Année universitaire 2024/2025

Mathématiques, Apprentissage, Sciences et Humanités - 2e année de Master

Crédits ECTS : 60

LES OBJECTIFS DE LA FORMATION

Une formation complète de "Data Scientist". L'objectif de ce Master est d'offrir à des étudiants et étudiantes issu(e)s d'un cursus mathématique une formation solide en apprentissage statistique dont les applications sont centrées sur l'économie numérique, les sciences et les humanités au sens large. Porté par la croissance exponentielle du flot de données générées par des applications aussi variées que la biologie, la médecine, le commerce en ligne, l'imagerie, la vidéo ou le traitement du langage.

Les objectifs de la formation :

- Maîtriser des fondations théoriques de l'apprentissage : méthodes de noyaux, apprentissage supervisé et non supervisé, optimisation, modèles graphiques, etc.
- Maîtriser des méthodes statistiques fondamentales : simulation, estimation, détection, etc.
- Ouvrir aux applications de l'apprentissage en marketing, santé, journalisme, politiques publiques, etc.
- Acquérir des compétences opérationnelles dans un certain nombre de langages informatiques clés : Python (notamment le package scikit-learn), HADOOP, R, MATLAB, Julia, etc.
- Développer un savoir-faire pratique dans la manipulation des jeux de données issus d'applications et de projets.

A compter de janvier 2025, les cours seront dispensés au 16 bis rue de l'Estrapade, 75005 Paris.

PRÉ-REQUIS OBLIGATOIRES

- Titulaires d'un diplôme BAC+4 (240 crédits ECTS) ou équivalent à Dauphine, d'une université, d'une école d'ingénieur ou d'un autre établissement de l'enseignement supérieur dans le domaine des mathématiques appliquées

POURSUITE D'ÉTUDES

La majorité des étudiants s'orientent vers une carrière professionnelle (informatique, téléphonie, nouvelles technologies, médias, marketing, aéronautique). Les étudiants du parcours MASH peuvent également s'orienter vers la recherche publique ou privée (financement universitaire ou industriel).

Débouchés : Data scientist, Ingénieux Recherche et Développement, Quantitative analyst, Associate, R&D - Data Scientist

PROGRAMME DE LA FORMATION

- Semestre 3
  - Cours introductifs
    - Introduction to R
    - Introduction to Bayesian Statistics
    - A review of probability theory foundations
    - Introduction to Python

- Cours fondamentaux
  - Optimization for Machine Learning
  - High-dimensional statistics
  - Foundations of machine learning
  - Graphical models
- Cours optionnels - 5 cours à choisir parmi :
  - Optimal transport
  - Computational methods and MCMC
  - Applied Bayesian statistics
  - Bayesian non parametric and Bayesian Machine Learning
  - Mixing times of Markov chains
  - Large Language Models
  - Reinforcement Learning
  - Kernel methods
  - Non-convex inverse problems
  - Mathematics of deep learning
  - Topological Data Analysis
  - Deep learning for image analysis
  - Dimension reduction and manifold learning
  - Bayesian asymptotics

- Semestre 4
  - Bloc mémoire
    - Mémoire de recherche

DESCRIPTION DE CHAQUE ENSEIGNEMENT

A review of probability theory foundations

ECTS : 0

Volume horaire : 15

Description du contenu de l'enseignement :

Outline :

1. Basics of measure theory and integration

2. Probability : random variables, independence

3. Convergence of random variables

4. Law of Large Numbers and Central Limit Theorem

5. Conditional expectations

6. Martingales in discrete time

7. Gaussian vectors

8. Brownian motion : definition, existence, first properties

Compétence à acquérir :

The aim of this class is to provide a quick review of the probability theory that is required to follow the 1st semester classes in MATH, MASEF and MASH.

Most of the content should already be familiar to students with a M1 in Mathematics.

Applied Bayesian statistics

ECTS : 4

Volume horaire : 18

Description du contenu de l'enseignement :

We shall put in practice classical models for statistical inference in a Bayesian setting, and implement computational methods. Using real data, we shall study various models such as linear regression, capture-recapture, and a hierarchical model. We shall discuss issues of model building and validation, the impact of the choice of prior, and model choice via Bayes Factors. The implementation shall use several algorithms: Markov Chain Monte Carlo, importance sampling, Approximate Bayesian Computation. The course is based on the free software R.
Practical information: Large portions of the course are devoting to students coding. Students should bring their own laptop, which must have R installed before the first session; I strongly suggest installing RStudio (free) as well.

