ENSAI

Course Catalog
MSc in Statistics for Smart Data

ACADEMIC YEAR 2021 / 2022

List of courses

General Presentation and Objectives

The world is producing previously unimaginable amounts of data every second. This data could help to understand and improve our society, to predict and prevent, to combat diseases and generally improve life. Extracting valuable information and creating knowledge from the massive and heterogeneous data require skills in statistical modelling, machine learning algorithms, as well as computer science. The synergy of these academic fields, oriented towards their application, is the guiding idea of the Master of Science “Statistics for Smart Data” at ENSAI.

ENSAI is part of the network of prestigious higher-education establishments in France known as *Grandes écoles*, or specialized graduate schools. ENSAI trains its students to become qualified, high-level specialists in information processing and analysis.

The graduates of this MSc will be capable of creating and implementing methodologies and algorithms for analyzing large flows of data arriving from different sources, of using statistical tools and machine learning algorithms to identify correlations, effects, patterns and trends in data, and of formalizing predictions. As such, they will be qualified for data scientist and artificial intelligence jobs in industry, marketing, banking and insurance, media, or further pursuing a PhD.

This Master's program is composed of 1 semester of coursework at ENSAI, followed by a four to six-month paid internship in France or abroad within the professional world, academia, or research laboratories.

Since this program welcomes students with varying academic levels and skills in Computer Science, Applied Mathematics and Statistics, preliminary coursework is put in place to bring all students to the same scientific level in these fields, with respect to their existing training, knowledge, and skills.

For the 2021-2022 academic year, most of the lectures are intended to take place at ENSAI, in live. Some of the lectures are scheduled online using Microsoft Teams. Depending on the pandemic evolution, more courses could move online.

Curriculum – Program Overview and Credits

	Semester Hours	ECTS Credits	Total in the block
UE–MSD01 – Statistical Models for Dependent Data			
Inhomogeneous Markov Models & Applications	30	2.5	5
Graphical Models & Dynamic Networks	18	1.5	
Dynamic Data Visualization	12	1	
UE–MSD02 – Machine Learning			
Machine Learning: Features Selection & Regularization Methods	18	1.5	5
Deep Learning	30	2.5	
Parallel Computing with R and Python	12	1	
UE–MSD03 – Smart Sensing			
Foundations of Smart Sensing	36	3	5
Advanced Topics in Smart Sensing	24	2	
UE–MSD04 – Models for Complex Data			
High-Dimensional Time Series	30	2.5	5
Functional Data Analysis	30	2.5	
UE–MSD05 – IT Tools			
IT Tools 1 (GNU Linux & Shell Scripting, Hadoop & Cloud Computing)	30	2.5	5
IT Tools 2 (NoSQL, Big Data Processing with Spark)	30	2.5	
UE–MSD06 – Challenges for Smart Societies			
Energy Transitions: Quantitative Aspects	12	1	5
Smart Data Project (approximately 8 weeks)	24	2	
Topics & Case Studies in Data Science (conferences)	24	2	
(UE-MSD07 - French as a Foreign Language*)			
(French Summer Program [July-August at CIREFE]	(intensive)		(8)
(Courses for foreigners: Written/Oral French language S1 [at CIREFE])	(22)		
* for foreign students as needed			
TOTAL Semester 1	360 H	30 credits	
UE-MSD08- Internship End-of-Studies Internship	(4 to 6 months)		30
TOTAL Semester 2		30 credits	
TOTAL Academic Year	360 H	60 credits	

Prior to the start of the first semester, the students will be given the opportunity to attend courses designed to reinforce different topics in Computer Science, Statistics, and Mathematics. The list of these courses for September 2021 is the following.

Statistical Languages – R, Python	21 h
Multivariate Data Exploration	12 h
Markov Chains	12 h
Simulation Based Inference	9 h
Topics in Bayesian Inference	9 h
Basics on Shiny	3 h

List of Professors and Lecturers

Code	Topic	Professor/Lecturer
Preliminary 1	Statistical Language: R	Matthieu MARBAC-LOURDELLE
Preliminary 2	Statistical Language: Python	Pierre NAVARO
Preliminary 3	Multivariate Data Exploration	Cesar SANCHEZ SELLERO
Preliminary 4	Markov Chains	Adrien SAUMARD
Preliminary 5	Simulation Based Inference	Valentin PATILEA Myriam VIMOND
Preliminary 6	Topics in Bayesian Inference	Myriam VIMOND
Preliminary 7	Basics on Shiny	Laurent ROUVIERE
MSD 01-1	Inhomogeneous Markov Models & Applications	Salima EL KOLEI Lionel TRUQUET
MSD 01-2	Graphical Models & Dynamic Networks	Eftychia SOLEA
MSD 01-3	Dynamic Data Visualization	Laurent ROUVIERE
MSD 02-1	Machine Learning: Features Selection & Regularization Methods	François PORTIER
MSD 02-2	Deep Learning	Pavlo MOZHAROVSKYI
MSD 02-3	Parallel Computing with R and Python	Matthieu MARBAC-LOURDELLE Pierre NAVARO
MSD 03-1	Foundations of Smart Sensing	Clément ELVIRA Cédric HERZET Claude PETIT
MSD 03-2	Advanced Topics in Smart Sensing	Romaric GAUDEL Adrien SAUMARD
MSD 04-1	High-Dimensional Time Series	Valentin PATILEA Romain TAVENARD
MSD 04-2	Functional Data Analysis	Valentin PATILEA Eftychia SOLEA
MSD 05-1	IT Tools 1 (GNU Linux & Shell Scripting, Hadoop & Cloud Computing)	Guillaume GRABE Shadi IBRAHIM
MSD 05-2	IT Tools 2 (NoSQL, Big Data Processing with Spark)	Nikolaos PARLAVANTZAS Hervé MIGNOT
MSD 06-1	Energy Transitions: Quantitative Aspects	Edouard CIVEL
MSD 06-2	Smart Data Project (8 weeks)	Industrial partners

Code	Topic	Professor/Lecturer
MSD 06-3	Topics & Case Studies in Data Science (conferences)	
	Bandit Theory	Romaric GAUDEL
	Is Data The New Currency of The Digital Economy?	Valeriu PETRULIAN
	New Trends in Cloud Computing	Shadi IBRAHIM
	Case Studies in Smart Data	Thomas ZAMOJSKI
MSD 07-1	French as a Foreign Language	Séverine BORDEAU
MBD 08-1	End-of-Studies Internship	

Preliminary Courses

Preliminary 1 – MSD - Before the start of the 1st Semester

Statistical Language: R

Professor	:	Matthieu MARBAC-LOURDELLE (ENSAI)
ECTS Credits	:	0 (preliminary course)
Estimated personal workload (beyond lecture and tutorial time)	:	9 hrs
Lectures and Tutorials	:	9 hrs (ENSAI)
Teaching language	:	English
Software	:	R
Course materials	:	Slides and tutorials on Moodle
Prerequisites	:	A laptop with R and RStudio installed

Learning Objectives

At the end of the lectures, the student will know the basic concepts of R programming.

Main Subjects covered

This course is organized in three parts:

- Introduction to the data analysis with R
- Presentation of the programming elements
- Performing a simulation with R

References

1. WINSTON CHANG, *R Graphics Cookbook*, O'Reilly, 2013.
2. CORNILLON P-A et al., *R for Statistics*, Chapman & Hall, 2012.
3. COTTON R, *Learning R*, O'Reilly, 2013.
4. GROLEMUND G., *Hands-On Programming with R*, O'Reilly, 2014.

Preliminary 2 – MSD - Before the start of the 1st Semester

Statistical Language: Python

Professor	:	Pierre NAVARO (Université Rennes 1)
ECTS Credits	:	0 (preliminary course)
Estimated personal workload (beyond lecture and tutorial time)	:	1 hour
Lectures and Tutorials	:	12 hrs (ENSAI) including 1 h of independent work on a small project to implement the linear regression model using a Python class
Teaching language	:	English
Software	:	miniconda (https://docs.conda.io/en/latest/miniconda.html)
Course materials	:	https://github.com/pnavaro/python-notebooks
Prerequisites	:	Experience in programming with another language

Learning Objectives

Python is a programming language used for many different applications. In this practical course, students will start from the very beginning, with basic arithmetic and variables, and learn how to handle data structures, such as Python lists, Numpy arrays. Students will learn about Python functions, control flow and data visualizations with Matplolib.
At the end of the lecture, the students are expected to know how to code with Python.

Main Subjects covered

- Setting up your Python environment
- Write functions using control flow tools and manage files input and output
- Introduction to object orienting programming.
- Jupyter Notebook
- NumPy
- Matplotlib
- Implementation of a simple regression model

References

1. Python documentation http://docs.python.org/
2. LUTZ M., ASCHER D., Learning Python, O'Reilly
3. LANGTANGEN H.P, Python Scripting for Computational Science, Springer
4. Python Data Science Handbook https://jakevdp.github.io/PythonDataScienceHandbook/
5. How to Think Like a Computer Scientist: Learning with Python
6. http://interactivepython.org/runestone/static/thinkcspy/

Preliminary 3 – MSD - Before the start of the 1st Semester

Multivariate Data Exploration

Professor	:	Cesar SANCHEZ SELLERO (Universidad de Santiago de Compostela)
ECTS Credits	:	0 (preliminary course)
Estimated personal workload (beyond lecture and tutorial time)	:	12 hrs
Lectures and Tutorials	:	12 hrs (online)
Teaching language	:	English
Software	:	R
Course materials	:	Slides on Moodle and scripts on R
Prerequisites	:	Basic knowledge of Statistics (notions of estimation, confidence intervals and hypothesis testing) and basic Algebra (vectors, matrices, scalar product, norms…)

Learning Objectives

This course provides an introduction to the main exploratory methods used to analyze multivariate data and to summarize their main characteristics. The concepts will be illustrated by applications using R packages. The contents are structured in the following chapters.
At the end of the lectures, the students are expected to know how to analyse and represent multivariate data and how to make groups in data.

