UAB
Universitat Autònoma de Barcelona

Machine Learning

Code: 102787
ECTS Credits: 6

2024/2025

Degree	Type	Year
2502441 Computer Engineering	OB	3
2502441 Computer Engineering	OT	4

Contact

Name: Jordi Gonzalez Sabate

Email: jordi.gonzalez@uab.cat

Teaching groups languages

You can view this information at the end of this document.

Prerequisites

It is recommended that in order to take this course, minimum competences had been achieved in the courses of Algebra, Calculus, Discrete Mathematics, Fundamentals of Computers, and Programming Methodology (first year), as well as of Artificial Intelligence, Statistics and Programming Lab (second year).

Objectives and Contextualisation

The Course on Machine Learning is embedded in the "Computing" mention, along with other subjects like "Knowledge, Reasoning and Uncertainty", "Computer Vision" and "Robotics, Language and Planning". Due to its contents, this subject is not only for students who follow the "Computing" mention, but indeed for any student of the Computer Engineering grade, since it is closely related to the subject of "Artificial Intelligence" in the second year. It is also highly recommended to have understood and feel manageable with the mathematical concepts explained in the subjects of "Calculus", "Algebra" and "Discrete Mathematics" of the first year, and "Statistics" of the second year, due to the strong mathematical content of this Course.

The course aims both to expand some of the topics developed during "Artificial Intelligence", and to introduce new problems associated with AI, mainly the learning of concepts and trends from data. It is about training students to be "data engineers/scientists", one of the occupations with the most brilliant future and most demanded by an increasing number of companies, including Facebook, Google, Microsoft and Amazon, to cite but a few. In fact, it is expected that the growth of the demand of these professionals in data engineering/science will be exponential at an international level, especially due to the growth in the generation of massive data. Thus, the main objective of the Course is to teach how to find a good solution (sometimes the best one is impossible) for different data analysis problems at different context,, based on identifying the best knowledge representation and applying the most appropriate technique to automatically generate good mathematical models that best explain the observed data with an acceptable deviation.

The contents taught in this Course are also given in the Universities of Stanford, Toronto, Imperial College London, MIT, Carnegie Mellon and Berkeley, to put just the most representative names. Therefore, on the one hand, thestudent gets an opportunity to achieve knowledge and skills comparable to those taught at the best universities. On the other hand, the student must be aware that this knowledge has an inherent mathematical difficulty, which involves considerable study and dedication. This is because in this Course not only the most

important contents to become a data engineer are taught, but also a curriculum line is formed to allow the student to expand the range of jobs available after the Career, as well as giving the necessary methodological bases for carrying out a Master degree in data engineering/science or artificial intelligence.

If you are looking for a Course to open an international labor market, and to learn the most used machine learning algorithms in not only the great technological companies mentioned above, but also in many data analysis SME and spin-offs in our country, this Course will not disappoint if you put both attitude and aptitude.

The objectives of the Course can be summarized in:

Knowledge:

- Describe the basic techniques of computer learning.

- List the essential steps of different machine learning algorithms

- Identify the advantages and disadvantages of the learning algorithms.

- Solve problems by applying different machinelearning techniques to find the optimal solution.

- Understand the results and limitations of each learning technique in different case studies.

- Know how to choose the most appropriate learning algorithm to solve contextualized problems.

Skills:

- Recognize situations in which the application of machine learning algorithms may be adequate

- Analyze the problem to solve and design the optimal solution applying the learned techniques

- Write technical documents related to the analysis and solution of a problem

- Program the basic algorithms to solve the proposed problems

- Evaluate the results of the implemented solution and propose possible improvements

- Defend and argue the decisions taken in the solution of proposed problems

Competences

Computer Engineering

- Acquire thinking habits.
- Have the capacity for in-depth knowledge of the fundamental principles and models of computation and know how to apply them to interpret, select, value, model and create new concepts, theories, uses and technological developments related with IT.
- Have the capacity to acquire, obtain, formalise and represent human knowledge in a computable form to solve problems by means of a computer system in any field of application, particularly related with aspects of computation, perception and performance in intelligent environments.
- Have the capacity to know and develop computational learning techniques and develop and implement applications and systems that use them, including those used for automatic extraction of information and knowledge from large volumes of data.
- Have the right personal attitude.
- Work in teams.

