Année universitaire 2026/2027

Intelligence Artificielle, Systèmes, Données (IASD) - Informatique - 2e année de Master en alternance

Responsable pédagogique : TRISTAN **CAZENAVE** - https://dauphine.psl.eu/recherche/cvtheque/tristan-cazenave

Crédits ECTS : 60

LES OBJECTIFS DE LA FORMATION

La **2e année du Master Informatique parcours Intelligence Artificielle, Systèmes, Données - Informatique en apprentissage**, formation d'excellence, offre de solides connaissances en mathématiques appliquées et conception de systèmes d'intelligence artificielle afin de couvrir l'ensemble des problématiques de traitement d'analyse des données massives que rencontrent les entreprises. Il met l'accent sur l'articulation entre apprentissage automatique, gestion et fouille de grandes masses de données, paradigmes du Big Data, représentation des connaissances, le traitement des données et sur les méthodologies récemment développées.

Les objectifs de la formation :

- Former des informaticiens capables de maîtriser les problèmes conceptuels, sémantiques et algorithmique soulevés par l'intelligence artificielle et la science des données ;
- Développer une compréhension générale et en profondeur des différentes facettes de l'IA ;
- Renforcer des connaissances théoriques des étudiantes et des étudiants et leur transmettre une expérience pratique de l'Intelligence Artificielle et des Sciences des Données.

Les cours sont dispensés sur le site de PariSanté Campus situé au 2-10 rue d'Oradour-sur-Glane, 75015 Paris.

MODALITÉS D'ENSEIGNEMENT

Les cours sont dispensés sur le site de PariSanté Campus situé au 2-10 rue d'Oradour-sur-Glane, 75015 Paris.

Les Modalités des Contrôles de Connaissances (MCC) détaillées sont communiquées en début d'année.

La formation démarre en septembre et la présence en cours est obligatoire. Les enseignements de la 2e année de master mention Informatique parcours Intelligence Artificielle, Systèmes, Données voie Informatique en apprentissage sont organisés en semestre 3 et semestre 4. Chaque semestre est constitué d'UE auxquelles s'ajoute un mémoire pour le semestre 4.

Le rythme d'alternance est de quatre semaines en entreprise et deux/trois semaines à l'université.

ADMISSIONS

- Titulaires d'un diplôme BAC+4 (240 crédits ECTS) ou équivalent à Dauphine - PSL, d'une université ou d'un autre établissement de l'enseignement supérieur dans les domaines suivants : informatique, mathématiques appliquées avec un attrait pour l'informatique et l'algorithmique ;
- Etudiantes et étudiants en dernière année d'école d'ingénieur (ou ayant obtenu un diplôme d'ingénieur) en lien avec les thématiques de la formation.

POURSUITE D'ÉTUDES

Le Master **Intelligence Artificielle, Systèmes, Données (IASD) - Informatique en alternance** est une formation d'excellence ouvrant vers des carrières de recherche et développement dans des laboratoires de recherche publics ou privés ou des entreprises innovantes.

PROGRAMME DE LA FORMATION

- Semestre 3 - 29 ECTS
  - UE fondamentales
    - Apprentissage par renforcement
    - Apprentissage profond
    - Apprentissage topologique
    - Bases de données avancées (SBGD non classiques)
    - Ethique et science des données
    - Fondamentaux de l'apprentissage automatique
    - IA générative
    - Optimisation pour l'apprentissage automatique
    - Systèmes, paradigmes et langages pour les big data
    - Traitement automatique des langues - NLP
- Semestre 4 - 24 ECTS
  - UE fondamentales
    - Agentic AI
    - Apprentissage profond pour l'analyse d'images
    - Flux de données
    - Machine Learning sur big data
    - Modélisation de problèmes
    - Projet Sciences des Données
    - Recherche Monte-Carlo et jeux
    - Visualisation de données
  - Bloc mémoire - 7 ECTS
    - Memoire

DESCRIPTION DE CHAQUE ENSEIGNEMENT

SEMESTRE 3 - 29 ECTS

UE fondamentales

Apprentissage par renforcement

ECTS : 3

Enseignant responsable : STEPHANE **AIRIAU** (https://www.lamsade.dauphine.fr/~airiau/)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

- Models: Markov decision processes (MDP), multiarmed bandits and other models
- Planning: finite and infinite horizon problems, the value function, Bellman equations, dynamic programming, value and policy iteration
- Basic learning tools: Monte Carlo methods, stochastic approximation, temporal-difference learning, policy gradient
- Probabilistic and statistical tools for RL: Bayesian approach, relative entropy and hypothesis testing, concentration inequalities
- Optimal exploration in multiarmed bandits: the explore vs exploit tradeoff, lower bounds, the UCB algorithm, Thompson sampling
- Extensions: Contextual bandits, optimal exploration for MDP

