

8INF867 – Fondamentaux de l'apprentissage automatique

Travail No. 1 - Programmation d'ID3

1. Travail à effectuer

Dans le cadre de ce travail pratique, vous devrez programmer votre propre version de ID3. Il existe plusieurs implémentations en ligne, mais je vous demande de tenter de le faire par vous-même. Par défaut, ID3 est un algorithme très simple et nous allons y faire quelques modifications. Vous ne pouvez évidemment pas utiliser d'IA générative pour programmer à votre place.

Votre algorithme doit être programmé en Python et supporter en entrée des ensembles de données variés au format CSV. Il devrait fonctionner avec un nombre arbitraire d'instances d'apprentissage, d'attributs et de classes. Ainsi, vous devez prévoir la possibilité d'avoir 2, 5, 20 attributs et de ne pas avoir simplement un problème de classification binaire.

Votre algorithme devra aussi supporter les attributs numériques. Gardez en tête que si vous avez 50 valeurs possibles, le facteur de ramification (*branching factor*) à chaque niveau de votre arbre pourrait atteindre 50. Bref, si un attribut a une possible infinité de valeurs, votre algorithme ne fonctionnera plus très bien. Pour y remédier, discrétisez automatiquement les attributs numériques avec la méthode de votre choix.

La construction de votre arbre se fait avec la fonction `fit` (e.g. : `monArbre = ID3().fit()`). Vous devez permettre de limiter la profondeur de l'arbre final (e.g. : `monArbre=ID3(depth_limit=2)`).

Enfin, vous devez implémenter une fonction d'élagage postérieur ou *post-pruning* (e.g. : `monArbre.Prune()`). Je vous laisse choisir vos propres critères. Vous devrez m'expliquer ce que vous avez choisi, mais ne vous inquiétez pas, vous ne serez pas évalués sur la performance finale.

2. Remise

Le travail devra être déposé sur Moodle dans la section « Travaux et dépôts » sous l'intitulé Travail pratique #1. Attention! tout travail en retard recevra une pénalité minimale de 20% jusqu'à possibilité d'obtenir zéro.

Vous devez me remettre une copie de votre code et un document contenant minimalement une page titre (vos noms, codes permanents et titre du cours), les contraintes de fonctionnement (librairies nécessaires), ainsi qu'une brève explication de votre travail.

Équipe : Équipe de 2 à 3 personnes

Date limite de remise : Vendredi le 18 octobre à 23h59 (EST).

3. Évaluation

Le travail vaut

Voici la grille d'évaluation que j'utiliserai pour vous noter. Il s'agit d'une évaluation *négative*, c'est-à-dire que vous démarrez avec 100% et que chaque élément peut vous faire perdre un certain nombre de points :

Dimension	Élément spécifique	Pénalité max
Code	Code difficile à lire	2
	Dépendances non justifiées	1
	Bogues	5
Fonctionnalités	ID3 et mesures	5
	CSV	5
	Attributs	2
	Limite de profondeur	1
	Élagage	2
Qualité de la remise	Présence des éléments demandés	2
	Qualité du français (ou de l'anglais)	2