

Week 1 – Information of the course: **Introduction to Machine Learning (IML)**

Course 2021-2022

Dr. Maria Salamó
maria.salamo@ub.edu

Dept. Mathematics and Informatics,
Faculty of Mathematics and Informatics,
University of Barcelona

Brief description

- This course provides an **introduction on Machine Learning**
- It gives an overview of many concepts, techniques and algorithms in machine learning, beginning with topics such as **classification** and **linear regression** and ending up with more recent topics such **support vector machines** and **recommender systems**

Brief description

- The course is divided into three main topics:
 - supervised learning, unsupervised learning, and machine learning theory
- Topics include:
 - (i) **Supervised learning** (linear decision, non linear decision)
 - (ii) **Unsupervised learning** (clustering, factor analysis, and visualization)
 - (iii) **Learning theory** (bias/variance theory, empirical risk minimization)
- The course will also draw from numerous case studies and applications, so that you'll also learn how to apply learning algorithms to computer vision, medical informatics, and signal analysis

Summary

Introduction to Machine Learning

Unsupervised
Learning

Supervised
Learning

Decision
Learning
Theory

Cluster
Analysis

Factor
Analysis

Visualization

Non Linear Decision

Linear
Decision

Basic
concepts of
Decision
Learning
Theory

K-Means,
Fuzzy C-means,
EM

PCA, ICA

Self Organized
Maps (SOM) ,
Multi-
Dimensional
Scaling

Lazy
Learning
(K-NN, IBL,
CBR)

Overfitting,
model selection
and feature
selection

Kernel
Learning

Ensemble
Learning
(Trees,
Adaboost)

Perceptron,
SVM

Bias/Variance,
VC dimension,
Practical advice
of how to use
learning
algorithms

Contents

1. Introduction to Machine Learning
2. Introduction to unsupervised learning
3. Cluster analysis
 - a) Classification of clustering algorithms
 - b) K-Means, Bisecting K-Means, Fuzzy C-means
 - c) EM. Introduction to Mixture of Gaussians
4. Factor analysis
 - a) Principal Components Analysis (PCA)
 - b) Independent Component Analysis (ICA)

5. Visualization

- a) Self-Organized Maps (SOM)
- b) Multi-dimensional Scaling (MDS)

6. A gentle introduction to supervised learning

- a) The linear regression model
- b) Descent optimization methods
- c) Application of the learning model
- d) Tour on Machine Learning terms

7. Lazy Learning

- a) Nearest Neighbour (NN) and kNN
- b) Instance-based Learning (IBL)
- c) Case-based Reasoning (CBR) foundations

8. Feature Selection

- a) Description of Wrappers, Filters, and embedded
- b) Feature Selection Perspectives (Search directions, Search Strategies)
- c) Measures for making the selection (based on information, distance, dependence, consistency, accuracy)

9. Model Selection

- a) Introduction to Model evaluation (performance metrics, confusion matrix, ROC curves, etc.)
- b) Model evaluation (hold-out, cross-validation, overfitting and underfitting, bias vs variance, regularization, etc.)

10. Kernel Learning

- a) Statistical learning theory
- b) Support Vector Machines (SVM)

11. Recommender Systems

- a) Introduction to recommendation techniques
- b) An overview of Collaborative Filtering
- c) An overview of Content-based Filtering
- d) Conversational Recommenders

12. Ensemble Learning

- a) Introduction to ensemble learning
- b) Additive model: Bagging
- c) Additive model: Boosting

Face-to-face scenario (expected model)

- **Theoretical teaching** (Tuesday 10:00 to 12:00 room B2 at UB)
 - Between 50% to 100% of in-person activities
- **Theory/Practical**
 - Between 50% to 100% of in-person activities
 - **Face-to-face Theory/Practical** (Tuesday 12:00 to 13:00 room ID **Group 1**)
 - **Face-to-face Theory/Practical** (Tuesday 13:00 to 14:00 room ID **Group 2**)

Face-to-face scenario (expected model)

Theory/Practical Sessions

1. Teacher Maria Salamó (maria.salamo@ub.edu)
2. The time will be devoted to explain the **description** and **support material** for doing the practical exercises and solving doubts about these exercises.
3. **IMPORTANT No recordings of the face-to-face class will be performed**

Theoretical Sessions

1. Teacher Maria Salamó (maria.salamo@ub.edu)
2. The time will be devoted to:
 1. Summarize (30 to 60 minutes) the theory blocks of this week
 2. Perform question and answer sessions about theory
3. **IMPORTANT the support material explained in the face-to-face class will not be recorded**
4. The student may ask questions or may formulate a specific doubt about a topic. The teacher may repeat specific explanations to clarify or to provide support.
5. Short videos or additional material will be shared in the campus virtual by the teacher in case of specific common doubts to all students

Face-to-face scenario (expected model)

Campus Virtual (<https://campusvirtual.ub.edu>)

1. Provisional calendar
2. Theory – Slides of the course
 - Slides with the theory
 - Videos in specific theory blocks if necessary
3. Practical - Exercise text and instructions
 - Support slides for the practical
 - Description of the practical exercise in a pdf
 - Deliverable link in which the student will upload the code and the report of the practical exercise
4. Forum
 - Tool in which the student will write questions to the teacher. It allows to share the doubt and the answer to all the students, to avoid to repeat explanations.
 - Do not hesitate to use it! Other students might have the same doubt as you.
 - In addition, you can also help your mates by answering their questions if you have the answer.