Main Subjects covered

- Principal Components Analysis: Algebraic derivation of the principal components of a random vector. Geometric properties of the principal components as a least squares approximation and comparison with regression. Rescaling principal components. Choosing the number of components. Interpreting the components. Simultaneous representation of individuals and variables: the biplot.
- Correspondence Analysis: Contingency tables. Chi-Squared statistic as a measure of the variability between conditional distributions. Decomposing the variability. Simultaneous representation of rows and columns in a contingency table.
- Hierarchical Clustering: Distances, similarities and hierarchical clustering. Agglomerative and divisive methods. Single, complete or average linkage methods. Ward's method. Representation of hierarchical clustering: the dendrogram.
- Non-hierarchical Clustering: K-means method. Clustering mixtures of Gaussian distributions.

References

1. EVERITT, B.S, An R and S-Plus companion to multivariate analysis, Springer, 2005.
2. EVERITT, B.S, Dunn, G. Applied multivariate data analysis. Hodder Education, 2001.
3. HUSSON, F., Le, S., PAGES, J. Exploratory multivariate analysis by example using R. CRC Press, 2011.
4. JOHNSON, R.A., WICHERN, D.W, Applied multivariate statistical analysis, Pearson Education, 2007.

Preliminary 4 – MSD - Before the start of the 1st Semester

Markov Chains

Professor	: Adrien SAUMARD (ENSAI)
ECTS Credits	: 0 (preliminary course)
Estimated personal workload : (beyond lecture and tutorial time)	12 to 24 hrs depending on the student's knowledge about Markov Chains
Lectures and Tutorials	: 12 hrs (ENSAI)
Teaching language	: English
Software	: N/A
Course materials	: Blackboard
Prerequisites	: Basic probability notions

Learning Objectives

Markov chains are a central family of random processes that naturally arises in various fields of application through modelisation. Markov chains allow also to describe a great variety of (stochastic) optimization techniques. It is thus very important to recall the basic notions related to Markov chains and to their long-time behavior, which is the primary goal of this course.

At the end of the lecture, the students are expected to be able to:

- Identify a Markov chain in a modelisation context and prove that a stochastic process is a Markov chain.
- Analyze the static structure of a Markov chain (i.e. establish the associated transition graph, identify the structure in communication classes, show the recurrence or transience of the states of the chain, calculate the periodicities of the classes).
- Describe the limit behaviour of an ergodic chain (i.e., by calculating the stationary, possibly reversible law; cite and apply the limit theorems of the course).

Main Subjects covered

- Basic definition, discrete state space
- Chapman-Kolmogorov equation and Markov properties.
- States classification, periodicity, recurrence and transcience.
- Stationary law and limit theorem (long time behavior)

References

1. NORRIS J.R., Markov Chains, Cambridge Series in Statistical and Probabilistic Mathematics, 1997.
2. GRIMMETT G.R. & STIRZAKER D.R., Probability and Random Processes, Oxford Sciences Publications, 1992 (2nd edition).
3. PARDOUX E., Processus de Markov et applications: Algorithmes, réseaux, génome et finance. Dunod, 2007.

Preliminary 5 – MSD - Before the start of the 1st Semester

Simulation Based Inference

Professor	:	Valentin PATILEA (ENSAI) Myriam VIMOND (ENSAI)
ECTS Credits	:	0 (preliminary course)
Estimated personal work-load (beyond lecture and tutorial time)	:	9 to 18 hrs
Lectures and Tutorials	:	9 hrs (ENSAI) including 1 h of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Handout
Prerequisites	:	Probability, Statistical Inference

Learning Objectives

The aim is to give a quick overview of the use of Monte Carlo experiments for statistical inference. The lecture covers Monte Carlo Integration, Monte Carlo for estimation and for hypothesis tests, the bootstrap.

Main Subjects covered

- Methods for generating random variables
- Monte Carlo Methods in Inference
- Bootstrap and Cross Validation

References

1. RIZZO, M. L, Statistical computing with R. Chapman and Hall/CRC, 2007.
2. ROBERT, C. P., & CASELLA, G., Monte Carlo statistical methods, 2005.

Preliminary 6 – MSD - Before the start of the 1st Semester

Topics in Bayesian Inference

Professor	:	Myriam VIMOND (ENSAI)
ECTS Credits	:	0 (preliminary course)
Estimated personal workload (beyond lecture and tutorial time)	:	9 to 12 hrs
Lectures and Tutorials	:	9 hrs (ENSAI) including 1 h of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Handout
Prerequisites	:	Probability, Statistical Inference

Learning Objectives

The aim is to provide an overview of Bayesian Inference and Bayesian computation.

Main Subjects covered

- Bayesian Statistics
- Markov Chain Monte Carlo
  - ü The Metropolis Hastings Algorithm
  - ü The Gibbs Sampler
  - ü Monitoring Convergence

References

1. RIZZO, M. L., Statistical computing with R. Chapman and Hall/CRC, 2007.
2. ROBERT, C. P., & CASELLA, G., Monte Carlo statistical methods, 2005.
3. NTZOUFRAS, I. Bayesian modeling using WinBUGS (Vol. 698), John Wiley & Sons, 2011.

Preliminary 7 – MSD - Before the start of the 1st Semester

Basics on Shiny

Professor	:	Laurent ROUVIERE (Université Rennes 2)
ECTS Credits	:	0 (preliminary course)
Estimated personal workload (beyond lecture and tutorial time)	:	2 hrs
Lectures and Tutorials	:	3 hrs (ENSAI)
Teaching language	:	English
Software	:	
Course materials	:	Slides and tutorials available at https://lrouviere.github.io/VISU/
Prerequisites	:	Basics on R

Learning Objectives

Shiny is an R package that makes it easy to build interactive web apps straight from R. You can host standalone apps on a webpage or embed them in R Markdown documents or build dashboards. The purpose of this course is to learn how to develop Shiny applications on R.

Main Subjects covered

- Structure of Shiny applications
- Integrating dynamic contents
- Reactive expressions
- Interactive maps

References

1. THIEURMEL, B., shiny tutorial: https://github.com/datastorm-open/tuto_shiny_rennes
2. BEELEY, C., Web Application Development with R Using Shiny, 2013.
3. http://shiny.rstudio.com/

First Semester

1st Semester

TEACHING UNIT MSD-01 :
STATISTICAL MODELS FOR DEPENDENT DATA

Supervisor	: Valentin PATILEA (ENSAI)
ECTS Credits	: 5
Estimated personal workload (beyond lecture and tutorial time)	: 60 to 80 hrs
Lectures and Tutorials	: 60 hrs

Learning Objectives of the Teaching Unit

In many applications, the interest lies in capturing and modeling the interactions and depencies between observation units. Such interactions could be in the dimension of time, or inside of a network. Another challenge for the data scientist is to produce useful and effective data visualization output.

Description

The unit is composed of three courses:

- Advanced Markov chains and applications;
- Modeling graphs and networks;
- Data visualization.

Acquired Skills

Modeling of time-dependent phenomena and networks. Conceive meaningful data visualization output.

Pre-requisites

Basics in probability theory, Markov chains, mathematical statistics, principal component analysis, programming with R

UE-MSD01 – Statistical Models for Dependent Data – MSD 01.1 - 1st Semester

Inhomogeneous Markov Models & Applications

Professors	:	Salima EL KOLEI (ENSAI)
	:	Lionel TRUQUET (ENSAI)
ECTS Credits	:	2.5
Estimated personal workload (beyond lecture and tutorial time)	:	35 hrs
Lectures and Tutorials	:	30 hrs (ENSAI) including 1.5 h of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Slides
Prerequisites	:	Probability theory, Markov Chains, Baysesian Statistics, Monte Carlo Methods, Statistical inference, Generalized Linear Models.

Learning Objectives

Homogeneous Markov chains are exploited in a broad range of applications. Nevertheless, in some situations homogeneous transition probabilities do not adequately model real processes. In these situations, Markov models with inhomogeneous rates, i.e., rates that are time-varying functions or that depend on covariates, could be more appropriate. In the first part of this course, the theory of these processes will be described and will be illustrated with applications in several areas (financial, aeronautics, meteorological…). The second part of this course will be devoted to Hidden Markov Models (HMM). After presenting the basic HMM, the framework is extended by considering nonhomogeneous hidden Markov models and Markov-switching models. Such models are used in finance, electricity prices, genomics... Bayesian methods, such as MCMC, Kalman and particle filters, forward backward, Vitterbi and the Expectation Maximization algorithms will be introduced and applied to infer in these models. Several real data applications will be considered to illustrate the methods.

At the end of the lectures, the student will be able to
- choose appropriate time-inhomogenous (hidden) Markov models to handle some nonstationary datasets,
- have a good understanding of the theory of estimation of such processes
- fit such models and interpret the outputs of statistical inference,
- perform covariate selection and calibrate some additional tuning parameters.