Compétence à acquérir :

Modelling and inference in a Bayesian setting

Bayesian asymptotics

ECTS : 4

Volume horaire : 18

Bayesian non parametric and Bayesian Machine Learning

ECTS : 4

Volume horaire : 18

Description du contenu de l'enseignement :

Bayesian nonparametrics:

- Basics: infinite mixture models and clustering
- Models beyond the Dirichlet process
- Posterior sampling
- Applications

Gaussian Processes

Bayesian Deep Learning

Compétence à acquérir :

Essential concepts of Bayesian nonparametrics

Essentials of Bayesian Deep Learning

Bibliographie, lectures recommandées :

- Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (Eds.). (2010). *Bayesian nonparametrics* (Vol. 28). Cambridge University Press.
- Orbanz, P., & Teh, Y. W. (2010). Bayesian nonparametric models. *Encyclopedia of machine learning*, *1*, 81-89.
- Müller, P., Quintana, F. A., Jara, A., & Hanson, T. (2015). *Bayesian nonparametric data analysis* (Vol. 1). New York: Springer.
- Ghosal, S., & van der Vaart, A. W. (2017). *Fundamentals of nonparametric Bayesian inference* (Vol. 44). Cambridge University Press.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT press.

Computational methods and MCMC

ECTS : 4

Volume horaire : 21

Description du contenu de l'enseignement :

Motivations
Monte-Carlo Methods
Markov Chain Reminders
The Metropolis-Hastings method
The Gibbs Sampler
Perfect sampling

Sequential Monte-Carlo methods

Compétence à acquérir :

This course aims at presenting the basics and recent developments of simulation methods used in statistics and especially in Bayesian statistics. Methods of computation, maximization and high-dimensional integration have indeed become necessary to deal with the complex models envisaged in the user disciplines of statistics, such as econometrics, finance, genetics, ecology or epidemiology (among others!). The main innovation of the last ten years is the introduction of Markovian techniques for the approximation of probability laws (and the corresponding integrals). It thus forms the central part of the course, but we will also deal with particle systems and stochastic optimization methods such as simulated annealing.

Deep learning for image analysis

ECTS : 4

Volume horaire : 24

Dimension reduction and manifold learning

ECTS : 4

Volume horaire : 24

Description du contenu de l'enseignement :

Modern machine learning typically deals with high-dimensional data. The fields concerned are very varied and include genomics, image, text, time series, or even socioeconomic data where more and more unstructured features are routinely collected. As a counterpart of this tendency towards exhaustiveness, understanding these data raises challenges in terms of computational resources and human understandability. Manifold Learning refers to a family of methods aiming at reducing the dimension of data while preserving certain of its geometric and structural characteristics. It is widely used in machine learning and experimental science to compress, visualize and interpret high-dimensional data. This course will provide a global overview of the methodology of the field, while focusing on the mathematical aspects underlying the techniques used in practice.

Compétence à acquérir :

- Curse of dimensionality, manifold hypothesis and intrinsic dimension(s)
- Multidimensional scaling
- Linear dimension reduction (random projections, principal component analysis)
- Non-linear spectral methods (kernel PCA, ISOMAP, MVU, Laplacian eigenmaps)
- Ad-hoc distance-preserving methods (diffusion maps, LLE)
- Probabilistic dimension reduction and clustering (SNE, UMAP)
- Neural network-based dimensionality reduction

Bibliographie, lectures recommandées :

- Ghojogh, B., M. Crowley, F. Karray, and A. Ghodsi (2023). Elements of dimensionality reduction and manifold learning
- Lee, J. A., M. Verleysen, et al. (2007). Nonlinear dimensionality reduction

Foundations of machine learning

ECTS : 5

Volume horaire : 24

Description du contenu de l'enseignement :

Typologie des problèmes d’apprentissage (supervisé vs. non-supervisé).
Modèle statistique pour la classification binaire : Approches génératives vs. discriminantes.
Algorithmes classiques : méthodes paramétriques, perceptron, méthodes de partitionnement.
Critères de performances : erreur de classification, courbe ROC, AUC.
Convexification du risque : Algorithmes de type boosting et SVM. Mesures de complexité combinatoires, métriques géométriques.
Sélection de modèle et régularisation.
Théorèmes de consistance et vitesses de convergence.

Compétence à acquérir :

Bases mathématiques pour la modélisation des problèmes d’apprentissage supervisé et l’analyse des algorithmes de classification en grande dimension. Il s’agit de présenter les bases mathématiques pour la modélisation des problèmes d’apprentissage supervisé et l’analyse des algorithmes de classification en grande dimension.

Graphical models

ECTS : 4

Volume horaire : 18

Compétence à acquérir :

Modélisation probabiliste, apprentissage et inférence sur les modèles graphiques. Les principaux thèmes abordés sont :
Maximum de vraisemblance.
Régression linéaire.
Régression logistique.
Modèle de mélange, partitionnement.
Modèles graphiques.
Familles exponentielles.
Algorithme produit-somme.
Hidden Markov models.
Inférence approximée
Méthodes bayésiennes.

