
Programme de Formation

Introduction au Machine Learning : méthodes statistiques de prédiction, quantification d'incertitude, mise en œuvre sous Python

(Lm^B)

Organisation

Durée : 21 heures

Mode d'organisation : Présentiel

Contenu pédagogique

Public visé

Techniciens supérieurs, ingénieurs ou chercheurs désireux de découvrir les notions essentielles des sciences des données et du Machine Learning et de savoir les mettre en œuvre sur des cas d'étude ; chargés de R&D souhaitant valoriser une base de données.

Objectifs pédagogiques

- Conceptualiser la notion de prédiction dans le cadre de l'apprentissage supervisé en régression et classification
- Comprendre les principaux algorithmes de prédiction en Machine Learning tels que Random Forest, Gradient Boosting et Neural Network
- Entraîner des modèles prédictifs, évaluer leurs performances et sélectionner le meilleur modèle
- Quantifier l'incertitude de prédiction grâce à la prédiction probabiliste et aux réseaux de neurones bayésiens
- Mettre en œuvre sur des cas concrets avec Python (Scikit Learn et Keras/TensorFlow)

Description

1er jour

- Motivation : lien entre problématique appliquée/question de Machine Learning
- Cadre conceptuel de l'apprentissage supervisé (régression, classification)
- Estimation d'erreur de prédiction (validation simple, validation croisée)
- Workflow du Machine Learning pour la recherche d'un bon modèle prédictif



- Algorithmes prédictifs basés sur les arbres : arbres CART, méthode d'ensemble (Bagging et Random Forest), méthodes de Boosting (Gradient Boosting, XGBoost, LightGBM)
- Interprétation, identification des variables importantes et de leur effet
- Mise en œuvre sur des études de cas avec Scikit-Learn

2ème jour

- Modèles simples de réseau de neurones (Multi Layer Perceptron)
- Entraînement par descente du gradient stochastique et calcul de rétropropagation du gradient
- Régularisation des réseaux de neurones : pénalisation, Dropout, Early Stopping, Batch Normalisation
- Réseaux de neurones pour les données structurées (Convolutional Neural Network pour les images,
- Recurrent Neural Nets pour les séries temporelles)
- Mise en œuvre sur des études de cas avec Keras/Tensorflow

3ème jour

- Quantification d'incertitude de prédiction
- Préviation probabiliste et régression distributionnelle
- Réseaux de neurones bayésiens
- Mise en œuvre sur des études de cas sous Python



Prérequis

Bases du langage de programmation Python, bases de l'algorithmique, notions élémentaires de statistiques.



Modalités pédagogiques

La formation contiendra environ 50% de cours et 50% de TP sous Python. Pour chaque demi-journée, 1h45 de cours et 1h45 de TP sont prévues. Formation disponible en hybride avec la possibilité d'accueillir simultanément des étudiants en présentiel ou à distance.



Moyens et supports pédagogiques

A distance : ordinateur avec les logiciels (libres) nécessaires à la formation préalablement installés (liste communiquée avant le début de la formation).

En présentiel : ordinateur personnel ou salle de TP du laboratoire de mathématiques de Besançon mise à disposition.

Transparents de cours et Notebook python corrigés (format PDF et Jupyter) remis aux participants à l'issue de la formation.



Modalités d'évaluation et de suivi

Un suivi individualisé par des évaluations formatives est assuré. Une attestation de fin de formation est délivrée à la fin du parcours.



Informations sur l'admission

L'admission à cette formation ne fait l'objet d'aucun examen, test ou sélection préalable ; l'inscription est validée après réception du dossier complet et confirmation par l'organisme de formation.



Informations sur l'accessibilité

Notre organisme s'engage à garantir l'accessibilité de ses formations à distance et en présentiel aux personnes en situation de handicap. Un référent handicap est mobilisable afin d'analyser les besoins spécifiques et de mettre en place, lorsque cela est possible, les adaptations pédagogiques, techniques ou



organisationnelles nécessaires.