Main Subjects covered

PART 1
1. Discrete Time Markov Model: Estimation of the transition matrix
2. Continuous Time Markov Model: Homogeneous and Inhomogeneous, Poisson Process, Queueing Process
3. Recent Developments and Applications

PART 2
4. Hidden Markov-Models: HMM architecture for discrete and continuous state space, Parameter Estimation, Applications
5. Non-Homogeneous Hidden Markov Model: NHMM architecture for discrete state space, Parameter estimation, Applications

Evaluation

An oral exam or a written exam (PART 1) and a project (PART 2)

References

1. CHING, HUANG, NG and SIU. “Markov chains: models, algorithms and applications” (2nd ed.). New York : Springer; 2013.
2. IVERSEN, MOLLER, MORALES and MADSEN “Inhomogeneous Markov models for describing driving patterns”. IEEE Transactions on Smart Grid, Vol. 8 (2), p 581-588; 2017.

3. DYMARSKI, PRZEMYSLAW, ed. “Hidden Markov Models: Theory and Applications”. InTech, 2011.
4. AILLIOT, P., and PENE F. Consistency of the maximum likelihood estimate for non-homogeneous Markov-switching models. ESAIM: Probability and Statistics, Vol. 19, p. 268-292; 2015.
5. Yaakov BAR-SHALOM, X. RONG LI, Thiagalingam KIRUBARAJAN, Estimation with Applications to Tracking and Navigation - 2001.

UE-MSD01 – Statistical Models for Dependent Data – MSD 01.2 - 1st Semester

Graphical Models & Dynamic Networks

Professor	: Eftychia SOLEA (ENSAI)
ECTS Credits	: 1.5
Estimated personal workload (beyond lecture and tutorial time)	: 1.5 h personal workload per 1h lecture
Lectures and Tutorials	: 18 hrs (ENSAI) including 1.5 h of independent work
Teaching language	: English
Software	: R
Course materials	: Lecture notes, textbook (see list of references below)
Prerequisites	: Basic knowledge in probability theory, mathematics & programming recommended

Learning Objectives

At the end of this course students will be familiar with the fundamentals of random graphs and graphical models. Graphs have been used in data science as an exploratory tool to investigate complicated inter-relationships in data structures. This course will cover the basic computational and theoretical tools for learning and inference in graphical models.
In particular, the course is divided into two thematic blocks: First, random graphs, which are used when the data itself is available as a graph whose nodes are fixed and edges random. In this case, the objective is to adjust a model to unravel some particular organization of the data. Second, probabilistic graphical models, which give a compact and analytically useful representations of joint distributions over a large number of variables, using graphs. Each graph represents a family of distributions – the nodes of the graph represent random variables, the edges encode independence assumptions. Large-scale of social or biological networks examples will illustrate the algorithms presented in this course.

Main Subjects covered

1. Introduction: Review of probability and statistics; Introduction to graphs.
   Tutorial: analysis of real networks with the R package igraph
2. Randoms graphs analysis: Spectral Clustering; Stochastic Block Model (SBM).
   Tutorial: Variational inference in the SBM
3. Probabilistic graphical models: Directed/undirected graphical models; Log-linear models; Gaussian graphical models (GGM).
   Tutorial: Sparse inference of high-dimensional GGM

Evaluation

Grades will be based on three components: Tutorial reports (30%); Midterm exam (30%); Final small seminar thesis; 15 minutes presentation (10 minutes talk, 5 minutes questions) (40%);

References

1. WAINWRIGHT, M.J., JORDAN, M.I. "Graphical models, exponential families, and variational inference." Foundations and Trends® in Machine Learning, Vol. 1, No 1-2: 1-305, 2008.
2. HøJSGAARD, S., EDWARDS, D., LAURITZEN, S. "Graphical Models with R. Springer, New York. 2012.
3. BISHOP, C. "Introduction to graphical modelling". 2nd edn. Springer, New York. 2000.
4. LAURITZEN, S.L. "Graphical models“. Clarendon Press, Oxford. 1996.
5. KOLACZYK, E.D., GABOR, C. “Statistical analysis of network data with R”. New York: Springer, 2014.
6. KOLACZYK, E.D. "Statistical Analysis of Network Data: Methods and Models”. Springer. 2009.
7. NEWMAN, M. "Networks: An Introduction”. Published to Oxford Scholarship Online. 2010.

UE-MSD01 – Statistical Models for Dependent Data – MSD 01.3 - 1st Semester

Dynamic Data Visualization

Professor	:	Laurent ROUVIERE (Université Rennes 2)
ECTS Credits	:	1
Estimated personal workload (beyond lecture and tutorial time)	:	10 hrs
Lectures and Tutorials	:	12 hrs (ENSAI)
Teaching language	:	English
Software	:	
Course materials	:	
Prerequisites	:	During this course, we will manipulate basic notions used in data science. A minimal knowledge of the basic tools used in data science, as well as in statistics is required such as: PCA, classification algorithms. Basics on R are also necessary.

Learning Objectives

Data visualization is a fundamental ingredient of data science as it “forces us to notice what we never expected to see” in a given dataset. In this course, we show through examples and case studies that graphical methods are powerful tools for revealing not only the structure of the data, but also patterns and (ir)regularities, groups, trends, outliers…
Dataviz is relevant both for data analysis, when the analyst wants to study data and, as any statistics, to question the data. It is also a tool for communication and, as such, is a visual language with a theory of the functions of signs and symbols used to encode the visual information. All along the course, we’ll focus on methods, tools and strategies to represent simple and then complex or high-dimensional datasets, highlighting the growing development of dynamic and interactive tools.

Main Subjects covered

- Data visualization for data sciences
- Classics in Data visualization
- Grammar of graphics with ggplot2
- Mapping with sf and leaflet
- Interactive and dynamic visualization

Evaluation

The evaluation consists on a data visualization project. The students will have to:
- deploy a shiny web application and to publish it on the web.
- write a markdown report to present the application
They will work in groups with two members.

References

1. BERTIN, J. 1983. Semiology of Graphics, translation from Sémilogie graphique . 1967.
2. TUFTE, E. R The Visual Display of Quantitative Information. 2 ed. Graphics Press. 2001.
3. http://ggplot2.org
4. https://ggplot2-book.org
5. https://statnmap.com/fr/2018-07-14-initiation-a-la-cartographie-avec-sf-et-compagnie/
6. https://rstudio.github.io/leaflet/
7. https://rmarkdown.rstudio.com/flexdashboard/

1st Semester

TEACHING UNIT MSD-02 : MACHINE LEARNING

Supervisor	: Valentin PATILEA (ENSAI)
ECTS Credits	: 5
Estimated personal workload (beyond lecture and tutorial time) /	: 30 to 40 hrs
Lectures and Tutorials	: 60 hrs

Learning Objectives of the Teaching Unit

Present fundamental modern machine learning approaches and provide computing tools for effective implementation. Topics in model/feature selection and regularization methods, regression trees, aggregation methods and support vector machine, as well as neural networks and deep learning concepts and algorithms will be presented. Parallel computing techniques are also presented. The students are expected to know the main up to date algorithms and to be able to implement them.

Description

The unit is composed of one course of regularization methods by penalization and nonlinear models, a second course on the main machine learning approaches and the up to date neural networks. A third course on parallel computing is expected to complete the panel of tools necessary for implementation.

Acquired Skills

Knowledge of a large panel of algorithms, use of modern machine learning approaches for complex data problems, implementation of algorithms using packages and notebooks.

Pre-requisites

Regression models, notions of probability theory, combinatorics and geometry, algorithm complexity.

UE-MSD02 – Machine Learning – MSD 02.1 - 1st Semester

Machine Learning:
Features Selection & Regularization Methods

Professor	:	François PORTIER (ENSAI)
ECTS Credits	:	1.5
Estimated personal workload (beyond lecture and tutorial time)	:	5 to 10 hrs
Lectures and Tutorials	:	18 hrs (ENSAI) including 1.5 h of independent work
Teaching language	:	English
Software	:	R & Python
Course materials	:	Online textbooks + slides
Prerequisites	:	Familiarity with linear algebra; a working knowledge of R or Python programming; familiarity with multiple linear regression.

Learning Objectives

Starting from classical notions of shrinkage and sparsity, this course will cover regularization methods that are crucial to high-dimensional statistical learning. The syllabus includes feature selection and model selection, linear and nonlinear techniques for regression and for classification. The course will focus on methodological and algorithmic aspects, while trying to give an idea of the underlying theoretical foundations. Practical sessions will give the opportunity to apply the methods on real data sets using either R or Python. The course will alternate between lectures and practical lab sessions (9h of lecture, 9h of computer lab sessions, one of the labs will be carried out in independently).

Upon completing this course, students should be able to: select the appropriate methods; implement these statistical methods; compare leading procedures based on statistical arguments; assess the prediction performance of a learning algorithm; apply these key insights into class activities using statistical software.

Main Subjects covered

1. Subset Selection 1.1. Best Subset Selection 1.2. Stepwise Selection 1.3. Choosing the Optimal Model	2. Shrinkage Methods 2.1. Ridge Regression 2.2. The Lasso 2.3. Lasso variants
3. Basis Expansions and Regularization 3.1. Smoothing Splines 3.2. Choosing the Smoothing Parameter	4. Generalized Additive Models 4.1. GAMs for Regression Problems 4.2. GAMs for Classification Problems

Evaluation

Laptop exam or/and project

References

1. HASTIE, T., TIBSHIRANI, R., & FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer. Free download. 2009.
2. JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. An Introduction to Statistical Learning. New York: Springer. Free download. 2013.

