COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

COURSE SYLLABUS
2021-2022

GENERAL INFORMATION

Course information	
Name	Machine Learning
Code	DOI-MIC-515
Degree	Máster Universitario en Ingeniería Industrial + Máster en Industria Conectada [1st year]
Semester	2nd (Spring)
ECTS credits	6.0
Type	Compulsory
Department	Industrial Organization
Coordinator	Antonio Muñoz San Roque

Lecturer	
Name	Antonio Muñoz San Roque
Department	Electronics, Control and Communications
Office	D-515 (Alberto Aguilera, 25)
e-mail	amunoz@comillas.edu
Phone	91.540.61.47
Office hours	Arrange an appointment through email.
Lab instructor	
Name	Fernando San Segundo Barahona
Department	Applied Mathematics
Office	D-201 (Alberto Aguilera, 25)
e-mail	fsansegundo@comillas.edu
Phone	91.542.28.00 – Ext. 2377
Office hours	Arrange an appointment through email.

DETAILED INFORMATION

Contextualization of the course

Contribution to the professional profile of the degree

The purpose of this course is to provide students with a fundamental understanding and an extensive practical experience of how to extract knowledge from an apparently unstructured set of data.

By the end of the course, students will:

- Understand the basic principles behind machine learning.
- Have practical experience with the most relevant machine learning algorithms.
- Have well-form criteria to choose the most appropriate techniques for a given application.

Prerequisites

Students willing to take this course should be familiar with linear algebra, basic probability and statistics, machine learning, and undergraduate-level programming. Previous experience with the R programming language is also desired although not strictly required.

CONTENTS

Contents
Theory and laboratory
Unit 1. Introduction
1.1 Machine learning 1.2 The learning process 1.3 Types of machine learning 1.4 Resources
Unit 2. Classification methods
2.1 The classification problem 2.2 Logistic regression 2.3 Discriminant analysis 2.4 K-nearest neighbors 2.5 Decision trees 2.6 Support vector machines 2.7 Multilayer perceptrons for classification
Unit 3. Regression methods
3.1 The regression problem 3.2 Linear regression. Model selection and regularization 3.3 Polynomial regression 3.4 Splines 3.5 Generalized additive models 3.6 Multilayer perceptrons for regression
Unit 4. Time series forecasting
4.1 Introduction 4.2 Decomposition methods 4.3 Stochastic processes 4.4 Exponential smoothing 4.5 ARIMA models 4.6 Dynamic regression models
Unit 5. Unsupervised learning
5.1 Probability density estimation 5.2 Dimensionality reduction methods 5.3 Clustering and vector quantization 5.4 Self-organizing feature maps

Competences and learning outcomes

Competences[1]

General competences

CG1.	Have acquired advanced knowledge and demonstrated, in a research and technological or highly specialized context, a detailed and well-founded understanding of the theoretical and practical aspects, as well as of the work methodology in one or more fields of study. *Haber adquirido conocimientos avanzados y demostrado, en un contexto de investigación científica y tecnológica o altamente especializado, una comprensión detallada y fundamentada de los aspectos teóricos y prácticos y de la metodología de trabajo en uno o más campos de estudio.*
CG2.	Know how to apply and integrate their knowledge, understanding, scientific rationale, and problem-solving skills to new and imprecisely defined environments, including highly specialized multidisciplinary research and professional contexts. *Saber aplicar e integrar sus conocimientos, la comprensión de estos, su fundamentación científica y sus capacidades de resolución de problemas en entornos nuevos y definidos de forma imprecisa, incluyendo contextos de carácter multidisciplinar tanto investigadores como profesionales altamente especializados.*
CG3.	Know how to evaluate and select the appropriate scientific theory and the precise methodology of their fields of study in order to formulate judgements based on incomplete or limited information, including, when necessary and pertinent, a discussion on the social or ethical responsibility linked to the solution proposed in each case. *Saber evaluar y seleccionar la teoría científica adecuada y la metodología precisa de sus campos de estudio para formular juicios a partir de información incompleta o limitada incluyendo, cuando sea preciso y pertinente, una reflexión sobre la responsabilidad social o ética ligada a la solución que se proponga en cada caso.*
CG4.	Be able to predict and control the evolution of complex situations through the development of new and innovative work methodologies adapted to the scientific/research, technological or specific professional field, in general multidisciplinary, in which they develop their activity. *Ser capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador, tecnológico o profesional concreto, en general multidisciplinar, en el que se desarrolle su actividad.*
CG5.	Be able to transmit in a clear and unambiguous manner, to specialist and non-specialist audiences, results from scientific and technological research or state-of-the-art innovation, as well as the most relevant foundations that support them. *Saber transmitir de un modo claro y sin ambigüedades, a un público especializado o no, resultados procedentes de la investigación científica y tecnológica o del ámbito de la innovación más avanzada, así como los fundamentos más relevantes sobre los que se sustentan.*
CG6.	Have developed sufficient autonomy to participate in research projects and scientific or technological collaborations within their thematic area, in interdisciplinary contexts and, where appropriate, with a high knowledge transfer component. *Haber desarrollado la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro de su ámbito temático, en contextos interdisciplinares y, en su caso, con una alta componente de transferencia del conocimiento.*
CG7.	Being able to take responsibility for their own professional development and their specialization in one or more fields of study. *Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio.*

