

## M1 de Mathématiques et Intelligence Artificielle

### M1IA-01 - Probabilités – S1. Volume horaire : 24h CM + 24h TD

Contenu : L'UE Probabilités vise à donner aux étudiants des bases théoriques d'éléments de probabilités nécessaires à la suite du parcours du Master. On rappellera brièvement quelques éléments de théorie de la mesure. On s'intéressera ensuite aux différents aspects du conditionnement : espérance conditionnelle, loi conditionnelle et densité conditionnelle. Enfin, on étudiera la théorie des chaînes de Markov, en particulier les aspects de convergence et leurs applications.

### M1IA-02- Optimisation numérique – S1. Volume horaire : 18h CM + 18h TD

Contenu : Le cours a comme objectif l'étude de l'optimisation convexe (sans et sous contraintes) en dimension finie en se focalisant ensuite sur les propriétés de convergence des méthodes numériques d'ordre 1 (simple à programmer) et les applications à des problèmes concrets en lien avec le machine learning. Les principaux thèmes qui seront abordés sont :

- L'optimisation convexe lisse sans contraintes,
- L'optimisation convexe non lisse par régularisation,
- L'optimisation avec contrainte,
- La dualité convexe,
- Les algorithmes de descente et algorithmes primaux-duaux.

Programmation en python.

### M1IA-3 - Modélisation statistique – S1. Volume horaire : 51h CI

Prérequis : Bases théoriques de l'inférence statistique (niveau L3)

Objectif :

- Définir une modélisation adaptée à un jeu de données,
- Construire un estimateur et étudier ses propriétés,
- Évaluer le risque de l'estimation statistique, estimer une régression linéaire ou logistique avec un logiciel (R).

Contenu : Concepts fondamentaux en statistique :

- Méthode des moindres carrés, du maximum de vraisemblance, propriétés des estimateurs, intervalles de confiance,
- Tests de Wald et de rapport de vraisemblance,
- Modèle linéaire gaussien, tests de Fisher, sélection de variables,
- Classification paramétrique : règle de Bayes, étude du risque,
- Modèle de régression logistique, tests de Wald et de déviance.

### M1IA-10 - Mathématiques pour l'Intelligence Artificielle 1 – S1. Volume horaire : 48h CM

Cette UE est mutualisée avec le M1 MFA

Contenu : Ce cours présente le formalisme et quelques outils mathématiques fondamentaux utilisés en apprentissage statistique, sur lequel se fondent les méthodes d'intelligence artificielle modernes qui sont toutes « entraînées sur des données ». Les thématiques abordées seront notamment :

- Théorie de la décision : fonction de prédiction, modélisation probabiliste, perte et risque, estimateurs, consistance d'un algorithme d'apprentissage,
- Panorama de méthodes de classification linéaire classiques,
- Étude de la méthode des plus proches voisins,

- Rudiments de la théorie du contrôle de l'erreur de généralisation,
- Méthodes à noyaux de Hilbert auto-reproduisants,
- Quelques éléments d'analyse des méthodes de réseaux de neurones.

**M1IA-11 - Approfondissement statistique et apprentissage non supervisé – S1.** Volume horaire : 24h CM + 12h TD

Contenu :

- 1<sup>ère</sup> partie : approfondissement statistique. Optimalité des tests de rapport de vraisemblance, étude de tests non-paramétriques. Estimation non-paramétrique des fonctions de distribution et de densité. Études de simulations et analyses de données avec R ;
- 2<sup>ème</sup> partie : introduction à l'apprentissage non-supervisé. Analyse exploratoire multidimensionnelle, méthodes factorielles (ACP, ...), Clustering. Analyses de données avec R.

**M1IA-04 - Apprentissage supervisé – S2.** Volume horaire : 24h CI

Contenu : Au 1<sup>er</sup> semestre l'accent a été mis sur la construction de modèles prédictifs linéaires (régression ou classification binaire). Dans cet enseignement, nous poursuivons avec des méthodes non-linéaires d'apprentissage supervisé.

- Méthodes de ré-échantillonnage bootstrap, validation croisée et évaluation du risque de généralisation,
- Méthodes de régularisation RIDGE et LASSO, introduction à des problèmes de grande dimension,
- Arbres de décision et forêts aléatoires,
- Analyses de données avec R.

**M1IA-05 - Introduction au Deep Learning – S2.** Volume horaire : 16h CM

Objectif : L'objectif est d'exposer les concepts fondamentaux de cette discipline, depuis ses origines jusqu'à ses avancées les plus récentes. L'approche adoptée se situe à l'interface entre les domaines mathématiques et informatiques, mettant l'accent sur la compréhension approfondie de la formulation théorique pour son application éclairée sur des problèmes concrets. Les étudiants seront immergés dans l'univers complexe du deep learning, acquérant ainsi une expertise tant théorique que pratique.

Contenu :

- Généralités sur le Deep Learning,
- Estimation d'un réseau de neurones,
- Sélection d'hyperparamètres,
- Réseaux convolutionnels,
- Réseaux génératifs,
- Réseaux récurrents,
- Perspectives.

**M1IA-06 - Data Challenge de modélisation prédictive.** Volume horaire : 24h CI

Contenu : Ce cours a pour but de donner les bases de la modélisation prédictive de données observées au cours du temps (séries temporelles multivariées). Nous aborderons les méthodes de régression linéaire, régression pénalisées, GAM et les méthodes d'ensemble (forêts aléatoires, boosting) et les mettrons en œuvre sur des données réelles dans le cadre d'un data challenge. Pour le travail de projet, les étudiants doivent

- Effectuer une analyse descriptive des données,
- Enrichir les données proposées par de l'open data : identifier les données pertinentes, nettoyer et fusionner les données,
- Développer un modèle de prévision,
- Choisir un modèle,
- Sélectionner/construire les variables pertinentes,
- Valider le modèle : simulation, bootstrap, public score,

- Obtenir le meilleur score (plus faible erreur de prévision) sur le private set, en respectant les contraintes imposées,
- Présenter ses résultats de manière intelligible et reproductible lors d'une soutenance.

## **M1IA-12 - Mathématiques pour l'Intelligence Artificielle 2.** Volume horaire : 48h CM

Cette UE est mutualisée avec le M1 MFA

### Contenu :

- Part 1 : Sequential learning : Complement on probability : Sub-Gaussian random vector, sub-exponential random variables, concentration inequalities. Sequential learning problems, Stochastic Gradient Optimisation, connection with classical optimisation problems. Learning with expert advices, adversarial environment, mirror descent. Learning in an unknown environment. Exploration/exploitation trade-off, Information-Theoretic lower bounds.
- Part 2 : Matricial problems in ML : dimension reduction, clustering, community detection. Singular value decomposition, matrix norms, perturbation bounds. Concentration of the operator norm of random matrices, application in ML.

## **M1IA-13 - Projet de Deep Learning – S2.** Volume horaire : 8h Projet + 16h CI

### Contenu :

- Analyser et présenter à l'oral un article récent de Deep Learning,
- Implémenter les algorithmes étudiés sous forme d'application Streamlit,
- Tester les performances sur données réelles ou simulées.

## **M1IA - 18 Stage – S2**

Contenu : Le stage de M1 permet de prendre un premier contact avec le monde professionnel, dans une entreprise ou un laboratoire de recherche, en France ou à l'étranger. Il doit durer au moins quatre mois, et peut s'étendre de la fin des cours (fin mars) à fin août.