

Présentation du cours

Introduction à l'apprentissage automatique – GIF-4101 / GIF-7005

Professeur : Christian Gagné

Semaine 1



UNIVERSITÉ
LAVAL

Introduction à l'apprentissage automatique

- GIF-4101 : cours optionnel, bacc. en génie informatique, informatique, statistique, mathématique-informatique, génie logiciel, génie mécanique, génie géomatique, etc.
 - Préablables
 - STT-2920 – Probabilités pour ingénieurs (ou équivalent)
 - MAT-2930 – Algèbre linéaire appliqué (ou équivalent)
- GIF-7005 : cours de 2e-3e cycles, maîtrise professionnelle en informatique – intelligence artificielle, maîtrise et doctorat en génie électrique, informatique
 - Accessible aux étudiants d'autres programmes de maîtrise et doctorat (sciences et génie)
 - Aucun préalable officiel, mais exige des connaissances de base en :
 - Algèbre linéaire
 - Probabilité et statistique
 - Calcul différentiel
 - Programmation et algorithmique

Introduction to Machine Learning

- GIF-7015 : official English version of GIF-7005 (**new** this year)
 - Slides and video capsules available in English
 - Howeworks, quizzes and exam provided in English
 - Bilingual text in monPortail (when possible)
 - Discussions on PAX forum in both languages
- In-class discussions (Wednesday morning) can be in English
 - By default, explanations will be given in French, but I will answer in the language the questions are asked
 - Do not hesitate to ask for repeating an answer in English when you do not get the French explanations
- Everything else is the same than GIF-7005

- Matière théorique présentées en capsules vidéo
 - Présentations détaillées de la matière dans diapositives
 - Capsules vidéos présentant la matière à chaque semaine
- Période de présence hybride synchrone, mercredi de 8h30 à 9h45
 - En présentiel au VCH-3820 (106 places disponibles)
 - Diffusion simultanée sur Zoom avec enregistrement des séances pour écoute ultérieure
 - Explications, présentations techniques, questions et réponses

Évaluations

- Cinq quiz à la fin des séances du mercredi (9h45-10h30), sur matière des 2-3 dernières semaines
- Un examen écrit sur matière théorique et développements mathématiques (9 novembre au soir)

	Date	Heure	Pond.	Format
Quiz 1	28 septembre 2022	9h45 à 10h30	4 %	Questionnaire objectif en-ligne
Quiz 2	12 octobre 2022	9h45 à 10h30	4 %	Questionnaire objectif en-ligne
Quiz 3	26 octobre 2022	9h45 à 10h30	4 %	Questionnaire objectif en-ligne
Quiz 4	16 novembre 2022	9h45 à 10h30	4 %	Questionnaire objectif en-ligne
Quiz 5	30 novembre 2022	9h45 à 10h30	4 %	Questionnaire objectif en-ligne
Examen	9 novembre 2022	18h30 à 20h20	20 %	Questionnaire papier avec réponses à développement

Travaux pratiques

- Cinq devoirs réalisés dans l'environnement PAX (notebook Jupyter)
 - Premier cycle : devoirs en équipe de 2 étudiants
 - Étudiants gradués : devoirs individuels
- Un projet libre en équipe
 - En équipe de 3 à 5 étudiants (mixité des groupes permise)
 - Présentation lors de séance d'affiches, 21 et 22 décembre en matinée
- Programmation en Python avec scikit-learn et PyTorch

Travail	Date	Heure	Pondération
Devoir 1	7 octobre 2022	12h00	8 %
Devoir 2	21 octobre 2022	12h00	8 %
Devoir 3	11 novembre 2022	12h00	8 %
Devoir 4	25 novembre 2022	12h00	8 %
Devoir 5	9 décembre 2022	12h00	8 %
Projet	21-22 décembre 2022	matinée	20 %

- GIF-4101 (1er cycle) : note de passage à **50 %**

A+ : [100, 90]	A :]90, 85]	A- :]85, 80]
B+ :]80, 76]	B :]76, 72]	B- :]72, 68]
C+ :]68, 64]	C :]64, 60]	C- :]60, 56]
D+ :]56, 53]	D :]53, 50]	E :]50, 0]

