Master in Computer Sciences, Nantes Université

Track : Data Sciences

Semester 9 (FALL : September to January)

There are ten teaching units during the ninth semester (second year of the Master, first semester), all of them are 24 hours long and provide 3 ECTS each.

Semester 10 (SPRING : February to July)

There is only one teaching unit during the tenth semester (second year of the Master, second semester) : the internship, which is 30 ECTS worth.

Timeline of First Semester

The first semester is divided into three periods of 6 weeks each. During each period, 3 courses are given. One course, RMCS, spans over the three periods due to the project.

Period 1 (W37-W42)	Period 2 (W43-W49)	Period 3 (W50-W5)
Visual Analytics	Advanced graphs and networks	Semantic AI
Deep learning[1]	Pattern mining	Reinforcement learning and recommender systems
Graphical Models	Data intensive processing	Ethics, Data and AI
Research methodology and case study[2]		

Hourly planning

Acronym	Teaching Unit	ECTS	Hours	In charge	Period
DL	*Deep learning*[1]	3	24	HM	1
SAI	Semantic AI	3	24	CM	3
AGN	Advanced Graphs and Networks	3	24	HLC	2
RRS	Reinforcement learning and recommender systems	3	24	MG	3
GM	Graphical Models	3	24	PHL	1
VA	Visual Analytics	3	24	YP	1
PM	Pattern Mining	3	24	FG	2
EDA	Ethics, Data and AI	3	24	GR	3
DIP	Data intensive Processing	3	24	JM	2
RM	*Research methodology and case study*[2]	3	24	HLC	1,2,3
		30	240		

[1] Shared with ATAL and VICO tracks
[2] Partly shared with ATAL and VICO tracks

Deep learning

Instructors : **Harold Mouchère**, Alexandre Bruckert

Description :

This course is an in-depth introduction to the principles and techniques of deep learning, a branch of artificial intelligence that has transformed many fields such as computer vision, natural language processing, and speech recognition. Students will learn to design, train, and evaluate various types of deep neural networks, with a focus on modern architectures and regularization techniques.

Syllabus :

- **Introduction to Neural Networks** :
    - History and evolution
    - Applications and use cases
- **Multilayer Perceptron (MLP)** :
    - MLP architecture
    - Activation functions
    - Training and optimization
- **Convolutional Neural Networks (CNN)** :
    - Convolutions and pooling
    - Classic architectures
    - Regularization techniques for CNNs
    - Applications in computer vision
- **Recurrent Neural Networks (RNN)** :
    - RNN architectures
    - Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)
    - Applications in natural language processing and time series
- **Practical Sessions** :
    - Implementation of simple models using a popular framework (PyTorch)
    - Practical case studies and projects

Acquired skills :

Upon completion, the student will be able to:

- Understand the fundamental principles of neural networks and deep learning techniques.
- Design and implement neural network architectures tailored to various problems.
- Evaluate the performance of deep learning models and apply regularization techniques to improve results.
- Use deep learning frameworks to develop solutions to real-world problems.

Prerequisites :

- Mathematics (linear algebra, calculus)
- Programming (Python)
- Basic knowledge in machine learning (Bayesian classifier, system evaluation)

Evaluation :

- Written exam (50%)
- Lab work and projects (50%)

Semantic AI

Instructors : Claudia Marinica, Mounira Harzallah

Description :

This course proposes to the students an extended introduction to semantic models and their use in AI techniques. To this end, the course will first present the ontology formalization in OWL and querying. Second, the course proposes an introduction to ontology building methodologies. Last, the course will study how to integrate semantic information in artificial intelligence techniques.

Syllabus :

- Introduction to semantic models
- Ontology and Applications
- Ontology Formalisation
    - Introduction to description logics
    - OWL/protege (LabWork)
- Ontology querying
- Ontology Learning
    - Ontology Building
    - Pipeline for ontology learning from texts
- Knowledge-driven AI
    - User Knowledge Formalisation (ontologies, rules, schemas, topologies, ...)
    - Unsupervised approaches driven by user knowledge
    - Supervised approaches driven by user knowledge
- Project (Ontology learning from texts + Ontology validation (competencies))

Acquired skills :

Upon completion, the student will be able to:

- Build manually an ontology in a given domain from text
- Build an ontology by using AI techniques from text
- Validate the ontology that was built manually or by using AI techniques
- Formalize user knowledge by means of ontologies and other formalism such as rules and schemas
- Propose a knowledge-drive AI approach to a given problem

Prerequisites :

- Programming (Python)
- Basic knowledge in machine learning

Evaluation :

- Written exam (40%)
- Project (60%)

Advanced Graphs and Networks

Instructors : **Hoel Le Capitaine**, François Queyroi

Description :

Complex data can be represented as a graph of relationships between objects. Such networks are a fundamental tool for modeling social, technological, and biological systems. This course focuses on the computational, algorithmic, and modeling challenges specific to the analysis of graphs. By means of studying the underlying graph structure and its features, students are introduced to machine learning techniques and data mining tools apt to reveal insights on a variety of networks.

