

# Contenu des enseignements

Les intitulés, le contenu et le nombre d'heures sont indicatifs et susceptibles de varier d'une année à l'autre.

## Module 1 - Analyse Appliquée, gestion UJM

*Ce module revient sur les bases et les outils d'analyse fonctionnelle, nécessaires à la compréhension des concepts utilisés dans ce Master. Les différents types d'équations aux dérivées partielles sont ensuite présentés et illustrés par des exemples issus de la mécanique des fluides des solides... Enfin, une présentation et manipulation des différentes méthodes de discrétisation suivant le type d'équations est faite.*

### 12h. Equations de la physique

*EDP de la Mécanique des fluides et solides*

### 12h. Analyse fonctionnelle

*Grands théorèmes pour l'existence et l'unicité des solutions ; Problèmes elliptiques, théorie variationnelle, théorie spectrale des problèmes aux limites ; Problèmes paraboliques, estimation d'erreur ; Problèmes hyperboliques, solutions faibles, non unicité, solutions physiques*

### 12h. Méthodes numériques

*Eléments finis, différences finies, volumes finis ; Spécificité des différentes méthodes*

## Module 2 - Modélisation Stochastique et Apprentissage Statistique = UP2 MSD

*Ce module propose les méthodes et outils mathématiques permettant d'appréhender la notion d'apprentissage avec modèle statistique pour la réduction (analyse factorielle), la régression (régression généralisée) et la classification (analyse discriminante, arbres de décision, boosting) dite supervisée. Ces notions seront abordées en théorie et en pratique sur des données réelles.*

### 15h. Analyse des données

*Analyse de données et réduction de dimension par analyses factorielle - analyse en composantes principales, analyse des correspondances, analyse factorielle discriminante*

### 12h. Arbres de décision et forêt aléatoires

*CART, Forêts aléatoires, Isolation forest*

### 9h. Régression avancée

*Régressions linéaires et non linéaires. régression logistique. Lasso, Ridge, parcimonie.*

### **Module 3 - Optimisation et Machine Learning = UP3 MSD**

*Ce module donne un aperçu des méthodes d'optimisation classiques et modernes, pour des applications en apprentissage automatique et en science des données dans un contexte supervisé et non supervisé. En particulier, l'évolutivité des algorithmes les plus populaires de Machine Learning, leur application à différents ensembles de données, sera discutée en théorie et en pratique.*

#### **9h. Optimisation classique**

*Optimisation numérique locale : gradient(s), Newton, méthodes duales (Uzawa). Application aux SVM et SVR.*

#### **9h. Réseaux de Neurones**

*Réseaux feedforward, Réseaux de neurones récurrents, Réseaux convolutifs, Rétropropagation de gradient*

#### **9h. Optimisation pour Machine Learning**

*Back-propagation. Méthodes de gradients déterministes et stochastiques, Robbins-Monro, Kiefer-Wolfowitz, applications aux réseaux de neurones*

#### **14,5h. Classification-clustering**

*Clustering (k-means, k-medoids, hiérarchique, DBSCAN), classification supervisée (k-NN, variantes), règles d'association (Apriori, ECLAT)*

### **Module 4 - Calcul intensif et simulation numérique, gestion UJM**

*Ce module présente les différents outils nécessaires au calcul parallèle : architecture hardware, cluster de calcul, bibliothèques spécifiques, ainsi que les méthodes mathématiques nécessaires à une telle approche parallèle : stockage efficace des données, algorithmes de décomposition de domaine avec ou sans recouvrement, méthode de Monte-Carlo*

#### **9h. Introduction aux architectures de calcul parallèle**

*Architectures de parallélisation, calcul parallèle en mémoire distribuée avec la bibliothèque MPI, calculs déportés sur un cluster de calcul sous linux*

#### **9h. Parallélisation : algèbre linéaire et décomposition de domaines**

*Stockage matrice creuse versus méthodes de résolution ; Décomposition de domaine avec et sans recouvrement dans le cas d'une équation aux dérivées partielles de type elliptique.*