Learning Outcomes

1. Accept and respect the role of the various team members, and its different levels of dependence.

2. Develop a capacity for analysis, synthesis and prospection.
3. Generate proposals that are innovative and competitive.
4. Identify, manage and resolve conflicts.
5. Know and apply the most suitable learning techniques in different case studies.
6. Know and understand techniques for the representation of human knowledge.
7. Resolve computational problems applying different necessary learning mechanisms to find the optimum solution.
8. Understand and evaluate the results and limitations of the most common learning techniques.

Content

UNIT 1: INTRODUCTION

1.1 Basic concepts

1.2 History of machine learning

UNIT 2: DATA REGRESSION

2.1 Linear regression and gradient descent

2.2 Regularization and polynomial regression

UNIT 3: DATA CLASSIFICATION

3.1 Logistic regression

3.2 Support vector machines

UNIT 4: BIOINSPIRED REGRESSION AND CLASSIFICATION

4.1 Multilayer Perceptron

4.2 backpropagation

UNIT 5: GROUPING DATA

5.1 Data memorization: lazy learning

5.2 Data clustering: k-means and Expectation-Maximization

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
MD0: Theoretical contents and seminars	12	0.48	5, 6, 8, 3, 7
MD1: Problems resolution	10	0.4	5, 2, 8, 7
MD2: Solution of Practical projects	6	0.24	1, 5, 6, 2, 8, 4, 7
MD3: Kaggle Case	12	0.48	5, 2, 8, 3, 7

Type: Supervised			
MD2: Project Programming	6	0.24	1, 5, 6, 2, 8, 4
MD3: Kaggle Case	12	0.48	1, 5, 6, 2, 8, 3, 4, 7
Type: Autonomous			
MD0: Individual study	12	0.48	5, 6, 2, 8, 3, 7
MD1: Resolving problems (individually)	24	0.96	5, 6, 2, 8, 3, 7
MD2: Solving practical cases (in group)	16	0.64	1, 5, 6, 2, 8, 4, 7
MD3: Practical description in python of a Machine Learning case	32	1.28	5, 2, 8, 7

All the subject information and related documents that students need can be found on the Caronte page (https://caronte.uab.cat/course/view.php?id=95), in the subject menu Machine Learning (102787). It will serve to be able to see the materials, manage the practice groups, make the corresponding deliveries, see the notes, communicate with the teachers, etc. To be able to use it, you must do the following steps:

- Register as a user by giving your name, NIU, and a passport photo in JPG format. If you have already registered for another subject, there is no need to do it again, you can go to the next step.
- Enroll in the "Machine Learning (102787)" teaching type, giving the subject code "apc2024" (without the quotes).

In the development of the subject, seven types of teaching activities can be distinguished:

MD0 Presentation of theory content: Presentation of the theoretical content to be worked on in the subject. These contents must have been prepared before class by reading texts, searching for information, etc. The contents presented will be directly related to the problems, projects and seminars proposed in other teaching activities, so that they will be the basis on which other activities of the course will be developed. The contents will be found on the Caronte page (presentations and videos) and will consist of two parts: a presentation where the main theoretical and mathematical concepts related to specific computational learning tasks are exposed (this syllabus will be the basis of the exam theory of the subject, see evaluation section of this teaching guide), and a second part of code in python on Jupyter notebooks that exemplify the details of coding and libraries to implement in a practical case the main concepts seen in the previous hour . The students will then be able to watch the videos of the classes, download the presentations and the python notebooks and test all the codes on their computer, to do the necessary tests and to be able to play with the various parameters to finish understanding the reasons for the different performances and precisions that are achieved in a specific database with specific configurations of the algorithms explained in the subject.

