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 que 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.

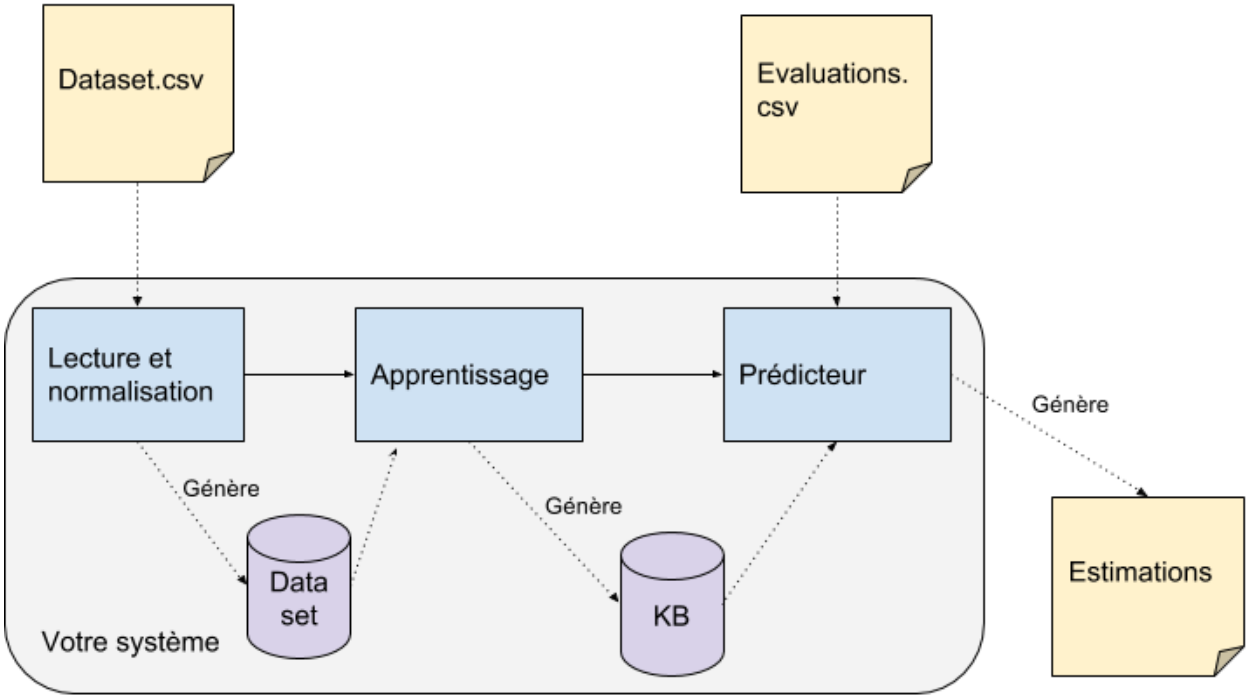
Liste d'algorithmes

- 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

*Auteurs : Richard Rail et Lévis Thériault; Mise à jour : 2017-07-06

†Métaheuristique

En deuxième partie, vous devez implémenter votre système d'apprentissage. Celui-ci doit prendre en entrée un fichier csv (Dataset) 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 peux implémenter plusieurs éléments différents tel que la réduction d'attributs, la normalisation des données, des métaheuristiques... 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.

*Auteurs : Richard Rail et Lévis Thériault; Mise à jour : 2017-07-06
†Métaheuristique

Mise en contexte

La compagnie ***sans20*** vous demande de les aider à prévoir la qualité de leur nouvelle cuvée de vin. Ils ont accumulé une très grande quantité de données de la qualité estimée par des spécialistes du vin. À l'aide de celle-ci, vous devez créer un logiciel qui permet de prévoir la qualité d'un vin selon les attributs fournis.

Un fichier de données est fourni (Dataset). Vous avez le choix de la séparation de ces données pour faire l'entrainement. Vous devez penser à avoir au minimum 2 groupes. Le premier pour l'entrainement et le deuxième pour la validation. Il est laissé à votre discrétion de réduire le nombre d'attributs, si certains ont une moins grande importance.

Les paramètres des données sont décrits dans le tableau suivant :

Paramètres	Description	Valeurs
fixed acidity	Niveau d'acide tartrique	g/l
volatile acidity	Niveau d'acide acétique	g/l
citric acid	Niveau d'acide citrique	g/l
residual sugar	Teneur en sucre	g/l
chlorides	Teneur en chlorures	g/l
free sulfur dioxide	Teneur en SO2	mg/l
total sulfur dioxide	Teneur totale en SO2	mg/l
density	Densité du vin	g/c^3
pH	pH du vin	Valeur du pH entre 0 et 14
sulphates	Teneur en sulfate de potassium	g/l
alcohol	Taux d'alcool	Pourcentage entre 0 et 100
quality	Qualité du vin	Entre 0 et 10

Évaluation

Système (40%)

Le système sera pondéré selon :

- Le système peut être entraîné
- Le système permet d'estimer une valeur selon des paramètres
- Affiche les prédictions de toutes les entrées dans l'ordre du fichier
- Le système utilise un algorithme pour l'entrainement (les algorithmes supplémentaire donnent des points pour la section *Points d'excellence*.
- 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
- 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 d'excellence (40%)

Le dernier 40% des points sont 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 chooser	1%
Normalisation des données	3%
Élimination des éléments anormaux	3%
Adaptation aux valeurs manquantes	3%
Réduction des attributs	5%
Séparation des données d'entraînement et test	3%
Sérialisation des données entraîné (possibilité de partir l'application sans faire d'entrainement)	3%
Utilisation d'algorithmes supplémentaires	20%
Utilisation de métaheuristiques	6%
Entrainement selon un facteur (epsilon de change sinon arrêt de l'entrainement)	5%
Validation croisée (Calcul de l'erreur moyenne)	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	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.

L'adresse courriel pour la remise est la suivante : **courslog635@gmail.com**

Le titre du courriel doit être : Log635-01-Lab3-EquipeX