Syllabus :

Introduction to network analysis :

- Relational data modelization
- Measures in networks (centrality, distances, etc.)
- Network clustering (spectral methods, approximations, …)

Graph representation learning :

- shallow methods: Node2vec, Deepwalk
- deep methods: graph neural networks
- Graph Transformers
- article reading from last LoG conference

Acquired skills :

Upon completion, the student will be able to :

- modelize network and graph data
- choose an adapted method to the data depending on their objectives
- give insights on algorithms bias
- embed and generate graph data using deep learning approaches
- perform standard prediction and classification tasks on graph data

Prerequisites : Graph theory, statistics, machine learning, Deep Learning

Evaluation :

- Written exam (60%)
- Lab work (40%)

Reinforcement learning and recommender systems

Instructors : **Marc Gelgon**, Hoel Le Capitaine

Description : Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward. In this course, we will study the elemental aspects needed to define RL environments, and the different approaches and algorithms that can be used. A second part of the course is dedicated to recommender systems. Recommender systems are algorithms and technologies designed to suggest relevant items to users based on their preferences and behaviors. These systems are widely used in various applications to help users discover products, services, or content that they might find interesting or useful.

Syllabus :

1. Introduction to reinforcement learning and Markov Decision Processes (MDPs)
2. Dynamic Programming and model-based reinforcement learning
3. Model-free reinforcement learning and deep reinforcement learning
4. Socio-economics motivating and characterizing recommendation systems
5. Models and algorithms for recommendation

Acquired skills :

Upon completion, the student will be able to :

- design a scientific solution based on reinforcement learning, suited to a given problem
- implement and characterize experimentally reinforcement learning techniques
- analyze a situation involving a recommendation task from socio-economics perspectives
- select, implement and evaluate experimentally a solution for recommendation, suited to a given problem

Prerequisites : Machine learning, statistics

Evaluation :

- Written exam (60%)
- Lab work (40%)

Graphical Models

Instructors : Philippe Leray, Hoel Le Capitaine

Description :

Graphical Models (GMs) are a branch of Machine Learning which uses a graph to represent a domain problem. Some examples of such Graphical Models in Machine Learning or Deep Learning models are Bayesian networks, Graphical Event Models or Restricted Boltzmann Machines.
In this course, we will focus on the definition of these models, and we will discuss how probabilistic inference is made and how these models are learned from data, using the two main paradigms of machine learning : supervised and unsupervised learning.

Syllabus :

- Introduction to GMs and Bayesian networks: definition and probabilistic inference
- Bayesian network learning: parameter learning and structure learning
- Another directed probabilistic GM : Graphical Event models
- Markov random field : Energy and probability, Boltzmann machine and Restricted Boltzmann machine (RBM)
- MCMC : Markov chain and its limiting probability distribution, Importance sampling, Gibbs sampling, Metropolis-Hasting algorithm
- Unsupervised feature learning : Contrastive divergence (CD), Stacked RBM, Deep belief network (DBN)

Acquired skills :

Upon completion, the student will

- model simple problems with probabilistic graphical models such as Bayesian networks, Graphical Event models, …
- understand probabilistic inference with an exact algorithm, or with an approximate one,
- understand parameter/Structure learning algorithm dedicated to such models

Prerequisites :

Linear algebra, probability theory, statistics, machine learning

Evaluation :

- Written exam (65%)
- Lab work (35%)

Visual Analytics

Instructors: **Yannick Prié,** Fabien Picarougne

Description:

Visual analytics is the process of analyzing data using both the strengths of humans (intuition, visual perception, etc.) and computers (rapid data manipulation and aggregation, dataviz generation, feature extraction algorithms, etc.). The module begins with the general principles of visual analytics, as well as the evaluation of visual analytics systems using quantitative and qualitative methods. Then we focus on three important sub-themes of visual analytics: multidimensional data, graph visualization and immersive analytics, each with a lecture, a lab session and several paper presentations by the students. At the end of the course, students will prepare an in-depth study on a specific topic, which they will present to their peers.

Syllabus:

- Visual analytics
- Evaluating visual analytics systems
- Multidimensional data VA
- Graph data VA
- Immersive Analytics
- In-depth study

Acquired skills:

- Describe an existing visual analytics system along several dimensions: interaction, visualizations and mappings, processes, automatization, etc.
- Analyze a visual analytics research article focusing on important domain characteristics: justification, contributions, evaluation, etc.

Prerequisites:

- Data analysis
- We consider the students already know the basics of data visualization: definition, mapping, psychovisual important characteristics, main data visualization (bar chart, line chart, scatter plots, choropleth maps, etc), as well as how to use tools such as Tableau.

Evaluation:

- Article presentations (25%)
- Project (25%)
- Written exam (50%)

Pattern Mining

Instructors : **Fabrice Guillet**, Julien Blanchard

Description :

Pattern Mining is a part of Data Sciences which is concerned with the discovery of interesting data patterns by machine learning. Patterns may be subsets, subsequences, subspaces, subgraphs that are both frequently occurring in the data, statically relevant, and significant for a user (i.e. decision maker). The teachings are organized in 3 parts : an introduction to the general principles of pattern mining and quality metrics, then a first extension to sequence mining, and at least an extension to graph mining. In addition, the principles of pattern mining are coded and experimented on data during an 8 hours project homework.

