

UE Modélisation et Machine Learning

Master 1 – Physique fondamentale et applications

Dernière édition: Septembre 2022

Auteurs: Alexandre Boucaud, Christophe Deroulers, Guillaume Rousseau, Marc-Antoine Verdier

Table des matières

- [Résumé](#)
- [Objectifs](#)
- [Connaissances développées](#)
- [Compétences développées](#)
- [Programme](#)
- [Références](#)
 - [Machine learning pour la physique et applications en sciences](#)
 - [Ressources de calcul GPU](#)

Résumé

- *Intitulé UE* : Modélisation et Machine Learning – MML
- *Semestre* : S2
- *Crédits ECTS* : 5 ECTS
- *Responsable de l'UE* : Alexandre Boucaud
- *Equipe pédagogique* : Alexandre Boucaud, Christophe Deroulers, Guillaume Rousseau, Marc-Antoine Verdier
- *Volume horaire*
 - Cours : 15h
 - TD/TP : 27h cours/TD sur ordinateur
- *Conditions particulières* : TD en salle informatique ou sur ordinateur personnel permettant la programmation (des ordinateurs portables peuvent être empruntés au SCRIPT)
- *Pré-requis*
 - programmation en Python
 - usage de bibliothèques de calcul scientifique
 - éléments d'algèbre linéaire.
- *Modalités d'évaluation* :
 - évaluations ponctuelles des rapports de TD
 - projet d'analyse de données à réaliser seul ou en binôme en fin de semestre
- *Effectif maximum* : 32 étudiants répartis en trois salles informatiques pour les TD

Objectifs

Introduire aux étudiants de M1 physiciens des approches d'analyse de données et des méthodes d'apprentissage automatique ("machine learning") en lien notamment avec des thématiques de physique. Différents types d'applications seront passés en revue durant le cours en explorant tout particulièrement des usages innovants dans le domaine de la physique, sans exclusion d'autres exemples sociétaux. L'enseignement présentera les fondements du "deep learning" dans la physique statistique et l'application des méthodes de "machine learning" en physique.

Connaissances développées

- compréhension d'algorithmes centraux en analyse de données et machine learning
- concepts de bases de statistiques, erreurs, variance, vraisemblance, corrélations
- connaissances de base de statistique Bayésienne
- méthodes d'optimisation, descente de gradient
- méthodes Monte Carlo, chaînes de Markov
- méthodes de régression et classification supervisées et non supervisées
- éléments sur les réseaux de neurones et l'apprentissage, backpropagation
- notions de calcul sur GPU

Compétences développées

Les étudiants apprendront à :

- manipuler des algorithmes disponibles pour différentes méthodes de "machine learning"
- ajuster des modèles à des données
- évaluer la validité des ajustements
- choisir parmi plusieurs modèles d'optimisation en fonction du problème
- effectuer des régressions à partir de jeux de données
- réaliser des classifications et des regroupements d'objets à partir de leurs attributs
- penser à des usages de méthodes de "machine learning" dans des applications physiques
- être capable d'étendre les connaissances acquises à d'autres exemples et situations
- utiliser des bibliothèques établies d'analyse de données et machine learning en Python

Programme

1. Introduction générale sur le "machine learning" suivi d'un panorama des usages en sciences et dans la vie de tous les jours
2. Notions de statistiques
 - rappels sur les distributions continues vs discrètes
 - notion d'incertitude dans les données et mesures sur jeux de données
 - traitement fréquentiste vs bayésien
3. Ajustement de modèles linéaires sur des données

- descente de gradient
 - chaînes de Markov, MCMC
 - inférence bayésienne
 - notion d'incertitude dans les modèles
4. Apprentissage non-supervisé
 - introduction de scikit-learn
 - clustering : regroupement automatique d'objets en différentes classes
 - réduction de dimension : détermination des variables dominantes dans un jeu de données et
 5. Apprentissage supervisé
 - classification : affectation à une catégorie d'un objet suivant ses caractéristiques
 - régression : prédiction d'une valeur associée à un objet
 - algorithmes classiques d'apprentissage supervisé (régression linéaire/logistique, arbres de décision, support vector machines, random forests, Gaussian mixture models)
 - prétraitement des données, cross-validation, selection des modèles, optimisation des paramètres
 6. Réseaux de neurones et introduction au deep learning
 - anatomie des réseaux de neurones
 - backpropagation
 - traitement d'image avec réseaux de convolution (CNN)
 - ouverture sur les réseaux génératifs de type GANs et VAE.

Références

Machine learning pour la physique et applications en sciences

- [Machine learning for physicists](#), Pankaj Mehta, Boston University
- [Learning from Data](#), Yaser Abu-Mostafa, Caltech
- [Data Analysis and Machine Learning](#), Morten Hjorth-Jensen, Michigan State University & Oslo University

Ressources de calcul GPU en ligne

- [Google Collaboratory](#)