
Laboratoire 3*

Introduction à l'apprentissage

Objectifs

L'objectif de ce laboratoire est de familiariser l'étudiant avec un système d'apprentissage de son choix. À la fin du laboratoire, il aura démontré qu'il peut :

- Rechercher des informations sur un algorithme connu;
- Développer un système d'apprentissage;
- Présenter un compte rendu de sa recherche.

Description

La première partie du laboratoire consiste à faire le choix de l'algorithme. Cet algorithme peut provenir de la liste proposée dans la section Liste d'algorithmes ou provenir d'une autre source telle qu'un livre, les diapositives du cours, l'Internet ou autre. Vous devez trouver suffisamment de documentation pour être capable de faire l'implémentation de l'algorithme choisi. De plus, vous devez être en mesure d'expliquer théoriquement cet algorithme.

Liste d'algorithmes

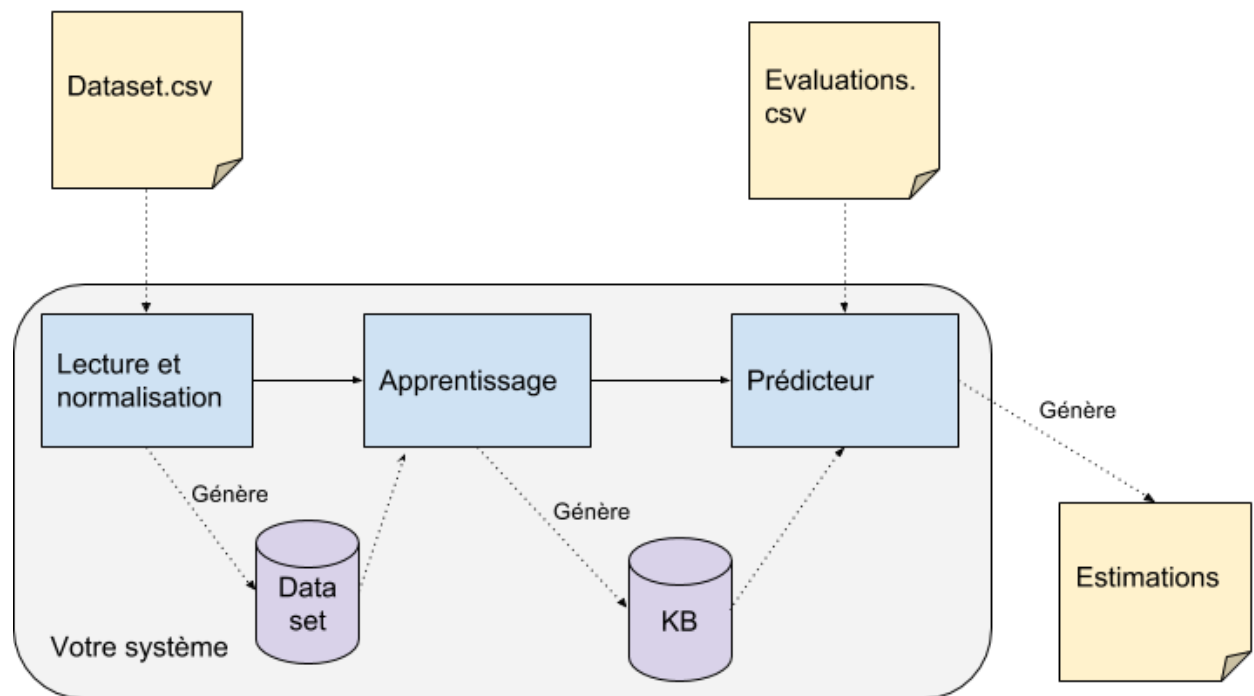
- Algorithme glouton[†]
- Algorithme génétique[†]
- Arbre de décision avec valeurs numériques
- Closest match
- Colonie de fourmis[†]
- Distribution gaussienne
- K plus proches voisins
- Méthode bayésienne naïve
- Recherche taboue[†]
- Recuit simulé[†]
- Réseau de neurones
- PCA
- Optimisation par essaims particuliers[†]

* Auteurs : Richard Rail et Lévis Thériault; Mise à jour : 2018-03-15

† Métaheuristique

En deuxième partie, vous devez implémenter votre système d'apprentissage. Celui-ci doit prendre en entrée un fichier csv (Dataset.csv) de données comme source d'entraînement. Il peut par la suite créer un fichier représentant sa base de connaissance ou simplement la garder en mémoire. Votre système peut implémenter plusieurs éléments différents tel que la réduction d'attributs, la normalisation des données, des métaheuristiques, etc. La section **Évaluation** de l'énoncé présente les points attribués aux différents éléments que vous pouvez ajouter à votre système.