Compétences à acquérir :

Reinforcement Learning (RL) refers to scenarios where the learning algorithm operates in closed-loop, simultaneously using past data to adjust its decisions and taking actions that will influence future observations. Algorithms based on RL concepts are now commonly used in programmatic marketing on the web, robotics or in computer game playing. All models for RL share a common concern that in order to attain one's long-term optimality goals, it is necessary to reach a proper balance between exploration (discovery of yet uncertain behaviors) and exploitation (focusing on the actions that have produced the most relevant results so far).

The methods used in RL draw ideas from control, statistics and machine learning. This introductory course will provide the main methodological building blocks of RL, focussing on probabilistic methods in the case where both the set of possible actions and the state space of the system are finite.

Pré-requis recommandés

Some basic notions in probability theory are required to follow the course. The course will imply some work on simple implementations of the algorithms, assuming familiarity with Python.

Bibliographie, lectures recommandées :

- Reinforcement Learning: An Introduction, Richard S. Sutton & Andrew G. Barto, Second Edition, MIT Press, 2018
- Bandit Algorithms, Tor Lattimore & Csaba Szepesvári, Cambridge University Press, 2020

Apprentissage profond

ECTS : 3

Enseignant responsable : TRISTAN **CAZENAVE** (https://dauphine.psl.eu/recherche/cvtheque/tristan-cazenave)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

This course is about using deep learning tools.

The objective of the course is to be able to design deep neural networks and to apply them to various problems. The language used for the course is Torch. It relies on the Lua scipting language augmented with tensor specific instructions. During the course, we will use simple examples to learn how to generate and transform data in Torch as well as how to

learn from this data. We will cover deep neural networks, deep convolutional neural networks and some optimizations of the architecture such as residual nets.

Apprentissage topologique

ECTS : 3

Enseignant responsable : MAIXENT **CHENEBAUX**

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

COURSE CONTENT

1. Topological Foundations

- **From distances to topology**: Metric spaces, open balls, neighborhoods, and the transition to abstract topological spaces.
- **Invariants and deformations**: Connectivity, holes, homotopy, and why topology is more fundamental than geometry for learning.
- **Discrete and continuous structures**: Graphs, hypergraphs, simplicial complexes, CW-complexes, and their geometric realizations.

2. Mathematical Tools for Topological Analysis

- **Combinatorial topology**: Simplices, complexes, cochains, and associated matrices (incidence, adjacency, Laplacian).
- **Algebraic topology**: Homology, cohomology, and applications in Topological Data Analysis (TDA).

3. Deep Learning on Topological Structures

- **Message passing**: Convolutions, attention, and diffusion on graphs and higher-order structures.
- **Spectral representations**: Normalized Laplacian, Fourier analysis on graphs, and implications for learning.
- **Limitations and extensions**: Oversmoothing, heterophily, sheaf theory, and Ricci curvature for more robust models.

4. Applications and Implementation

- **Graphs and hypergraphs**: Social networks, knowledge graph embeddings (TransE), community detection.
- **Simplicial complexes**: 3D mesh processing, molecular classification, and topological signal processing.
- **Practical tools**: Hands-on experience with libraries like TopoNetX to prototype topological models.

5. Critical Review and Perspectives

- **State of the art**: GCN, GAT, Simplicial Neural Networks, and foundational graph models.
- **Open challenges**: Topological identification capacity, spatio-temporal structures, and integration with generative AI.

Compétences à acquérir :

- **Understand and apply** core topological concepts (metric spaces, topological spaces, topological invariants) to data science and machine learning problems.
- **Model and analyze** discrete (graphs, hypergraphs, simplicial complexes) and continuous (manifolds, abstract topological spaces) structures within a unified framework.
- **Design and interpret** topological data representations (embeddings, Laplacian matrices, homology) for deep learning tasks.
- **Implement** message-passing algorithms (GNNs, attention mechanisms, diffusion) and spectral analysis on topological structures.
- **Evaluate** the limitations of existing models (oversmoothing, heterophily) and propose solutions inspired by algebraic or differential topology (sheaf theory, Ricci curvature).
- **Apply** these tools to real-world problems: social networks, 3D mesh processing, molecular classification, dimensionality reduction, and structured data generation.
- **Engage with** recent literature in topological learning, including advances in Topological Data Analysis (TDA), manifold learning, and topological neural networks.