High-dimensional statistics

ECTS : 5

Volume horaire : 24

Description du contenu de l'enseignement :

Fléau de la dimension et hypothèse de parcimonie pour la régression gaussienne, les modèles linéaires généralisés et les données de comptage.

Ondelettes et estimation par seuillage.

Choix de modèles et sélection de variables.

Estimation par pénalisation convexe : procédure Ridge, lasso, group-lasso… Liens avec l’approche bayésienne.

Tests multiples : procédures FDR, FWER.

Données fonctionnelles

Compétence à acquérir :

L’objectif de ce cours de statistique est de présenter les outils mathématiques et les méthodologies dans la situation où le nombre de paramètres à inférer est très élevé, typiquement beaucoup plus important que le nombre d’observations.

Mode de contrôle des connaissances :

Examen sur table

Bibliographie, lectures recommandées :

Wasserman, L. (2005) All of statistics. A concise course in statistical inference. Springer

---

Introduction to Bayesian Statistics

ECTS : 0

Volume horaire : 3

Introduction to Python

ECTS : 0

Volume horaire : 3

Description du contenu de l'enseignement :

Dans ce cours de 3h, nous voyons (ou re-voyons) la base de Python, et l'utilisation des notebooks. Il est illustré par 3 notebooks. Le premier rappelle les bases générales de Python. Le second porte sur l'utilisation du module *pandas*, et le dernier sur un problème simple d'optimisation de portfolio.

Compétence à acquérir :

- Installer Python sur sa machine
- Utiliser un notebook
- Savoir lire la documentation de Python, et écrire des codes simples

---

Introduction to R

ECTS : 0

Volume horaire : 3

Description du contenu de l'enseignement :

Introduction to the R programming language: loading data, writing simple functions, producing standard plots.

Compétence à acquérir :

Programming in R

Mode de contrôle des connaissances :

No evaluation

Kernel methods

ECTS : 4

Volume horaire : 18

Description du contenu de l'enseignement :

Reproducing kernel Hilbert spaces et le “kernel trick”
Théorème de représentation
Kernel PCA
Kernel ridge regression
Support vector machines
Noyaux sur les semigroupes
Noyaux pour le texte, les graphes, etc.

Compétence à acquérir :

Présenter les bases théoriques et des applications des méthodes à noyaux en apprentissage.

Large Language Models

ECTS : 4

Volume horaire : 24

Description du contenu de l'enseignement :

The course focuses on modern and statistical approaches to NLP.

Natural language processing (NLP) is today present in some many applications because people communicate most everything in language : post on social media, web search, advertisement, emails and SMS, customer service exchange, language translation, etc. While NLP heavily relies on machine learning approaches and the use of large corpora, the peculiarities and diversity of language data imply dedicated models to efficiently process linguistic information and the underlying computational properties of natural languages.

Moreover, NLP is a fast evolving domain, in which cutting-edge research can nowadays be introduced in large scale applications in a couple of years.

The course focuses on modern and statistical approaches to NLP: using large corpora, statistical models for acquisition, disambiguation, parsing, understanding and translation. An important part will be dedicated to deep-learning models for NLP.

- Introduction to NLP, the main tasks, issues and peculiarities
- Sequence tagging: models and applications
- Computational Semantics
- Syntax and Parsing
- Deep Learning for NLP: introduction and basics

- Deep Learning for NLP: advanced architectures
- Deep Learning for NLP: Machine translation, a case study

Compétence à acquérir :

- Skills in Natural Language Processing using deep-learning
- Understand new architectures

Mathematics of deep learning

ECTS : 4

Volume horaire : 24

Mixing times of Markov chains

ECTS : 4

Volume horaire : 24

Description du contenu de l'enseignement :

How many times must one shuffle a deck of 52 cards? This course is a self-contained introduction to the modern theory of mixing times of Markov chains. It consists of a guided tour through the various methods for estimating mixing times, including couplings, spectral analysis, discrete geometry, and functional inequalities. Each of those tools is illustrated on a variety of examples from different contexts: interacting particle systems, card shuffling, random walks on groups, graphs and networks, etc. Finally, a particular attention is devoted to the celebrated cutoff phenomenon, a remarkable but still mysterious phase transition in the convergence to equilibrium of certain Markov chains.

Compétence à acquérir :

See the webpage of the course.

Mode de contrôle des connaissances :

Final written exam, in class.

Bibliographie, lectures recommandées :

See the webpage of the course.