En troisième partie, votre système doit être capable de prendre une série de données d'entrées et de prédire la sortie. Votre estimation de la sortie doit être sensiblement réaliste. Cette sortie peut être simplement affichée dans la console. Le fichier Evaluations.csv est similaire à celui du Dataset mais il manque la colonne à évaluer. C'est le but final du travail, donner des valeurs que votre système estime être les plus probables.



Mise en contexte

La mise en contexte se trouve dans le fichier **MiseEnContexte.pdf**. Il décrit les attributs que vous devez analyser ainsi que la valeur qui doit être estimé (ou prédite).

Évaluation

Système (40%)

Le système sera pondéré selon :

- Le système peut être entraîné;
- Le système permet d'estimer ou prédire une valeur selon des paramètres d'entrées;
- Affiche les prédictions de toutes les entrées dans l'ordre du fichier;
- Le système utilise un algorithme pour l'entraînement (les algorithmes supplémentaire donnent des points pour la section **Points à choix**;
- Les prédictions du système sont suffisamment bonnes
10% seront attribués à la qualité de vos estimations lors de la correction interactive.

Rapport (20%)

Le rapport (environ 4 pages) doit contenir les informations suivantes :

- Choix de l'algorithme
 - Quels ont été votre raisonnement?
 - Pourquoi ne pas avoir utilisé un autre algorithme?
- Fonctionnement de l'algorithme
 - Explication d'environ 1 page du fonctionnement de l'algorithme;
 - Si vous avez plusieurs, choisissez celui qui vous semble le plus performant en lien avec sa complexité, son efficacité, son implémentation, etc.
- Problèmes rencontrés et améliorations possibles;
- Résultats expérimentaux
 - Présentation des forces/faiblesses de votre système appuyé par des valeurs quantitatives;
 - Utilisation de mesures telles que : précision, rappel, F-mesure et courbe ROC.

Points à choix (40%)

Le dernier 40% des points sont attribués selon ce que vous implémentez dans votre système. Voici une liste d'éléments communs. Vous pouvez valider avec le chargé de laboratoire pour les points attribués pour des éléments qui ne sont pas contenue dans cette liste (soyez créatif). Les points sont attribués selon le niveau d'implémentation. Une même fonctionnalité ne donne pas de points dans deux sections en même temps.

Éléments supportés par votre système	Points max
File choser	1%
Normalisation des données ou standardisation <ul style="list-style-type: none">- Ajuster toutes les valeurs entre 0 et 1- Ajuster toutes les valeurs avec une moyenne de 0 et un écart type de 1	3%
Élimination des éléments anormaux/aberrants <ul style="list-style-type: none">- Gestion des valeurs qui n'ont pas de sens : une personne qui a 300 ans	3%
Adaptation aux valeurs manquantes <ul style="list-style-type: none">- Ajustement quand il n'y a pas de valeur	3%
Réduction des attributs <ul style="list-style-type: none">- Éliminer ou fusionner des colonnes	5%
Séparation des données d'entraînement et de test	3%
Sérialisation des données entraînées <ul style="list-style-type: none">- Possibilité de partir l'application sans faire d'entraînement	3%
Utilisation d'algorithmes supplémentaires <ul style="list-style-type: none">- Les points sont attribués selon la complexité de tous les algorithmes	20%
Utilisation d'au moins une métaheuristique <ul style="list-style-type: none">- Ils sont utilisés pour aider l'entraînement des algorithmes	6%
Entraînement selon un facteur <ul style="list-style-type: none">- On arrête l'entraînement quand il n'y a pas assez de changement (epsilon)	5%
Validation croisée	5%
Parallélisation	5%
Si plusieurs algorithmes, combiner les prédictions pour en donner une seule	3%
Calcule et affichage du taux de confiance de l'estimation <ul style="list-style-type: none">- Par réponse ou en générale	3%
Graphique ou présentation intéressante	3%
Autres éléments	

Pénalités de correction

Les pénalités suivantes ainsi que leur valeur sont applicables comme suit :

- Fautes de français et erreurs de rédaction : -0.5 % par erreur, maximum 20 fautes (-10 %)
- Professionnalisme et qualité du document (document incomplet, mise en page non professionnelle, langage parlé, manque de nomenclature, texte à la première personne (singulier/pluriel)) : jusqu'à -10 %
- Non-respect des normes de remise : note de zéro automatique et non négociable

Consignes de remise

La remise comprend :

- Le code source de l'application;
- Le rapport en format PDF.

La remise doit se faire via **Moodle**.