



Département d'informatique
IFT 603 — Techniques d'apprentissage
Plan de cours
Hiver 2018

Enseignant :

Pierre-Marc Jodoin

Courriel : pierre-marc.jodoin@USherbrooke.ca

Local : D4-1016-1

Téléphone : 819-821-8000 (62025)

Site : <http://info.usherbrooke.ca/pmjodoin/cours/IFT603/index.html>

Disponibilité : Jeudi de 9h à 17h

Auxiliaire

Carl Lemaire

Courriel : carl.lemaire@USherbrooke.ca

Responsable(s) : Pierre-Marc Jodoin, André Mayers, Shengrui Wang**Horaire :**

| | | | |
|--------------------------|----------|---------------|--------------------|
| Exposé magistral : | Mardi | 10h30 à 12h20 | Salle à déterminer |
| | Vendredi | 13h30 à 14h20 | Salle à déterminer |
| Exercices/laboratoires : | Vendredi | 13h30 à 14h20 | Salle à déterminer |

Description officielle de l'activité pédagogique¹

Cibles de formation : Introduire aux techniques d'apprentissage automatique en intelligence artificielle.

Contenu : Concepts fondamentaux de l'apprentissage automatique, formulation mathématique probabiliste, régression linéaire, classification linéaire, astuce du noyau, machine à vecteurs de support, apprentissage bayésien, modèle de mélange de gaussiennes, algorithme EM, réduction de dimensionnalité et combinaison de modèles.

Crédits 3

Organisation 3 heures d'exposé magistral par semaine
0 heure d'exercices par semaine
6 heures de travail personnel par semaine

Particularités Aucune

¹ <https://www.usherbrooke.ca/fiches-cours/ift603>

1 Présentation

Cette section présente les objectifs spécifiques et le contenu détaillé de l'activité pédagogique. Cette section, non modifiable sans l'approbation d'un comité de programme du Département d'informatique, constitue la version officielle.

1.1 Mise en contexte

L'apprentissage automatique ou l'apprentissage par machine (Machine Learning) s'intéresse à la conception, l'analyse, l'implémentation et l'application de programmes d'ordinateur capables de s'améliorer, au fil du temps, soit sur la base de leur propre expérience, soit à partir des données antérieures fournies par d'autres programmes. De nos jours, l'apprentissage automatique joue un rôle essentiel dans de nombreux domaines d'applications, tels que la vision par ordinateur, le traitement automatique du langage, la reconnaissance vocale, les systèmes tutoriels intelligents, la modélisation de l'utilisateur, la robotique, la bio-informatique, les finances, le marketing, les jeux vidéo, la télédétection, etc. En fait, la plupart des programmes de l'intelligence artificielle contiennent un module d'apprentissage. Presque tous les systèmes de reconnaissances de formes sont basés sur des techniques d'apprentissage.

1.2 Objectifs spécifiques

Ce cours vise à introduire les concepts fondamentaux de l'apprentissage automatique et à présenter une vaste gamme d'approches différentes utilisées dans des systèmes d'apprentissage automatique :

1. Comprendre les notions fondamentales de l'apprentissage automatique;
2. Comprendre la formulation probabiliste de l'apprentissage automatique;
3. Maîtriser des algorithmes de régression et classification en apprentissage supervisé;
4. Comprendre comment utiliser l'astuce du noyau;
5. Maîtriser les fondements des réseaux de neurones et leur apprentissage par rétro propagation;
6. Maîtriser des algorithmes de combinaison de modèles.

1.3 Contenu détaillé

| Thème | Contenu | Heures | Objectifs |
|-------|---|--------|-----------|
| 1 | Concepts fondamentaux : <ul style="list-style-type: none"> - apprentissage supervisé vs. non-supervisé - performance d'entraînement vs. généralisation - régularisation - sélection d'hyper-paramètres, validation croisée - exemple de la régression polynomiale | 3 | 1 |
| 2 | Formulation probabiliste : <ul style="list-style-type: none"> - théorie des probabilités (cas discret et continu) - formulation probabiliste de la régression polynomiale - théorie de l'information (entropie, information mutuelle, divergence de Kullback-Leibler) - Bayes et probabilités jointes, conditionnelles, a priori - Introduction au maximum de vraisemblance - Introduction au maximum a posteriori | 5 | 1,2 |
| 3 | Régression linéaire : <ul style="list-style-type: none"> - apprentissage par maximum de vraisemblance - apprentissage par maximum a posteriori | 2 | 3 |
| 4 | Classification linéaire : <ul style="list-style-type: none"> - Classification comme un problème de régression - Méthode de Fisher 2 classes - Méthodes génératives *Perceptron binaire vs. Multiclasse | 6 | 1,3,5 |

