ESG UQÀM
Département des sciences économiques
École des sciences de la gestion
Université du Québec à Montréal

Syllabus	X
TP	
Autre	

Sigle	*Groupe*	*Trimestre*
ECO930J		**Hiver 2025**
Titre	**Données massives et apprentissage automatique avec applications en économie**	
Enseignant	Philippe Goulet Coulombe, DS-5975, goulet_coulombe.philippe@uqam.ca Disponibilité sur rendez-vous	
Horaire	Jeudi, 9h30-12h30, DS-5650	

Ce cours vise à familiariser les étudiants avec les méthodes d'apprentissage automatique (on dira ML pour Machine Learning) qui sont maintenant utilisées par les firmes qui ont accès au « Big Data » pour prédire les comportements d'achat des consommateurs. On pense à Google, Amazon, Facebook. Elles sont aussi utilisées par les compagnies d'assurance pour développer des indices de risque pour les individus et par les pharmaceutiques pour prédire certaines maladies, en particulier le plus rares. Il y aussi des applications en finance et en prévision macroéconomique. Puisque ces méthodes, dans bien des cas, sont fondamentalement des méthodes de régression pénalisée, elles tombent naturellement dans le champ de l'économétrie.

Dans un premier temps, les grands principes de ces méthodes seront présentés, en particulier les techniques de validation. Puis les méthodes les plus courantes seront présentées en classe, avec exemple numérique et code informatique. On peut classifier ces méthodes en deux grands types, régression (LASSO, Deep Learning), ou par construction d'arbre (Bagging, Random Forest, Boosting). Différentes méthodes de réduction de dimension seront également considérées. Une des particularités de ces méthodes est le choix d'hyperparamètres (paramètres fixés avant la régression).

Il est pris pour acquis que les étudiant(e)s sont familiers avec les concepts de base d'économétrie. Vous êtes responsable de revoir et (ré)apprendre ces concepts.

L'évaluation se fera sur la base un travail pratique en équipe, des présentations, et un travail personnel. Tout retard implique la note 0 pour le travail concerné.

Les absences pour raison de maladie devront être justifiées par un billet du médecin stipulant explicitement qu'en raison de votre maladie vous n'avez pas pu vous présenter ou accomplir la tâche demandée.

Tous les travaux pratiques devront être exécutés à l'aide d'un logiciel économétrique tel que **R ou Python**.

Sommaire

Voici le plan de match provisoire. Les changements peuvent se produire durant la session. Dépendant des intérêts des étudiants et de leur provenance (maitrise ou doctorat), nous pourrons modifier certaines activités/matières du cours. La majorité des applications seront macroéconomiques, donc sur les séries temporelles, mais les étudiants pourront aussi explorer les méthodes avec les données microéconomiques.

Les thèmes suivants pourraient être couverts dans l'ensemble du cours :

1. Apprentissage Statistique (JWHT chap 2)
2. Régression linéaire et apprentissage (révision rapide) (JWHT chap 3)
3. Économétrie des séries temporelles
4. Théorie de la prévision
5. (Probit Revisité) Variables dépendantes qualitatives (JWHT chap 4)
6. Bootstrap et validation croisée (JWHT chap 5)
7. Méthodes et choix de modèles linéaires et le concept de régularisation (JWHT chap 6)
    a. Ridge
    b. Lasso
    c. Elastic-net
8. Modèle à facteurs et analyse par composantes principales (JWHT chapitres 6 et 10, HTF chapitre 14)
9. Modèles non-linéaires (JWHT chap 7)
10. Méthodes d'arbres (JWHT chap 8)
    a. Pruning
    b. Bagging
    c. Random Forest
11. Boosting (HTF, chapitre 10)
12. Deep Learning (HTF, chapitre 11)
13. (Si le temos le permet) Support vector machine regression (JWHT chapitre 9, HTF chapitre 12)
14. Algorithmes ML pour applications économiques (effets causaux hétérogènes, modèles macroéconométrique à base de Random Forest ou Deep Learning)

Sources de données macroéconomiques et financières :

- Federal Reserve Economic Data (données américaines)
- Statistique Canada et Institut de la statistique du Québec (données canadiennes et québécoises)
- EuroStat (données européennes)
- OECD (données mondiales)
- Site Web de Kenneth French (données financières américaines et canadiennes)
- Real-Time Data Set for Macroeconomists (données historiques américaines en temps réel)

Notes : Ce sont seulement quelques sources principales.

