

MA431 : Mathématiques appliquées à la sécurité

Présentation du cours

D. Barcelo

Grenoble INP ESISAR

2022/2023

- 1 Modalités
- 2 Applications
- 3 Programme

Cours et intervenants

Intervenants : Didier Barcelo et Richard Jarry

- 9 CM (D. Barcelo)
- 3 TDM de 1h30 : D. Barcelo en R
- 2 TDM de 1h30 : R. Jarry en Python
- 2 TP de 3h (R. Jarry) en Python

Utilisations des statistiques

Besoin de notions de statistiques pour :

- trier les données, savoir y repérer des liens et des aberrations,
- les modèles et techniques de data science,
- la validation des modèles choisis.

Certaines notions seront abordées dans le cours, d'autres (tests) seront à revoir en amont des cours.

Pour se rafraichir la mémoire sur les tests statistiques, consultez les vidéos (liens sur Chamilo) et effectuez le QCM (sur Chamilo). A faire avant le CM2.

Applications de la data science

Data Science : Science des données, objectif d'extraction d'informations ou de modèle à partir de grandes bases de données.

Applications à la sécurité ?

- Détection d'anomalies,
- Analyse de liens,
- Classification,
- Prédiction.

Les algorithmes étudiés pourront être utilisés dans d'autres cadres que la sécurité (maintenance préventive par exemple).

Applications de la data science

Data Science appliquée :

- **au test d'intrusion**

- techniques qui aident à automatiser le process de détection d'intrusion,
- techniques de classification qui permettent de réduire les erreurs dans l'algorithme de détection d'intrusion.

- **à la détection d'anomalie**

- bâtir un modèle (profil) de données saines,
- détecter les déviations du modèle.

- **analyse de firewall**

- **Traffic mining**

- **Worm détection**

Programme de cours

- Les données :
 - Analyse de données et réduction de la dimension : ACP.
 - Exploration d'une base de données,
 - Données manquantes, etc.,
 - Construction de bases tests et de bases d'apprentissage
 - Les indicateurs.
- Les régressions :
 - Regression linéaire et modèle linéaire,
 - Regression logistique.
- Apprentissage supervisé :
 - Principe de l'apprentissage supervisé,
 - Techniques de classification : k-NN, SVM, voire d'autres modèles.
- Apprentissage non supervisé :
 - Principe de l'apprentissage non supervisé,
 - Techniques de clustering : k-means, CAH, Dbscan, voire d'autres modèles.
- Bilan et extension

Programme de TDM/TP

- TP1 : Analyse exploratoire de données (RJ)
- TDM1 : Regressions (DB)
- TDM2 : Classification (DB)
- TDM3 : Lancement du DM (RJ)
- TDM4 : Clustering (DB)
- TDM5 : Clustering (RJ)
- TP2 : TP final (RJ)

Remarques

- Méthodes statistiques faisant appel à l'algèbre linéaire, utilisation de la statistique inférentielle pour la validation des modèles,
- Techniques utilisées en machine learning et Big data
- **Machine learning \subset Data Mining \subset Statistiques**
- Vous ne deviendrez pas des experts ...
- Module "bref" de **sensibilisation et présentation des possibilités**, les méthodes ne sont pas étudiées de manière approfondie
- Notions importantes si vous êtes intéressés par le **machine learning** ou l'IA, mais il y en a beaucoup d'autres, ceci sera complété par un cours en 5A (machine learning)
- Illustration des applications en sécurité (si possible).

Evaluations

- TDM avec D. Barcelo évalué, travail sous R,
- un DM sera donné par R. Jarry, travail sous Python,
- DS : de petits exercices, voire QCM, et du cours. Apprenez votre cours, des exos seront traités en CM et en TDM.

Modalités d'évaluation prévues : TDM 1/3 DM 1/3 examen 1/3