Bases de données avancées (SBGD non classiques)

ECTS : 3

Enseignant responsable : LYES **ATTOUCHE** (https://dauphine.psl.eu/recherche/cvtheque/attouche-lyes)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

Le cours a pour objectif d'apprendre aux étudiants les aspects fondamentaux des différents types bases de données qu'elles soient basées sur le SQL, le NoSQL (Not Only SQL) ou récemment le NewSQL.

Le cours s'articule en trois parties. Dans la première partie, l'accent est mis sur les bases de données relationnelles : leurs avantages et leurs inconvénients, ainsi que la correspondance objet-relationnel (Object-Relationnel Mapping -ORM) avec la norme JPA. La deuxième partie présentera les différents modèles noSQL (clé-valeur, document, graphe), les notions de disponibilité et de partitionnement à la cohérence (propriétés BASE, théorème CAP), les différents systèmes NoSQL (MongoDB, Cassandra, CouchBase, ...), les avantages et les inconvénients du NoSQL. La troisième partie sera consacrée aux bases NewSQL : leur définition et leurs caractéristiques, les nouvelles architectures et la notion de DBaaS (Database as a service), leurs avantages et leurs inconvénients. Les notions apprises seront mises en pratique dans un projet où les étudiants devront manipuler différents types de bases de données afin de les comparer.

Ethique et science des données

ECTS : 2

Enseignant responsable : OLIVIA **TAMBOU** (https://dauphine.psl.eu/recherche/cvtheque/tambou-olivia)

Langue du cours : Français

Volume horaire : 12

Description du contenu de l'enseignement :

The course will be the occasion, for future data scientists, and for students in general, to question the benefits and risks of science. The course will permit them to approach from a pragmatic viewpoint questions they may have to face some day, and issues such as the various facets of privacy, the fairness of automatic decisions, the transparency of algorithmic processes, their explainability.

Compétences à acquérir :

Les étudiants devront être capable de comprendre les principaux enjeux éthiques en matière de données et d'intelligence artificielle tout en ayant un aperçu des différentes obligations juridiques existantes et en cours de création à l'échelle de l'Union européenne

---

Fondamentaux de l'apprentissage automatique

ECTS : 3

Enseignant responsable : YANN **CHEVALEYRE** (https://dauphine.psl.eu/recherche/cvtheque/chevaleyre-yann)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

The aim of this course is to provide the students with the fundamental concepts and tools for developing and analyzing machine learning algorithms. The course will introduce the theoretical foundations of machine learning, review the most successful algorithms with their theoretical guarantees, and discuss their application in real world problems. The covered topics are: - Introduction to the different paradigms of ML and applications - Computational learning theory - PAC model - VC-dimension - Rademacher complexity,… - Supervised learning - Logistic regression and beyond - Perceptron - SVM - Kernel methods - Decision trees and Random Forests - Ensemble methods: bagging and boosting - Unsupervised learning - Dimensionality reduction: PCA, ICA, Random Projections, Kernel PCA, ISOMAP, LLE - Density estimation - EM - Spectral clustering - Online learning - Multiclass and ranking algorithms - Practical sessions

Bibliographie, lectures recommandées :

References: - Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). Foundations of machine learning. MIT press. - Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press. - Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media. - Bishop Ch. (2006). Pattern recognition and machine learning. Springer - Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York, NY, USA:: Springer series in statistics. - James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical

IA générative

ECTS : 3

Langue du cours : Français

Volume horaire : 21

Optimisation pour l'apprentissage automatique

ECTS : 3

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

L'optimisation numérique est un composant fondamental de la science des données, qui se retrouve à la fois en apprentissage statistique (régression linéaire, SVM), mais aussi dans les progrès récents en apprentissage profond en tant que technique d'entraînement de réseaux de neurones. Dans ce cours, on étudiera les fondements mathématiques et algorithmiques de l'optimisation en se concentrant sur les applications en science des données.

Ce cours consiste en plusieurs séances accompagnées d'illustrations en Python, notamment via la bibliothèque PyTorch. On commencera par poser les fondements mathématiques de l'optimisation (convexité, conditions d'optimalité), puis on s'intéressera aux algorithmes essentiels en apprentissage tels que le gradient stochastique et ses variantes. On abordera ensuite des problématiques plus avancées à la fois sur le plan théorique (régularisation, problèmes matriciels) et sur le plan algorithmique (différentiation automatique, optimisation distribuée).

Compétences à acquérir :

- Identifier les caractéristiques d'un problème d'optimisation à partir de sa formulation.

- Connaître les propriétés théoriques et pratiques des algorithmes d'optimisation les plus courant en apprentissage.

- Déterminer et tester l'algorithme le plus approprié pour un problème donné.

