

Machine Learning

Code: 102787
ECTS Credits: 6

| Degree | Type | Year | Semester |
|------------------------------|------|------|----------|
| 2502441 Computer Engineering | OB | 3 | 1 |
| 2502441 Computer Engineering | OT | 4 | 1 |

Contact

Name: Jordi Gonzalez Sabaté
Email: jordi.gonzalez@uab.cat

Use of Languages

Principal working language: catalan (cat)
Some groups entirely in English: No
Some groups entirely in Catalan: Yes
Some groups entirely in Spanish: No

Other comments on languages

In the case of Erasmus students or from outside Catalonia, please consult the English Version

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, the student 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 machine learning 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

Methodology

This year there is a special itinerary for international students, or alternatively those students who want to follow the subject in the English language. 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.

Anyways, all the information of the subject and the related documents that the students need will be found on the Caronte page (<https://caronte.uab.cat/course/index.php?categoryid=2>), the menu of the subject Machine Learning (102787). It will serve to see the materials, manage the practice groups, make the corresponding deliveries, see the notes, communicate with the teachers, etc. In order to use it, it is necessary to do the following steps:

- Register as a user giving the name, NIU, and a photo ID in JPG format. If you have already registered for some other subject, you do not have to do it again, you can go to the next step.
- Enroll in the type of teaching "Machine Learning (102787)", giving the subject code "apc2022" (without quotation marks).

In the development of the subject for the English itinerary, the following types of teaching activities will be distinguished:

MD0 Theory content exposition: Presentation of the theoretical contents to work on the subject. These contents must have been prepared before class from 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 course activities will be developed. The contents (slides and videos) will be found on Caronte's page 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 theoretical examination of the subject, see evaluation section of this teaching guide), and a second part of python code on Jupyter notebooks that exemplify the coding details of libraries to implement in a practical case the main concepts seen in the previous hour. The student will then be able to 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 reached in a specific database with specific configurations of the algorithms explained in the course.

MD1 Realisation of the online Coursera course on Machine Learning: with the prior approval of the teacher, the student who follows the English itinerary should complete the online Machine Learning course at the Coursera educational platform (<https://es.coursera.org/learn/machine-learning>).

MD2 Didactic development of a practical case of machine learning: each student will make a jupyter notebook where they will explain the different steps taken to solve a specific Machine Learning problem. The projects will be applied to selected databases of the Kaggle platform

(<https://www.kaggle.com/search?q=machine+learning>), and a report will be delivered which consists of three parts: an explanation of the most important attributes of the database data and attribute predict / classify; brief description of the applied computational learning method, together with the chosen parameters; and a presentation of the results that have been obtained. Examples of jupyter notebooks can be found in the following repository: <https://datauab.github.io/>

MD3 Consultations and doubts: Hours freely available by the student for consultations and tutorials on aspects in which they need additional help from the teaching staff. All inquiries will be made online, through the subject forum, or emails to teachers, for example. It will be appreciated that students answer the doubts of their classmates as well as that in these answers they provide information that helps in understanding the content of teaching activities.

MD4 Assessment activities: for each of the activities described above. See evaluation section of this teaching guide.

In the case of repeat students, if the responsible teacher is notified, they will be validated the notes of any teaching activities that they have passed the previous course.

Transversal Competences

-T01 Habits of thought (T01.02 Developing the capacity for analysis, prospective synthesis): in autonomous and supervised activities (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.

Activities

| Title | Hours | ECTS | Learning Outcomes |
|---|-------|------|------------------------|
| Type: Directed | | | |
| MD0: Theoretical contents and seminars | 22 | 0.88 | 5, 6, 8, 3, 7 |
| MD1: Problems resolution | 8 | 0.32 | 5, 2, 8, 7 |
| MD2: Project Programming | 16 | 0.64 | 1, 5, 6, 2, 8, 4, 7 |
| Type: Supervised | | | |
| MD3: Solution of Practical projects | 16 | 0.64 | 1, 5, 6, 2, 8, 3, 4, 7 |
| Type: Autonomous | | | |
| MD0: Individual study | 10 | 0.4 | 5, 6, 2, 8, 3, 7 |
| MD1: Resolving problems (individually) | 18 | 0.72 | 5, 6, 2, 8, 3, 7 |
| MD2: Preparation, programming, documentation and presentation of practical projects | 22 | 0.88 | 1, 5, 6, 2, 8, 4, 7 |
| MD3: Practical description in python of a Machine Learning case in a jupyter notebook | 18 | 0.72 | 5, 2, 8, 7 |

