense of responsibility, taking commitments taking into account available resources.

CT7. Third language. Know a third language, preferably English, with an adequate oral and written level and in line with the needs of graduates.

Full-or-part-time: 26h

Self study: 20h

Laboratory classes: 6h



GRADING SYSTEM

The subject is evaluated through a partial exam, a final exam and a practical work in which a real problem is attacked, writing the corresponding report.

The final grade is calculated as:

$$\text{Grade} = 0.4 * \text{Work} + 0.4 \text{ final} + 0.2 \text{ mid-term}$$

For those students who can and want to attend re-evaluation, the re-evaluation exam grade will replace mid-term and final exams.

BIBLIOGRAPHY

Basic:

- Bishop, C.M. Pattern recognition and machine learning. Springer, 2006. ISBN 0387310738.
- Murphy, K.P. Machine learning: a probabilistic perspective. Cambridge, Mass: MIT Press, 2012. ISBN 9780262018029.
- Cherkassky, V.S.; Mulier, F. Learning from data: concepts, theory, and methods. 2nd ed. John Wiley, 2007. ISBN 0471681822.
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer, 2009. ISBN 0387848576.