- GIF-7005 et GIF-7015 (2e-3e cycles) : note de passage à **60 %**

A+ : [100, 90]	A :]90, 85]	A- :]85, 80]
B+ :]80, 76]	B :]76, 72]	B- :]72, 68]
C+ :]68, 64]	C :]64, 60]	E :]60, 0]

Contenu (1/2)

- **Introduction (2h)** : types d'apprentissage ; optimisation ; régularisation ; généralisation ; méthodologie.
- **Méthodes paramétriques (6h)** : fonctions discriminantes ; théorème de Bayes ; méthodes paramétriques ; estimation par maximum de vraisemblance ; compromis biais/variance ; méthodes multivariées ; densité mélange ; régression linéaire ; régression multivariée.
- **Méthodes non paramétriques (3h)** : fonctions de densité ; estimation de densité par noyau ; classement par les plus proches voisins ; mesures de distance.
- **Discriminants linéaires (3h)** : descente du gradient ; séparation à plusieurs classes ; algorithme du Perceptron ; méthode des moindres carrés ; régression logistique.
- **Méthodes à noyaux (3h)** : fonctions noyaux ; maximisation des marges géométriques ; séparateurs à vastes marges ; exemples d'autres méthodes à noyaux.
- **Perceptron multicouche (3h)** : modèle de réseaux de neurones ; algorithme de rétropropagation des erreurs.

Contenu (2/2)

- **Apprentissage profond (6h)** : apprentissage de représentations ; techniques pour l'apprentissage de réseaux profonds ; exemples d'applications.
- **Méthodes par ensembles (3h)** : votes à majorité ; codes à correction d'erreurs ; mixtures d'experts ; Bagging ; Boosting ; arbres de décision ; forêts aléatoires.
- **Prétraitements et configuration de modèles (3h)** : analyse en composantes principales ; sélection vorace avant/arrière de caractéristiques ; données hétérogènes ; variables manquantes ; imputation ; validation croisée ; optimisation d'hyper-paramètres.
- **Clustering (3h)** : algorithme K-means ; algorithme espérance-maximisation ; clustering hiérarchique ; apprentissage de variétés ; positionnement multidimensionnel.
- **Conception et analyse d'expérimentation en apprentissage (3h)** : plans d'expériences ; matrice de confusion ; courbes ROC ; Bootstrap ; estimation de l'erreur ; tests statistiques ; comparaison d'algorithmes.

- Plusieurs références disponibles pour approfondissement
 - Pas de livre obligatoire
 - La plupart des références sont disponibles en format numérique

📖 Ethem Alpaydin. *Introduction to Machine Learning*. 4e édition, MIT press, 2020.

📖 Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
<https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/>

📖 Trevor Hastie, Robert Tibshirani et Jerome Friedman. *The elements of statistical learning : data mining, inference, and prediction*. 2e édition, Springer, 2009.
<https://web.stanford.edu/~hastie/ElemStatLearn/>

📖 Ian Goodfellow, Yoshua Bengio et Aaron Courville. *Deep Learning*. MIT press, 2016.
<http://www.deeplearningbook.org/>

📖 Andriy Burkov. *The Hundred-Page Machine Learning Book en français*. 2019.
<http://thtmlbook.com/wiki/doku.php>

- Nom : Christian Gagné
- Bureau : PLT-1138-F
- Courriel : christian.gagne@gel.ulaval.ca
- Web : <https://vision.gel.ulaval.ca/~cgagne>
- Aide pour le cours :
 - Questions et discussions d'intérêt général sur la matière et les évaluations : forum du cours sur PAX
 - Considérations plus spécifiques ou personnelles : courriel au professeur

- Site monPortail :

<https://sitescours.monportail.ulaval.ca/ena/site/accueil?idSite=146163>

- Plans de cours
- Capsules vidéo et présentations
- Quiz en-ligne
- Notes des travaux et examens

- Site PAX : <https://pax.ulaval.ca/GIF-4101-7005/A22/>

- Devoirs (notebook Jupyter)
- Forum de discussions

- Chaîne YouTube :

https://www.youtube.com/channel/UCkMXqYn0zhI1RseU0gq8_Xw

- Capsules vidéos du cours