UE-MSD02 – Machine Learning – MSD 02.2 - 1st Semester

Deep Learning

Professor	:	Pavlo MOZHAROVSKYI (Telecom ParisTech)
ECTS Credits	:	2.5
Estimated personal workload (beyond lecture and tutorial time)	:	15 hrs
Lectures and Tutorials	:	30 hrs (ENSAI) including 3 hrs of independent work
Teaching language	:	English
Software	:	R, Python
Course materials	:	Slides, lab subjects and codes for practical sessions
Prerequisites	:	Regression analysis, gradient descent, (matrix) algebra, R, Python (basics).

Learning Objectives

The course starts with a general introduction to machine learning and deep learning giving a brief overview of the problems which may be addressed using different kind of approaches. For the machine learning part we begin with a review of classification and regression trees and then focus on aggregation methods like bagging, random forests (RF), and boosting (AdaBoost algorithm and the gradient boosting optimisation). Support vector machine (SVM) are also discussed. The second part of the course is devoted to neural network (NN) architectures and their extension known as deep learning. Beforehand, the stochastic gradient descent algorithm and the back-propagation - its application to feedforward neural networks - are introduced to be further used as the learning basis. This is followed by the study of most spread NN architectures for regression and classification. Among those, convolutional neural networks (CNN) are investigated in detail and other structures like Restricted Boltzmann machines (RBM) and the contrastive divergence algorithm (CD-k) are examined. Further practical aspects will be addressed about the usage of Deep Learning to resolve typical problems like pattern recognition or object detection/tracking. Presented material shall be motivated by the theoretical background together with real data illustrations. There will be specific labs for each topic majorly held in R, along with sparse Python illustrations. At the end of the course, the student will know the theoretical basics and how to apply in practice large-scale statistical learning techniques and deep neural networks.

Main Subjects covered

- Introduction to machine learning.
- Decision trees, bagging, and random forests.
- Boosting classifiers.
- Support vector machines.
- Stochastic gradient descent and the back-propagation algorithm.
- Neural networks for regression and classification.
- Convolutional neural networks, Restricted Boltzman Machines.
- Applications: Pattern recognition, object detection

Evaluation

Written Exam + Lab

References

1. HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. The Elements of Statistical Learning. Springer-Verlag. 2009.
2. HAYKIN, S.O. Neural Networks and Learning Machines. Pearson. 2008.
3. SCHAPIRE, R.E., FREUND Y. Boosting: Foundations and Algorithms. The MIT Press. 2012.
4. VAPNIK, V.N. Statistical Learning Theory. Wiley-Blackwell. 1998.

UE-MSD02 – Machine Learning – MSD 02.3 - 1st Semester

Parallel Computing with R & Python

Professor	:	Matthieu MARBAC-LOURDELLE (ENSAI) - lectures on "R" Pierre NAVARO (Université Rennes 1) - lecture on "Python"
ECTS Credits	:	1
Estimated personal workload (beyond lecture and tutorial time)	:	12 hrs
Lectures and Tutorials	:	12 hrs (ENSAI)
Teaching language	:	English
Software	:	R and Python
Course materials	:	Material on Moodle for R and https://github.com/pnavaro/big-data for Python
Prerequisites	:	Knowledge of R and Python

Learning Objectives

- Detecting the slow parts of a script by using graphical tools for code profiling. Students will be able to detect the parts of a script where the code should be improved and where the memory allocations should be reduced.
- Improving the code performances using CPU parallel computation. Students will be able to use both of the forking and socket methods of parallel computation.

Main Subjects covered

First, an introduction of code profiling is proposed (micro and macro profiling, memory monitoring). Then, the two standard methods for CPU parallel computations are presented (forking and socket).
With Python, we will cover basic ideas and common patterns in parallel computing, including embarrassingly parallel map, unstructured asynchronous submit, and large collections.

Evaluation

Lab 2 hrs (the report is written at home)

References

1. https://www.r-project.org (R-packages Rmpi, RHadoop assembly, gpuR).
2. https://wiki.python.org/moin/ParallelProcessing (ScientificPython library).
3. https://computing.llnl.gov/tutorials/mpi/
4. DEAN, J., GHEMAWAT, S. MapReduce: simplified data processing on large clusters. Proceedings of OSDI’04. 2004.
5. https://www.khronos.org/opencl/

1st Semester

TEACHING UNIT MSD-03 :
SMART SENSING

Supervisor	: Valentin PATILEA (ENSAI)
ECTS Credits	: 5
Estimated personal workload (beyond lecture and tutorial time)	: 60 to 80 hrs
Lectures and Tutorials	: 60 hrs

Learning Objectives of the Teaching Unit

A main objective of modern data analysis approaches is to detect patterns, structures in complex high-dimensional data. The aim of smart sensing is to identify such patterns and structures in data at the acquisition step, and thus to avoid handling large datasets with low value. Many data sources produce signals (curves, images...) and most of them have a low dimension representation which preserves the essential information carried by the signal. The aim of the unit is to present approaches for identifying parsimonious representations of the signals.

Description

The first course is more methods oriented, while in the second the focus will be on some applications and some up to date problems.

Acquired Skills

Modern data/signals compression approaches.

Pre-requisites

Basics in Fourier analysis, matrix algebra, probability theory and mathematical statistics.

UE-MSD03 – Smart Sensing – MSD 03.1 - 1st Semester

Foundations of Smart Sensing

Professor	Clément ELVIRA (Centrale-Supélec) Cédric HERZET (INRIA, Centre Rennes-Bretagne Atlantique) Claude PETIT (INSEE)
ECTS Credits	3
Estimated personal workload (beyond lecture and tutorial time)	24 hrs
Lectures and Tutorials	36 hrs (ENSAI) including 3 hrs of independent work
Teaching language	English
Software	Matlab/Python
Course materials	Slides + Textbook
Prerequisites	Basic linear algebra ; basic matrix algebra; basic Fourier analysis ; basic statistical foundations

Learning Objectives

Although most signals (audio, images,…) of interest belong to a very large dimensional ambient space, many of them possess some structure that contains the useful information and makes them intrinsically low dimensional or compressible. Such structures can be modeled and exploited to reduce the cost of their acquisition and their processing. Sparse representations are a powerful tool to express and represent such signals, typically by a small number of nonzero coefficients in an appropriate basis or dictionary. Combined with some appropriate design of the sensing device, they allow to acquire and to reconstruct signals with an extremely reduced number of measurements. This course will present theoretical and algorithmic frameworks and tools for low-dimensional representations of large-scale data and their compressed acquisition and recovery.

At the end of the lecture, the student will know the main ingredients of the theory of sensing. He/she will understand the concepts of "Shannon sampling" and "compressive sampling". In particular, she/he will be able to implement the main algorithms of sparse representations and understand their theoretical properties.

Main Subjects covered

Introduction:

- From "classical" to "smart" sensing: elements of digital signal processing (Fourier analysis, the Shannon-Nyquist theorem...), examples and limitations.
- Main definitions, NP-Hardness of the sparse recovery problem and the need for efficient algorithms, introduction to geometrical interpretations, brief history and overview of the field and its applications.

Sparse Representations :

- Sparsity, sparse approximations, sparsity measures
- Sparse recovery algorithms: greedy algorithms, thresholding algorithms and proximal operators, convex relaxation
- Theoretical guarantees of recovery (deterministic point of view) : Restricted Isometry Property, Null Space Property, Exact Recovery Conditions, mutual coherence
- Dictionaries: fixed dictionaries and sparsifying transforms, overcomplete dictionaries, dictionary learning

Applications: Sparse Component Analysis, image denoising and compression…

Compressive sensing

- Presentation and motivation of compressive sensing
- Measurement matrix design (i.e. how many measurements and constraints) and major inequalities
- Random matrices and theoretical guarantees (probabilistic point of view) (Restricted Isometry property)
- Applications: compressed MRI acquisition and reconstruction, digital communications, image inpainting
- and interpolation, hyperspectral images, audio and acoustic applications.

Evaluation

Written exam + 1 project

References

1. FOUCART S., RAUHUT H. A mathematical introduction to compressive sensing. Birkhauser. 2013.
2. CANDES E.J., WAKIN M.B. An Introduction To Compressive Sampling. IEEE - Signal Processing Magazine, March. 2008.
3. KUTYNIOK G., Theory and Applications of Compressed Sensing, GAMM Mitteilungen 36, 79-101. 2013.
4. ELAD M., Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer. 2010.
5. ELDAR Y. and KUTYNIOK G., Compressed Sensing: Theory and Applications. Cambridge University Press. 2012.
6. BOCHE H., CALDERBANK R, KUTYNIOK G., VYBIRAL J. (editors). Compressed Sensing and its Applications. Birkhauser. 2015.