Face-to-face scenario (expected model)

Common doubts?

1. Am I obliged to attend the face-to-face session?
 - No, you are not obliged, however it is recommended
 - The support material will be available in the campus virtual and is sufficient.
 - Beware! No face-to-face video will be shared, and forum will answer to specific doubts
 - A week questionnaire will be solved in these sessions

2. Am I obliged to attend practical sessions?
 - No, you are not obliged, however it is recommended
 - The teaching material will be available in the campus virtual and is sufficient.
 - Beware! No theory video will be shared every week, and practical sessions and forum will answer to specific doubts (question/answer) of the practicals

Organization of the course

Online scenario (in case of COVID restrictions)

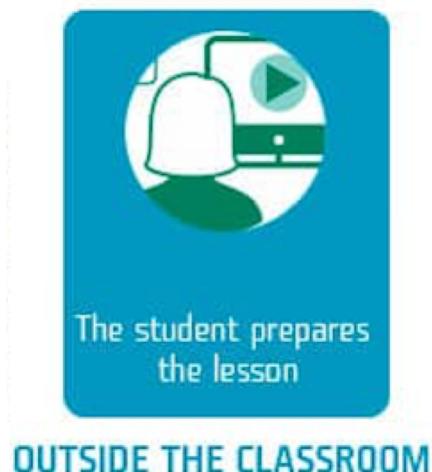
- **Online Theory/Practical** (Tuesday 12:00 to 13:00 **Group 1**) Synchronous
- **Online Theory/Practical** (Tuesday 13:00 to 14:00 **Group 2**) Synchronous
- **Online Session** (Tuesday 10:00 to 12:00 – **Group 1 and Group 2**)

Organization of the course

Methodology

- Every week the corresponding theoretical lesson will appear in several blocks in campus virtual
 - Every Tuesday you will have the new lesson available on campus virtual
- During one week students prepare the lesson
- Next Tuesday **the theoretical session** will be devoted to summarize, discuss the contents of the lesson and question/answering
- A theoretical questionnaire will be done in the face-to-face theoretical class
- The practical exercises are done in the **face-to-face theoretical/practical classes** and are related to several lessons

**It is very important
your active role
and to follow every
week the
theoretical lessons**



Activities – work deliveries

- Work 1 – (W1)
 - Clustering exercise
- Work 2 – (W2)
 - Factor Analysis exercise
- Work 3 – (W3)
 - Lazy Learning or kNN Recommenders
- Work 4 – (W4)
 - Kernels exercise (mandatory)



*This year work
is done in
groups of three*

Evaluation

- The course is divided into two parts:
 - **Exam**: an exam at the end of the term
 - **Work**: Work deliveries during the semester



Mark = a x Exam + b x Work if exam>=3,5 and Work>= 4,5

This year **a** and **b** will be established as: $a = 0,4$ and $b = 0,6$

Exam = an exam at the end of the term (11th January 2022)

Work = 0,30 x W1 + 0,2 x W2 + 0,30 x W3 + 0,2 x Tests

- Some Works may have an exam after the delivery. The mark will be a part of the Wx.
- **Tests** are the weekly theoretical questionnaires

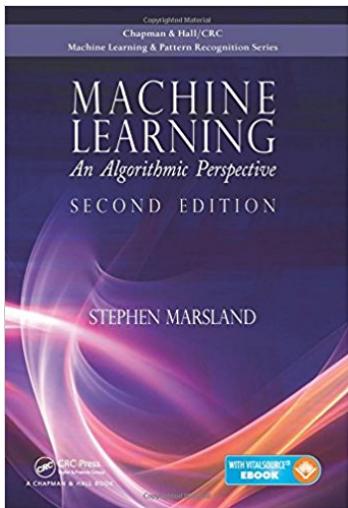
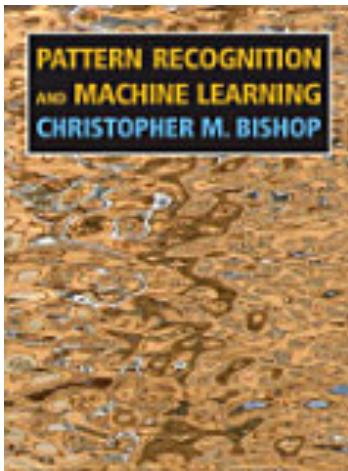
Calendar