Mémoire de recherche

ECTS : 20

Non-convex inverse problems

ECTS : 4

Volume horaire : 18

Description du contenu de l'enseignement :

An inverse problem is a problem where the goal is to recover an unknown object (typically a vector with real coordinates, or a matrix), given a few ``measurements'' of this object, and possibly some information on its structure. In this course, we will discuss examples of such problems, motivated by applications as diverse as medical imaging, optics and machine learning. We will especially focus on the questions: which algorithms can we use to numerically solve these problems? When and how can we prove that the solutions returned by the algorithms are correct? These questions are relatively well understood for convex inverse problems, but the course will be on non-convex inverse problems, whose study is much more recent, and a very active research topic.

The course will be at the interface between real analysis, statistics and optimization. It will include theoretical and programming exercises.

Compétence à acquérir :

Understand what is a non-convex inverse problems; get some familiarity with the most classical algorithms to solve them.

Optimal transport

ECTS : 4

Volume horaire : 18

Description du contenu de l'enseignement :

Optimal transport (OT) is a fundamental mathematical theory at the interface between optimization, partial differential equations and probability. It has recently emerged as an important tool to tackle a surprisingly large range of problems in data sciences, such as shape registration in medical imaging, structured prediction problems in supervised learning and training deep generative networks.
This course will interleave the description of the mathematical theory with the recent developments of scalable numerical solvers. This will highlight the importance of recent advances in regularized approaches for OT which allow one to tackle high dimensional learning problems.

The course will feature numerical sessions using Python.

- Motivations, basics of probabilistic modeling and matching problems.
- Monge problem, 1D case, Gaussian distributions.
- Kantorovitch formulation, linear programming, metric properties.
- Shrödinger problem, Sinkhorn algorithm.
- Duality and c-transforms, Brenier’s theory, W1, generative modeling.
- Semi-discrete OT, quantization, Sinkhorn dual and divergences

---

Optimization for Machine Learning

ECTS : 6

Volume horaire : 48

Description du contenu de l'enseignement :

This course delves into the mathematical underpinnings and algorithmic strategies essential for understanding and applying Machine Learning techniques. Central to the course is the exploration of optimization, a pivotal element in contemporary advancements in machine learning. This exploration encompasses fundamental approaches such as linear regression, SVMs, and kernel methods, and extends to the dynamic realm of deep learning. Deep learning has become a leading methodology for addressing a variety of challenges in areas like imaging, vision, and natural language processing. The course content is structured to provide a comprehensive overview of the mathematical foundations, algorithmic methods, and a variety of modern applications utilizing diverse optimization techniques. Participants will engage in both traditional lectures and practical numerical sessions using Python. The curriculum is divided into three parts: The first focuses on smooth and convex optimization techniques, including gradient descent and duality. The second part delves into advanced methods like non-smooth optimization and proximal methods. Lastly, the third part addresses large-scale methods such as stochastic gradient descent and automatic differentiation, with a special focus on their applications in neural networks, including both shallow and deep architectures.

Detailed Syllabus:
1. Foundational Concepts in Differential Calculus and Gradient Descent:
- Introduction to differential calculus
- Principles of gradient descent
- Application of gradient descent in optimization

2. Automatic Differentiation and Its Applications:
- Understanding the mechanics of automatic differentiation
- Implementing automatic differentiation using modern Python frameworks

3. Advanced Gradient Descent Techniques:
- In-depth study of gradient descent theory
- Accelerated gradient methods
- Stochastic gradient algorithms and their applications

4. Exploring Convex and Non-Convex Optimization:
- Fundamentals of convex analysis
- Strategies and challenges in non-convex optimization

5. Special Topics in Optimization:
- Introduction to non-smooth optimization methods
- Study of semidefinite programming (SDP)
- Exploring interior points and proximal methods

6. Large-Scale Optimization Methods and Neural Networks:
- Techniques in large-scale methods, focusing on stochastic gradient descent
- Applications of automatic differentiation in neural networks
- Overview of neural network architectures: shallow and deep networks

Bibliographie, lectures recommandées :

- Theory and algorithms: Convex Optimization, Boyd and Vandenberghe
- Introduction to matrix numerical analysis and optimization, Philippe Ciarlet
- Proximal algorithms, N. Parikh and S. Boyd
- Introduction to Nonlinear Optimization - Theory, Algorithms and Applications, Amir Beck
- Numerics: Python and Jupyter installation: use only Python 3 with Anaconda distribution.
- The Numerical Tours of Signal Processing, Gabriel Peyré
- Scikitlearn tutorial, Fabian Pedregosa, Jake VanderPlas

Reinforcement Learning

ECTS : 4

Volume horaire : 24

Topological Data Analysis

ECTS : 4

Volume horaire : 18