UE-MSD03 – Smart Sensing – MSD 03.2 - 1st Semester

Advanced Topics in Smart Sensing

Professor	:	Romaric GAUDEL (ENSAI) Adrien SAUMARD (ENSAI)
ECTS Credits	:	2
Estimated personal workload (beyond lecture and tutorial time)	:	50 hrs
Lectures and Tutorials	:	24 hrs (ENSAI)
Teaching language	:	English
Software	:	Python
Course materials	:	Slides (Romaric Gaudel) ; Blackboard (Adrien Saumard)
Prerequisites	:	Basic linear algebra; basic matrix algebra; basic statistical foundations; foundations of smart sensing

Learning Objectives

This course collects some advanced topics in compressive sensing and related techniques. Although self-contained, a prerequisite to this course is the course of foundations of smart sensing, that focuses on the notion of vector sparsity. In many applications, signals are structured in many ways that go beyond the basic notion of sparsity for vectors. We will thus introduce notions of sparsity for matrices, focusing on low rank matrices. We will also describe some emerging techniques of compressive learning, where data compression is oriented toward a subsequent learning task. In this context, data compression is also called sketching. The example of sketched PCA will be presented. This course will present theoretical and algorithmic frameworks and discuss applications that are in the scope of the methods, including collaborative filtering based recommender systems.

At the end of the lectures the student will be expected to:

- know the definition of sparsity for matrices based on the rank and to be able to calculate it using SVD.
- understand the three main optimization techniques to perform a sparse matrix recovery and to be able to use them in practice.
- understand the theoretical underlying of matrix recovery, such as the structure of low rank matrices, the Null Space Property and the notion of coherence of a family of vectors.
- understand the collaborative filtering: identify the setting, choose an algorithm, measure the quality of the recommendation
- be aware of the possibility of data compression for statistical learning, in particular using random moments and in the context of the PCA technique.

Main Subjects covered

- Compressed sensing with the Fourier basis
- Sparsity and compressive sensing with matrices
- Sketched learning.

Evaluation

Written exam + 1 project

References

1. FOUCART S., RAUHUT H. A mathematical introduction to compressive sensing. Birkhauser. 2013.
2. LEE K., BRESLER Y. ADMiRA: atomic decomposition for minimum rank approximation. IEEE Trans. Inform. Theory 56 (9): 4402–4416. 2010.
3. RECHT B., A simpler approach to matrix completion. JMLR. 12:3413--3430. 2011.

1st Semester

TEACHING UNIT MSD-04 :
MODELS FOR COMPLEX DATA

Supervisor	: Valentin PATILEA (ENSAI)
ECTS Credits	: 5
Estimated personal workload (beyond lecture and tutorial time)	: 60 to 70 hrs
Lectures and Tutorials	: 60 hrs

Learning Objectives of the Teaching Unit

Modern applications produce data under various complex forms. This unit presents, on one hand, methods for multivariate, possibly high-dimensional, time-series, and, on the other hand, functional data analysis (FDA) approaches. Quite often, data consists of several time-series which interact. Specific methods allow the identification of the joint dynamics of the series making reliable predictions. The first course presents statistical and machine learning approaches for these purposes. The second course is a first course in FDA. Many existing data are composed of curves and images for which the time-series paradigm is not applicable or relevant. One would like to summarize by some data-driven techniques. Such techniques as well as some applications will be proposed in the second course of this unit.

Description

The unit is organized into two distinct modules.

Acquired Skills

Model complex multivariate data using advanced methods.

Pre-requisites

Gaussian vectors, principal component analysis, univariate time-series, basic machine learning algorithms.

UE-MSD04 – Models for Complex Data – MSD 04.1 - 1st Semester

High-Dimensional Time Series

Professors	:	Valentin PATILEA (ENSAI) Romain TAVENARD (Université Rennes 2)
ECTS Credits	:	2.5
Estimated personal workload (beyond lecture and tutorial time)	:	20 hrs
Lectures and Tutorials	:	30 hrs (ENSAI) including 6 hrs of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Slides, codes, articles, book chapters
Prerequisites	:	Standard background in probability theory. Gaussian vectors. Variance-Covariance matrices. Linear projections in L^2. Basic notions of univariate time series: autocorrelation function, ARMA processes, least squares method. Principal component analysis. Basic regularization and classification methods.

Learning Objectives

Time Series Analysis accounts for the fact that data points (numbers or vectors) observed over time have a specific structure (such as autocorrelation or seasonal variation). Time series models aim revealing such structures, making inference, building forecasts, ... The standard concepts and models for univariate and multivariate time series will be reviewed in the first part of the lectures. Next, the factor model will be presented. Motivated by the increasing amount of available information, such models are a versatile approach to summarize information contained in large vectors of data. The fundamental factor models and the common inferences approaches will be presented and illustrated with real datasets.

The second part of this course will deal with machine learning models for time series. More specifically, the use of alignment-based methods in traditional machine learning models will be discussed. Recurrent neural network models will also be tackled. All these models will be illustrated on real datasets.

After the first part of the lectures, the students will know and will be able to apply the main diagnosis tools and models to analyze and forecast 1- or multi-dimensional time series. After the second part, they will be able to choose an adequate machine learning model and apply it for a given time series task.

Main Subjects covered

- Univariate Time series: stationarity, autocorrelations, basic models
- Multivariate autocorrelation function. Vector autoregressive models. Stationarity and statistical inference of VAR models. Factor models.
- Machine learning for time series, Recurrent neural networks

Evaluation

Written exam + report on a real-data analysis

References

1. GOOFELLOW, I., BENGIO, Y., COURVILLE, A. (2016). Deep learning. MIT Press.
2. LUTKEPOHL, H. New introduction to multiple time series analysis. Springer. 2005.
3. TSAY, R.S. Multivariate time series analysis: with R and financial applications. Wiley. 2014.
4. STOCK, James H., and Mark W. Watson. "Dynamic Factor Models." In The Oxford Handbook of Economic Forecasting. Oxford University Press. 2011.

UE-MSD04 – Models for Complex Data – MSD 04.2 - 1st Semester

Functional Data Analysis

Professors	:	Valentin PATILEA (ENSAI) Eftychia SOLEA (ENSAI)
ECTS Credits	:	2.5
Estimated personal workload (beyond lecture and tutorial time)	:	1.5 h personal workload per 1h lecture
Lectures and Tutorials	:	30 hrs (ENSAI) including 3 hrs of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Lecture notes and textbook (see below)
Prerequisites	:	Statistical inference and methods, Multivariate statistical analysis

Learning Objectives

This course aims to provide an introduction to functional data analysis. The fundamental statistical tools for modeling and analyzing such data will be explored. This course introduces ideas and methodology in functional data analysis (FDA) as well as the use of software. Students will learn the idea of different methods and the related theory, and also the numerical and estimation routines to perform functional data analysis. Students will also have an opportunity to learn how to apply FDA to a wide array of application areas. The course will demonstrate applications where FDA techniques have clear advantage over classical multivariate techniques. Some recent development in FDA will also be discussed.

Main Subjects covered

Chapter 1. Introduction.
Chapter 2. Representing functional data and exploratory data analysis. Including: basic expansions, FPCA, derivatives, penalised smoothing, registration, fda package.
Chapter 3. Elements of Hilbert space theory and random functions.
Chapter 4. Estimation and inference from a random sample. Including, estimation of functional principal component analysis (FPCA), testing hypothesis about the mean.
Chapter 5. Functional Linear regression models. Including: Functional linear regression models with scalar or functional response variable (function-on-scalar, scalar-on-function and function-on-function models).
Chapter 6. Functional Generalised Linear Models.
Chapter 7. Analysis of functional time series and the ftsa package.
Chapter 8. Further problems.

Evaluation criteria

The final grade will be determined by three criteria: Homework (20 %), Project (35%), Final exam (45%)

References

1. RAMSAY, J.O. and SILVERMAN, B. W. Functional Data Analysis. Springer. 2005.
2. RAMSAY, J.O., HOOKER, G. and GRAVES, S. Functional Data Analysis in R and Matlab. Springer. 2009.
3. SHI, J. Q. and CHOI, T. Gaussian Process Regression Analysis for Functional Data. Chapman & Hall/CRC Press. 2011.
4. HORMANN, S. and KIDZINSKI, L., HALLIN, M. Dynamic Functional Principal Components. JRSSB, Vol. 77, No. 2, pp. 319-348. arXiv 1210.7192v5. 2015.
5. SHANG, H. L. ftsa: An R package for analysing functional time series. The R journal, 64-72. 2013.
6. HORVATH, L. and KOKOSZKA, P. Inference for Functional Data with Applications. Springer Series in Statistics, Volume XIV. 2012.
7. KOKOSZKA, P and REIMHER, M. Introduction to Functional Data Analysis. Chapman & Hall/CRC, Texts in Statistical Science. 2017.

1st Semester

TEACHING UNIT MSD-05 :
IT TOOLS

Supervisor	:	Valentin PATILEA (ENSAI)
ECTS Credits	:	5
Estimated personal workload (beyond lecture and tutorial time)	:	30 to 40 hrs
Lectures and Tutorials	:	60 hrs

Learning Objectives of the Teaching Unit

This unit presents a panorama of modern computer/cloud tools for processing massive amounts of complex data.

Description

Courses of Linux, NoSQL, Hadoop and Spark are proposed.

Acquired Skills

Using the most recent computer/cloud computing tools for data processing.

Pre-requisites

Basics in programming and databases: Java, Python, R, Linux, SQL.

