

Fondements du Machine Learning

En bref

➤ **Langue de cours:** Anglais

Présentation

Objectifs d'apprentissage

La multiplication de capteurs et de systèmes de collectes et d'enregistrements de données est telle que la quantité d'informations accessible aux utilisateurs est trop importante pour pouvoir être traitée sans l'aide d'outils informatiques et d'algorithmes performants. Parmi les algorithmes indispensables au traitement de données, ceux de classification sont largement utilisés, soit pour agréger des données au sein de groupes cohérents (clustering ou classification non supervisée), soit pour décider de l'affectation automatique de nouvelles données dans des groupes déjà constitués (classification supervisée).

Les objectifs d'apprentissage de cette UE sont de :

- connaître et savoir mettre en œuvre les principaux algorithmes de classification supervisée et non supervisée,
- connaître les conditions de mise en œuvre de ces algorithmes et le prérequis d'éventuels prétraitements des données,
- savoir évaluer quantitativement la qualité de ces algorithmes.
- Filtrage et prédiction de séquences temporelles.

Description du programme

- *Classification supervisée* : gestion des données (création d'ensembles d'entraînement, de validation et de test). Métriques en classification supervisée (rappel, précision, courbes ROC et aire sous la courbe, matrices de confusion). Détails des principaux algorithmes de classification supervisée : K-plus proches voisins (KNN : k-nearest neighbors), séparateurs à vaste marge (SVM linéaire et à noyaux), arbres et forêts aléatoires, réseaux de neurones.
- *Classification non supervisée* : prétraitement des données (réduction de dimensions). Détails des méthodes de classifications hiérarchiques ascendantes (étude des critères de dissimilarité), des k-moyennes (k-means) et des mélanges de gaussiennes (algorithme EM).
- *Séries temporelles* : Modèles statistiques, Modèles Autorégressifs (statsmodels, ARIMA, ARIMAX, SARIMA...)

Implémentation et manipulation de ces méthodes à l'aide de la bibliothèque *sklearn* de python.

Compétences et connaissances scientifiques et techniques visées dans la discipline

- prétraitement de données
- choix et évaluation d'un algorithmes de classification
- présentation des résultats de classification (présentation de métriques et/ou représentation graphique)
- bonne connaissance de la librairie sklearn

Modalité de contrôle des connaissances

Mises en œuvre sur des données réelles (données *iris*, *moon*, *mnist*, *telecom churn* et les données de prédiction de maladies cardiovasculaires) ou simulées (données (non)linéairement séparables, données issues de lois (multi)-normales) des algorithmes de classification et analyse critique des résultats.

Bibliographie

- *Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow*, 2nd edition, Aurélien Géron, O' Reilly Media, 2019, 600 pp., ISBN: 978-1-492-03264-9
- Vapnik, V. Statistical Learning Theory. Wiley-Interscience, New York, (1998)
- Dinov, ID. "[Expectation Maximization and Mixture Modeling Tutorial](#)". [California Digital Library](#), Statistics Online Computational Resource, Paper EM_MM,

Equipe pédagogique

- Valeriya STRIZHKOVA

Total des heures		22h
CM	Cours Magistral	15h
TD	Travaux Dirigés	7h