

## Computational Learning

Code: 104403  
ECTS Credits: 6

Degree	Type	Year	Semester
2503740 Computational Mathematics and Data Analytics	OB	3	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, additional classes will be given in the common language between student and teacher so that the student can follow the subject without linguistic problems

### Prerequisites

A good knowledge of the contents of the subjects studied during the first year as Probability and Calculus and of the second year as Artificial Intelligence and Modeling and Inference is considered very important. Although some theoretical part of the contents has already been treated in those subjects from a mathematical point of view, in this subject we will focus on their implementation in Python and their application in multiple real cases.

### Objectives and Contextualisation

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 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**

- Make effective use of bibliographical resources and electronic resources to obtain information.
- Solve problems related to the analysis of large volumes of data through the design of intelligent systems and computational learning.
- Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
- Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
- Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
- Using criteria of quality, critically evaluate the work carried out.
- Work cooperatively in a multidisciplinary context assuming and respecting the role of the different members of the team.

## **Learning Outcomes**

1. Identify and define computational solutions in multiple domains for decision making based on exploring alternatives, uncertain reasoning and task planning.
2. Learn and apply the most appropriate learning techniques for solving computational problems in distinct case studies.
3. Make effective use of bibliographical resources and electronic resources to obtain information.

4. Students must be capable of applying their knowledge to their work or vocation in a professional way and they should have building arguments and problem resolution skills within their area of study.
5. Students must be capable of communicating information, ideas, problems and solutions to both specialised and non-specialised audiences.
6. Students must have and understand knowledge of an area of study built on the basis of general secondary education, and while it relies on some advanced textbooks it also includes some aspects coming from the forefront of its field of study.
7. Understand and evaluate the results and limitations of the most common learning techniques.
8. Using criteria of quality, critically evaluate the work carried out.
9. Work cooperatively in a multidisciplinary context, taking on and respecting the role of the distinct members in the team.

## **Content**

### UNIT 1: INTRODUCTION

- 1.1 Basic concepts
- 1.2 History of machine learning

### UNIT 2: DATA REGRESSION

- 2.1 Gradient descent
- 2.2 Regularization

### UNIT 3: DATA CLASSIFICATION

- 3.1 Regularized 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**

All the information about the Course and related documents that students need will be found on the Caronte page (<https://caronte.uab.cat/course/index.php?categoryid=42>), on the subject's menu Aprenentatge Computacional (104403). This site will be used to view the materials, manage the practices groups, make the corresponding deliveries, see the notes, communicate with the teachers, etc. To be able to use it you have to do the following steps:

1. Register as a user giving the name, NIU, and a photo ID in JPG format. If you have already registered for another subject, you do not need to do it again, you can go to the next step.
2. Register for the type of teaching "Aprenentatge Computacional (104403)", giving as a subject code "apc2021" (without quotes).

In the development of the subject, five types of teaching activities can be differentiated:

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 Solving numerical problems: Solving a set of 3 problems proposed to students. All theory topics will be accompanied by a list of problems that the student must solve and submit. These activities must allow the student to deepen understanding and personalize theoretical knowledge in a specific numerical case. Therefore, some examples of data will be presented that require the design of a solution that uses the methods seen in the theory classes. It is impossible to follow the problem classes without following the contents of the theory classes. The result of these sessions is to achieve the necessary skills for solving problems that must be delivered according to the specific delivery mechanism that will be indicated on the subject's website (Caronte space).

MD2 Implementation of concise machine learning projects: Carrying out 2 coding projects to delve into applied aspects of the theory. The practical part of the course will be completed with practical sessions, where students must solve specific problems of a certain complexity implemented in python. These projects will be solved in small groups of 2-3 people, each member of the group must do a part and share it with the rest to have the final solution. These working groups must be maintained until the end of the course and must self-manage: role distribution, work planning, assignment of tasks, management of available resources, conflicts, etc. Although the teacher will guide the learning process, their intervention in group management will be minimal. To develop the project, the groups will work autonomously and the practical sessions will be dedicated mainly by the teacher to monitor the status of the project, indicate errors to correct, propose improvements, etc. The doubts that may arise for the realization of the practices will be transmitted through the Charon forum, where the other students will be able to answer them.