UE-MSD05 – IT Tools – MSD 05.1 - 1st Semester

IT Tools 1 (GNU Linux & Shell Scripting, Hadoop & Cloud Computing)

Professors	:	Guillaume GRABE (Orange Cyberdefense) Shadi IBRAHIM (INRIA – Rennes)
ECTS Credits	:	2.5
Lectures and Tutorials (total)	:	30 hrs (12+18)
Teaching language	:	English

GNU Linux & Shell Scripting

Professor	:	Guillaume GRABE (Orange Cyberdefense)
Estimated personal workload (beyond lecture and tutorial time)	:	5 to 7 hrs
Lectures and Tutorials	:	12 hrs (ENSAI) including 2 to 3 hrs of independent work
Software	:	Linux + Shell (installed during the lecture)
Course materials	:	A computer (lent by ENSAI)
Prerequisites	:	A computer + internet connection + VirtualBox

Learning Objectives

This class teaches students the concepts that they should understand before they start working with GNU/Linux. During this course, students will install a distribution on their computer and learn how to interact with the shell, from basic tasks (navigation, file edition, network configuration) to more advanced operations with shell scripting.
GNU/Linux is essential in particular when using and developing Big Data technologies.

Main Subjects covered

1. GNU/Linux
- Introduction to GNU/Linux
- Installing a distribution
- The shell
- Users, groups, permissions
- Packages management
- Network management

2. Shell scripting
- Shell scripting principles
- Variables in the shell, operations on variables
- Conditional expressions, basic statements, functions
- Regular expressions

Evaluation

Written evaluation or project

References

1. https://wiki-dev.bash-hackers.org/
2. http://tldp.org/index.html
3. B. FOX and C. RAMAY, Bash Reference Manual, Free Software Foundation

Hadoop & Cloud Computing

Professor	:	Shadi IBRAHIM (INRIA – Rennes)
Estimated personal workload (beyond lecture and tutorial time)	:	9 to 15 hrs
Lectures and Tutorials	:	18 hrs (ENSAI) including 2 hrs of independent work
Software	:	Hadoop, Virtual Machine Mangers (e.g., Virtual Box, VMware-Player, VMware Fusion, etc)
Course materials	:	All course materials presentations, tutorials and hand-ons, libraries and codes will be available online on the course website in pdf and zip format
Prerequisites	:	Familiar with Linux command-line Familiar with Java/Python

Learning Objectives

At the end of the lectures, the student will realize the potential of Big Data and will know the main tools to process this tsunami of data at large-scale. In particular, the students will understand the main features of MapReduce programming model and its open-source implementation Hadoop, and will be able to use Hadoop and test it using different configurations.

Data volumes are ever growing, for a large application spectrum going from traditional database applications, scientific simulations to emerging applications including Web 2.0 and online social networks. To cope with this added weight of Big Data, we have recently witnessed a paradigm shift in computing infrastructure through Cloud Computing and in the way data is processed through the MapReduce model. First promoted by Google, MapReduce has become, due to the popularity of its open-source implementation Hadoop, the de facto programming paradigm for Big Data processing in large-scale infrastructures. On the other hand, cloud computing is continuing to act as a prominent infrastructure for Big Data applications.

The goal of this course is to give a brief introduction to Cloud Computing: definitions, types of cloud (IaaS/PaaS/Saas, public/private/hybrid), challenges, applications, main cloud players (Amazon, Microsoft Azure, Google etc.), and cloud enabling technologies (virtualization). Then we will explore data processing models and tools used to handle Big Data in clouds such as MapReduce and Hadoop. An overview on Big Data including definitions, the source of Big Data, and the main challenges introduced by Big Data, will be presented. We will then present the MapReduce programming model as an important programming model for Big Data processing in the Cloud. Hadoop ecosystem and some of major Hadoop features will then be discussed.

Main Subjects covered

Throughout the course we will cover the following topics:

- Cloud Computing: definitions, types, Challenges, enabling technologies, and examples (2.25 hrs)
- Big Data: definitions, the source of Big Data, challenges (1.5 hrs)
- Google Distributed File System (1.5 hrs)
- The MapReduce programming model (1.5 hrs)
- Hadoop Ecosystem (2.25 hrs)
- Practical sessions on Hadoop (7 hrs)
  - ü How to use Virtual Machines and Public Cloud Platforms
  - ü Starting with Hadoop
  - ü Configuring HDFS
  - ü Configuring and Optimising Hadoop
  - ü Writing MapReduce applications

Independent work (tentative): Students will be divided into groups where each group will do a 15 - 20 min presentation on one of the main subjects or a life demonstration on one of the practical sessions (2 hrs)

Evaluation

Written exam

References

1. JIN Hai, IBRAHIM Shadi, BELL Tim, GAO Wei, HUANG Dachuan, WU Song. Cloud Types and Services. Book Chapter in the Handbook of Cloud Computing, Springer Press, 26 Sep 2010.

2. JIN Hai, IBRAHIM Shadi, BELL Tim, LI QI, HAIJUN Cao, WU Song, XUANHUA Shi. Tools and technologies for building the Clouds. Book Chapter in Cloud Computing: Principles Systems and Applications, Springer Press, 2 Aug 2010.

3. ARMBRUST Michael, FOX Armando, GRIFFITH Rean,. JOSEPH Anthony D, KATZ Randy, KONWINSKI Andy, LEE Gunho, PATTERSON David, RABKIN Ariel, STOICA Ion, and ZAHARIA Matei. 2010. A view of cloud computing. Commun. ACM 53, 4 - April 2010.

4. GHEMAWAT Sanjay, GOBIOFF Howard, and LEUNG Shun-Tak. The Google file system. In SOSP '03.

5. DEAN Jeffrey, GHEMAWAT Sanjay, OSDI, MapReduce: Simplified Data Processing on Large Clusters. 2004.

6. JIN Hai, IBRAHIM Shadi, LI QI, HAIJUN Cao, WU Song, XUANHUA Shi. The MapReduce Programming Model and Implementations. Book Chapter in Cloud Computing: Principles and Paradigms.

7. VAVILAPALLI Vinod Kumar, MURTHY Arun C., DOUGLAS Chris, AGARWAL Sharad, KONAR Mahadev, EVANS Robert, GRAVES Thomas, LOWE Jason, SHAH Hitesh, SETH Siddharth, SAHA Bikas, CURINO Carlo, O'MALLEY Owen, RADIA Sanjay, REED Benjamin, and BALDESCHWIELER Eric. Apache Hadoop YARN: yet another resource negotiator. In SOCC '13.

UE-MSD05 – IT Tools – MSD 05.2 - 1st Semester

IT Tools 2 (Big Data Processing with Spark, NoSQL)

Professors	:	Nikolaos PARLAVANTZAS (Irisa Rennes) - NoSQL Hervé MIGNOT (Equancy) – Big Data Processing with Spark
ECTS Credits	:	2.5
Lectures and Tutorials (total)	:	30 hrs (12+18)
Teaching language	:	English

NoSQL

Professor	:	Nikolaos PARLAVANTZAS (Irisa Rennes) - NoSQL
Estimated personal workload (beyond lecture and tutorial time)	:	14 hrs
Lectures and Tutorials	:	12 hrs (ENSAI) including 1,5 h of independent work
Software	:	
Course materials	:	
Prerequisites	:	Computer systems, architecture and databases basic knowledge SQL language practice

Learning Objectives

Understand the fundamentals of NoSQL databases and the features and specific challenges NoSQL databases are addressing, compared to classic SQL databases.Get some introduction to-Gain hands-on experience in deploying and using NoSQL databases, such as MongoDB or Cassandra.

Main Subjects covered

- NoSQL origins (history & players)
- Key concepts of databases:
  - ü CAP Theorem
  - ü ACID transactions
  - ü BASE capabilities
- SQL / NoSQL high level comparison
- NoSQL databases architecture
- NoSQL databases overview & comparison (MongoDB, Cassandra, Neo4j, Redis, ElasticSearch…)
- Neo4j introduction + lab
- Cassandra introduction + lab
- ElasticSearch introduction + lab

Evaluation

Questionnaire and Project

References

Many online resources are available

Big Data Processing with Spark

Professor	:	Hervé MIGNOT (Equancy)
Estimated personal workload (beyond lecture and tutorial time)	:	
Lectures and Tutorials	:	18 hrs (ENSAI) including 3 hrs of independent work
Software	:	
Course materials	:	
Prerequisites	:	Familiar with Linux command-line; Familiar with Python

Learning Objectives

Companies & organizations are collecting massive amounts of various data, making distributed storage & processing key technological challenges. Over the last decade, several cornerstone systems have been released to address these topics, such as Apache Hadoop and more recently Apache Spark. Promoted by a vivid open source world, distributed storage projects are blooming while Apache Spark is becoming a de facto standard for data processing. Beyond data processing and transformation, building data science applications using statistical and machine learning is now the new challenge, requiring both distributed learning and prediction engines. Also, dealing with streaming data for near real-time data processing is getting momentum as companies move to more event processing oriented architecture to cope with the data deluge they are facing.

The goal of this course is to understand key concepts of distributed data processing frameworks and get practice with Apache Spark. All steps of a typical data science project using large volumes of data will be covered: accessing data sources, preparing and processing data, storing them, but also using distributed machine learning libraries such as Apache Spark MLlib to train and apply models. Emphasis will be set on practice & hands-on sessions.

