Universitat Autònoma de Barcelona

2019/2020

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

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 and bioinspired paradigms

1.2 History of computer learning

UNIT 2: REGRESSION AND CLASSIFICATION

2.1 Regression of numerical data: gradient descent

2.2 Regularization and logistic regression

2.3 Classification of numerical data: support vector machines

2.4 Artificial neural networks: backpropagation algorithm

UNIT 3: CLUSTERING AND SEARCH

3.1 Memorization: lazy learning

3.2 Recommender systems: Content-based vs. Collaborative filtering

3.3 Clustering: k-means and Expectation-Maximization

3.4 Genetic algorithms

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=2), on the subject's menu Computational Learning (102787).

The different activities that will be carried out in the subject are organized as follows:

Theory classes

The main concepts and algorithms of each theoretical subject will be exposed for two hours per week. These subjects represent the starting point in the work of the subject and the classes are composed of two parts: during the first hour, the main theoretical and mathematical concepts related to specific tasks of machine learning will be explained (this will be the basis of the theoretical exam of the course, which will be compulsory and recoverable), and during the second hour, the code in python on Jupyter notebooks will be explained to exemplify the coding details to implement in a practical case the main concepts seen at the previous hour. The students will then be able to download the python notebooks and try all the codes on their computers, to perform the necessary tests and be able to play with the various parameters to fully understand the reasons for the different performances and accuracies achieved in a specific database with specific configurations.

Problem seminars

There will be classes of small groups of students to facilitate interaction. In these sessions, practical cases will be considered that require the design of a solution in which the methods seen in theory classes. It is impossible to follow the classes of problems if the contents of the theory classes have not been followed. The result of these sessions is to achieve the necessary competencies for the resolution of up to 4 problems that must be delivered mandatory at the 2nd week after its explanation at the problem seminar. The specific mechanism for deliver the solved exercises will be indicated on the web page ofthe subject (Caronte).

The delivery of the solved exercises will be compulsory and recoverable (problems can be delivered again the same day of the second part of the theoretical exam in January).

Laboratory work

Lab working groups will be formed by 2 students during the second week of the course. Work groups must be maintained until the end of the course and they must be self-managed: role play, work planning, assignment of tasks, management of available resources, conflicts, etc should be attained autonomously. 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 should be devoted mainly to the teacher in monitoring the project status, indicating errors to be corrected, proposing improvements, etc. The doubts that may arise for the realization of the practices will be transmitted through the forum of Caronte, where both students and teachers will be able to answer them.

During 6 sessions throughout the course, 3 different databases will be worked on 2 projects (regression/classification and neural networks), for which the 2 students will define their roles (there will be a coordinator for each of the 2 projects). During the semester, students will work cooperatively and must analyze each project to be solved, will design implement solutions based on different algorithms of machine learning seen in class, analyze the results obtained by each of the 3 databases, apply a methodology to justify the best parameters chosen, explain and justify the obtained performance, etc.

The first follow-up session will be INFORMATIVE of the first project, which will focused on applying 2 machine learning tasks (regression and classification) to 2 bases of specific data (the professor will assign the regression and classification databases to each group in the week 4 of the course). The second and fifth follow-up sessions will be of CONTROL of project 1 (regression and classification) and project 2 (neural networks) respectively, where the groups optionally present that the features of each database, and show that some minimum functionalities have already been achieved, that there is a strategy to get to the end of each project. Finally there will be the other 3 sessions of EVALUATION (the third, fourth and sixth sessions for the regression/classification project and the neural network project), in which a ZIP with the results of the projects will be delivered, both the Python code and also a report in PDF, toghether with a public presentation during the session of the work done, when the teacher can ask questions to the project members to assess the work done.

Assistance to information and control sessions is OPTIONAL, while the assistance in the sessions of evaluation is COMPULSORY for the 2 components of the group

The source code of the projects will be Python, the realization of the projects is mandatory and not recoverable.

Scientific work seminars

In the last two Fridays of December (during the two hours of problems), the work groups will have 10 minutesfor presenting a scientific project of machine learning that researchers from around the world have developed in open source code shared on the GitHub platform. The groups will choose a project from the following website:

https://medium.mybridge.co/amazing-machine-learning-open-source-tools-projects-of-the-year-v-2019-95d772e4e985

and send the chosen project to the professor before week 11 of the Course.