MD1 Computational problem solving: Delivery of up to a maximum of 3 problems implemented in a Jupyter Notebook. All the theory topics will be accompanied by a list of notebooks, from which the student will have to work on the problem sessions and hand in optionally. These activities must allow the student to deepen their understanding and personalize the theoretical knowledge in a specific numerical case. Some examples of data that require the design of a solution in which the methods seen in the theory classes are used will be considered. It is impossible to follow the problem classes if you do not follow the contents of the theory classes. The result of these sessions is to achieve the necessary skills for solving problems that will have to be delivered according to the specific delivery mechanism that will be indicated on the subject's website (Caronte area).

MD2 Implementation of a short guided group project: Realization of 1 guided practice to deepen the applied aspects of the theory. The practical part of the subject will be completed with practical sessions, where the students will have to solve specific problems of a certain complexity implemented in python. These projects will be solved in small groups of 2-4 people, and where each member of the group will have to do a part and share it with the rest to have the solution end These working groups must be maintained until the middle of the course and must be self-managed: distribution of roles, work planning, assignment of tasks, management of

available resources, conflicts, etc. Although the teacher will guide the learning process, his intervention in the management of the groups will be minimal. To develop the project, the groups will work independently and the practice sessions must be dedicated mainly by the teacher to monitoring the status of the project, indicating errors to be corrected, proposing improvements, etc. Doubts that may arise regarding the implementation of the practicals will be transmitted through the Caronte forum, where other students can answer them.

MD3 Solution of a Kaggle practical case: each group of 1-2 persons will create a jupyter notebook where the various steps taken to solve a Computational Learning problem should be explained. The projects will be applied to selected databases from the Kaggle platform (https://www.kaggle.com/search?q=machine+learning), and will consist of three parts: an explanation of the most important attributes of the database and of the attribute to predict/classify; brief description of the computational learning method applied, along with the chosen parameters; and a presentation of the results obtained. Examples of jupyter notebooks can be found in the following repository: https://datauab.github.io/

MD4 Consultations and doubts: Free hours for the student for consultations and tutorials on aspects in which he needs additional help from the teaching staff. All inquiries will be made online, through the subject's forum, or emails to teachers, for example. It will be appreciated that the students answer the doubts of their colleagues as well as that in these answers they provide information that helps in understanding the content of the teaching activities.

MD5 Evaluation activities: for each of the activities described above. See the assessment section of this teaching guide.

In the case of repeaters, if the teacher in charge is asked, the grades for the teaching activities they took the previous year will be validated, if they havepassed. Repeaters must retake the individual theoretical tests (MD0).

This year there is a special itinerary for international students. In this case, students should contact at the start of the semester with the responsible professors who will describe the methodology followed in the English itinerary, which is described in this section.

Transversal Competences

-T01 Habits of thought (T01.02 Developing the capacity for analysis, prospective synthesis): in autonomousand supervisedactivities (study of the MD0 theory, realization of the MD2 practices, realization of the MD1 problems, and description of an MD3 practical case)

- T03 Teamwork (T03.02 Assume and respect the role of the various team members, as well as the different levels of dependence on it; T03.03 Identify, manage and resolve conflicts): in MD2 practices, as an autonomous activity in its preparation and delivery, and as a supervised activity in its preparation and presentation in a seminar.

- T06 Personal attitude (T06.03 Generate innovative and competitive proposals in the professional activity): in autonomous activities (study of MD0 theory, participation in the subject forum in Caronte MD4), directed (resolution of practical MD2 projects) and supervised (analysis of a MD3 case study).

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes

Delivery of problems	25%	0	0	5, 6, 2, 8, 3, 7
Individual theory tests	20%	4	0.16	5, 6, 2, 8, 3, 7
Written documentation, implementation and presentation of the Kaggle case	35%	2	0.08	1, 5, 6, 2, 8, 4, 7
Written documentation, implementation and presentation of the practical project	20%	2	0.08	1, 5, 6, 2, 8, 4, 7

Activities and evaluation instruments:

a) Scheduled evaluation process and activities

The subject consists of the following assessment activities:

- MD0: Theoretical exams, where for each exam the student will have to answer individually and in writing 5 questions (to be developed in one full sheet of DIN-A4 paper) on computational learning concepts seen in the theory classes. It represents 20% of the final grade, is optional, individual and recoverable (there will be two partials and their respective recoveries, without penalty).

