



ÉCOLE NATIONALE
SUPÉRIEURE
D'INFORMATIQUE
POUR L'INDUSTRIE
ET L'ENTREPRISE

FICHE TECHNIQUE

MASTÈRE SPÉCIALISÉ CALCULS HAUTES PERFORMANCES / DONNÉES MASSIVES / INTELLIGENCE ARTIFICIELLE

OPTIMISATION EN GRANDES DIMENSIONS (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

Très souvent, les problèmes d'optimisation sont des problèmes d'optimisation discrète où les variables modélisent des entités indivisibles : un nombre de voitures ou d'avions à construire par exemple, ou bien des décisions à prendre, choix d'un coup ou d'une stratégie dans un jeu. Un problème d'optimisation discrète revient à trouver dans un ensemble discret (les solutions réalisables), la meilleure solution relativement à une fonction objective donnée.

Les problèmes d'optimisation discrète sont, la plupart du temps, difficiles et il est important de savoir les résoudre efficacement.

COMPÉTENCES ATTENDUES

Avoir des notions de complexité des algorithmes, connaître les outils efficaces pour résoudre les problèmes d'optimisation discrète de façon exacte ou approchée, savoir modéliser un problème et utiliser un solveur (commercial ou libre).

MACHINE LEARNING (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

Le machine Learning, apprentissage machine en français, regroupe un ensemble de méthodes mathématiques et algorithmiques permettant d'apprendre une fonction à partir d'exemples. Ces méthodes sont devenues aujourd'hui incontournables pour l'exploitation des bases de données numériques, de plus en plus nombreuses. La session proposera à la fois des cours théoriques sur les enjeux mathématiques de ces méthodes et des travaux pratiques qui permettront d'utiliser les méthodes présentées sur des problématiques de régression ou de classification en utilisant le logiciel R.

COMPÉTENCES ATTENDUES

1. Introduction au machine learning, à l'apprentissage supervisé et aux modèles prédictifs.
2. Modèles de régression pénalisée. Critères de performances. Sélection de modèles.
3. Modèles de classification. Courbes ROC.
4. Modèles d'ensemble. Arbres de classification et de régression. Bagging. Random Forest.
5. Agrégation de modèles. Boosting

PROGRAMMATION PARALLÈLE (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

Présentation des différentes briques fonctionnelles intervenant lors du développement d'un code de calcul scientifique, de son exploitation en phase de production, et enfin lors de la phase d'extraction et de présentation des résultats. Mise en pratique des bonnes pratiques de développement, de production et de post-processing, tout en découlant les contraintes spécifiques du calcul scientifique hautes performances. La maîtrise des outils utilisés sera également très profitable aux étudiants se destinant à une carrière incluant du développement non exclusivement scientifique.

CONNAISSANCES

Être capable de structurer un projet de code scientifique, de lancer ou de rejoindre un développement collaboratif, de déployer le code sur un super-calculateur et d'extraire et visualiser les données résultantes. Une partie de ce module sera consacrée à l'utilisation du langage Python comme « glue » entre différents langages de plus bas niveau, et les élèves seront amenés à écrire des interfaces et à échanger des données depuis et vers Python. Une introduction à la visualisation scientifique sera également proposée.

COMPÉTENCES ATTENDUES

Outils de développement, python comme glue, portage / déploiement sur super calculateur, debug / profiling, stockage de résultats, post-processing, extraction de données, visualisation de données.

MISE EN PRATIQUE SOUS PYTHON (CERTIFICATION)

DURÉE 20H (1 UE = 2 ECTS)



CONTACTS

PIERRE DOSSANTOS-UZARRALDE

pierre.dossantos-uzarralde@ensiie.fr

SEYDINA M. NDIAYE

seydina.ndiaye@uvs.edu.sn





ÉCOLE NATIONALE
SUPÉRIEURE
D'INFORMATIQUE
POUR L'INDUSTRIE
ET L'ENTREPRISE

FICHE TECHNIQUE

MASTÈRE SPÉCIALISÉ CALCULS HAUTES PERFORMANCES / DONNÉES MASSIVES / INTELLIGENCE ARTIFICIELLE

MANAGEMENT DE PROJET ET APPLICATION TRANSVERSALE

DURÉE 40H (2 UE = 4 ECTS)

Maîtriser la méthode SCRUM qui est une méthode de gestion de projet agile. Comprendre les fondements de l'agilité pour bien comprendre et appliquer correctement les méthodes qui en découlent.

OBJECTIFS

Connaître les différentes méthodes qui gravitent autour de l'agilité et de SCRUM.

Connaître autant les concepts que savoir les appliquer justement. Travail en groupe encadré par des professionnels. À travers la réalisation d'un cahier des charges et du développement d'un prototype, mise en œuvre des notions acquises dans le module.

CONNAISSANCES

Gestion de projet Agile - Scrum (Certification).

COMPÉTENCES ATTENDUES

Capacité à travailler en équipe en mode projet.

MISE EN PRATIQUE SOUS HADOOP ET SPARK (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

APPRENTISSAGE (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

SIMULATION NUMÉRIQUE (CERTIFICATION)

DURÉE 40H (2 UE = 4 ECTS)

Ce cours présente les bases des méthodes de simulation utilisées en statistique notamment en statistique bayésienne, en particulier les méthodes de calcul de maximisation et d'intégration en dimension élevée qui sont nécessaires pour traiter les modèles complexes utilisés dans les domaines tels que l'économétrie, la finance, la génétique, l'écologie ou la physique.

CONNAISSANCES

Méthodes de Monte Carlo, méthodes de Quasi Monte Carlo, méthodes MCMC (Algorithme de Metropolis-Hastings, le recuit simule...), Bootstrap, inférence bayésienne.

OPEN DATA CAMP

PROJET EN DATA SCIENCE

ORGANISÉ PAR LA SOCIÉTÉ GAINDE2000

DURÉE 30H (1,5 UE = 3 ECTS)

PROJET EN DATA SCIENCE

ORGANISÉ PAR LA SOCIÉTÉ BAAMTU

DURÉE 30H (1,5 UE = 3 ECTS)

PROJET EN DATA SCIENCE

ORGANISÉ PAR LA SOCIÉTÉ ATOS

DURÉE 40H (2 UE = 4 ECTS)

Durant une semaine, porteurs de projets et (ré)utilisateurs de données se retrouveront pour poursuivre ou proposer des initiatives élaborées à partir de données ouvertes, afin d'analyser un sujet, de développer un service ou de résoudre un problème. L'objectif est de faire communiquer les étudiants, professeurs et professionnels de la donnée autour d'un problème unique permettant de faire converger à la fois différentes compétences et surtout différentes visions de ce que l'on peut faire avec les données.

SYSTÈMES À BASE DE CONNAISSANCES

APPRENTISSAGE PAR RENFORCEMENT