Pré-requis recommandés

Bases d'algèbre linéaire. Expérience de programmation en Python.

Mode de contrôle des connaissances :

Examen écrit + Projet.

Bibliographie, lectures recommandées :

J. Wright and Y. Ma, High-Dimensional Data Analysis with Low-Dimensional Models, 2022.
S. J. Wright and B. Recht. Optimization for Data Analysis, 2022.

Systèmes, paradigmes et langages pour les big data

ECTS : 3

Enseignant responsable : DARIO **COLAZZO** (https://dauphine.psl.eu/recherche/cvtheque/dario-colazzo)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

The main aim of this course is to give students a deep and solid understanding of the state of the art of Big Data systems and programming paradigms, and to enable them to devise and implement efficient algorithms for analysing massive data sets.

The focus will be on paradigms based on distribution and shared-nothing parallelism, which are crucial to enable the implementation of algorithms that can be run on clusters of computers, scale as the size of input data increases, and can be safely executed even in the presence of system failures.

Lectures will give articular emphasis to the MapReduce paradigm and the internal aspects of its related runtime support Hadoop, as well as to MapReduce-based systems, including Spark and Hive, that provide users with powerful programming tools and efficient execution support for performing operations related to complex data flows. The attention will be then given to mechanisms and algorithms for both iterative and interactive data processing in Spark. A particular attention will be given to SQL-like data querying, graph analysis, and the development of machine learning algorithms.

A large part of the course consists of lab-sessions where students develop parallel algorithms for data querying and analysis, including algorithms for relational database operators, matrix operations, graph analysis, and clustering. Lab-sessions rely on the use of both desktop computers and Hadoop clusters on the Amazon WS cloud.

Program:

1. Introduction to massive data management and processing.

2. A data operating system for distributed data management, Hadoop.

3. MapReduce paradigm, algorithm design, implementation and optimisation.

4. iterative and interactive massive data processing, algorithm design, implementation and optimisation in Spark

5. large scale data-warehouse in Hive

References:

Mining of Massive Datasets. Jure Leskovec, Anand Rajaraman, Jeff Ullman http://www.mmds.org/#top

Data-Intensive Text Processing with MapReduce. Jimmy Lin and Chris Dyer. Mogan & Claypool Publishers

Hadoop: The Definitive Guide - Tom White. O'Reilly.

Apache Hadoop Yarn - Arun C.Murty, Vinod Kumar Vavilapalli, et al. Addison Wesley

Programming Hive. Edward Capriolo, Dean Wampler, Jason Rutherglen. O'Reilly.

Big Data Analytics with Spark.

Traitement automatique des langues - NLP

ECTS : 3

Langue du cours : Français

Volume horaire : 24

SEMESTRE 4 - 24 ECTS

UE fondamentales

Agentic AI

ECTS : 3

Enseignant responsable : LOUIS **FONTAINE**

Langue du cours : Anglais

Volume horaire : 24

Apprentissage profond pour l’analyse d’images

ECTS : 3

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

Deep learning has achieved formidable results in the image analysis field in recent years, in many cases exceeding human performance. This success opens paths for new applications, entrepreneurship and research, while making the field very competitive.

This course aims at providing the students with the theoretical and practical basis for understanding and using deep learning for image analysis applications.

The course will be composed of lectures and practical sessions. Moreover, experts from industry will present practical applications of deep learning. Lectures will include: • Introduction to machine learning • Artificial neural networks, back-propagation algorithm • Convolutional neural network • Design and optimization of a neural architecture • Successful architectures (AlexNet, VGG, GoogLeNet, ResNet) • Analysis of neural network function • Image classification and segmentation • Auto-encoders and generative networks • Current research trends and perspectives

During the practical sessions, the students will code in Python, using Keras and Tensorflow. They will be confronted with the practical problems linked to deep learning: architecture design; optimization schemes and hyper-parameter selection; analysis of results.

Prerequisites: Linear algebra, basic probability and statistics

---

Flux de données

ECTS : 3

Enseignant responsable : KODJO **KLOUVI** (https://dauphine.psl.eu/recherche/cvtheque/klouvi-kodjo)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

Ce cours a pour objectif de décrire les principes des systèmes capables de traiter les grandes masses de données en temps réel ou en temps quasi-réel et d'expliquer les apports des architectures microservices dans ce contexte. Ce cours est découpé en trois parties : -Streaming des données : Présentation des différentes architectures et frameworks permettant de capturer, traiter, analyser et visualiser les données massives en temps réel -Architectures microservices : Principes de découpage des systèmes en services simples, facilement couplés assurant l'agilité du système global ainsi que les technologies et les pratiques de développement associés seront traités dans cette partie du cours. -Projet : Mise en pratique avec Java d'une application mettant en œuvre Spark Streaming et les microservices en REST.