This presentation is optional and not recoverable.

In the case of repeaters, it is not necessary to re-do the projects nor the problems: when requested to the professor, these activities will be validated by considering the grade of last year, in case they had been approved (note> = 5.0). Repeaters will have to re-do the individual theoretical examen tests.

Transversal competences

- T01 Habits of thought (T01.02 Develop the capacity for analysis, prospective syntesii): in autonomous and supervised activities (study of the theory, realization of laboratory practices and realization of problems)

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

- T06 Personal attitude (T06.03 Generate innovative and competitive proposals in the professional activity): in the autonomous activities (study of the theory, participation in the forum of the subject in Caronte), directed (interaction in blackboard classes) and supervised (delivery of problems).

Activities

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Theory classes	26	1.04	5, 6, 8, 3, 7
Type: Supervised			
Laboratory work	12	0.48	1, 5, 6, 2, 8, 4, 7
Seminars of problems and oral defence of a scientific work	10	0.4	5, 6, 2, 8, 3, 7
Type: Autonomous			
Individual Study	40	1.6	5, 6, 2, 8, 3, 7
Preparation, programming, documentation and presentation of practical projects and scientific work	42	1.68	1, 5, 6, 2, 8, 4, 7

Assessment

Evaluation activities and instruments:

a. Scheduled process and evaluation activities

The subject consists of the following evaluation activities:

- Theoretical exams, where for each exam the student must answer individually by answering 5 questions (each one developed in a maximum of one sheet of paper) on concepts of machine learning seen in the theory classes. It represents 40% of the final qualification, it is obligatory and recoverable (there will be two partial and their respective recoveries on January 2020).

- Delivery of solved exercises, i.e. a written report of up to 4 solved problems seen in classes of problems (regression, backpropagation, memorization, and recommenders). It represents 10% of the final qualification, it is obligatory and recoverable (it can be delivered the day of the second partial of the Course on January).

- Resolution of practices with delivery of report explaining the resolution and the results of each practice, where each group composed of two people will deliver the code python for each one the two projects (regression/Classification; Neural networks) applied to 3 different databases, as well as a report where the students will describe the database, the strategy they have used to analyze their data, as well as the experimental results with different values of the parameters they have tested and the performances they have obtained with the best configuration possible. It represents 40% of the final qualification, it is obligatory and is not recoverable.

- Oral defense in gorup for 10 minutes of a scientific project developed by a group of researchers or engineers, as open source code in the Github platform. It represents 10% of the final qualification, it is not obligatory and is not recoverable.

Next, the procedure for passing the Course with continuous evaluation is described:

-Individual theoretical exams

The final note of theory shall be calculated from two partial examinations:

- Grade theory = (0.5 * Parcial1) + (0.5 * Parcial2)

Parcial1 is done in the middle of the semester and serves to eliminate some of the matter if approved. Parcial2 is done at the fiend of the school semester and serves to eliminate the part of the syllabus that comes after Parcial1.

These tests aim for an individualized evaluation of the student with her/his abilities to answer 5 long questions about the techniques explained in class, as well as to evaluate the level of conceptualization that the student has made of the techniques seen.

To pass the theory part of the Course, two requirements will have to be fulfilled:

- It will be necessary for the grades of the Partials 1 and 2 to be equal to or greater than 4.0 (in both partials). If less than a 4.0 is obtained in any of the two partials, the corresponding partial must be re-done during the recovery period in January-February 2020.

- The final grade of theory should be greater than or equal to 4.0. In the event that the final theory grade is not equal to or greater than 4.0, students may apply to the recovery exam to be evaluated for all the contents seen in the subject.

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

-Individual delivery of a report with solved exercises

The exercises are aimed at causing the student to enter with the contents of the subject on a continuous basis and, from small problems, to be familiarized directly in the application of the theory. As evidence of this work is asked for the compulsory presentation of a portfolio in which he has been keeping the problems that he will have been doing (T06 competition).

- Grade problems = Evaluation of the portfolio with a maximum of 4 problems resolved (according to the timetable indicated at Caronte).

One has to deliver a minimum of 2 exercises to pass this part. There will be recovery of the problems (delivering the problems not delivered during the courseon the day of the second partial of theory), since it is necessary to take into account inthe case of not delivering 2 minimum problems, it is considered the suspended subject.