Main Subjects covered

Course description (throughout the course we will cover the following topics)

- Distributed Storage & Computing: key concepts, origins, challenges, and examples. Introduction to Apache Spark, pySpark & SparkR (3 hrs)
- Apache Spark first hands-on session (3 hrs)
- Data sources and data manipulation (3 hrs)
- SparkSQL hands-on session (3 hrs)
- Distributed Machine Learning and pipelines (3 hrs)
- SparkMLlib, pipelining hands-on session (3 hrs)

Evaluation

Project evaluation and written exam

References

1. CHAMBERS Bill, ZAHARIA Matei. O'REILLY Media. Spark: The Definitive Guide. February 2018.
2. KARAU Holden, WARREN Rachel. O'REILLY Media. High Performance Spark. June 2017.
3. RYZA Sandy, LASERSON Uri, OWEN Sean & WILLS Josh. O'Reilly Media. Advanced Analytics with Spark. 2nd Edition. March 2017.
4. Many articles on Databricks Blog: https://databricks.com/blog

1st Semester

TEACHING UNIT MSD-06 :
CHALLENGES FOR SMART SOCIETIES

Supervisor	:	Valentin PATILEA (ENSAI)
ECTS Credits	:	5
Estimated personal workload (beyond lecture and tutorial time)	:	
Lectures and Tutorials	:	60 hrs

Learning Objectives of the Teaching Unit

This unit is a complement for the more methods and tools-oriented Teaching Units. It is designed to further encourage the curiosity and interest of the MSc students for applications. It is also designed as a link between the mathematical modeling and computer science knowledge and the professional world.

Description

A module on energy problems and another one on special topics and case studies allow students to become familiar with problems and modern solutions for current societal problems. A module on the business aspects initiate the MSc students to business environment and issues. A project completes the teaching unit. It is expected to gather students in groups of 2 or 3. Each group must solve a real data problem proposed by a partner company or institute, under the supervision of an external specialist. A report and an oral defense of the project culminate this unit.

Acquired Skills

Make connections between models, algorithms, and concrete problems.

Pre-requisites

Complete the previous Teaching Units.

UE-MSD06-Challenges for Smart Societies – MBD 06.1 - 1st Semester

Energy Transitions: Quantitative Aspects

Professor	:	Edouard CIVEL – Ecole Polytechnique & Climate Economics Chair (Paris Dauphine University)
ECTS Credits	:	1
Estimated personal workload (beyond lecture and tutorial time)	:	3 hrs
Lectures and Tutorials	:	12 hrs (ENSAI) including 3 hrs of independent work
Teaching language	:	English
Software	:	R
Course materials	:	Energy and Economic datasets
Prerequisites	:	Data Analysis with R

Learning Objectives

Today's energy transition raises multiple issues, questioning political choices but also industrial firms' strategy and consumers' behavior. This course aims at illustrating the challenges of the upcoming energy transition using empirical data explored through different econometric strategies.
At the end of the lectures, the student will know how to collect and analyze various types of energy data through different econometric strategies.

Main Subjects covered

A - The industrial revolution: a major energy transition (application: the price of light)
B - Energy: markets, prices and dynamics (application: interactions between oil, natural gas and coal)
C - Externality pricing and the low-carbon transition (European carbon market)

Evaluation

Practical Exercises

References

1. GRUBLER, A. Energy transitions research: Insights and cautionary tales. Energy Policy, 50, 8-16. 2012.
2. FOUQUET, R. The slow search for solutions: Lessons from historical energy transitions by sector and service. Energy Policy, 38(11), 6586-6596. 2010.
3. DINDA, S. Environmental Kuznets curve hypothesis: a survey. Ecological economics, 49(4), 431-455. 2004.
4. BEINE, M., BOS, C. S., & COULOMBE, S. Does the Canadian economy suffer from Dutch disease?. Resource and Energy Economics, 34(4), 468-492. 2012.
5. KEPPLER, J.H., & MANSANET-BATALLER, M. Causalities between CO2, electricity, and other energy variables during phase I and phase II of the EU ETS. Energy Policy, 38(7), pp. 3329-3341. 2010.
6. JAFFE, A. B., & STAVINS, R. N. The energy-efficiency gap- What does it mean? Energy policy, 22(10), 804-810. 1994.
7. APERGIS, N., & PAYNE, J. E. Renewable energy consumption and growth in Eurasia. Energy Economics, 32(6), 1392-1397 - 2010.
8. JOUVET, P. A., & SOLIER, B. An overview of CO 2 cost pass-through to electricity prices in Europe. Energy Policy, 61, 1370-1376 - 2013.
9. DE PERTHUIS, C., & JOUVET, P. A. Green Capital: A New Perspective on Growth. Columbia University Press. 2015.

UE-MSD06-Challenges for Smart Societies – MBD 06.2 - 1st Semester

Smart Data Project

Supervisors : Several industrial partners

Learning Objectives

The main part of courses focuses on studying several facets of statistics, mathematics and computer sciences, according to the Big/Smart Data paradigm. One of the main objectives of this project is to apply this new knowledge learned among the 1st semester into a unique application. This project puts into practice theoretical methods studied in different courses and starts with project management.

The learning objective is not limited to putting the theory learned in other courses into practice, but aims to raise awareness of other aspects linked to project management among students, such as communication (between students and also with the client that proposed the project).

This project should provide additional support, be carried out by an expert of the field, according to the needs of students. The expert is expected to provide

- Supervising at start for requirement
- Distant supervising on technical queries
- Technical supervising during implementation phase
- Help for defense preparation

Main Subjects covered

The topic of the Smart Data project could be related to any type of application requiring advanced data science tools.

Evaluation

The evaluation is two-fold:

1 - a report written by all students of each project team, eventually supervised by the external organism.
2 - a project defense in front of a jury

UE-MSD06-Challenges for Smart Societies – MBD 06.3 - 1st Semester

Topics & Case Studies in Data Science

Professors		Romaric GAUDEL (ENSAI) Shadi IBRAHIM (INRIA - Rennes) Valeriu PETRULIAN
	:	Thomas ZAMOJSKI (DATASTORM)
ECTS Credits	:	2
Lectures and Tutorials (total)	:	24 hrs (Ensai)
Teaching language	:	English

Bandit Theory

Professor	:	Romaric GAUDEL
Estimated personal workload (beyond lecture and tutorial time)	:	2 hrs
Lectures and Tutorials	:	6 hrs (ENSAI)
Software	:	Python (and Python notebook)
Course materials	:	Slides
Prerequisites	:	Being confortable with classes and objects in Python - Not required, but recommended: basic knowledge about Machine Learning / Statistical Learning

Learning Objectives

At the end of the lectures, the student will be able to:

- Identify settings requiring exploration
- Propose an alternative framework to A-B testing
- Develop simple Bandits algorithms

Main Subjects covered

Nowadays, more and more decisions are made by computers. While some of these decisions arise from programs written by humans, some of them are the result of data analysis: the algorithm looks at the past to identify efficient decisions, and repeats these decisions in the present. In such a context, the decisions have a short term impact (they can be good or bad), but they also have a long term impact: the results of these decisions will be used to take future decisions.

During this course, we will focus on a "simple" setting (the Multi-Armed Bandit) and show that to be optimal in the long run, an algorithm has to act in a counter-intuitive way: from time to time, the algorithm as to take a decision which is sub-optimal given the past data (aka. "explore").

The course presents the setting, gives a sketch-proof for the need for exploration and contains a practical session to implement optimal algorithms.

Evaluation: Quizz at the end of practical session - Pratical session notebook

References

1. BUBECK Sébastien, CESA-BIANCHI Nicolò. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. Foundations and Trends in Machine Learning 5, 1-122.
2. LATTIMORE Tor and SZEPESVARI Csaba. Bandit Algorithms.Cambridge University Press.
   - free pdf version : https://tor-lattimore.com/downloads/book/book.pdf
   - corresponding blog : http://banditalgs.com/
   - tutorial slides related to the book : http://banditalgs.com/2018/02/09/bandit-tutorial-slides-and-update-on-book/

Is Data the New Currency of the Digital Economy?

Professor	:	Valeriu PETRULIAN
Estimated personal workload (beyond lecture and tutorial time)	:	1.5 – 2 hrs (reading prior to coming to class)
Lectures and Tutorials	:	6 hrs including 1 h of independent work
Course materials	:	.ppt slides provided at the end of the course
Prerequisites	:	See below reading material prior to coming to class

Learning Objectives

Welcome to the Digital Economy! “Fuel” for some, “currency” for others, Data has become in just few years a core component of our modern societies and economic systems and its importance is still growing, especially in conjunction with Artificial Intelligence (AI). This lecture will provide an economic, business and management perspective on Data & AI, as fundamental trends in today’s economy. Using several multidisciplinary concepts, the lecture will illustrate how Data, Analytics and AI are about to transform organizations and how they impact our professional lives.
The “Data as Currency” course aims at attaining, for prospective students, the following specific objectives:

- Understand the present challenges of the “Digital Revolution” and its implications for the business environment
- Explore the “game-changing” nature of selected digital technologies, specifically Big Data and AI
- Assess, concretely, the changing potential of digital technologies on existing business models Be able to articulate, by the end of the course, a “Digital Transformation” personal vision

Main Subjects covered

Part 1:

- Welcome to the digital economy! A perspective on how Digital Technologies have become central to today’s economy
- Group Discussion/Team Work 1: Innovation & Technology adoption

Part 2:

- The Digital (r)Evolution : At the heart of the Digital Transformation of our economic systems, Data and AI change the way we work and how companies do business
- Group Discussion/Team Work 2: Digital Platforms

Part 3:

- Digital Transformation of Industries - Illustrations of Data & AI usages and applications across several corporate functions (for example: marketing & sales, operations, etc.)
- Group Discussion/Team Work 3: Applications of Data and AI in the workplace
- Individual Pitches: As a Data Specialist, I will apply for a job, what will it be?

Evaluation: Team work and individual pitches will be awarded grades.