MD3 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 will consist 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/>

MD4 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.

MD5 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 the teaching activities that they have done the previous course, if they have passed. The repeaters will have to redo the individual theory tests (MD0).

Transversal competences

The transversal competences T01, T02 and T04 are worked on and evaluated at the home of the course in the following activities

- T01 Evaluate critically and with quality criteria the work done: in activities MD0 (study of theory), MD1 (completion of problems), MD2 (completion of laboratory practices) and MD3 (explain a practical case of computer learning)
- T02 Work cooperatively in a multidisciplinary context assuming and respecting the role of the different Team Members: in the practices of MD2 and the analysis of projects in MD3.
- T04 Effectively use the bibliography and electronic resources to obtain information: in the preparation of the MD0 theory material, and the preparation of the description of the practical case in MD3.

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	22	0.88	8, 2, 7, 6, 3
MD1: Problems resolution	8	0.32	8, 2, 1, 6, 4
MD2: Solution of Practical projects	16	0.64	8, 2, 7, 1, 6, 4, 9
Type: Supervised			
MD2: Project Programming	16	0.64	8, 2, 7, 1, 6, 5, 9
Type: Autonomous			
MD0: Individual study	10	0.4	8, 7, 1, 6, 5, 4, 3
MD1: Resolving problems (individually)	18	0.72	8, 2, 7, 1, 6, 5
MD2: Solving practical cases (in group)	22	0.88	8, 7, 1, 6, 5, 9, 3
MD3: Practical description in python of a Machine Learning case in a jupyter notebook	18	0.72	8, 2, 7, 6, 5, 4, 3

## Assessment

Evaluation activities and instruments:

a) Process and scheduled evaluation activities

The course consists of assessment activities:

- MD0: Theoretical exams, 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 compulsory, individual and recoverable (there will be two partials and their respective recoveries).

- MD1: Report delivery with solved problems, where each student will individually deliver a written report of up to 3 solved problems seen in classes. It represents 20% of the final grade, it is optional, individual and recoverable.

- MD2: Resolution of practices with delivery of report explaining the resolution and results of each practice, where each group made up of two people will deliver the python code for each of the two projects (regression and classification ) applied to 2 different databases, as well as a report of up to 30 pages where they will describe each database, the strategy they have used to analyze their data, as well as provide them with different parameter values they have tried and the results they have obtained with the best possible configuration. A presentation of each project will be made online. Represents 40% on the final grade, is mandatory, group and not recoverable.

- MD3: Preparation of a jgithub repository 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](https://www.kaggle.com/search?q=machine+learning)), and a presentation of the code, 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 optional, individual and recoverable (penalization of 20% is the presentation is not done).

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:

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 Partials, 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: Individual delivery of a report with resolved problems

The problems are intended to cause the student to enter the subject content continuously and, from small problems, to become directly familiar with the application of the theory. As evidence of this work, the obligatory presentation of a portfolio in which the student will have kept the problems that she will have been carrying out is required (competence T04).

Note Problems = Portfolio evaluation with 3 solved problems (according to the calendar indicated to Caronte).

A minimum of 2 problems must be submitted to pass this part. There will be recovery from problems (delivering the problems not delivered during the course the day of the second theory part).

- MD2: Resolution of group project

The evaluation of each of the 2 internship projects will include:

- Joint evaluation of each project (competence T02): single note 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 evaluation (competence T01): the individual work will be valued 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 brief confidential form will be evaluated, qualifying the contribution of each group partner to the final result.

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

$$\text{Practical Note} = (0.5 * \text{Project Note 1 Regression}) + (0.5 * \text{Project Note 2 Classification})$$

$$\text{Note 1 and 2 Projects} = (0.9 * \text{Group Note}) + (0.1 * \text{Individual Note})$$

$$\text{Group Note} = (0.3 * \text{Program}) + (0.1 * \text{Presentation}) + (0.6 * \text{Documentation})$$

There is no recovery of the practices: in case of not presenting a delivery or considering it copied, if the Final Project Grade does not exceed 5.0, the subject is considered NotPassed.

