

00

Introduction au cours

420-A58-SF — Algorithmes d'apprentissage non supervisé — Été 2021
Spécialisation technique en intelligence artificielle — M. Swawola, M.Sc.

**NOUS ÉCLAIRONS.
VOUS BRILLEZ.**

FORMATION CONTINUE
ET SERVICES AUX ENTREPRISES

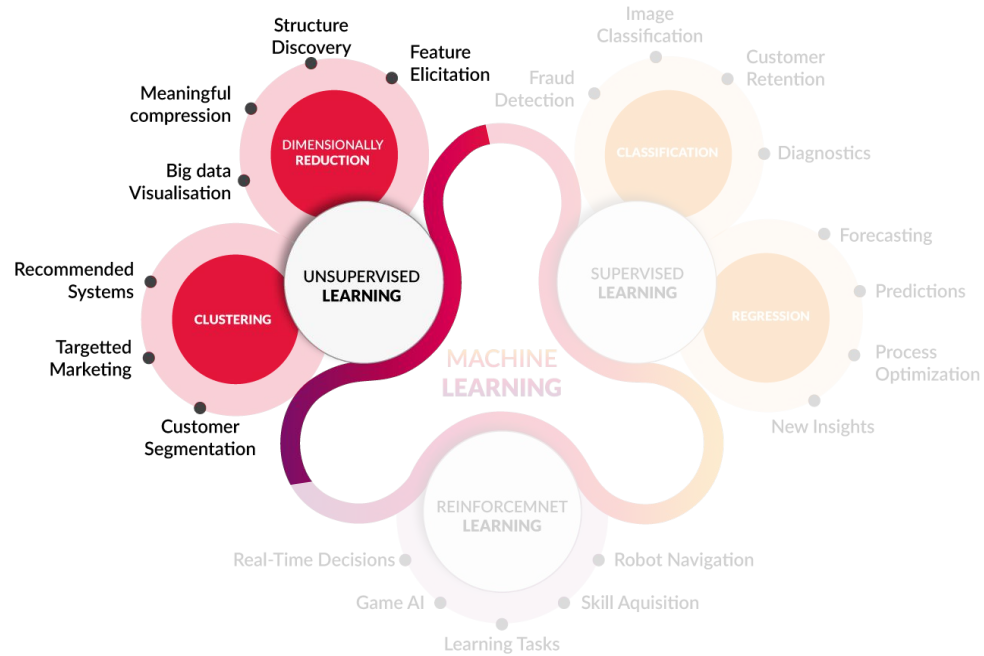


Sommaire

1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée

Sommaire

1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée



Extrait du cours "Introduction à l'apprentissage automatique" de François Laviolette, CRDM

Exemples d'applications (1/3)

Optimisation de portfolio



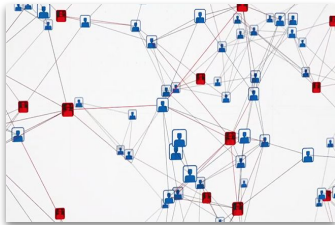
Centres de données



Astrophysique



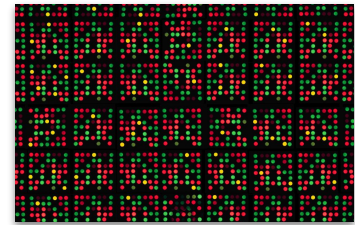
Réseaux sociaux



Marketing



Génétique

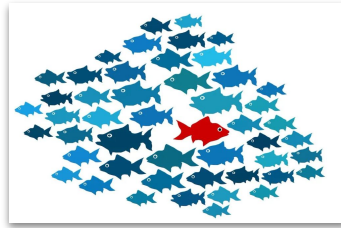


Exemples d'applications (2/3)

Cybersécurité



Détection d'anomalies



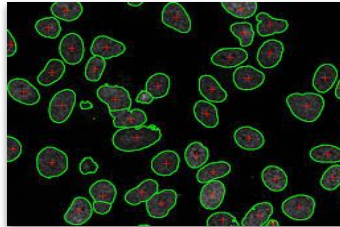
Commerce de détail



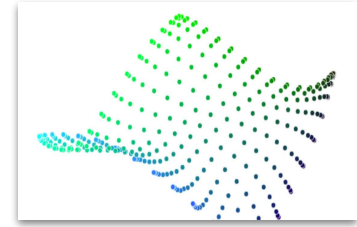
Recommandation



Vision numérique



Visualisation



Exemples d'applications (3/3)

Télédétection



Ingénierie de données



Recherche de documents



Et bien d'autres ...