- MD1: Delivering a report with up to 3 solved problems, where each student will individually deliver a written report on up to 3 problems seen in classes (regression, classification, memorization, and clustering). It represents 25% of the final grade, is optional, individual and recoverable (without penalty).

- MD2: Resolution of a practice with delivery of a report explaining the resolution and the results, where each group composed of two to four people, will deliver the python code, as well as a report of up to 30 pages in which they will describe the databases, the strategy they used to analyze the data, as well as the tests with different parameter values they tried and the results they obtained with the best possible configuration. There will be an optional presentation of each project. It represents 20% (code 5% + report 10% + presentation 5%) on the final grade, it is optional, group and recoverable.

- MD3: Individual development of a project with the description of the python code and the results of a specific case of machine learning (either regression, classification, clustering or memorization, chosen from here: https://www.kaggle.com /search?q=machine+learning), and making an optional presentation describing the data in code, the models used (with the parameters that work best forthe data),and the resultsof the chosen problem. Examples of jupyter notebooks applied to concrete cases can be found here https://datauab.github.io/. It represents 35% of the final grade (code 10% + report 15% + github 5% + presentation 5%), it is optional, group and NOT recoverable.

Below is a description of how to pass the subject with continuous assessment:

- MD0: Individual theoretical exams

The final theory grade will be calculated from two partial exams:

Theory Grade = (0.5 * Partial1) + (0.5 * Partial2)

Partial1 is done in the middle of the semester and is used to eliminate part of the subject if it is approved. Partial2 is done at the end of the academic semester and serves to eliminate the part of the syllabus that comes after Partial1.

These exams seek an individualized assessment of the student with their ability to answer 5 long questions (develop until they occupy a maximum sheet of paper) about the techniques explained in class, as well as evaluate the level of conceptualization that the student has done regarding the techniques seen in the lectures.

In order to pass the theory part of the subject by taking face-to-face exams, two requirements must be met:

- it will be necessary that the grades of partials 1 and 2 be equal to or higher than 4.0 (in both partials). If less than a 4.0 is taken in any of the two Partials, the corresponding partial must be retaken during the make-up exam.
- the final theory grade must be greater than or equal to 4.0. In the event that the final theory grade is not equal to or higher than 4.0, students can take the make-up exam to be assessed on all the content seen in the subject.

Recovery exam (end of January or beginning of February). In this exam, you can recover the partial(s) that did not exceed4.0, orrecovertheentire syllabus in the event that the final theory gradedoes not exceed 4.0.

– MD1: Individual delivery of a report with up to 3 solved problems

The aim of the problems is to cause the student to engage with the contents of the subject continuously and, based on small problems, to become familiar directly with the application of the theory. As evidence of this work, the mandatory presentation of a portfolio in which you will have kept the problems you have been working on (competence T04) is requested.

Note Problems = Portfolio evaluation with 3 solved problems out of the 5 seen in class. They can be recovered by handing in for Caronte during the week before Christmas, without penalty.

– MD2: Resolution of 1 group practice

The evaluation of the 2 internship project will include:

– Joint evaluation of each project (T02 competence): single grade for all members of the working group that will assess the overall result of the project, the quality of the code, the general structure of the final presentation and the documents delivered throughout the project.

- Individual assessment (T01 competence): the individual work will be assessed based on the answers to the questions in the online control sessions, the final presentation of the online project and mainly the active participation in the Caronte forums. In the cases required by any group (in cases of incidents between colleagues), a short confidential form will be evaluated qualifying the contribution of each group member to the final result.

The grade of the project will be calculated according to the formula:

Practice Grade = (0.2 * Program) + (0.2 * Presentation) + (0.6 * Documentation)

In very justified cases (e.g. PIUNE,for work, family or health issues, ...), groupscan be 1person.

MD3: Realization of a specific case of computational learning on the Kaggle platform

The evaluation will be based on the explanation iof the code and results that should be placed in a github repository. The kaggle case grade will be calculated according to the formula:

Note Case Kaggle = (0.1 * Introduction to the database) + (0.25 * Analysis of attributes, correlations,...) + (0.25 * Description of the method used, how to find the best parameters, comparison of methods...) + (0.3 * Deciption of results, confusion matrices, graphs of models and data, examples of false positives/negatives, ROC curves, ...) + (0.1 * Github repository presentation)

Examples of jupyter notebooks applied to concrete cases can be found here https://datauab.github.io/.

