GENERAL INFORMATION

Course information	
Name	Machine Learning
Code	DOI-MIC-515
Degree	MIC, MII, MIT
Year	
Semester	Spring
ECTS credits	6 ECTS
Type	Elective
Department	DOI
Area	
Coordinator	Antonio Muñoz

Lecturer	
Name	Antonio Muñoz
Department	Electronics, Control Engineering and Communications
Area	
Office	D-513
e-mail	amunoz@comillas.edu
Phone	915 406 147
Office hours	Ask for an appointment by email.
Lecturer	
Name	José Portela
Department	Quantitative Methods
Area	
Office	IIT-D302
e-mail	jportela@comillas.edu
Phone	
Office hours	Ask for an appointment by email.
Lecturer	
Name	Guillermo Mestre
Department	
Area	
Office	IIT-P304
e-mail	gmestre@comillas.edu
Phone	
Office hours	Ask for an appointment by 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-formed 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, and undergraduate-level programming. Previous experience with R programming language desired although not strictly required.

CONTENTS

Contents

CHAPTER 1: INTRODUCTION

1.1. Data Mining & Machine Learning
1.2. The learning process
1.3. Types of Machine Learning

CHAPTER 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

CHAPTER 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
3.7. Radial Basis Function Networks

CHAPTER 4: TIME SERIES FORECSTING

4.1. Stochastic Processes
4.2. Exponential Smoothing
4.3. Decomposition methods
4.4. ARIMA models
4.5. Dynamic Regression models
4.6. GARCH models
4.7. Advanced methods for forecasting

CHAPTER 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 and Learning Outcomes	
Competences	
General Competences	
CG3.	The capability of adapting to new theories, methods and changing engineering situations based on a sound technical training.
CG4.	The capability of solving problems with personal initiative, efficient decision making, critical reasoning and transmitting technical information in the engineering world.
CG5.	The capability of conducting measurements, calculations, assessments, studies, reports, planning, etc.
CG10.	The ability to work in a multilingual and multidisciplinary environment.
Basic Competences	
Specific Competences	

Learning outcomes
RA1. The student understands the basic principles behind Machine Learning. RA2. The student has a practical experience with the application of the most relevant Machine Learning algorithms. RA3. The student has 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 teacher will explain the basics of the subject and will go in depth in the more important issues with illustrative examples. The students will be grouped in pairs in order to put in practice the proposed methods and techniques using software tools in a collaborative way.

In-class activities

1. **Lectures and problem-solving sessions (28 hours)**: The lecturer will introduce the fundamental concepts of each chapter, 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.
2. **Lab sessions (28 hours):** Under the instructor's supervision, students, divided in small groups, will apply the concepts and techniques covered in the lectures to real problems and will become familiar with the practical application of the most relevant algorithms using software tools and libraries.
3. **Assesment (4 hours)**

Off-class activities

1. **Personal study** of the course material and resolution of the proposed exercises (60 hours)
2. **Lab session** preparation, analysis of results and reporting (60 hours).

ASSESSMENT AND GRADING CRITERIA

Assessment activities	Grading criteria	Share
Mid-term 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:
  - Mid-term: 15%
  - Final exam: 35%

- **Lab** will account for the remaining 50%

In order to pass the course, the mark of the final exam must be greater or equal to 4 out of 10 points.

Retakes

Lab practice 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 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[1]

In and out-of-class activities	Date/Periodicity	Deadline
• Mid-term exam	Session 15	-
• Final exam	Last week	-
• Lectures + Lab sessions	Weekly	-
• Review and self-study of the concepts covered in the lectures	Weekly	-
• Lab preparation and reporting	Weekly	One week after the end of each lab session

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

BIBLIOGRAPHY

Basic

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

Complementary

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

[1] 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	Date	h/s	SESSION	THEORY	LAB	ASSESMENT
1	09-jan	2	Introduction I	Introduction to Machine learning	Lab Practice 1.1: Introduction to R for Machine Learning	
2	11-jan	2	Introduction II		Lac Practice 1.2: Introduction to R for Machine Learning	
3	16-jan	2	Classification I	The classification problem. Logistic regression.		
4	18-jan	2	Classification II		Lab Practice 2.1	
5	23-jan	2	Classification III	Discrimninant analysis. KNN	Lab Practice 2.2	
6	25-jan	2	Classification IV	Decision trees	Lab Practice 2.3	
7	30-jan	2	Classification V	SVM	Lab Practice 2.4	Assignment 1
8	01-feb	2	Classification VI	MLP	Lab Practice 2.5	
9	06-feb	2	Classification VII	MLP	Lab Practice 2.6 (hackathon)	
10	08-feb	2	Regression I	The regression problem. Linear Regression.	Lab Practice 3.1	Assignment 2
11	13-feb	2	Regression II	Model selection and Regularization	Lab Practice 3.2	
12	15-feb	2	Regression III	Polynomial Regression, Splines, GAMs	Lab Practice 3.3.	
13	20-feb	2	Regression IV	MLP, SVM. Ejemplo sintético generado a partir de los datos de mercado para que vean el efecto de las no linealidades	Lab Practice 3.4	
14	22-feb	2	Regression V		Lab Practice 3.5 (assignment)	Assignment 3
15	27-feb	2	Mid-term exam I			Mid-term exam
16	01-mar	2	Forecasting I	Stochastic Processes. Decomposition methods	Lab Practice 4.1	
17	06-mar	2	Forecasting II	Exponential Smoothing	Lab Practice 4.2	
18	08-mar	2	Forecasting III	ARMA	Lab Practice 4.3	
19	13-mar	2	Forecasting IV	ARIMA	Lab Practice 4.4	
20	15-mar	2	Forecasting V	SARIMA	Lab Practice 4.5	
21	20-mar	2	Forecasting VI	Dynamic Regression models I	Lab Practice 4.6	Assignment 4
22	22-mar	2	Forecasting VII	Dynamic Regression models II	Lab Practice 4.7	
23	03-apr	2	Density estimation I	Parametric & Non-parametric methods	Lab Practice 5.1	Assignment 5
24	05-apr	2	Density estimation II	NN for density estimation	Lab Practice 5.2	
25	10-apr	2	Dimensionality reduction	PCA. ICA.	Lab Practice 5.3	
26	12-apr	2	Clustering I	Hierarquical & partitional clustering	Lab Practice 5.4	
27	17-apr	2	Clustering II	Vector Quantization. Neural Gas	Lab Practice 5.5	
28	19-apr	2	Self Organising Maps	SOM	Lab Practice 5.6	
29	24-apr	2	Course summary			Assignment 6
30	26-apr	2	Final exam			Final exam