

TRAVAUX PRATIQUES – ATELIER N° 02

Objectifs de l'atelier

Découvrir un environnement d'apprentissage automatique

Classification avec Weka:

L'onglet "Classify .": Pour lancer une opération de classification.

Weka implémente les modèles les plus populaires de machine learning tels que : Les Réseaux de neurones, Forêt d'arbres décisionnels, arbre de décision, Réseau de bayes, support vector machine, Logiboost, Adaboost, JR48... ,

On essaie de tester 3 algorithmes qui permettent (en général) d'obtenir de très bonne performance de classification. Il s'agit de l'arbre de décision JR48, le Random Forest et le perceptron multicouches . On note aussi les deux remarques suivantes :

- Avec l'arbre de décision JR48, nous obtenons une performance de 96% d'exemples correctement classés. C'est un bon taux de classification (accuracy)
- Weka est par défaut configuré sur la technique « Fold-cross-validation » c'est pourquoi pour commencer le test en cross-validation, il suffit de cliquer sur start et le test se lance automatiquement.

(Interprétation des résultats affichés après application de l'algorithme DT-JR48)

Appliquons maintenant un autre algorithme appelé « Random Forest. » :

- Nous obtenons une performance moins bonne qu'avec l'arbre de décision classique.
- Le résultat obtenu est un peu surprenant quand on sait qu'un « RandomForest est un ensemble d'arbres de décision, et que le principe reposant derrière ce modèle est de combiner plusieurs arbres décisionnels différents afin de prendre une meilleure décision. C'est la raison pour laquelle, Weka n'affiche pas le modèle, car il est trop difficile d'afficher 100 arbres de décision.

Application de l'algorithme « perceptron multicouches ».

Choose → Functions → MultilayerPerceptron

- En utilisant le perceptron multi-couches, nous obtenons un taux de classification de 97.33%
- C'est un résultat supérieur au résultat obtenu par l'arbre de décision et par le Random Forest. Par conséquent, c'est le modèle qui sera retenu pour ce jeu de données.

Clustering avec Weka (Onglet cluster)

En plus des algorithmes de classification nommés ci-dessus, Weka propose les différents algorithmes de clustering suivants: Cobweb, espérance-maximisation, k-moyennes, DBSCAN, Hierarchical Clustering, Density Based Clusterer, ...

Onglet Cluster → choose → simple-k-means(algorithm)

Sélection de caractéristiques avec Weka:

Weka offre en plus des onglets, preprocessing, classifications et clustering, ... un onglet feature selection. Cet onglet permet de sélectionner les caractéristiques du jeu de données étudiées les plus pertinentes pour le traitement que l'utilisateur souhaite faire.

Bon Courage