- Evaluation of transversal skills

The partial exams will allow you to evaluate your acquisition of personal thinking and work habits (T01 Evaluate critically and with quality criteria the work done, Theory Note). With the Internship Project Note,

teamwork will also be assessed (T02 Work cooperatively in a multidisciplinary context assuming respect for the role of the different members of the team, Group Note). By completing the problems and completing a Kaggle case, the acquisition of habits to solve a predetermined task with completely different data values than those seen in class will be assessed (T04 Effectively use the bibliography and electronic resources to get information, Note Issues and Kaggle Case).

The final grade of the subject is obtained by combining the evaluation of these 4 activities as follows:

Final Grade = (0.20 * Theory) + (0.20 * Practical) + (0.25 * Problems) + (0.35 * Kaggle Case)

Conditions to approve:

All assessment activities are optional, to pass the subject it is necessary that the sum of the assessment of each of the activities exceeds 5 points. In the event of not passing the subject, the numerical grade of the file will be the lower value between 4.5 and the weighted average of the grades obtained:

The final MD0 theory grade must be greater than or equal to 4.0 to be able to add the theory part to the subject's final grade.
The grade of the MD2 project must be greater than or equal to 5.0 to be able to add the part of the practice to the final grade of the subject.
The final grade of the subject must be greater than or equal to 5.0 in order to pass the subject.

If the student takes an exam or completes a practice, he can no longer be assessed as "Not Assessable" in the event that he does not take any of the other assessments, but the final grade will be calculated from of those continuous evaluations that have been presented.

b) Programming of evaluation activities

The dates of continuous assessment and assignment of assignments will be published in Caronte (http://caronte.uab.cat/), in the area of this subject and may be subject to schedule changes for reasons of adaptation to possible incidents ; caronte.uab.cat will always be informed about these changes since this platform will become the usual mechanism for exchanging information between professors and students.

c) Recovery process

The student can present himself for recovery as long as he has presented himself to a set of activities that represent a minimum of two thirds(4evaluation activities out of 7 total: 2 partial theory exams; 1 practice evaluation; 3 problem submissions; and 1 kaggle case submission) of the subject's total grade.

Of these, those students whohave a grade above 3.0 as an average for all the activities of the subject will be able to submit a retake.

It should be borne in mind that the Resolution ld the Kaggle Case (MD3) is not recoverable.

d) Qualification review procedure

For each individual theoretical exam, a review place, date and time will be indicated in which the student can review the activity with the teacher. In this context, claims can be made about the grade of the activity, which will be evaluated by the teaching staff responsible for the subject. If the student does not appear for this review, this activity will not be reviewed later.

e) Qualifications

Honors Matriculations: Honors Matriculations will be granted at the discretion of the teaching staff responsible for the subject, up to five percent or a fraction of the students enrolled in all the teaching groups of the subject. UAB regulations indicate that MH can only be awarded to students who have obtained a final grade equal to or higher than 9.00.

Non-evaluable: A student will be considered non-evaluable (NA) if he/she has not appeared in any of the partial exams or in any of the 2 practical assessments.

f) Irregularities on the part of the student, copying and plagiarism

Without prejudice to other disciplinary measures deemed appropriate, irregularities committed by the student that may lead to a change in the grade of an assessment act will be graded with a zero. Therefore, copying, plagiarism, deception, allowing copying, etc. in any of the assessment activities will involve failing it with a zero. Assessment activities qualified in this way and by this procedure will not be recoverable. If it is necessary to pass any of these assessment activities to pass the subject, this subject will be suspended directly, with no opportunity to recover it in the same course. These irregularities include, amongothers:

thetotal or partial copy of a practice, report, or any other assessment activity;
let copy;
present group work not done entirely by group members (applied to all members, not just those who have not worked);
present as own materials prepared by a third party, even if they are translations or adaptations, and in general works with non-original and exclusive elements of the student;
have communication devices (such as mobile phones, smart watches, pens with cameras, etc.) accessible during individual theoretical-practical assessment tests (exams);
talk with peers during individual theoretical-practical assessment tests (exams);
copy or attempt to copy from other students during theoretical-practical assessment tests (exams);
use or try to use writings related to the subject during the theoretical-practical assessment tests (exams), when these have not been explicitly allowed.