-Lab Project Resolution in Group

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

- Joint evaluation of each project (T03 competence): A single note for all members of the Working Group that will assess the overall outcome of the project, the quality of the code, the overall structure of the fipresentationand the documents Delivered throughout the project.

- Individual evaluation (T01 competence): Individual work will be assessed from the answers to the questions in the control sessions, the final fipresentationof the project and mainly the active participation in the Charon forums. In the cases required by any group (in cases of incidents between peers), a short form will be assessed withfidential CalifiCando The contribution of each group partner to the result final.

The project note will be calculated according to the formula (standardized by a 60% Project 1 and 40% Project 2):

- Practical Grade = (0.3 * Grade Project 1 regression) + (0.3 * Grade Project 1 Classification) + (0.4 * Grade Project 2 neural networks)
- Grade Projects 1 and 2 = (0.9 * Group Grade) + (0.1 * Individual Grade)
- Group Grade = (0.3 * code) + (0.1 * presentation) + (0.6 * report)

As for Project 2 of Neural networks, the database will be the same for all groups, who will participate in a competition in the platform CODALAB and where the 3 lab groups that get the best performance scores will receive an additional point in the note grade of the Course.

There is no recovery of the practices: in case of not submitting a delivery or consider it copied, if the note project final does not exceed 5.0, the Course is considered the suspended.

In very exceptional cases (e.g. inability to attend to the project sessions due to work issues,...), instead of carrying out these 2 projects, the student may carry out the so-called itinerary Cousera: Previous approval of the professor, the student who requested it can deliver the practices of the online course of computational learning of the educational platform Coursea (https://es.coursera.org/learn/machinelearning)

-Oral defense in group of a scientific work

The oral defense or seminary aims to cause the group to see and understand how another group of engineers or scientificfihave addressed a computational learning problem, since there is a shared source Code on theGitHub platform. For 10 minutes, the group (which may be the same as practices, or NOT) will explain the algorithm that has been used, the strategy of how the data analysis has been addressed, and the results. As evidence of this work, the non-compulsory presentation of the chosen project (T06 competition) is requested.

Grade Oral Defense = presentation of a chosen project (from the Web https://medium.mybridge.co/amazing-machine-learning-open-source-tools-projects-of-the-year-v-2019-95d772e4).

There is no recovery of oral defense, nor is there a minimum note in this activity to approve the subject.

-Evaluation of transversal competences

The partial examinations will allow to evaluate their acquisition of habits of thought and of personal work (T01 habits of thought, note theory). With the grade of the projects, the teamwork (T03 teamwork, group note) and the individual presentation (T01 habits of thought, Individual note) will be also evaluated. With the realization of

the problems and the realization of a seminar of a scientific project, the acquisition of habits to solve a predetermined task with data totally different from those seen in class will be evaluated (T06 Personal attitude, note problems and seminary).

The final note of the Course is obtained by combining the evaluation of these 4 activities in the following way:

Final Grade = (0.4 * theory) + (0.4 * project) + (0.1 * problems) + (0.1 * seminar)

Conditions to pass the Course:

To pass the Course, it is necessary that the evaluation of each of the obligatory parts exceeds the minimum demanded and that the total evaluation exceeds the 5 points. In case of not exceeding the threshold, the numerical note assigned to the Course will be the lower value between 4.5 and the weighted average of the grades:

- The theoretical grade must be greater than or equal to 4.0 in order to pass the theory part.
- The project grade must be greater than or equal to 5.0 in order to pass the project part.
- Two minimum problems must be delivered to pass the exercise part.
- The final grade of the Course must be greater than or equal to 5.0 in order to approve the Course.

In case of not reaching the minimum required in any of the obligatory activities of evaluation (theory, problems or practices), if the calculation of the final grade of the Course were equal to or greater than 5, a final grade 4.5 would be assigned to the Course.

In case of not passing the Course because one of the obligatory evaluation activities does not reach the minimum required grade (4.0 or 5.0), the numerical note of the dossier shall be the lowest value between 4.5 and the weighted average of the grades.

