



POLYTECHNIQUE
MONTREAL

LE GÉNIE
EN PREMIÈRE CLASSE

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

PLAN DE COURS - MTH6312 (3,0,6)

MÉTHODES STATISTIQUES D'APPRENTISSAGE

AUTOMNE 2016

Site Web du cours (moodle) : <http://moodle.polymtl.ca>

Professeur : Luc Adjengue (luc.adjengue@polymtl.ca)

Bureau : A-520.33

Cours : jeudi 13h45, 14h45, 15h45 Local C-539.6

1. BUT DU COURS

Le cours MTH6312 vise à communiquer une bonne connaissance des récentes méthodes statistiques d'apprentissage. Les principes de base des méthodes d'apprentissage supervisé et non supervisé y sont solidement établis et une importance est accordée aux contraintes rencontrées avec les données en pratique.

2. DESCRIPTION

Le cours traite des sujets suivants : introduction aux méthodes statistiques à l'aide d'un logiciel de traitement de données massives ; apprentissage supervisé, semi-supervisé et non supervisé ; régression linéaire ; régression logistique ; critères de sélection de variables d'un modèle ; validation croisée ; régression pénalisée ; régression non paramétrique ; méthodes de classification ; mélange de Gaussiennes et algorithme espérance-maximisation (EM) ; apprentissage automatique ; modèles additifs ; arbres de régression ; apprentissage bayésien.

3. OBJECTIFS GÉNÉRAUX

- Utiliser le logiciel R pour analyser des données massives ;
- Utiliser la méthode d'apprentissage appropriée de prévision, de classification, et de clustering, pour chaque ensemble de données ;
- Estimer les paramètres et choisir le meilleur modèle ;
- Combiner différentes méthodes d'apprentissage pour une meilleure performance ;
- Appliquer les méthodes d'apprentissage d'ensemble pour augmenter la précision ;
- Rédiger des fonctions en R pour de nouvelles méthodes d'apprentissage, et combiner celles-ci avec des langages de bas niveau (tels que C++).

4. PROGRAMME DU COURS

- Introduction : définitions et rappels sur la statistique (échantillons, estimation du maximum de vraisemblance, test du maximum de vraisemblance).
- Régression linéaire
- Régression logistique et analyse discriminante
- Classification (classificateurs bayésiens, linéaires, quadratiques, Knn)
- Méthodes de rééchantillonnage : validation croisée et bootstrap
- Sélection de modèles et régularisation
- Régression par spline adaptative
- Modèles additifs et arbres de régression
- Machines à vecteurs de support
- Apprentissage non supervisé :
 - composantes principales
 - clustering (K-means ; regroupement hiérarchique)

5. RÉFÉRENCES

- P. Dalgaard, *Introductory Statistics with R*, Statistics and Computing Series, Springer, Second Edition, 2008.
- G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer Texts in Statistics, 2013.
- T. Hastie, R. Tibshirani, J. Friedman *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, Springer Series in Statistics, Second Edition, 2011.
- K.P. Murphy, *Machine Learning : A Probabilistic Perspective*, MIT Press, Cambridge Massachussets, 2012.

6. ÉVALUATION

Type	Nombre	Pondération
Devoirs	5	50 %
Projet de session	1	15 %
Examen final	1	35 %