

**Sigle : INF6243 Gr. 01**

**Titre : Techniques d'apprentissage**

**Session : Automne 2019 Horaire et local**

**Professeur : Allili, Mohand Saïd**

**1. Description du cours paraissant à l'annuaire :**

**Objectifs**

Permettre aux étudiants de maîtriser les concepts fondamentaux de l'apprentissage automatique et d'appliquer ces notions à des problèmes concrets. Leur faire acquérir des connaissances sur les techniques d'apprentissage supervisé et non supervisé, les techniques d'apprentissage pour les données textuelles, les algorithmes de classement des pages Web.

**Contenu**

Concepts d'apprentissage supervisé : classification et régression, frontière de décision et fonctions discriminantes; Arbres de décision et techniques de traitement du sur-apprentissage (overfitting); Apprentissage par ensemble : (bagging), (boosting) et forêt d'arbres; Machine à noyaux : dimension VC et machines à supports vectorielles; Apprentissage non supervisé : (clustering), les mélanges de loi de distribution statistique, carte de Kohonen et algorithme SOFM; Apprentissage de données multidimensionnelles : techniques de réduction de la dimension, classification non supervisée dans les sous-espaces de dimension (subspace clustering); Fouille de données textuelles : modèle TF-IDF et analyse sémantique latente; Prospection du Web : algorithmes HITS et PageRank.

Descriptif – Annuaire

**2. Objectifs spécifiques du cours :**

Au terme de cette activité, l'étudiant(e) aura acquis des connaissances sur :

- Les principes de l'apprentissage automatique;
- Les principes de l'apprentissage supervisé et non supervisé;
- Les méthodologies de classification et de validation;
- Les techniques de classification et de groupement de données;
- Les applications de l'apprentissage automatique.

**3. Stratégies pédagogiques :**

Les formules pédagogiques suivantes seront utilisées :

- Cours magistraux (3 heures par semaine)
- Présentation de la théorie et des exemples

**4. Heures de disponibilité ou modalités pour rendez-vous :**

**Bureau :** B-2022

**Téléphone :** 819-595-3900, poste 1601

**Courriel :** [Mohandsaid.allili@uqo.ca](mailto:Mohandsaid.allili@uqo.ca)

**5. Plan détaillé du cours sur 15 semaines :**

Semaine	Thèmes	Dates
1	<b>Introduction à l'apprentissage par ordinateur</b> <ul style="list-style-type: none"> <li>• Données massives (Big data) et apprentissage.</li> </ul>	06 sept. 2019

	<ul style="list-style-type: none"> <li>• Analyse de données et applications.</li> <li>• Rappel sur le calcul des probabilités et statistique.</li> </ul>	
2	<b>Concepts et types d'apprentissage</b> <ul style="list-style-type: none"> <li>• Concepts d'apprentissage supervisé.</li> <li>• Modèles linéaires pour la régression.</li> <li>• Modèles linéaires pour la classification.</li> <li>• Sur-apprentissage et validation d'apprentissage.</li> </ul>	13 sept. 2019
3	<b>Algorithmes de classification supervisée I</b> <ul style="list-style-type: none"> <li>• Frontières de décision et fonctions discriminantes.</li> <li>• Algorithme des K plus proches voisins (KPPV).</li> <li>• Classification par arbres de décision.</li> <li>• Classification naïve de Bayes.</li> </ul>	20 sept. 2019
4	<b>Algorithmes de classification supervisée II</b> <ul style="list-style-type: none"> <li>• Analyse discriminante linéaire.</li> <li>• Analyse discriminante non-linéaire.</li> <li>• Classification par régression logistique.</li> </ul>	27 sept. 2019
5	<b>Algorithmes de classification supervisée III</b> <ul style="list-style-type: none"> <li>• Réseaux de neurones.</li> <li>• Apprentissage profond.</li> <li>• Machine à vecteurs de support (MVS).</li> </ul>	04 oct. 2019
6	<b>Examen de mi-session</b>	11 oct. 2019
7	<b>Semaine d'études</b>	18 oct. 2019
8	<b>Algorithmes de regroupement de données</b> <ul style="list-style-type: none"> <li>• Algorithme des K-moyennes.</li> <li>• Modèle de mélange Gaussiens.</li> <li>• Regroupement hiérarchique de données.</li> <li>• Nombre optimal de groupes.</li> </ul>	25 oct. 2019
9	<b>Réseaux Bayésiens</b> <ul style="list-style-type: none"> <li>• Dépendances entre variables aléatoires.</li> <li>• Modèles graphiques orientés et non-orientés.</li> <li>• Estimation par maximum de vraisemblance.</li> <li>• Exemples d'applications des Réseaux Bayésiens.</li> </ul>	01 nov. 2019
10	<b>Techniques de réduction de dimensions</b> <ul style="list-style-type: none"> <li>• Concept de réduction de dimensions.</li> <li>• Analyse en composantes principales.</li> <li>• Techniques de sélection de caractéristiques.</li> <li>• Classification dans les sous-espaces de dimensions.</li> <li>• Les auto-encodeurs.</li> </ul>	08 nov. 2019
11	<b>Analyse et classification de données textuelles et Web</b>	15 nov. 2019