Syllabus :

- introduction to pattern mining
  - patterns
  - mining algorithms, quality metrics, (discriminant patterns)
- sequence and episode patterns
  - principles
  - sequence models
  - applications
- graph patterns
  - principles
  - extension to GNN and application to molecular mining
- homework project

Acquired skills :

- understand what are patterns for sets, sequences, graphs
- understand the quality measures relevant for these patterns
- apply mining algorithm to extract patterns
- interpret the results
- implement the pattern extraction with python/R libraries

Prerequisites :

- Basics in Machine learning, graphs, probabilities, statistics
- python programming

Evaluation :

- Project Homework (50%)
- Written exam (50%)

Ethics, Data and AI

Instructors : **Guillaume Raschia**, Marc Gelgon, Hoel Le Capitaine

Description : Privacy of individuals is paramount in any Data Science project and ML workflow as well. We focus in this course on anonymization techniques that allow to balance privacy preservation vs. data utility in various scenarios such as data publishing, data mining and machine learning. We will also present the notion of explainability from the perspective of different end users, discuss different classes of interpretable models and post hoc explanations (e.g., rule-based and prototype-based models, feature attributions, counterfactual explanations, mechanistic interpretability), and explore the connections between explainability and fairness, robustness, and privacy.

Syllabus :

- Privacy & Anonymization
    - Privacy issues, laws and rules (GDPR)
    - k-Anonymity principle and limitations, derivative models
    - Privacy vs. utility trade-off, generalization lattice and algorithms
    - Differential Privacy (DP) model, sensitivity, privacy budget
    - Practical DP: Laplacian mechanisms, exponential mechanism
    - DP Programming frameworks
- Fairness
    - legitimacy, stakes and law of automated decision
    - fairness criteria and their properties
    - algorithms and techniques for implementation of criteria
- Interpretability and explainability
    - Human factors
    - Interpretable models
    - Post hoc explanations
    - Impacts on fairness and privacy
    - Opening to L(L)Ms

Acquired skills :

Upon completion, students will be able to:

- identify privacy requirements in data science projects
- decide for the relevant privacy model and its parameters
- perform anonymization within a dedicated programming toolbox
- characterize situations where automated decisions affects humans, from a non-technical perspective
- evaluate technical solutions for implementing fairness
- build technical solutions for implementing fairness

Prerequisites :

Statistics, Algorithmics, Python Programming

Evaluation :

- Written exam (70%)
- Lab work (30%)

Data-intensive Processing

Instructors : **José Martinez**, Guillaume Raschia

Description :

More and more data is produced and has to be processed. As it becomes hardly possible to store it all, e.g., in a database, for postponed processing, we have to process it on-the-fly. The summarized results can possibly be stored or more often further pushed into another processing step. This leads to distributed and pipelined processing architectures.

Syllabus :

- [Database] Data Models and Query Languages
  - Limitations of the Relational Model
  - Nested Relations and Algebra
  - Document Data Model, Encoding and Querying
  - Graph Data Models
  - All in all, the NoSQL Landscape
- [Data Stream] Data Stream Processing and Complex Event Processing
  - Models, Timing
  - Operators: Filtering, Join, Pattern Matching, Windowing
  - Data Stream – Space-Efficient – Algorithms: Counting, Rank and Order Statistics
  - Time Window Aggregation
  - Architectures and Practical Frameworks
- [High Performance Computing] Introduction to Distributed Computing
  - Differences and Similarities between Distributed and Parallel Systems
  - Benefits and Limitations: Scalability, Speed-up
  - Centralised vs. Distributed Control: From Multi-cores to Edge Computing
  - Case Study on Processing Data Streams

Acquired skills :

Upon completion, the student will be able to:

- Design and implement space-efficient algorithms as building blocks for many data science problems
- Deploy a fully-fledged stream data pipeline
- Possibly design a software and hardware architecture in order to process data from sources to users

Prerequisites :

Algorithmics and Complexity Analysis, Database, Networks

Evaluation :

- Written exam (67%)
- Lab work (33%)

Research methodology and case study

Instructors :
Yannick Prié, Nicolas Hernandez, **Hoel Le Capitaine**

Description :
This teaching unit is composed of three different blocks.
The first block is dedicated to understanding what scientific knowledge production means, how scientific knowledge develops with scientific publishing, what scientific disciplines are and to discuss the relations between science and society. Then we focus on the content of scientific documents, how to read them, and how to write scientific material, before addressing the topic of how to search for scientific material with the tools at hand and organize them in a bibliography.
The second block deals with open science, sustainability and general ethics.
The third block is a small scale research project driven by a supervisor from the LS2N laboratory. The project typically consists in addressing and solving a research problem by reading state-of-the-art papers, making a research proposal and evaluating it through experiments. This project spans the entire semester.

Syllabus :

Acquired skills :
Upon completion, the student will

Prerequisites : none

Evaluation :