Description du cours

Objectifs

Dans ce cours, nous visons à introduire des méthodes qui permettront aux étudiantes et étudiants d'étudier un jeu de données de "haute dimension" (ici, "haute" est pris dans le sens où l'on ne peut pas faire un simple graphique de l'ensemble des variables pour toutes les observations) sans avoir recours à un modèle probabiliste. Les techniques que l'on y enseigne servent à réduire la dimension des données, identifier certains liens entre les variables, visualiser les données ou à diviser le jeu de données en groupes/classes.

Sans négliger la théorie, l'accent sera mis sur l'aspect pratique de l'analyse des données et l'utilisation d'un langage de programmation, que ce soit R, Python ou autre.

Place du cours dans le programme

Ce cours est généralement suivi par des étudiants du baccalauréat en statistique lors de leur deuxième année de formation. C'est aussi un cours à option en actuariat, en mathématique ainsi que dans certains programmes de génie et d'administration.

Tous les étudiants doivent s'assurer d'avoir suivi au minimum un cours d'algèbre (ex. MAT-1200) et un cours de statistique de base (ex. STT-1000), car la majorité des méthodes d'analyse de données s'appuient sur ces notions.

Les étudiants devraient aussi avoir une certaine familiarité avec les concepts algorithmiques de base, ainsi qu'avec (au moins) un langage de programmation.

Objectifs spécifiques

À la fin du cours, l'étudiant ou l'étudiante devrait être capable :

- de comprendre et décrire succinctement les fondements théoriques des méthodes d'analyse de données étudiées ;
- d'identifier correctement les situations où l'emploi de ces méthodes est indiqué ;

— d'utiliser efficacement un langage de programmation pour mettre en oeuvre ces méthodes ;
— d'analyser et d'interpréter judicieusement les résultats découlant de l'analyse ;
— de formuler par écrit les conclusions de l'analyse, dans le respect des limites de la méthodologie.

Équipement informatique

Vous pourriez avoir besoin d'un ordinateur, de haut-parleurs ou d'un casque d'écoute, d'un microphone, d'une webcam et d'une connexion Internet avec fil à large bande ou sans fil. Pour vérifier les paramètres de configuration minimaux selon le système d'exploitation, nous vous invitons à visiter cette page.

De plus, ce cours peut nécessiter des besoins logiciels particuliers qui seront alors décrits dans d'autres sections du plan de cours, le cas échéant.

Fonctionnement

Les cours et examens auront lieu en présentiel les mardi et vendredi matin, mais le matériel sera disponible en ligne. De façon générale, la séance du vendredi sera une séance magistrale et la séance du mardi sera dédiée à des exercices pratiques réalisés de façon autonome par les étudiants avec le soutien de l'enseignant.

Approches pédagogiques

L'approche pédagogique privilégiée est l'exposé interactif (vendredi) en alternance avec des périodes de laboratoires (mardi). L'approche pédagogique est très axée sur l'apprentissage actif et exige un engagement soutenu des étudiants tout au long de la session.