[1] Competences in English are a free translation of the official Spanish version.

Specific competences	
CE3.	Be able to design and train systems that learn automatically, mastering both supervised and unsupervised learning techniques. Understand the potential application of these systems in the improvement of industrial processes, relations with clients, etc. *Ser capaces de diseñar y entrenar sistemas que aprendan de manera automática, dominando tanto las técnicas de aprendizaje supervisado como no supervisado. Entender el potencial de aplicación de estos sistemas en la mejora de procesos industriales, relación con clientes, etc.*
Learning outcomes	
By the end of the course students should:	
RA1.	Understand the basic principles behind machine learning.
RA2.	Have practical experience with the application of the most relevant machine learning algorithms.
RA3.	Have well-formed criteria to choose the most appropriate techniques for a given application.

TEACHING METHODOLOGY

General methodological aspects	
Each session will combine theory and practice. The instructor will explain the basics of the subject and will go into depth in the more important issues with illustrative examples. Students will be grouped in pairs in order to put the proposed methods and techniques in practice in a collaborative way.	
In-class activities	**Competences**
▪ **Lectures:** The lecturer will introduce the fundamental concepts of each unit, along with some practical recommendations, and will go through worked examples to support the explanation. Active participation will be encouraged by raising open questions to foster discussion and by proposing short application exercises to be solved in class either on paper or using a software package.	CG1, CG3, CG7, CE3
▪ **Lab sessions:** Under the instructor's supervision, students, divided in small groups, will apply the concepts and techniques covered in the lectures and will become familiar with the practical application of the most relevant algorithms using software tools and libraries.	CG1, CG2, CG3, CG4, CG5, CG6, CG7, CE3
▪ **Tutoring** for groups or individual students will be organized upon request.	–
Out-of-class activities	**Competences**
▪ Personal study of the course material and resolution of the proposed exercises.	CG1, CG3, CG7, CE3
▪ Lab session preparation, analysis of the results and report writing.	CG1, CG2, CG3, CG4, CG5, CE3

ASSESSMENT AND GRADING CRITERIA

Assessment activities	Grading criteria	Weight
Midterm exam	▪ Understanding of the theoretical concepts. ▪ Application of these concepts to problem-solving. ▪ Critical analysis of numerical exercises' results.	15%
Final exam	▪ Understanding of the theoretical concepts. ▪ Application of these concepts to problem-solving. ▪ Critical analysis of numerical exercises' results.	35%
Lab sessions and reports	▪ Application of theoretical concepts to real problem-solving. ▪ Ability to use and develop data mining and machine learning software. ▪ Attitude and effort: Initiative and proactive work will be encouraged ▪ Written communication skills.	50%

GRADING AND COURSE RULES

Grading

Regular assessment

- **Theory** will account for 50%, of which:
  - Midterm: 15%
  - Final exam: 35%
- **Lab** will account for the remaining 50%

In order to pass the course, the weighted average mark must be greater or equal to 5 out of 10 points, and the mark of the final exam must be greater or equal to 4 out of 10 points. Otherwise, the final grade will be the lower of the two marks.

Retake

Lab marks will be preserved. In addition, all students will take a final exam. The resulting grade will be computed as follows:

- Final exam: 50%
- Lab practices: 50%

As in the regular assessment period, in order to pass the course, the weighted average mark must be greater or equal to 5 out of 10 points, and the mark of the final exam must be greater or equal to 4 out of 10 points. Otherwise, the final grade will be the lower of the two marks.