Assessment

Evaluation activities and instruments:

a) Process and scheduled evaluation activities

The course consists of the following assessment activities:

- MD0 (20%): Theoretical exams (in English), where for each exam the student will have to answer 5 questions individually and in writing (to develop in a maximum level) about concepts of computer learning seen in theory classes. It represents 20% of the final grade, it is individual and can be retaken (there will be two partials and their respective retakes).

- MD1 (50%): Realisation of the online Coursera course on Machine Learning: the mark obtained at this course will constitute the 50% on the final grade, is mandatory, and can not be retaken.

- MD2 (30%): Preparation of a jupyter notebook with the description in python code of a specific case of computational learning (either regression, classification, clustering or memorization, for example chosen here: <https://www.kaggle.com/search?q=machine+learning>), and the notebook will code the data, the models used (with the parameters that work best for the data), and the results of the chosen problem. Examples of jupyter notebooks applied to specific cases can be found here <https://datauab.github.io/>. Represents 20% of the final grade, is individual and can be retaken (online delivery the day of the second retake of the theoretical exam).

The following describes how to pass the course with continuous evaluation:

- MD0: Individual theory exams

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

$$\text{Theory Note} = (0.6 * \text{Partial1}) + (0.4 * \text{Partial2})$$

Partial1 is done in the middle of the semester and serves to eliminate part of the subject if it is approved.

Partial2 is done at the end of the school semester and serves to eliminate the part of the agenda that comes after Partial1.

These exams aim at an individualized evaluation of the student with their abilities to answer 5 long questions (develop to occupy a maximum page length) about the techniques explained in class, as well as to evaluate the level of conceptualization that the student has not made of the techniques. views.

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 marks of the partials 1 and 2 are equal to or greater than 4.0 (in both partials). In case that less than 4.0 is achieved in one of the two Partial, the corresponding partial must be done again during the recovery exam.
- the final theory grade must be greater than or equal to 4.0. In case the final theory grade is not equal to or greater than 4.0, students can take the recovery exam to be evaluated of all the contents seen in the subject.

Recovery exam (end of January or beginning of February). In this exam, the student can recover the partial (s) that have not exceeded 4.0, or recover the syllabus in the event that the final theory grade does not exceed 4.0.

- MD1: Coursera online course on Machine Learning

The evaluation of this part will be the mark obtained at the online course
<https://es.coursera.org/learn/machine-learning>.

This course can not be retaken: in case of not completing the course, the mark will not add up to the final mark of the subject

- MD2: Making a jupyter notebook in python describing a specific case of machine learning

The evaluation will be based on the python code and the explanation of the code that will be found in the jupyter notebook that will be delivered no later than the day of the second part of the course. The note of the notebook will be calculated according to the formula:

$$\begin{aligned} \text{Note Notebook} = & (0.1 * \text{Introduction to the database}) + (0.25 * \text{Analysis of the attributes, correlations, ...}) + \\ & (0.25 * \text{Description of the method used, howto find the best parameters, method comparison ...}) + (0.3 * \\ & \text{Described the results, confusion matrices, graphs of the models and the data, examples of false positives /} \\ & \text{negatives, ROC curves, ...}) + (0.1 * \text{Presentation of the Jupyter Notebook}) \end{aligned}$$

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

- Evaluation of transversal competences

Partial exams MD0 will allow you to evaluate your acquisition of thinking habits and personal work (T01 Thinking habits, Note Theory). With the completion of the online course MD1 and the implementation of a Kaggle project MD3, the acquisition of habits will be evaluated to solve a predetermined task and data values totally different from those seen in class (T06 Personal attitude).