Le “gâteau” de Yann LeCun



Yann LeCun

Apprentissage
non supervisé

■ “Pure” Reinforcement Learning (cherry)

- ▶ The machine predicts a scalar reward given once in a while.
- ▶ **A few bits for some samples**

■ Supervised Learning (icing)

- ▶ The machine predicts a category or a few numbers for each input
- ▶ Predicting human-supplied data
- ▶ **10→10,000 bits per sample**

■ Unsupervised/Predictive Learning (cake)

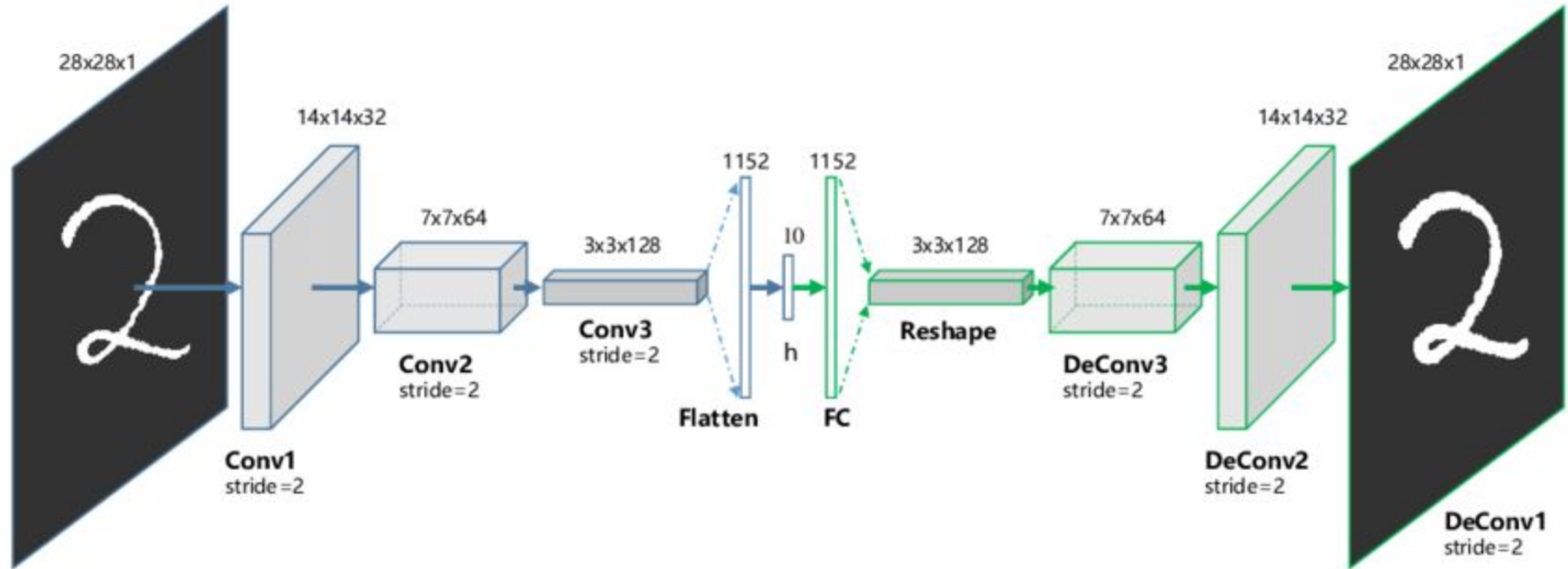
- ▶ The machine predicts any part of its input for any observed part.
- ▶ Predicts future frames in videos
- ▶ **Millions of bits per sample**

■ (Yes, I know, this picture is slightly offensive to RL folks. But I’ll make it up)



L'apprentissage non supervisé attend toujours sa révolution, son “deep learning” ...