#### **9h. Algorithmes stochastiques et calcul parallèle**

*Méthodes de Monte-Carlo (échantillonnage aléatoire) pour la résolution des équations et calculer des intégrales sans solution explicite connue. Exemples pratiques (TP) pour résoudre des équations aux dérivées partielles et calculer des constantes universelles.*

#### **9h. Application à la simulation numérique – étude de cas**

## **Module 5 - Méta-modèles et optimisation globale = UP4 MSD**

*Ce module propose des outils mathématiques et numériques permettant d'apprendre des résultats d'expériences numériques ou réelles dans un modèle statistique, ou méta-modèle, en l'occurrence un processus gaussien. Celui-ci pourra ensuite être utilisé pour choisir les expériences les plus importantes ou optimiser des systèmes réalistes. Les techniques enseignées sont : les plans d'expérience, la modélisation par processus gaussiens, l'analyse de sensibilité globale, l'optimisation globale et en particulier stochastique.*

### **6h. Plans d'expériences**

*Plans associés à la régression linéaire, plans factoriels, optimalités, critères géométriques minimax, maximin, discrédance, suites halton, sobol...*

### **6h. Krigeage**

*BLUP, krigeages simple, ordinaire, universel, fonctions de covariance et estimation*

### **9h. Analyse de sensibilité globale**

*décomposition ANOVA de fonction, indices de Sobol, méthodes de screening*

### **9h. Optimisation globale**

*algorithmes restarted Gradient, recuit simulé, CMA-ES, optimisation Bayésienne*

### **3h. Examen transversal ou défi**

*examen transversal écrit (QCM) sur l'ensemble des matières du module*

## **Module 6 - Modélisation Statistique avancée**

*Ce module propose des modélisations avancées en statistiques. D'une part, des techniques modernes qui visent à gérer les gros volumes de données (grand nombre d'observation et grande dimension) : sont abordées ainsi des techniques utilisées en machine learning, agrégation d'experts et autoencodeurs. D'autre part, les liens entre modèles stochastiques et déterministes sont étudiés avec détail, à travers un cours liant processus stochastiques et équations aux dérivées partielles.*

*- Les deux premiers cours sont empruntés au défi Big Data du cycle ICM -*

### **6h. Méthodes d'agrégation**

*Techniques d'agrégations classiques (agrégation d'experts): produit d'experts (PoE, GPoE), Bayesian Committee Machine (BCM, rBCM), BLUP, applications aux cas spatial et Krigeage*

### **6h. Auto-encodeurs et réduction de dimension**

*Auto-encodeurs (AE), principe de la réduction de dimension, Apprentissage d'un AE par retro-propagation, lien avec ACP, généralisations.*

### **24h. Processus Stochastiques et EDP**

*Lien probabilités et EDP, équation de la chaleur et mouvement Brownien, calcul stochastique et EDS. Applications à l'apprentissage statistique à partir d'EDP.*

## **Cours (facultatifs pour non ICM): Fondements probabilistes = UP1 MSD**

*Ce module apporte les bases théoriques du calcul des probabilités et de la statistique. En particulier, est abordée l'étude des vecteurs gaussiens qui joue un rôle central en statistique. Le conditionnement en probabilités est présenté en liaison avec la statistique bayésienne, formant ainsi un socle pour l'apprentissage statistique et le machine learning. Comme applications, les techniques de régression et de séries chronologiques sont introduites.*

*- Ces cours facultatifs pour les non-ICM correspondent à l'UP1 de la Majeure Sciences des Données, il est possible de les suivre en auditeur libre, sous réserve de compatibilité des emplois du temps -*

**15h. Probabilités – optionnel pour les maea non icm**

*Statistique et probabilités, estimation. Statistique bayésienne, lois a priori et a posteriori, inférence*

**12h. Régression – optionnel pour les maea non icm**

*Statistique linéaire, vecteurs Gaussiens et conditionnement. Modèle linéaire de régression - Estimation et validation du modèle - Prédiction*

**12h. Séries chronologiques – optionnel pour les maea non icm**

*Analyse exploratoire, modèles ARMA et extensions, inférence*