The final grade for the course is obtained by combining the evaluation of these activities as follows:

$$\text{Final Note} = (0.2 * \text{Theory}) + (0.5 * \text{Online course}) + (0.3 * \text{Kaggle Case})$$

Conditions to approve:

To pass, it is necessary that the evaluation of each of the activities exceeds the minimum required and that the total evaluation exceeds 5 points. In case of not passing the subject, the numerical note of the file will be the weighted average of the marks that exceed the minimum thresholds:

- The final grade of theory MD0 must be greater than or equal to 4.0 to contribute to the final mark.
- The grade of the MD1 and MD2 must be greater than or equal to 5.0 in order to contribute to the final mark.
- The final grade for the course must be greater than or equal to 5.0 to pass the course.

If the student attends an exam or if s/he submits a deliverable of MD1 or MD2, s/he can no longer be evaluated as "Not Evaluable" in case s/he does not take any of the other evaluations, but the final grade will be calculated from those continuous evaluations to which it has been submitted.

b) Scheduling of evaluation activities

The dates of continuous evaluation and delivery of papers will be published in Caronte (<http://caronte.uab.cat/>), in the space of this subject and may be subject to programming changes for reasons of adaptation to possible problems; caronte.uab.cat will always be informed of these changes as this platform will become the usual mechanism for the exchange of information between teacher and students.

The following schedule is foreseen (week 1 corresponds to the week of September 12th, 2022):

- Individual theoretical exams: weeks 8 and 17-19 of the subject.
- Submission of the mark obtained in the online course: weeks 17-19 of the subject.
- Submission of the Python notebook: weeks 17-19 of the subject.

c) Retake procedure

The student can retake the theoretical exam (MD0).

Keep in mind that the Online Course (MD1) and the Kaggle Notebook (MD2) can not be retaken.

d) Grade review procedure of MD0

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

e) Ratings

Enrollment of Honor: Enrollment of Honor will be granted at the decision of the teaching staff responsible for the subject, up to five percent or fraction of the students enrolled in all 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 greater than 9.00.

Not evaluable: A student will be considered non-evaluable (NA) if he has not taken any evaluation.

f) Student perpart irregularities, copying and plagiarism

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

- the total or partial copy of a practice, report, or any other evaluation activity;
- let copy;
- present a group work not done entirely by group members (applied to all members, not just those who have not worked);
- present their own material and work by a third party, whether 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 a camera, etc.) accessible during the course of individual theory-practical assessment(exams);
- talk to colleagues during individual theory-practical assessment tests (exams);
- copy or try to copy from other students during the theoretical-practical assessment tests (exams);
- use or try to use writing related to the subject during the theoretical-practical evaluation tests (exams), when these have not been explicitly allowed.

The numerical grade of the record will be the lesser value between 3.0 and the weighted average of the grades in the event that the student has committed irregularities in an act of evaluation (and therefore it will not be possible to pass the compensation). In future editions of this subject, the student who has committed irregularities in an act of evaluation will not be validated any of the evaluation activities carried out.

In summary: copying, copying or plagiarizing (or trying to) in any of the assessment activities is equivalent to a SUSPENSE, not compensable or recoverable and without validation of parts of the subject in subsequent courses.

g) Evaluation of repeating students

From the second enrollment, the evaluation of the subject will consist of the marks corresponding to the evaluations passed the first time the student enrolled in the course, provided that the marks of practice are higher or equal to 4.0 in case of the theory, or 5.0 in case of the online course and the kaggle notebook.

In order to be eligible for this differentiated assessment, the repeating student must ask the teacher at the latest until week 8.

Assessment Activities

| Title | Weighting | Hours | ECTS | Learning Outcomes |
|---|-----------|-------|------|---------------------|
| Delivery of problems | 20% | 4 | 0.16 | 5, 6, 2, 8, 3, 7 |
| Individual theory tests | 20% | 4 | 0.16 | 5, 6, 2, 8, 3, 7 |
| Jupyter notebook in python of a Machine Learning case | 20% | 4 | 0.16 | 5, 2, 8, 7 |
| Programming of code projects | 10% | 2 | 0.08 | 1, 5, 6, 2, 8, 4, 7 |
| Written documentation, presentation, follow-up practical projects | 30% | 6 | 0.24 | 1, 5, 6, 2, 8, 4, 7 |

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 bibliography:

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

Complementary bibliography

- 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.