

Introduction à l'apprentissage automatique

P. BOILY | UNIVERSITÉ D'OTTAWA | FACULTÉ DES SCIENCES | DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE
DATA ACTION LAB | IDLEWYLD ANALYTICS

AVEC LA CONTRIBUTION DE **J. SCHELLINCK** | SYSABEE | DATA ACTION LAB

Instructeur - Patrick Boily

Emploi

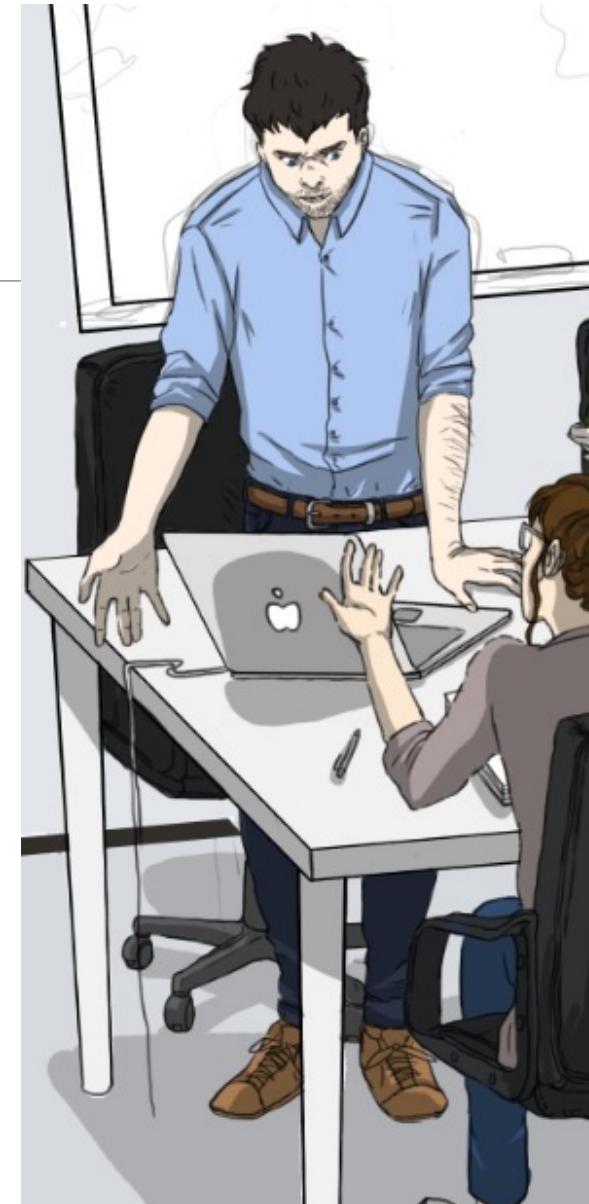
- Professeur Math/Stat [depuis '19, uOttawa]
- Président et conseiller principal [depuis '16, Idlewyld Analytics]
- Directeur et consultant principal ['12 - '19, CQADS, Carleton]
- Fonctionnaire GdC ['08 - '12, ASFC | StatCan | TC | TPSGC]
- 70+ cours universitaires ; 300+ jours d'atelier

Projets

- GAC ; NWMO ; CATSA ; etc.
- 40+ projets

Spécialisation

- Visualisation des données ; nettoyage des données (... malheureusement)
- Application d'un large éventail de techniques à tous les types de données
- Modélisation mathématique/statistique



Matériel de cours

Page web du cours :

<https://data-action-lab.com/103-iml>

Notes de cours :

<https://www.practicedataviz.com>

Contact :

pboily@uottawa.ca

Espace de travail Slack :

<https://dspdi.slack.com>

Description du cours

Ce cours amène les participants à analyser et à discuter des tâches et des problèmes généraux de l'apprentissage statistique (apprentissage automatique), ainsi que de leurs pièges.

Les participants seront initiés à des algorithmes simples d'extraction de règles d'association, de classification et de regroupement.

Après le cours, les participants ont la possibilité de travailler sur un projet guidé, en recevant les commentaires de l'instructeur.

Informations complémentaires

Les participants sont censés être familiers avec les concepts introduits dans les cours *Les principes fondamentaux de la science des données* (préparation des données, nettoyage des données), et *La visualisation des données et les tableaux de bord* (exploration des données), ainsi qu'avec leurs prérequis. Une bonne connaissance des méthodes d'optimisation serait bénéfique mais n'est pas nécessaire.

Les participants sont tenus d'apporter un ordinateur portable ou personnel sur lequel la version actuelle de R/RStudio (Posit) est installée (pour laquelle ils peuvent avoir besoin d'une autorisation administrative pour installer des librairies).

Les participants au projet guidé doivent être familiers avec R, le tidyverse et/ou Python.

Objectifs d'apprentissage

A l'issue de ce cours, les participants seront en mesure :

- différencier les situations qui nécessitent une approche d'apprentissage supervisé de celles qui nécessitent une approche non supervisée (ou une combinaison des deux)
- identifier les stratégies utilisées pour surmonter les problèmes et les défis courants de l'apprentissage statistique dans le monde réel
- reconnaître la variété des algorithmes d'apprentissage automatique à leur disposition
- mettre en œuvre des algorithmes simples d'apprentissage automatique pour fournir des informations exploitables
- construire un pipeline simple d'analyse de données intégrant des composants d'apprentissage automatique

Plan du Cours

Apprentissage Statistique

1. Types d'apprentissage ;
tâches d'apprentissage automatique

Règles d'association

2. Aperçu ;
Étude de cas : données médicales
3. Concepts

Classification

4. Aperçu ;
Étude de cas : audits fiscaux
5. Arbres de décision et autres algorithmes
6. Évaluation de la performance

Session 1

Session 2

Session 3

Session 4

Plan du Cours

Regroupement

- 7. Aperçu ;
Étude de cas : Livehoods
- 8. k -moyennes et autres algorithmes
- 9. Validation et commentaires

Enjeux et défis

- 10. Données pourries et données massives
- 11. Ajustement et transférabilité
- 12. Faits divers

Session 1

Session 2

Session 3

Session 4