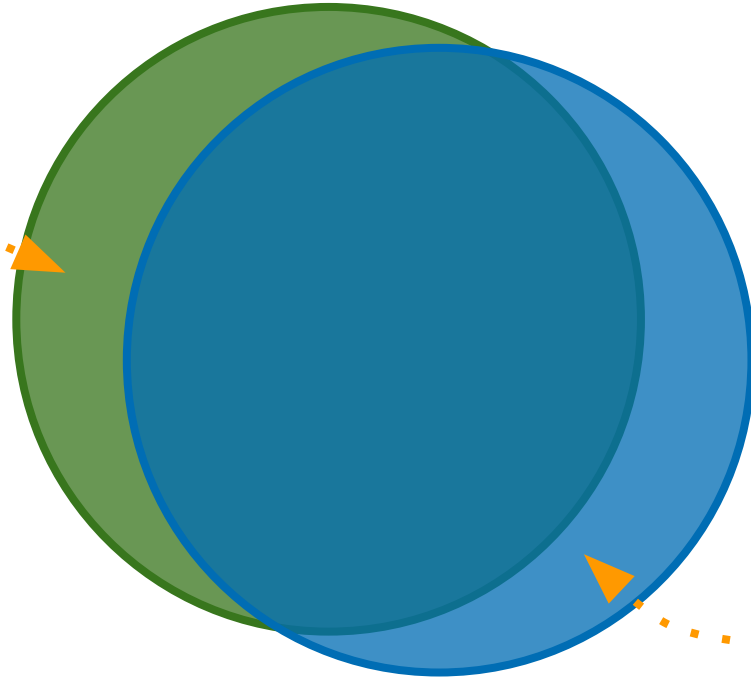
Deep learning “non supervisé”



The structure of proposed Convolutional AutoEncoders (CAE) for MNIST

Apprentissage non supervisé vs fouille de données

Apprentissage non
supervisé



Fouille de
données

Pourquoi l'apprentissage non supervisé est difficile ?

- Comme son nom l'indique, le caractère non-supervisé signifie l'absence de variable dépendante (y)
- Les données sont donc non étiquetées (X)
- Il n'y a pas de bonne réponse à un problème et la subjectivité est souvent de mise
- Étant donné l'absence d'étiquettes, il est difficile, voir parfois impossible, d'évaluer la performance des algorithmes
- Les résultats doivent être interprétés en se basant sur des connaissances reliées au domaine d'application
- Mais tout ceci ne fait que rendre l'apprentissage non supervisé encore plus **passionnant** !

Sujets non couverts par le cours

- **Réduction de dimension (ACP, t-SNE, ...)**

→ En partie couvert par Pierre-Marc Juneau (420-A55-SF)

- **Détection d'anomalie**

→ En partie couvert par Amor Amami (420-A56-SF)

- Les compétences correspondant à ces sujets (exemple: ACP) sont considérées acquises par l'étudiant et pourront être requises lors des examens à venir

Sommaire

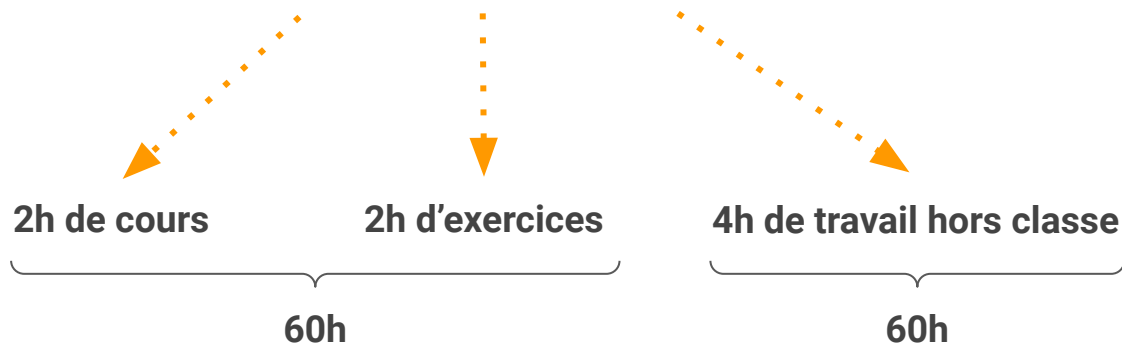
1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée

Organisation du cours

- Il s'agit d'un cours de 60 heures réparties sur 8 semaines en séances de 4h
- Sauf mention contraire, les séances ont lieu sur Microsoft Teams de 18h00 à 22h00
- Entre 30 minutes et 40 minutes de **pause** sont réparties sur la durée de chaque séance
- Il n'y a pas de séances prévues les fins de semaines
- En dehors des cours, entre-aidez vous sur le salon de clavardage Teams. Restez polis et courtois ;-)
- Vos **commentaires** (positifs, constructifs) sur le cours sont les bienvenus !