Machine Learning sur big data

ECTS : 3

Enseignant responsable : DARIO **COLAZZO** (https://dauphine.psl.eu/recherche/cvtheque/dario-colazzo)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

This course focuses on the typical, fundamental aspects that need to be dealt with in the design of machine learning algorithms that can be executed in a distributed fashion, on Hadoop clusters, in order to deal with big data sets, by taking into account scalability and robustness.

Machine learning algorithms are more and more used today, and there is an ever increasing demand of machine learning algorithms that scales over massives data sets. This course focuses on the typical, fundamental aspects that need to be dealt with in the design of machine learning algorithms that can be executed in a distributed fashion, on Hadoop clusters, in order to deal with big data sets, by taking into account scalability and robustness. So the course will focus on a bunch of main-stream, sequential machine learning algorithms, by taking into account the following crucial and complex aspects. The first one is the re-design of algorithms by relying on programming paradigms for distribution and parallelism based on

map-reduce, to this end Spark will be used. The second aspect is experimental analysis of the Spark implementation of designed algorithms in order to test their scalability and precision. The third aspect concerns the study and application of optimisation techniques in order to overcome lack of scalability and to improve execution time of designed algorithm.

The attention will be on machine learning technique for dimension reduction, clustering and classification, whose underlying implementation techniques are transversal and find application in a wide range of machine learning algorithms. For some of the studied algorithms, the course will present techniques for a from-scratch implementation in Spark core, while for other algorithms Spark ML will be used and end-to-end pipelines will be designed. In both cases algorithms will be analysed and optimised on real life data sets, by relaying on a local Hadoop cluster, as well as on a cluster on the Amazon WS cloud.

Bibliographie, lectures recommandées :

References: - Mining of Massive Datasets http://www.mmds.org - High Performance Spark - Best Practices for Scaling and Optimizing Apache Spark Holden Karau, Rachel Warren O'Reilly

---

Modélisation de problèmes

ECTS : 3

Enseignants : TRISTAN **CAZENAVE**, ARNAUD **LALLOUET**
https://dauphine.psl.eu/recherche/cvtheque/tristan-cazenave

Langue du cours : Français

Volume horaire : 24

Projet Sciences des Données

ECTS : 3

Enseignant responsable : Eric **BEN HAMOU**

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

The goal of this module is to provide students with a hands-on experience on a novel data-science/AI challenge using state-of-the-art tools and techniques discussed during other classes of this master.

Students enrolled in this class will form groups and choose one topic among a list of proposed topics in the core areas of the master such as supervised or unsupervised learning, recommendation, game AI, distributed or parallel data-science, etc. The topics will generally consist in applying a well-established technique on a novel data-science challenge or in applying recent research results on a classical data-science challenge. Either way, each topic will come with its own novel scientific challenge to address. At the end of the module, the students will give an oral presentation to demonstrate their methodology and their findings. Strong scientific rigor as well as very good engineering and communication skills will be necessary to complete this module successfully.

Recherche Monte-Carlo et jeux

ECTS : 3

Enseignant responsable : TRISTAN **CAZENAVE** (https://dauphine.psl.eu/recherche/cvtheque/tristan-cazenave)

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

Ce cours est une introduction aux méthodes dites de Monte-Carlo. Ces méthodes sont utilisées pour calculer des espérances, et par extension, des intégrales par simulation. L'objectif de ce cours est non seulement de fournir les bases théoriques des méthodes de Monte-Carlo, mais aussi de fournir les outils permettant leur utilisation pratique à travers des TP.

Le cours couvre les sujets suivants : -introduction de la méthode de Monte-Carlo -techniques de réduction de variance - introduction aux suites à discrétion faible

Visualisation de données

ECTS : 3

Langue du cours : Français

Volume horaire : 24

Description du contenu de l'enseignement :

Ce cours a pour objectif de décrire les démarches, méthodes et outils utilisés pour représenter les données complexes et multiples issues des grandes masses de données en visuels simples à comprendre et à interpréter, notamment pour les utilisateurs métier et pour les décideurs.

Le cas de l'apport de la visualisation des données sous différentes formes graphiques lors de la préparation des données en amont de la modélisation et de l'utilisation des modèles et algorithmes de Machine Learning sera développé. Le cours s'appuie sur de nombreux exemples puisés dans les domaines de la finance, de la santé, du marketing et des travaux pratiques sur des cas concrets sont également prévus.

Bloc mémoire - 7 ECTS

Memoire

ECTS : 7

Langue du cours : Français