In very justified cases (eg for work, family or health issues, ...), instead of carrying out these 2 projects, the student will be able to carry out the so-called Coursera itinerary: with the prior approval of the teacher, the student who If you request it and justify it, you will be able to deliver the practices requested in the online computer learning course of the Coursera educational platform (<https://es.coursera.org/learn/machine-learning>). In this case, the maximum mark of practices that the student will be able to reach is 7 instead of 10 (since there is no report or presentation, only code is given in this itinerary).

- MD3: Developing a repository in python at Github describing a specific case of machine learning

The evaluation will be based on the python code at the Github platform 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 Kaggle case will be calculated according to the formula:

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

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 Using criteria of quality, critically evaluate the work carried out, Note Theory). With a PracticeProject Note MD2, teamwork will also be evaluated (T02 Work cooperatively in a multidisciplinary context assuming and respecting the role of the different members of the team, Group Note). With the completion of the problems 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 (T04 Make effective use of bibliographical resources and electronic resources to obtain information, Note Problems and Kaggle case).

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

$$\text{Final Note} = (0.2 * \text{Theory}) + (0.4 * \text{Project}) + (0.2 * \text{Problems}) + (0.2 * \text{Kaggle Case})$$

Conditions to approve:

To pass, it is necessary that the evaluation of each of the two compulsory activities (MD1 and MD3) 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 lower value between 4.5 and the weighted average of the notes:

- The final grade of theory MD0 must be greater than or equal to 4.0 to pass the theory part.
- The grade of the MD2 project must be greater than or equal to 5.0 in order to pass the practical part.
- The final grade for the course must be greater than or equal to 5.0 to pass the course.

In the case of not reaching the minimum required in any of the compulsory assessment activities (MD0 and MD2), if the calculation of the final grade for the course is equal to or greater than 5, a final grade of 4.5 will be placed. the subject to the file.

In case of not passing the subject because some of the compulsory evaluation activities do not reach the minimum final grade required (5.0), the numerical grade of the record will be the lower value between 4.5 and the weighted average of the grades.

If the student attends an exam or if s/he submits a practice, 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.

#### c) Recovery process

The student can take the recovery as long as they have been submitted to a set of activities that represent a minimum of two thirds (5 evaluation activities out of 8 total: 2 partial exams; 2 practice evaluations; 3 problem deliveries; 1 delivery of a notebook in python) of the total qualification of the subject.

Of these, students who have on average of all the activities of the subject a grade higher than 3.0 may be presented in the recovery.

Keep in mind that the Practice Resolution (MD2) is not recoverable.

#### d) Grade review procedure

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 of the partial exams and in none of the 2 evaluations of the practices.

#### 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 individual theoretical exam MD0, plus the marks corresponding to the practices of MD1, MD2 and MD3 obtained the first time the student has enrolled in the course, provided that the marks of practice are higher or equal to 5.0.

In order to be eligible for this differentiated assessment, the repeating student must ask the teacher at the latest until week 5.

## Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Delivery of problems	20%	4	0.16	8, 7, 1, 6, 4, 3
Individual theory tests	20%	4	0.16	7, 1, 6, 5, 4, 3
Jupyter notebook in python of a Machine Learning case	20%	4	0.16	8, 2, 1, 6, 5, 3
Programming of code projects	10%	2	0.08	8, 7, 1, 6, 4, 9
Written documentation, presentation, follow-up practical projects	30%	6	0.24	8, 2, 6, 5, 9, 3

## 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/35jalzm>

#### Basic bibliography:

- S. Russell, P. Norvig. Artificial Intelligence: A Modern Approach. Ed. Prentice Hall, Second Edition, 2003. (Existeix traducció al castellà: Inteligencia artificial: Un Enfoque Moderno)

#### 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 bibliography

- 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 transcend 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