Week	Month	Tue	Theory Block	Theoretical class	Evaluation	Theoretical/Practical class
0	Sep	21	T1. Introduction to ML	none	none	none
1		28	T2. Clustering Part 1	Information of the course and T1	none	W1. Clustering (Session 1)
2	Oct	5	T3. Clustering Part 2	T2	T1	W1. Clustering (Session 2)
		12	BANK HOLIDAY			BANK HOLIDAY
3		19	T4. Factor Analysis (PCA) and (ICA)	T3	T2	W1. Clustering (Session 3)
4		26	T5. Visualization (SOM) and (MDS)	T4	T3	W1. Clustering (Session 4) (deliver W1)
5	Nov	2	T6. Introduction to Supervised Learning	T5	T4	W2. A gentle introduction to PCA and visualization (Session 1)
6		9	T7. Lazy Learning	T6	T5	W2. A gentle introduction to PCA and visualization (Session 2) (deliver W2)
7		16	T8. Feature Selection and model selection	T7	T6	W3. Lazy learning and FS (Session 1)
8		23	T9. Model Selection	T8	T7	W3. Lazy learning and FS (Session 2)
9		30	T10. SVM and Kernels	T9	T8	W3. Lazy learning and FS (Session 3)
	Dec	7	BANK HOLIDAY			BANK HOLIDAY
10		14	T11. Recommender Systems	T10	T9	W3. Lazy learning and FS (Session 4)
11		21		T11	T10	W4. A guided introduction to kernels and SVM (mandatory exercise in Lab) (deliver W3)
12	Jan	11	EXAM		EXAM	

Remarks

- All the information of the course will be in Campus Virtual
<https://campusvirtual.ub.edu>
 - You have access or you will obtain soon, meantime, I will use racó <https://raco.fib.upc.edu/>
- Work is in groups of 3 but the score is individual for each student
- Bring your own laptop computer (face-to-face classes)
 - Install previously Python 3.7 and Pycharm IDE
- **Deliveries should be in Python in a PyCharm project**
 - **Late deliveries:** work or projects submitted late will mean the deduction of 1 point per day (out of 10) from the final mark
 - **Copy of deliveries:** all the groups involved will obtain 0 points

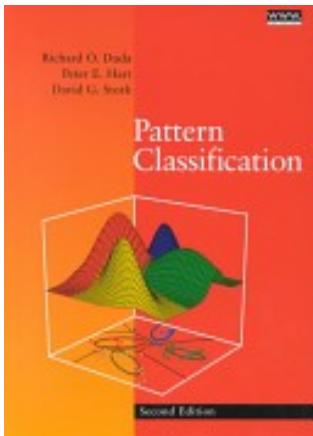
Bibliography



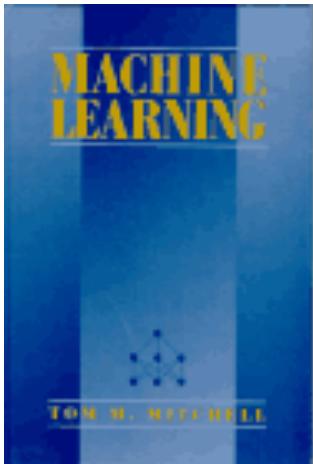
- Bishop, Christopher M., *Pattern Recognition and Machine Learning*. Springer. p. 738. [ISBN 978-0-387-31073-2](#).
- Marsland, Stephen, *Machine Learning: An algorithmic Perspective*, 2nd ed. CRC Press, 2015. [ISBN: 978-1-466-58328-3](#)



Bibliography



- Duda, Richard; Hart, Peter; and Stork, David, *Pattern Classification*, 2nd ed. John Wiley&Sons, 2001. [ISBN: 978-0-471-05669-0](#)



- Tom Mitchell, *Machine Learning*. McGraw-Hill. [ISBN 0-07-042807-7](#)

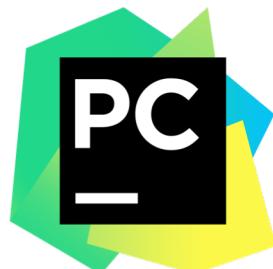
Tools



- Python

<https://www.python.org/>

<http://scikit-learn.org/>



- PyCharm

<https://www.jetbrains.com/pycharm/>

Survey

I want to know you a bit,

PLEASE fill this survey

- It is not anonymous,
- It is not an exam, so there are not good or bad answers
- It is intended to just know more about yourself, what do you like/love, and what do you expect from this course
- According to your interests I can adapt much better the course for you

<https://forms.gle/1eJyLjxbzkUWepMBA>