

Algorithmes d'apprentissage non supervisé

Mercredi 1^{er} septembre 2021

420-A58-SF — Algorithmes d'apprentissage non supervisé — Été 2021
Spécialisation technique en intelligence artificielle — M. Swawola, M.Sc.

**NOUS ÉCLAIRONS.
VOUS BRILLEZ.**

FORMATION CONTINUE
ET SERVICES AUX ENTREPRISES



Bienvenue à tous !

Agenda cours #1 - mercredi 1^{er} septembre 2021

- Mot de bienvenue
- 00 - Introduction au cours
- 01-01 - Introduction au partitionnement de données
- 01-02 - Partitionnement en K-moyennes (K-means clustering)
- Ateliers 01-02-A1 et 01-02-A2
- Description du projet

Description du projet

Projet

Démonstration d'une méthode de partitionnement de données appliquée à une problématique choisie par l'étudiant

- Pondération: 30%
- Date: rendre le projet au plus tard sur Lea le dimanche 19 septembre 2021 à 23h59
Aucun retard ne sera accepté, sauf cas de force majeure. Tout travail envoyé après ce délai sera refusé
- Format: capsule vidéo diffusée lors de la séance du lundi 20 septembre 2021 à 18h00 (séance non enregistrée)

Réalisation

- Présentation de 10 minutes sous forme d'une capsule vidéo pré-enregistrée
- Choix de la problématique libre
- L'étudiant devra:
 - motiver le choix de la problématique et expliquer l'intérêt d'appliquer le partitionnement de données à celle-ci
 - lister les étapes de préparation des données
 - justifier la méthode de partitionnement utilisée
 - expliquer la mise en oeuvre de la méthode de partitionnement
 - commenter et analyser les résultats obtenus
 - fournir le code documenté de la démonstration

Réalisation

- Concernant le choix de la problématique:
 - Celle-ci doit être originale (données collectées et préparées par l'étudiant)
 - Les données simulées sont permises, seulement si celles-ci montrent un réel degré de réalisme par rapport à la problématique
 - Ne doit pas être inspirée d'un blog, d'un article, ou d'une quelconque source sur Internet
 - Le code de la démonstration doit être intégralement écrit par l'étudiant
- Une attention toute particulière sera portée au plagiat de sources externes au cours
- La méthode de partitionnement utilisée est libre et peut correspondre à des algorithmes non vus en cours (exemple: CURE, FCM, OPTICS, etc...)

Critères d'évaluation

- Respect du sujet
- Motivation argumentée du choix de la problématique
- Choix juste et argumenté des algorithmes sélectionnés
- Mise en oeuvre des méthodes de partitionnement
- Qualité de l'analyse des résultats
- Fonctionnalité de la démonstration
- Qualité du code documenté et de la vidéo
- Respect du temps de 10 minutes (+/- 2 minutes)
Une pénalité de 10% par minute manquante ou excédentaire sera appliquée

Support

- L'entraide est permise sur le salon de clavardage du cours uniquement
- Il s'agit d'un travail individuel
- Un tutorat de 15 minutes sera accordé à chaque étudiant en faisant la demande pour répondre à des questions bien précises. Les questions devront démontrer une démarche de travail de la part de l'étudiant et la demande devra être faite par message Mio du 6 au 13 septembre

Exemple de questions ...

■ ... qui resteront sans réponse !

- « Est-ce que l'algo K-moyennes fonctionne sur mon jeu de données ? »
- « Je ne sais pas quoi choisir comme problématique, peux-tu trouver à ma place ? »
- « J'ai rien compris, peux-tu me réexpliquer le cours ? »
- « J'obtiens ces trois clusters, t'en penses quoi ? »
- « Ça marche pas », « Ça plante », « Je peux pas accéder à Jupyter », etc ...

■ ... qui seront répondues avec plaisir !

- “J'applique l'algorithme DBSCAN sur mon jeu de données, mais les clusters obtenus semblent incohérents avec les résultats attendus. Je suspecte la présence de clusters de densité différentes. L'ACP peut-elle aider à effectuer cette vérification ?”

Conseils

- Commencez le projet dès demain ! Ça va aller vite 😊
- Choisissez une problématique simple pour laquelle les résultats attendus sont prévisibles
- Préférez les jeux de données ouverts et inspirez-vous des problématiques de votre profession. Simulez vos données au besoin
- Évitez les données hautement complexes et les thèmes de recherche actuels. Il s'agit d'un projet de moins de 18 jours, pas d'une thèse de doctorat ;-)

Suggestion d'organisation

- **Ce découpage du travail est donné à titre indicatif**
 - **Cette semaine:** choix de la problématique et collecte / simulation des données
 - **Semaine du 6 septembre:** préparation des données et sélection de la méthode de partitionnement
 - **Semaine du 13 septembre:** validation des résultat et montage de la capsule vidéo

Réalisation de la capsule vidéo

- La capsule vidéo devra être au format MP4. Tout autre format sera refusé
- Celle-ci peut-être enregistrée à l'aide du logiciel [OBS Studio](#)
- La taille de la vidéo doit être approximativement 50 Mo. Au besoin compresser la vidéo à l'aide de l'outil [FFmpeg](#)
- Utilisez un logiciel pour le montage (ex. [Kdenlive](#))
- Soyez attentif à la résolution de la vidéo et à la qualité audio
 - **Résolution:** 720p ou 1080p
 - **Audio:** votre voix doit être claire et enregistrée avec un volume suffisant pour être entendue à mi-volume

Des questions ?

Bon succès à tous !