Pondération

2 - 2 - 4



Évaluations


- Projet - **30%** → **Commence dès aujourd'hui !**
- Examen de mi-session (2 heures) - **20%**
- Examen de fin de session (4 heures) - **40%**
- Évaluations formatives (4) - **10%**

Consignes sur le respect de la vie privée

- Sauf lors ses examens, vous êtes en droit de refuser d'activer votre caméra
- En cas de refus de votre part d'activer votre caméra le jour de l'examen, la possibilité de venir passer l'examen au Cégep avec caméra vous est offerte
- Lors des examens, le partage d'écran lorsque demandé est obligatoire
- La diffusion des supports de cours (slides) non modifiés en dehors de la classe est permise
- Il est strictement interdit de télécharger et diffuser les vidéos des cours

Plan de cours

- Le **plan de cours** détaillé comprenant le calendrier est disponible sur **Lea**



Cégep
de
Sainte-Foy
Formation continue et services aux entreprises

Spécialisation technique en Intelligence
Artificielle – AEC (LEA.D1)

Plan de cours
Algorithmes d'apprentissage non supervisé
420-A58-SF

Pondération : 2-2-4
Été 2021 – UBR 4126 – Groupe 14032

Professeur :	Mikael Swawola, M.Sc.
Bureau :	N/A
Téléphone (Cégep) :	N/A
Courriel :	mswawola@csoy.ca
Conseillère pédagogique :	Françoise Cloutier
Bureau :	P-271
Téléphone (Cégep) :	418-659-6620, poste 5043
Courriel :	fcloutier@csoy.ca

1

Structure du cours

- Le cours est divisé en 4 parties
 1. Partitionnement de données
 2. Fouille de données
 3. Systèmes de recommandation
 4. Séries temporelles

Sommaire

1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée

Ajustement de la pédagogie

- La **classe inversée** (vidéos des cours postées en ligne avant la séance) ne sera pas renouvelée cette session
- Les ateliers seront réalisés et corrigés en classe. Les fichiers contenant les solutions ne seront pas fournis
- Il sera possible d'envoyer le travail effectué sur les ateliers à l'enseignant pour correction et rétroactions
- Les séances seront toujours enregistrées et disponibles sur **Microsoft Stream**
- Les **supports de cours** seront postés sur le dépôt GitHub du cours au début de chaque séance

Ajustement de la pédagogie

- Le mode d'enseignement reste **synchrone**, c'est-à-dire:
 - Sauf exception, la présence aux séances est **obligatoire**
 - Les séances sont enregistrées, mais ne devrait servir qu'à s'y référer pour consolider les connaissances acquises après avoir assisté aux différentes séances
 - Il est important de ne pas raisonner dans l'autre sens, à savoir remettre à plus tard l'écoute d'un cours sachant que celui-ci est enregistré. Ce n'est pas un MOOC ! De plus, rattraper un cours de 4h nécessite généralement beaucoup plus que 4 heures !
 - **Ces ajustements n'ont qu'un seul objectif, augmenter vos chances de réussite au cours et vous permettre de développer de solides compétences (très recherchées) en intelligence artificielle !**

Sommaire

1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée

Langages et IDEs

- Voici ci-dessous les langages de programmation et IDEs qui seront utilisés lors du cours
 1. Partitionnement de données (Python / Jupyter)
 2. Fouille de données (Python et R / Jupyter et RStudio)
 3. Systèmes de recommandation (Python / Jupyter)
 4. Séries temporelles (R / RStudio)

- **À ce stade de la formation, il est important que vous soyez en mesure d'installer et configurer ces environnements de manière autonome. Aucun temps de cours ne sera consacré à ces aspects !**

Dépôt du cours

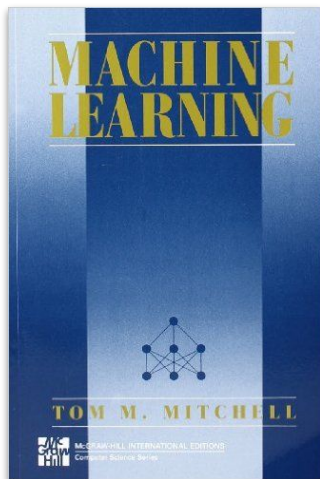
- Tous les exercices et jeux de données seront accessibles sur le **dépôt GitHub du cours**
<https://github.com/mswawola-cegep/420-a58-sf.git>
- Utilisez le fichier **requirements.txt** pour assurer votre environnement d'une complète compatibilité avec les ateliers du cours