Course rules

- Class attendance is mandatory according to Article 93 of the General Regulations (Reglamento General) of Comillas Pontifical University and Article 6 of the Academic Rules (Normas Académicas) of the ICAI School of Engineering. Not complying with this requirement may have the following consequences:
  - Students who fail to attend more than 15% of the lectures may be denied the right to take the final exam during the regular assessment period.
  - Regarding laboratory, absence to more than 15% of the sessions can result in losing the right to take the final exam of the regular assessment period and the retake. Missed sessions must be made up for credit.
- Students who commit an irregularity in any graded activity will receive a mark of zero in the activity and disciplinary procedure will follow (cf. Article 168 of the General Regulations (Reglamento General) of Comillas Pontifical University).

WORK PLAN AND SCHEDULE[2]

In and out-of-class activities	Date/Periodicity	Deadline
Midterm exam	Session 15	–
Final exam	Last week	–
Lectures and lab sessions	Twice a week	–
Review and self-study of the concepts covered in the lectures	After each lesson	–
Lab preparation	Before every lab session	–
Lab report writing	–	One week after the end of each session

STUDENT WORK-TIME SUMMARY			
IN-CLASS HOURS			
Lectures	Lab sessions		Assessment
28	28		4
OUT-OF-CLASS HOURS			
Self-study		Lab preparation and reporting	
60		60	
		ECTS credits:	6 (180 hours)

BIBLIOGRAPHY

Basic bibliography

- Slides prepared by the lecturer (available in Moodlerooms).
- G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, 2013.

Complementary bibliography

- M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, 2013.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, 2nd Ed., Springer, 2009.
- E. Alpaydin, *Introduction to Machine Learning*, 3rd Ed., MIT Press, 2014.
- S. Marsland, *Machine Learning: An Algorithmic Perspective*, 2nd Ed., Chapman & Hall/CRC Machine Learning & Pattern Recognition, 2015.
- T. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd Ed., Wiley-Interscience, 2000.
- C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.
- S. Haykin, *Neural Networks. A Comprehensive Foundation*, 2nd Ed., Pearson, 1999.
- W. Wei, *Time Series Analysis. Univariate and Multivariate Methods*, 2nd Ed., Addison-Wesley, 2006.

[2] A detailed work plan of the subject can be found in the course summary sheet (see following page). Nevertheless, this schedule is tentative and may vary to accommodate the rhythm of the class.

	In-class activities				
Session	**Time [h]**	**Session**	**Theory**	**Laboratory**	**Assessment**
1	2	Introduction I	Introduction to machine learning	Lab 1.1: Introduction to R for machine learning	
2	2	Introduction II		Lab 1.2: Introduction to R for machine learning	
3	2	Classification I	The classification problem	Intro assignment	
4	2	Classification II	Logistic regression	Lab 2.1	
5	2	Classification III	Discriminant analysis. KNN	Lab 2.2	
6	2	Classification IV	Decision trees	Lab 2.3	
7	2	Classification V	SVM	Lab 2.4	
8	2	Classification VI	MLP	Lab 2.4	
9	2	Classification VII		Hackathon	Assignment 1
10	2	Regression I	The regression problem. Linear regression	Lab 3.1	
11	2	Regression II	Model selection and regularization	Lab 3.2	
12	2	Regression III	Polynomial regression, splines, GAMs	Lab 3.3	
13	2	Regression IV	MLP, SVM	Lab 3.4	
14	2	Regression V		Hackathon	
15	2	Midterm exam			Midterm exam
16	2	Forecasting I	Decomposition methods Stochastic processes.	Lab 4.1	
17	2	Forecasting II	ARIMA	Lab 4.2	
18	2	Forecasting III	ARIMA	Lab 4.3	
19	2	Forecasting IV	SARIMA	Lab 4.4	
20	2	Forecasting V	Dynamic regression models I	Lab 4.5	
21	2	Forecasting VI	Dynamic regression models II	Lab 4.6	
22	2	Forecasting VII	Nonlinear models	Lab 4.7	
23	2	Forecasting VIII		Hackathon	Assignment 2
24	2	Density estimation	Parametric and non-parametric methods	Lab 5.1	
25	2	Dimensionality reduction	PCA, ICA	Lab 5.2	
26	2	Clustering I	Hierarchical and partitional clustering	Lab 5.3	
27	2	Clustering II	Vector quantization. Neural gas	Lab 5.4	
28	2	Self-organizing maps	SOM	Lab 5.5	
29	2	Course summary			
30	2	Final exam			Final exam