| | | | |
|----|---|---|-------|
| | *Régression logistique binaire vs multiclasse * descente de gradient * Sigmoid vs softmax * Distribution de Bernouilli vs entropie croisée | | |
| 5 | Méthode à noyau : - Représentation duale et régression à noyau - Construction de noyau | 3 | 4 |
| 6 | Méthode à noyau parcimonieuse : - Classifieur à marge maximale - Machine à vecteurs de support | 3 | 3,4 |
| 7 | Réseaux de neurones multi-couches - Approche linéaire vs non linéaire - Classification binaire et multiclassés - Fonction d'activation - Couche d'entrée vs couche cachée vs couche de sortie - Rétro propagation | 6 | 5 |
| 8 | Combinaison de modèles : - <i>Bootstrap</i> - <i>Boosting</i> | 3 | 6 |
| 9 | Théorie de la décision : - Décomposition biais-variance - Théorie de la décision | 2 | 1,2,3 |
| 10 | Mélange de gaussiennes (si le temps le permet): - Partitionnement de données (<i>clustering</i>) - Algorithme EM - Fenêtre de Parzen | 4 | 2 |

2 Organisation

Cette section propre à l'approche pédagogique de chaque enseignante ou enseignant présente la méthode pédagogique, le calendrier, le barème et la procédure d'évaluation ainsi que l'échéancier des travaux. Cette section doit être cohérente avec le contenu de la section précédente.

2.1 Méthode pédagogique

Une semaine comprend trois heures de cours magistraux. La plupart des présentations en classe se feront à l'aide de diapositives disponibles sur le site du cours au format pdf. Tout au long de la session, l'étudiant devra implémenter les notions vues en classe à travers cinq travaux pratiques. Des instructions particulières seront données pour chacun des travaux et les périodes de consultation seront déterminées en classe. En ce qui concerne les examens, seules des notes manuscrites non photocopiées sont admises, tout manuel et notes de cours imprimées étant interdits. L'examen final est récapitulatif.

2.2 Calendrier du cours approximatif

| | Semaine du | Chapitre | Travaux |
|---|------------|---------------------------------------|------------|
| 1 | 8 janvier | 1 (1h au lab) | |
| 2 | 15 janvier | 1-2 | |
| 3 | 22 janvier | 2 | |
| 4 | 29 janvier | 3 (1h au lab) | |
| 5 | 5 février | 4 | Remise tp1 |
| 6 | 12 février | 4 | |
| 7 | 19 février | Examen périodique (date à déterminer) | |

| | | | |
|----|------------|------------------------------------|------------|
| 8 | 26 février | Relâche des activités pédagogiques | |
| 8 | 5 mars | 5 (1h au lab) | |
| 10 | 12 mars | 5-6 | Remise tp2 |
| 10 | 19 mars | 6-7 | |
| 11 | 26 mars | 7 (1h au lab) | Remise tp3 |
| 12 | 2 avril | 7-8 | |
| 13 | 9 avril | 8 (1h au lab) | |
| 14 | 16 avril | 9 (10) | Remise tp4 |
| | XYZ avril | Examen final (date à déterminer) | |

Note : ce calendrier est approximatif et sujet à modification.

2.3 Évaluation

| | |
|----------------|------|
| Devoirs (4) : | 40 % |
| Examen intra : | 20 % |
| Examen final : | 40 % |

Conformément aux articles 36, 37 et 38 du règlement facultaire d'évaluation des apprentissages² l'enseignant peut retourner à l'étudiante ou à l'étudiant tout travail non conforme aux exigences quant à la qualité de la langue et aux normes de présentation.

Le plagiat consiste à utiliser des résultats obtenus par d'autres personnes afin de les faire passer pour sien et dans le dessein de tromper l'enseignant. Si une preuve de plagiat est attestée, elle sera traitée en conformité, entre autres, avec l'article 8.1.2 du Règlement des études³ de l'Université de Sherbrooke. L'étudiant ou l'étudiante peut s'exposer à de graves sanctions dont automatiquement un zéro (0) au devoir ou à l'examen en question.

Ceci n'indique pas que vous n'avez pas le droit de coopérer entre deux équipes tant que la rédaction finale des documents et la création du programme reste le fait de votre équipe. En cas de doute de plagia, l'enseignant peut demander à l'équipe d'expliquer les notions ou le fonctionnement du code qu'il considère comme étant plagié. En cas de doute, ne pas hésiter à demander conseil et assistance à l'enseignant afin d'éviter toute situation délicate par la suite.

