

# Machine Learning

## Apprentissage supervisé : arbres de décision

### Objectifs

- Familiarisation avec les arbres de décision

### Enoncé

Le choix du langage de programmation est libre. Dans cet exercice il s'agit de programmer la construction d'un arbre de décision binaire (cf. Chap. 5). Pour que ce ne soit pas trop compliqué on utilisera des données discrètes pour des problèmes à deux classes.

Les critères pour déterminer une feuille sont :

- profondeur de l'arbre ;
- proportion des classes dans tous les nœuds.

On définira un paramètre pour chacun de ces critères. En outre, il faudra définir le paramètre régissant le nombre minimal d'exemples toléré dans une feuille. A la fin, il faudra afficher l'arbre créé.

L'algorithme devra être testé sur un jeu de données en essayant plusieurs valeurs des paramètres. Quatre fichiers ont été mis à disposition (dans un répertoire *data*) :

- monks-1.train
- monks-1.test
- monks-2.train
- monks-2.test

Les données contiennent 8 composantes. La première représente la classe (0 ou 1) et la 8<sup>ème</sup> correspond à un identifiant qui doit être ignoré. Entre ces deux composantes nous avons 6 attributs (de a1 à a6) qui prennent entre 2 et 4 valeurs discrètes. Plus particulièrement les composantes sont :

- class : 0, 1
- a1 : 1, 2, 3
- a2 : 1, 2, 3
- a3 : 1, 2
- a4 : 1, 2, 3
- a5 : 1, 2, 3, 4
- a6 : 1, 2
- Identifiant

Ce travail n'est pas noté, mais il faudra le rendre sur *Cyberlearn* au plus tard le **lundi 28 octobre**.