

Introduction à l'intelligence artificielle - groupe A

[Accueil](#) / [Mes cours](#) / [Accès aux cours](#) / [STN](#) / [Informatique](#) / [Licence Informatique](#) / [IIA-A](#) / 09/11/2022 TP6

/ [Arbres de décision](#)

Arbres de décision

L'objectif de ce TP est de programmer un arbre de décision.

Les données à traiter sont disponibles ici : archive.ics.uci.edu/ml/datasets/Breast+Cancer.

Notez qu'on voudra pouvoir appliquer la même méthode à d'autres datasets trouvés ici : archive.ics.uci.edu/ml/datasets.php?att=cat.

Pour mémoire, l'algorithme de construction d'un arbre de décision consiste à :

1. Pour chaque attribut A_j , calculer $H(C | A_j)$
2. Choisir l'attribut A_j qui optimise $H(C | A_j)$
 - ajouter un nœud à l'arbre de décision pour A_j
3. Partitionner la base d'apprentissage, selon les modalités d' A_j
4. Itérer sur les attributs restants

et donc à optimiser un critère. Ici, on programmera l'indice de Gini et l'entropie de Shannon, à fin de comparaison, deux critères à minimiser.

Les indices étaient données ainsi dans le cours :

Mesures de discrimination classiques

- Utilisation de l'indice de Gini :

$$I_G(p) = \sum_{i=1}^n p_i \sum_{k \neq i} p_k$$

- Mesure adaptée du coefficient de Gini, mesure économique

$$H_G(C|A) = \sum_{a_i \in A} P(a_i) \times \left(1 - \sum_{c_k \in C} P(c_k|a_i)^2\right)$$

A. Provost d'Altenros

FA - 1

Mesures de discrimination classiques

- Utilisation de l'entropie de Shannon :

$$H_S(C|A) = - \sum_i P(v_i) \sum_k P(c_k|v_i) \log_2(P(c_k|v_i))$$

- Mesure issue de la théorie de l'information
 - initiée par C. E. Shannon en 1948
- Mesure un taux de désordre
 - ⇒ à minimiser

A. Provost d'Altenros

FA - 1

Statut de remise