Modes d'évaluation

1. Travail personnel de session. 40 points (un rapport +une présentation de 20 minutes en cours pour 15 points sur 40)
2. Travail pratique en équipe de 2 (simulations et empirique). 30 points.

3. Critique d'article : présentation. 15 points. Individuel.
4. Participation. 15 points.

Toutes les dates de remises des travaux et présentations seront décidées en classe. Un travail remis en retard obtient une pénalité de 40 % pour le premier jour (applicable dans son entièreté dès l'heure de tombée) et 100% après.

Références:

- James, Witten, Hastie, Tibshirani (JWHT), *An Introduction to Statistical Learning*, Springer texts, http://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf

- Hastie, Tibshirani, Friedman, (HTF), *The Elements of Statistical Learning*, Springer Series in Statistics, https://web.stanford.edu/~hastie/Papers/ESLII.pdf
- Athey, S. and G. Imbens (2019) "Machine Learning Methods Economists Should Know About", working paper, arXiv:1903.10075.
- Aiello, Eckstrand, Fu, Landry, Lanford, Aboyoun, *Machine Learning with R and H20*¸ H20 booklet, http://h2o-release.s3.amazonaws.com/h2o/master/3233/docs-website/h2o-docs/booklets/R_Vignette.pdf

- Nykodym, Rao, Wang, Krajevic, Lanford, Hussami, *Generalized Linear Modeling with H20*¸ H20 Booklet, https://h2o-release.s3.amazonaws.com/h2o/rel-slater/9/docs-website/h2o-docs/booklets/GLM_Vignette.pdf

- Malohlava, Candel, *Gradient Boosting Machine with H20*, H20 Booklet, http://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/GBMBooklet.pdf

- Candel, Landford, LeDell, Parmar, Arora, *Deep Learning with H20*, H20 Booklet, https://h2o-release.s3.amazonaws.com/h2o/rel-slater/9/docs-website/h2o-docs/booklets/DeepLearning_Vignette.pdf

- Bengio, Yoshua (2012), "Practical recommendations for gradient-based training of deep architectures", https://arxiv.org/abs/1206.5533
- Effets de traitement hétérogènes:
    - Athey, S., and G. Imbens (2016) "Recursive Partitioning for Heterogeneous Causal Effects" , Proceedings of the National Academy of Science 113 (27) 7353-7360
    - Athey, S., and S. Wager (2018) "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests", Journal of the American Statistical Association, 113 (523), 1228-1242.
    - Athey, S., J. Tibshirani and S. Wager (2019) "Generalized Random Forests," Annals of Statistics, 47 (2), 1148-1178, 2019. http://arxiv.org/abs/1610.01271
    - Athey, S., and S. Wager (2019) "Estimating Treatment Effects with Causal Forests: An Application". Forthcoming, Observational Studies. https://arxiv.org/abs/1902.07409

- Références additionnelles pour Support Vector Machine Regression:

- http://kernelsvm.tripod.com
- Smola, A. et B. Scholkopf (2004), « A tutorial on support vector regression », *Statistics and Computing* 14, 199-222.