If the student only applies to one of the two partial exams or to any of the projects, s/he can no longer be evaluated as "non-evaluative", but the final grade will be computed based on the submitted evaluations.

b. Scheduling evaluation Activities

The dates of continuous evaluation and delivery of works will be published in Caronte (http://caronte.uab.cat/), in the space of this Coyrse and may be subject to changes in programming for reasons of adaptation to possible incidents; Caronte.uab.cat will always be updated about any change, and this platform will constitute the usual mechanism of information exchange between students and professor.

The following scheduling is envisaged:

- Individual theoretical exams: Weeks 11 and 16-17 of the Course.
- Individual delivery of a report with the problems resolved: Delivery weeks 7, 9, 12 and 15 of the Course.
- Resolution in practiceGroup: Evaluation weeks 8, 10, and 14 of the subject. Group oral Defense of a work scientificfiCo: Weeks 14 and 15 of the Course.

C. Recovery process

The student may apply for a re-do procedure if s/he has delivered a minimum of two thirds of the evaluatio activities (6 evaluation activities of 10 total: 2 partial examinations; 3 project evaluations; 4 deliveries of exercises and 1 oral defense) of the total qualification of the Course.

Of these, students who have as a mean of all the activities of the subject a Calification higher than 3.0 can be presented in the recovery.

It should be kept in mind that the resolution of practices and the Oral defense in group of the scientific project are not recoverable.

d. Procedure for therevision of the Qualifications

For each individual theoretical examination, a place, date and time of review will be indicated in which the student will be able to review the activity with the teacher. In this context, claims may be made on the note of the activity, which will be assessed by the faculty responsible for the subject. If the student is not present in this review, this activity will not be reviewed later.

e. Qualifications

Outstanding grades (MH): Honor grades will be awarded up to five percent or fraction of the students enrolled in all the teaching groups of the Course. The regulations of the UAB indicate that MH can only be granted to students who have obtained a final qualification equal to or greater than 9.00.

Non-evaluative: A student shall be deemed not to be assessed (NA) if s/he has not delivered any partial exams nor any of the 3 deliveries of the projects.

f. Irregularities by the student, copying and plagiarism

Without prejudice to other disciplinary measures that are deemed opportune, the irregularities committed by a student, that can lead to a variation of the qualification during the evaluation of an activity, will be qualified as zero. Therefore, copying, plagiarism, deception, copying, etc. in any of the evaluation activities will mean suspending the activity with a zero. The evaluation activities qualified in this way and by this procedure will not be recoverable. If itis necessary to pass any of these evaluation activities to pass the Course, the Course will be suspended directly, without opportunity to recover it in the same course. These irregularities include, but are not limited to: the total or partial copying of a project code, report, or any other evaluation activity; let copy; to present a group work not entirely done by the members of the group (applied to all members, not only those who have not worked); To present as own materials produced 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, camera pens, etc.) accessible during the theoretical individual evaluation tests; talk to peers during the individual theoretical tests (exams); Copying or attempting to copy from other students during the theoretical-practical evaluations (tests); use or attempt to use material-related writings during the course of the theoretical tests (examinations), where these have not been explicitly permitted.

The numerical note of the dossier shall be the lowest value between 3.0 and the weighted average of the notes if the student has committed irregularities in an evaluation act. For the rest of the Course, the student who has committed irregularities will not be evaluated at any of the remaining evaluation activities to be carried out.

In summary: Copy, let copy or plagiarize (or the intent of) in any of the activities of evaluation equals a SUSPENSe, not compensable or recoverable and without convalidations of parts of the subject in later courses.

g. Evaluation of repeating students

For repeaters, the evaluation of the Course will consist of the individual theoretical examination and the delivery of problems, plus the grade corresponding to the projects obtained the first time that the student enrolled in the Course, provided that the grade of the projects is greater than or equal to 5.0.

In order to be eligible for this differentiated assessment, repeater students must ask the professor via e-mail no later than October 1st, 2019.

Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Delivery of problems	10%	2	0.08	5, 6, 2, 8, 3, 7

Individual theory tests	40%	6	0.24	5, 6, 2, 8, 3, 7
Presentation of a scientific work	10%	2	0.08	1, 5, 2, 8, 3, 7
Programming of code projects	20%	5	0.2	1, 5, 6, 2, 8, 4, 7
Written documentation, presentation, follow-up practical projects	20%	5	0.2	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/

Basic bilbiography:

- 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 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