The numerical grade on the file will be the lower value between 3.0 and the weighted average of the grades in the event that the student has committed irregularities in an evaluation act (and therefore the approved by compensation will not be possible). In future editions of this subject, the student who has committed irregularities in an assessment act will not be validated for any of the assessment activities carried out.

Summary: copying, allowing copying or plagiarism (or the attempt to) in any of the assessment activities is equivalent to a SUSPENSION, not compensable or recoverable and without validation of parts of the subject in subsequent courses.

g) Evaluation of repeat students

From the second enrolment, the evaluation of the subject will consist of the individual theoretical exam, adding the marks corresponding to the MDs obtained the first time the student has registered for the subject, as long as the marks of MD2 practices are greater than or equal to5.0.

To be able to opt for this differentiated assessment, the repeating student must ask the teacher until week 5 at the latest.

h) Single evaluation

The single evaluation of the subject will consist of the following evaluation activities that will be necessary to hand in on the day of the second part of the theory exam:

- MD0, evaluation of the theory, 30% of the final grade, recoverable.

- MD1, delivery of problems, 30% of the final grade, recoverable.

- MD3, Kaggle case submission (github repository), 40% over final grade, non-retrievable.

For MD0 and MD1, the same recovery system will be applied as for the continuous evaluation: the theory recovery exam can be taken, and problems can be handed in through Caronte on the same day as the theory recovery.

In addition, the review of the final grade follows the same procedure as for the continuous assessment.

Bibliography

Web links

- Caronte: http://caronte.uab.cat

- Artificial Intelligence: A Modern Approach. http://aima.cs.berkeley.edu/

- Web of the UAB Library Catalogue: https://bit.ly/3xdcdFB

Basic bilbiography:

- S. Russell, P. Norvig. Artificial Intelligence: A Modern Approach. Ed. Prentice Hall, Second Edition, 2003.

Complementary bilbiography

- L. Igual, S. Seguí. Introduction to Data Science. Ed. Springer, 2017

- Bishop, Pattern Recognition and Machine Learning, 2007.

- Duda, Hart, and Stork, Pattern Classification, 2nd Ed., 2002.

- Marlsand, Machine Learning: an Algorithmic Perspective, 2009

- Mitchell, Machine Learning, 1997

- Ripley, Pattern Recognition and Neural Networks, 1996.

Related bilbiography

- Eberhart, Shi, Computational Intelligence: Concepts to Implementations, 2007

- Friedman, Tibshirani, The Elements of Statistical Learning, 2009.

- Gilder, Kurzweil, Richards, Are we spiritual machines? Ray Kurzweil vs. the Critics of Strong AI, 2011

- Kurzweil, The Singularity is Near: When Humans trascend Biology, 2006

- Rosen, Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life (Complexity in Ecological Systems), 2005

- Witten,Frank, Hall, Data Mining: Practical Machine Learning Tools and Techniques, 2011

Software

The software required will be the Python programming language, a programming environment (such as Spyder, Pycharm or Visual Studio Code), the Jupyter Notebook web application, and the libraries needed for data analysis: scipy (contains NumPy, matplotlib, pandas), sklearn and Seaborn.

Language list

Name	Group	Language	Semester	Turn
(PAUL) Classroom practices	441	Catalan	first semester	morning-mixed
(PAUL) Classroom practices	442	Catalan	first semester	morning-mixed
(PLAB) Practical laboratories	441	Catalan	first semester	morning-mixed
(PLAB) Practical laboratories	442	Catalan	first semester	morning-mixed
(PLAB) Practical laboratories	443	Catalan	first semester	morning-mixed
(PLAB) Practical laboratories	444	Catalan	first semester	afternoon
(TE) Theory	440	Catalan	first semester	morning-mixed