270217 - AA1 - Machine Learning 1

Coordinating unit: 270 - FIB - Barcelona School of Informatics
Teaching unit: 723 - CS - Department of Computer Science
Academic year: 2019
Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Teaching unit Compulsory)
ECTS credits: 6 Teaching languages: Catalan

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.
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

El temario se expone en las clases de teoría de forma muy motivada (por qué se explica) y motivadora (por qué es importante conocerlo) complementado con muchos ejemplos.

Las clases de teoría introducen todo los conocimientos, las técnicas, conceptos y resultados necesarios para alcanzar un nivel bien fundamentado y entenidor. Estos conceptos se ponen en práctica en las clases de laboratorio. En estas se proporciona código R que permite resolver ciertos aspectos de un problema de análisis de datos con la o las técnicas correspondientes al tema en curso. Este laboratorio también sirve de guía para la parte correspondiente de la práctica, que desarrollan los alumnos a lo largo del curso.

Hay un trabajo práctico evaluable, que trabaja un problema real a elegir por el propio estudiante y que recoge e integra los conocimientos y las competencias de todo el curso. También se evalúa mediante el trabajo práctico la competencia genérica de comunicación eficaz escrita.

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

Total learning time: 150h	Hours large group:	30h	20.00%
	Hours small group:	30h	20.00%
	Guided activities:	0h	0.00%
	Self study:	90h	60.00%

Content

Introduction to Machine Learning

Degree competences to which the content contributes:

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

Degree competences to which the content contributes:

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

Degree competences to which the content contributes:

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

Degree competences to which the content contributes:

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

Degree competences to which the content contributes:

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

Feed-forward shallow neural networks

Degree competences to which the content contributes:

Description:
Feed-forward shallow neural networks (one hidden layer). Activation functions. Multilayer perceptron with one hidden layer and RBF (radial basis function network) and their training algorithms.

Recurrent shallow neural networks

Degree competences to which the content contributes:

Description:
Recurrent shallow neural networks: Hopfield networks and their training algorithms. Applications in associative memories and combinatorial optimization problems.

Kernel based learning methods

Degree competences to which the content contributes:

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

Ensemble methods

Degree competences to which the content contributes:

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

Reinforcement learning

Degree competences to which the content contributes:

Description:
Description of reinforcement learning. Markov processes. Bellman Equations. Values and methods of temporal differences. Q learning and the Sarsa algorithm. Applications.

Planning of activities

Development of topic 1	Hours: 5h 18m Theory classes: 2h Practical classes: 0h Laboratory classes: 0h Guided activities: 0h Self study: 3h 18m
Specific objectives: 1	

Development of topic 2	Hours: 12h 36m Theory classes: 4h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 6h 36m
Specific objectives: 1, 3, 7	

Development of topic 3	Hours: 18h Theory classes: 6h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 10h
Specific objectives: 1, 4	

Development of topic 4	Hours: 15h 18m Theory classes: 5h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 8h 18m
Specific objectives: 1, 2, 4	

Development of topic 6	Hours: 20h 36m Theory classes: 7h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 11h 36m
Specific objectives: 1, 2, 5	

Development of topic 7	Hours: 9h Theory classes: 3h Practical classes: 0h Laboratory classes: 1h Guided activities: 0h Self study: 5h
Specific objectives: 1, 2, 5	

Development of topic 8	Hours: 20h 36m Theory classes: 7h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 11h 36m
Specific objectives: 1, 6	

Development of topics 5 and 9	Hours: 23h 18m Theory classes: 8h Practical classes: 0h Laboratory classes: 2h Guided activities: 0h Self study: 13h 18m
Specific objectives: 1, 9	

Development of topic 10	Hours: 9h Theory classes: 3h Practical classes: 0h Laboratory classes: 1h Guided activities: 0h Self study: 5h
Specific objectives: 1, 8	

Control session for the practical work	Hours: 2h Guided activities: 2h Self study: 0h
Specific objectives: 1, 2, 3, 4, 5, 6, 7, 8, 9	

Delivery of the practical work	Hours: 3h Guided activities: 3h Self study: 0h
Specific objectives: 1, 2, 3, 4, 5, 6, 7, 8, 9	

Qualification 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:

Grade = 0.4 * Work + 0.6 * max (Final, 1/3 * Partial + 2/3 * Final)

For those students who can and want to attend re-evaluation, the re-evaluation exam grade will replace max (Final, 1/3 * Partial + 2/3 * Final).

270217 - AA1 - Machine Learning 1

Bibliography

Basic:

Bishop, C.M. Pattern recognition and machine learning. Springer, 2006. ISBN 0387310738.

Murphy, K.P. Machine learning: a probabilistic perspective [on line]. Cambridge, Mass: MIT Press, 2012 [Consultation: 16/10/2019]. Available on: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=3339490>. ISBN 9780262018029.

Cherkassky, V.S.; Mulier, F. Learning from data: concepts, theory, and methods [on line]. 2nd ed. John Wiley, 2007Available on: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=313393>. ISBN 0471681822.

Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction [on line]. 2nd ed. Springer, 2009Available on: <http://dx.doi.org/10.1007/978-0-387-84858-7>. ISBN 0387848576.