References

The following is a list of reading material. Please read items 1 and 2 prior to the lecture:

1. BERNSTEIN Amy, RAMAN Anand. The Great Decoupling: An Interview with Erik Brynjolfsson and Andrew McAfee. Harvard Business Review, June 2005 Issue
2. DAVENPORT Thomas H., RONANKI Rajeev. Artificial Intelligence for the Real World. Harvard Business Review, January–February 2018 Issue
3. BEAN Randy. How Companies Say They’re Using Big Data. Harvard Business Review, April 28, 2017.
4. WORLD ECONOMIC FORUM, Centre for the New Economy and Society. Data Science in the New Economy. A new race for talent in the Fourth Industrial Revolution. July 2019.
5. CHRISTENSEN C. & al. Big Idea: What is disruptive innovation? Harvard Business Review, Dec. 2015.

Some Recent Advances for Big Data Processing in the Cloud

Professor	:	Shadi IBRAHIM (INRIA – Rennes)
Estimated personal workload (beyond lecture and tutorial time)	:	3 to 5 hrs
Lectures and Tutorials	:	6 hrs (ENSAI) including 1 h of independent work
Software	:	Hadoop
Course materials	:	All course materials presentations, tutorials and hand-ons, libraries and codes will be available online on the course website in pdf and zip format.
Prerequisites	:	Attend the course: Big Data processing in Clouds: Hadoop

Learning Objectives

At the end of the lectures, the student will be able to identify the main performance bottlenecks when running Big data applications in Clouds and will know how the performance of Hadoop can be improved, accordingly.
During this conference, we will discuss several approaches and methods used to optimise the performance of Hadoop in the Cloud. We will also discuss the limitations of Hadoop and introduce state-of-the-art resource management systems and job schedulers for Big data applications including Mesos, Delay scheduler, and ShuffleWatcher.

Main Subjects covered

Approaches to optimize Hadoop in clouds (2.5 hrs)
Resource management and job scheduling for Big data applications: Mesos, Delay scheduler, ShuffleWatcher, etc (2.5 hrs)

Independent work (tentative): Students will be assigned to groups where each group will do a 15 -20 min presentation (1 hr)

Evaluation

During the session and/or a technical report to be submitted after the session

References

1. Apache Hadoop YARN: yet another resource negotiator. VAVILAPALLI Vinod Kumar, MURTHY Arun C., DOUGLAS Chris, AGARWAL Sharad, KONAR Mahadev, EVANS Robert, GRAVES Thomas, LOWE Jason, SHAH Hitesh, SETH Siddharth, SAHA Bikas, CURINO Carlo, O'MALLEY Owen, RADIA Sanjay, REED Benjamin, and BALDESCHWIELER Eric. In SOCC '13.
2. IBRAHIM Shadi, PHAN Tien-Dat, CARPEN-AMARIE Alexandra, CHIHOUB Houssem-Eddine, MOISE Diana, ANTONIU Gabriel. Governing energy consumption in hadoop through cpu frequency scaling: An analysis. In FGCS 2016.
3. PHAN Tien-Dat, IBRAHIM Shadi, ANTONIU Gabriel, BOUGE Luc. On Understanding the energy impact of speculative execution in Hadoop. In GreenCom2015.
4. YILDIZ Orcun, IBRAHIM Shadi, ANTONIU Gabriel. Enabling fast failure recovery in shared Hadoop clusters: Towards failure-aware scheduling.In FGCS 2016.
5. HINDMAN Benjamin, KONWINSKI Andy, ZAHARIA Matei, GHODSI Ali, JOSEPH Anthony D., KATZ Randy, SHENKER Scott, and STOICA Ion. Mesos: a platform for fine-grained resource sharing in the data center. In NSDI'11.

6. ZAHARIA Matei, BORTHAKUR Dhruba, SEN SARMA Joydeep, ELMELEEGY Khaled, SHENKER Scott, STOICA Ion. Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling. In EuroSys'10.
7. ZAHARIA Matei, KONWINSKI Andy, JOSEPH Anthony D., KATZ Randy, STOICA Ion. Improving MapReduce performance in heterogeneous environments. In OSDI'08.
8. AHMAD Faraz, CHAKRADHAR Srimat T., RAGHUNATHAN Anand, VIJAYKUMAR T. N. Shufflewatcher: Shuffle-aware scheduling in multi-tenant mapreduce clusters. In USENIX ATC 2014

Case Studies in Smart Data
MLOps: Machine Learning in a production environment

Professor	:	Thomas ZAMOJSKI (DATASTORM)
Estimated personal workload (beyond lecture and tutorial time)	:	1.5 – 2 hrs
Lectures and Tutorials	:	6 hrs (ENSAI) including 1 h of independent work
Software	:	Python, Docker
Course materials	:	
Prerequisites	:	Basic knowledge of Python Programming Language

Learning Objectives

At the end of the lecture, the student will know:

- What are the challenges in deploying and maintaining a machine learning model in operation.
- What are some best practices addressing these concerns.
- How to create a Docker image and run a container.
- How to serve a model as a service in python.
- Statistical methods for online and offline model monitoring.

Main Subjects covered

Machine Learning models are notoriously hard to put and maintain in production. But why is it so and what can we do about it?

In this course, we will explore the very latest trends in MLOps. We will learn about technologies such as Docker containers, FastAPI and MLFlow. We will also learn statistical methods to intelligently automate model monitoring and we will see how to put them in action via implementations in python packages such as scikit-multiflow and ruptures.

Evaluation

60% In-class exercises,
30% Code quality and clarity,
10% Participation.

References

1. TRUONG C., OUDRE L., VAYATIS N., Selective review of offline change point detection methods, Signal Processing, September 2019.
2. JAMES N.A., KEJARIWAL A., MATTESON D.S., Leveraging cloud data to mitigate user experience from ‘breaking bad’, 2016 IEEE International Conference on Big Data (Big Data).
3. WEB REFERENCES - 12 factors app: https://12factor.net

1st Semester

TEACHING UNIT MSD-07 :
FRENCH AS A FOREIGN LANGUAGE

Supervisor	: Todd DONAHUE (ENSAI)
ECTS Credits	: (8)
Estimated personal workload (beyond lecture and tutorial time)	:
Lectures and Tutorials	:

Learning Objectives of the Teaching Unit

Give students practical written and/or oral French skills, necessary for practical life in France.

Description

Weekly evening courses

Acquired Skills

Pre-requisites

UE-MSD07 - French as a Foreign Language –MSD07.1 -1st Semester – For foreign students as needed

French: Language & Civilization

Professor	:	Séverine BORDEAU
ECTS Credits	:	
Estimated personal workload (beyond lecture and tutorial time)	:	
Lectures and Tutorials	:	
Teaching language	:	
Software	:	
Course materials	:	
Prerequisites	:	

Learning Objectives

French Language & Civilization courses allow students to develop and hone their knowledge of the language and culture of the country in which they are studying. These courses are focused on giving the students the linguistic skills they need for their daily life in France and for their integration into the ENSAI student body.

Main Subjects covered

Designed for foreign students who are following a full-time academic program in Rennes, these weekly evening courses give students practical written and/or oral French skills, necessary for practical life in France.

Course Evaluation

Quizz and Exams

References

Various textbooks and authentic audiovisual documents will be used.

Second Semester

2nd Semester

TEACHING UNIT MSD-08 : INTERNSHIP

Supervisor	: Valentin PATILEA (ENSAI)
ECTS Credits	: 5
Working time	: Full time internship, for a period between 4 and 6 months

Learning Objectives of the Teaching Unit

The internship is the main bridge between, on one hand, the scientific courses, tutorials and labs and, on the other hand, the world of work. It has two major objectives. First, consolidate students' ability to choose appropriate models, algorithms and computer resources to address real data applications and case studies, to realize proof of concepts and/or develop user solutions, and, finally, explain and provide appropriate arguments for the choices made. Second, place the students in total immersion in a professional environment, in autonomy, as part of a team, in interaction with specialists from the same or complementary fields.

Description

The MSc students are expected to work on topics defined in the internship agreement, under the supervision of a senior professional from the internship unit (private or public company, labs, research institutes…). Each MSc student will have an Ensai adviser who can be contacted for advice.

Acquired Skills

Become a highly skilled specialist in data science able to address complex tasks using up to date modeling tools and computer resources.

Pre-requisites

Complete the previous teaching unit from the MSc program.

UE-MSD08 - Internship – MSD08.1

End-of-Studies Internship

4-6 months from March to August

Objectives

This final phase of the MSc in Statistics for Smart Data program involves a four to six-month paid internship, which can take place either in France or abroad, in either the professional world or academic/research laboratories.

Students should be proactive and begin the search for an internship as early as possible to increase the chances of finding an interesting and relevant internship. Finding an internship is the exclusive responsibility of the student. ENSAI provides assistance in the search process.

This experience should allow for the student to apply the data-science and computer science theory and methods that they have learned during the 1st semester of coursework. Internship topics that are exclusively or almost exclusively oriented towards computer science tools will not be accepted.

The internship should allow students to meet at least two objectives:

- A technical objective: a task is given and, applying theoretical knowledge and skills, the student attempts to complete the task using to the best of his/her ability the resources at his/her disposal.

- A professional objective: the student is immersed in a professional context and must use the internship period to become more knowledgeable and at ease in such an environment, developing professional and personal skills to become a part of the team.

Evaluation

During the internship, students will write a master’s thesis that will be examined by the jury and defended by the student in September.