Sommaire

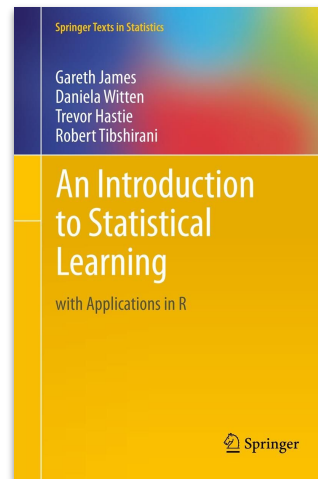
1. Introduction à l'apprentissage non supervisé et ses principaux défis
2. Organisation du cours
3. Ajustement de la pédagogie
4. Environnements logiciels du cours
5. Littérature recommandée

Littérature recommandée

- **Machine Learning**, Tom Mitchell
McGraw-Hill Higher Education

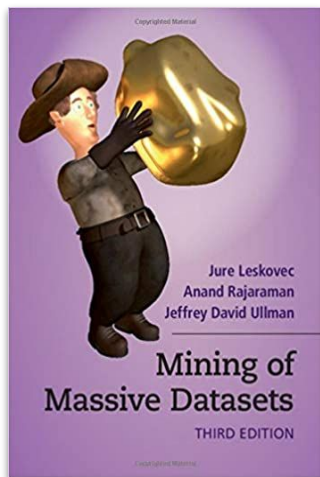


- **An Introduction to Statistical Learning with Applications in R**, G. James, D. Witten, T. Hastie, R. Tibshirani

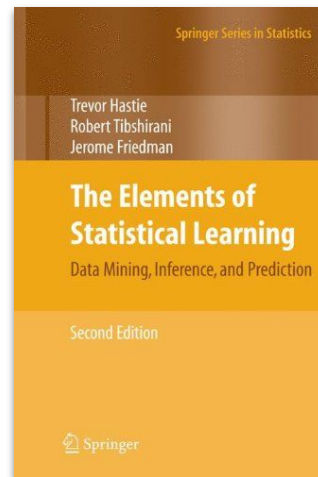


Littérature recommandée

- **Mining of Massive Datasets**, J. Leskovec, A. Rajaraman, J.D. Ullman

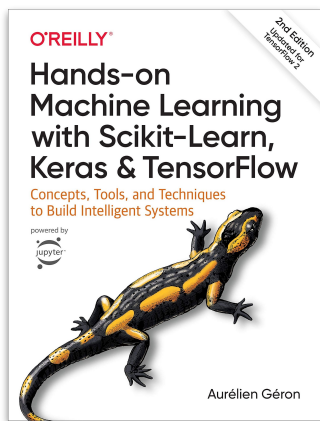


- **Elements of Statistical Learning Second Edition**, T. Hastie, R. Tibshirani, J. Friedman

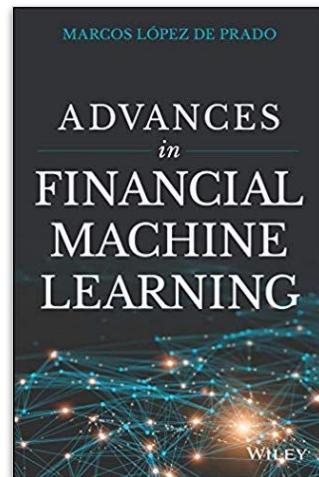


Littérature recommandée

- **Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Aurélien Géron**

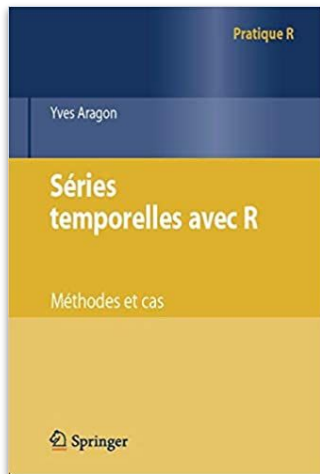


- **Advances in Financial Machine Learning, Marcos Lopez De Prado**



Littérature recommandée

- **Séries temporelles avec R**, Yves Aragon



- **Time Series : Theory and Methods**, Peter J. Brockwell et Richard A. Davis