2.4 Échéancier des travaux

| Devoir | Thème (sujet à modification) | Pondération | Date de remise |
|--------|---|-------------|----------------|
| 1 | Régression linéaire (ch. 3) | 10 % | 9 février |
| 2 | Classification linéaire (ch.4) | 10 % | 16 mars |
| 3 | Méthodes à noyau – classification non-linéaire (ch.5) | 10 % | 30 mars |
| 4 | Réseaux de neurones (ch.7) | 10 % | 16 avril |

Directives particulières

- Les devoirs doivent être effectués **par équipe de 2;**
- L'implémentation d'algorithmes dans le cadre des devoirs doit se faire dans le langage de programmation Python. Le code soumis doit être compatible avec (c'est-à-dire exécutable sous) la version de Python installée dans les laboratoires, sous Ubuntu;

2 <http://www.usherbrooke.ca/sciences/intranet/informations-academiques/reglement-devaluation/>

3 <http://www.usherbrooke.ca/programmes/references/reglement/>

- La qualité du français et de la présentation peut être considérée lors de l'évaluation des travaux;
- Toute soumission en retard entraîne une perte de 10 points par jour (incluant samedi et dimanche), sauf celles motivées par des raisons valables et conformes au règlement des études (par exemple, maladie avec attestation d'un médecin).

2.5 Utilisation d'appareils électroniques et du courriel

Selon le règlement complémentaire des études, section 4.2.3⁴, l'utilisation d'ordinateurs, de cellulaires ou de tablettes pendant une prestation est interdite à condition que leur usage soit explicitement permise dans le plan de cours.

Dans ce cours le règlement 4.2.3 s'applique à moins d'avoir obtenu personnellement l'autorisation du professeur. Cette permission peut être retirée en tout temps, si l'appareil n'est pas utilisé uniquement à des fins d'apprentissage.

Tel qu'indiqué dans le règlement universitaire des études, section 4.2.3⁵, toute utilisation d'appareils de captation de la voix ou de l'image exige la permission du professeur.

Note : Je ne réponds à aucun courriel.

3 Matériel pour le cours

- Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007.

Disponible en ligne sur Amazon, et en réserve à la bibliothèque et en format pdf sur le site web du cours.

4 <http://www.usherbrooke.ca/sciences/intranet/informations-academiques/reglement-devaluation/>

5 <http://www.usherbrooke.ca/programmes/references/reglement/>

L'intégrité intellectuelle passe, notamment, par la reconnaissance des sources utilisées. À l'Université de Sherbrooke, on y veille!

Extrait du Règlement des études

8.1.2 Relativement aux activités pédagogiques

L'expression **délit** désigne d'abord tout acte ou toute manœuvre visant à tromper quant au rendement scolaire ou quant à la réussite d'une exigence relative à une activité pédagogique.

Sans restreindre la portée générale de ce qui précède, est considéré comme un délit :

- a) la substitution de personnes ou l'usurpation d'identité lors d'une activité évaluée ou obligatoire;
- b) le **plagiat**, soit le fait, dans une activité évaluée, de faire passer indûment pour siens des passages ou des idées tirés de l'œuvre d'autrui;
- c) l'obtention par vol ou par toute autre manœuvre frauduleuse de document ou de matériel, la possession ou l'utilisation de tout matériel non autorisé avant ou pendant un examen ou un travail faisant l'objet d'une évaluation;
- d) le fait de fournir ou d'obtenir toute aide non autorisée, qu'elle soit collective ou individuelle, pour un examen ou un travail faisant l'objet d'une évaluation;
- e) le fait de soumettre, sans autorisation préalable, une même production comme travail à une deuxième activité pédagogique;
- f) la falsification d'un document aux fins d'obtenir une évaluation supérieure dans une activité ou pour l'admission à un programme.

Par plagiat, on entend notamment :

- Copier intégralement une phrase ou un passage d'un livre, d'un article de journal ou de revue, d'une page Web ou de tout autre document en omettant d'en mentionner la source ou de le mettre entre guillemets
- Reproduire des présentations, des dessins, des photographies, des graphiques, des données... sans en préciser la provenance et, dans certains cas, sans en avoir obtenu la permission de reproduire
- Utiliser, en tout ou en partie, du matériel sonore, graphique ou visuel, des pages Internet, du code de programme informatique ou des éléments de logiciel, des données ou résultats d'expérimentation ou toute autre information en provenance d'autrui en le faisant passer pour sien ou sans en citer les sources
- Résumer ou paraphraser l'idée d'un auteur sans en indiquer la source
- Traduire en partie ou en totalité un texte en omettant d'en mentionner la source ou de le mettre entre guillemets
- Utiliser le travail d'un autre et le présenter comme sien (et ce, même si cette personne a donné son accord)
- Acheter un travail sur le Web ou ailleurs et le faire passer pour sien
- Utiliser sans autorisation le même travail pour deux activités différentes (autoplagiat)

Autrement dit : mentionnez vos sources.
