

Introduction to Machine Learning

Structure : EUR LIFE	Code de l'UE : SMUSVPI	Lieu d'enseignement : Valrose
Niveau du cours : M1/M2	Semestre : Impair	Langue : Anglais

Professors :

- Edoardo SARTI (Edoardo.SARTI@univ-cotedazur.fr)
- Ercan SECKIN (ercan.SECKIN@univ-cotedazur.fr)

Description :

The purpose of this course is to provide the student with the mathematical framework and technical toolkit needed in order to understand advanced topics in machine learning and be able to adapt and refine reference methodologies to meet their scientific needs.

Given its brevity, the course does not attempt to provide an exhaustive overview of the wide variety of ML sub-disciplines, but to give the student the tools to autonomously delve deeper in these subjects without losing track of their scientific aim.

Warning : if you are looking for a discursive and exhaustive overview of Machine Learning and/or Deep Learning techniques, this course is not for you : EFELIA offers other resources covering the aforementioned subjects without delving too much into the mathematical background.

Upon successful completion of this course, students will be awarded **3 ECTS**.

This course was co-developed by the EUR LIFE and the EFELIA team (École Française de l'Intelligence Artificielle).

Public :

Students in the first or second year of one of the following Masters :

Life Sciences Major : Bioinformatics and Computational Biology, MemBioMed, ImAGEin

MSc BOOST: BOOST - Biocontrol for Plant Health

Prerequisites :

The course assumes a good expertise in Python programming (up and included the use of classes and objects), as well as a good understanding of multivariate calculus, linear algebra, and probability. Topics such as – but not limited to - partial derivatives, matrix diagonalization, expectation values will be used with little to no introduction. A self-evaluation test will be available for the student to decide whether or not take this course.

Learning outcomes :

At the end of this course, learners will be able to...

- *Clearly explain both the logic and the mathematical framework behind some of the most renown ML methodologies*
- *Detail the mathematical derivation of some ML algorithms*
- *Use off-the-shelf implementations of ML algorithms in diverse biomedical contexts*
- *Correct, modify, adapt, refine existing ML algorithms to meet specific scientific needs*
- *Analyze, preprocess, and transform data for ensuring an effective and correct use of a target ML method*

Teaching methods :

- 1h30 Theory (blackboard + slides)
- 0h30 Exercises (on paper)
- 2h00 Practicals (on machine)

Organisation :

- Présentiel :
 - 12 heures (CM)
 - 20 heures (TP)

Programme :

Session	Date	Duration	Professor	Topics
1			Edoardo Sarti Ercan Seçkin	Complexity and other core concepts ML [1h] - What is Machine Learning? - Essential vocabulary: supervised/unsupervised, regression/classification, etc. - How to approach the study of Machine Learning Towards DL [0.5h] - Bias, variance, and their tradeoff law - Double descent (might not exist)
2			Edoardo Sarti Ercan Seçkin	Linearity is not what it seems ML [1h] - Linear regression, revisited - Polynomial regression Towards DL [0.5h] - Regularizations
3			Edoardo Sarti Ercan Seçkin	Building Neurons ML [1h] - Logistic regression, revisited - Strategies for more than two labels - Separability and linear trick Towards DL [0.5h] - Perceptrons and dense layers

4			Edoardo Sarti Ercan Seçkin	SVMs and kernels ML [1h] - Support vector machines - Soft margin Towards DL [0.5h] - Kernel trick
5			Edoardo Sarti Ercan Seçkin	Generating with Bayes ML [1h] - Introduction to Bayesian probability - Naïve Bayes and LDA - Recursion: conjugate PDFs Towards DL [0.5h] - Generative Bayesian models
6			Edoardo Sarti Ercan Seçkin	The curse of dimensionality ML [1h] - High-dimensional data - Linear dimension reduction: PCA - Dimension reduction: t-SNE, UMAP, etc. Towards DL [0.5h] - Embeddings
7			Edoardo Sarti Ercan Seçkin	Clustering ML [1h] - Structured data - K-means - DB scan Towards DL [0.5h] - Expectation-maximization
8			Edoardo Sarti Ercan Seçkin	Trees, forests, and crowds ML [1h] - Decision trees - Random forests

				Towards DL [0.5h] - Bootstrap aggregation
--	--	--	--	---

Assessment : Theoretical questions (50%, 1h) and practical exercise (50%, 2h). More information will be provided at the beginning of the course

Equipment :

- A personal computer is recommended but not required

Bibliography :

- **Mehta et al. 2018** - *High-bias, low-variance introduction to Machine Learning for physicists* [REVIEW]
- **Goodfellow et al. 2016** - *Deep Learning* [TEXTBOOK]
- **Murphy 2012** - *Machine Learning : a probabilistic perspective* [TEXTBOOK]
- **Bishop 2006** - *Pattern recognition and Machine Learning* [TEXTBOOK]
- **Tripathy et al. 2021** - *Unsupervised learning approaches for dimensionality reduction and data visualization* [TEXTBOOK]
- **Friedman et al. 2001** - *The elements of statistical learning* [TEXTBOOK]

Resources for success :

- [TUT'TOP](#) : peer tutoring on methodological, social, administrative or logistical issues.
- [écri+](#) : to improve your written French.
- [Centre de ressources en langues](#) : to improve your foreign language skills (French or other).
- [METODA](#) : to improve your documentary research skills.
- [S'orienter / Se réorienter](#) : to be advised by the university's career counsellors.
- [Centre de santé et aide sociale](#) : to look after your physical and mental health, and to seek support in the event of social hardship.
- [Cellule Handicap](#) : support for students with disabilities.
- [Plateforme de signalement](#) : to report acts of violence, harassment or discrimination (sexual and gender-based violence, LGBTphobia, racism, xenophobia, etc.) you have witnessed or experienced at the university, and to get support.

Important : This syllabus has no binding value. Its content may change during the course of the year.

Ce travail a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche (ANR) au titre de France 2030 pour le projet EFELIA Côte d'Azur portant la référence ANR-22-CMAS-0004.