- Références additionnelles pour Dynamic Factor Model:
  - Stock, J.H. et M.H. Watson (2016), "Factor Models and Structural Vector Autoregressions in Macroeconomics" in *Handbook of Macroeconomics,* Vol2A, John B. Taylor and Harald Uhlig (eds), 2016, Chapter 8, pp 415-526
  - Bai, J. et Ng, S. (2008), "Forecasting economic time series using targeted predictors", *Journal of Econometrics* 146, 304–317
  - Bai, J. et Ng, S. (2008), "Boosting Diffusion Indices", *Journal of Applied Econometrics*, 24:4, 607-629
- Références additionnelles pour la prévision:
  - Leroux, M., Kotchoni, R. et D. Stevanovic (2017), « Macroeconomic forecasts accuracy in data-rich environment ».
  - Diebold, F.X., *Forecasting*, http://www.ssc.upenn.edu/~fdiebold/Textbooks.html
  - Stock, J.H. et M.H. Watson (2017) "Twenty Years of Time Series Econometrics in Ten Pictures" *Journal of Economic Perspectives*, Vol. 31 No. 2.
  - https://arxiv.org/abs/2008.12477 How is ML useful for Macro Forecasting? P. Goulet Coulombe, Leroux, M., S. Suprenant, et D. Stevanovic
  - https://arxiv.org/abs/2008.01714 Macro data transformations matter. P. Goulet Coulombe, Leroux, M., S. Suprenant, et D. Stevanovic
  - https://arxiv.org/abs/2103.01201 Can Machine Learning Catch the COVID-19 Recession? Philippe Goulet Coulombe, Massimiliano Marcellino, Dalibor Stevanovic

Des notes de cours additionnelles seront envoyées par le courriel.

Étudiants ayant une déficience de type visuelle, auditive, motrice, trouble d'apprentissage, trouble envahissant du développement et trouble de santé mentale
Les étudiants qui ont une lettre *d'Attestation des mesures d'aménagements académiques* obtenue auprès d'un conseiller de **l'Accueil et soutien aux étudiants en situation de handicap (ASESH)** doivent rencontrer leur professeur au début de la session afin que des mesures d'aménagements en classe ou lors des évaluations puissent être mises en place. Ceux qui ont une déficience ou une incapacité mais qui n'ont pas cette lettre doivent contacter l'**ASESH** au 514-987-3148 ou se présenter au local AB-2300 le plus tôt possible.

Page d'information

PLAGIAT - Règlement no 18 sur les infractions de nature académique

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes ;
- l'utilisation totale ou partielle du texte d'autrui en le faisant passer pour sien ou sans indication de référence ;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant ;
- l'obtention par vol, manœuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée ;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé ;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne ;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle ;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances ;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-création, un rapport de stage ou un rapport de recherche.

Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18

Pour plus d'information sur les infractions académiques et comment les prévenir : www.integrite.uqam.ca

Politique 16 sur le harcèlement sexuel

Le harcèlement sexuel se définit comme étant un comportement à connotation sexuelle unilatéral et non désiré ayant pour effet de compromettre le droit à des conditions de travail et d'études justes et raisonnables ou le droit à la dignité.

La Politique 16 identifie les comportements suivants comme du harcèlement sexuel :

1. Manifestations persistantes ou abusives d'un intérêt sexuel non désirées.
2. Remarques, commentaires, allusions, plaisanteries ou insultes persistants à caractère sexuel portant atteinte à un environnement propice au travail ou à l'étude.
3. Avances verbales ou propositions insistantes à caractère sexuel non désirées.
4. Avances physiques, attouchements, frôlements, pincements, baisers non désirés.
5. Promesses de récompense ou menaces de représailles, implicites ou explicites, représailles liées à l'acceptation ou au refus d'une demande d'ordre sexuel.
6. Actes de voyeurisme ou d'exhibitionnisme.
7. Manifestations de violence physique à caractère sexuel ou imposition d'une intimité sexuelle non voulue.
8. Toute autre manifestation à caractère sexuel offensante ou non désirée.

Pour plus d'information :
http://www.instances.uqam.ca/ReglementsPolitiquesDocuments/Documents/Politique_no_16.pdf
Pour rencontrer une personne ou faire un signalement : Bureau d'intervention et de prévention en matière de harcèlement : 514-987-3000, poste 0886
http://www.harcelement.uqam.ca