	<ul style="list-style-type: none"> <li>• Analyse de documents textuels.</li> <li>• Modèle TF-IDF et classification de documents textuels.</li> <li>• Modèle sac-à-mots et N-grams.</li> <li>• Modèle Word2Vec</li> <li>• Recherche et organisation de documents textuels.</li> </ul>	
12	<b>Analyse et classification de données visuelles</b> <ul style="list-style-type: none"> <li>• Description des images et vidéos.</li> <li>• Extraction de caractéristiques visuelles.</li> <li>• Classification et reconnaissance de formes.</li> <li>• Réseaux de neurones convolutionnels (CNN).</li> </ul>	22 nov. 2019
13	<b>Présentation des projets</b>	29 nov. 2019
14	<b>Présentation des projets</b>	06 déc. 2019
15	<b>Examen final</b>	13 déc. 2019

## 6. Évaluation du cours :

L'étudiant(e) dans ce cours sera évalué(e) par les examens de mi-session et final, ainsi que par des projets de session. La pondération de la note finale se fera comme suit :

- **Devoirs : 15 %**
- **Projets : 25 %**
- **Examen de mi-session : 30 %**
- **Examen final : 30 %**

Pour les projets, l'évaluation sera répartie comme suit :

- **Devoirs : 15 points**

Il s'agit de tester sur la plateforme Matlab des applications basées sur des techniques d'apprentissage. La plateforme Matlab est disponible dans les laboratoires de l'UQO ou téléchargée sur le Web.

- **Projets : 25 points**

Il s'agit de tester/développer des projets sur l'apprentissage par ordinateur. Ces derniers peuvent être dans l'analyse de données multimédias : texte, son, image ou Web. Quelques projets seront fournis par le professeur. Néanmoins, les étudiants auront la liberté de choisir des projets dans des domaines d'applications qu'ils souhaitent.

L'évaluation se fera sur deux volets :

- Un rapport entre 10 et 15 pages. Les normes de présentation de travaux (p. ex. page de garde, marge d'un pouce, interligne à 1,5, taille des caractères de 12 points) doivent être **absolument** respectées.

- Une présentation de 15 à 20 minutes sera faite en classe. Les projets se feront individuellement ou en équipes de deux si le nombre d'étudiants inscrits dépasse 10.

Une moyenne générale inférieure à **64 %** est éliminatoire et conduit automatiquement à l'échec de l'étudiant(e).

Tout retard dans la remise d'un travail entraîne une pénalité de **15 %/jour** sur la note attribuée à ce travail, **jusqu'au maximum d'une semaine**. La qualité du français sera considérée lors de la correction des travaux.

## 7. Politiques départementales et institutionnelles :

- Politique du département d'informatique et d'ingénierie relative à la tenue des examens
- Note sur le plagiat et sur la fraude
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQO
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

• **SANS OUI  
C'EST NON!**

Travaillons ensemble pour développer une culture du respect ! La communauté universitaire de l'UQO se mobilise et lance un message haut et fort de **tolérance zéro en matière de violence à caractère sexuel** (pour de plus amples renseignements, veuillez consulter la page Web : [uqo.ca/sansouicestnon](http://uqo.ca/sansouicestnon)).

## 8. Principales références :

1. S. Rogers et M Girolami. A first Course in Machine Learning, CRC press, 2012.
2. C. Bishop. Pattern Recognition and Machine Learning. Springer 2006.
3. R. Duda, P. Storck et D. Hart. Pattern Classification. Prentice Hall, 2002.
4. T. Hastie and R. Tibshirani and J. Friedman. The Elements of Statistical Learning Theory. Springer, 2009.
5. C.D. Manning, P. Raghavan et H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.
6. S. Ceri, A. Bozzon, M. Brambilla, E. Della Valle, P. Fraternali et S. Quarteroni. Web Information Retrieval, Springer, 2013.
7. E. Alpayedin. Introduction to Machine Learning. MIT Press, 2<sup>nd</sup> Edition, 2010.

<http://ciml.info/>

[http://www.sciencemag.org/site/feature/data/compsci/machine\\_learning.xhtml](http://www.sciencemag.org/site/feature/data/compsci/machine_learning.xhtml)

<http://archive.ics.uci.edu/ml/>

<https://work.caltech.edu/telecourse.html>

<http://mlg.eng.cam.ac.uk/>

## 9. Page Web du cours :

<https://moodle.uqo.ca>