

Correction TD : Les arbres de décisions

Exercice1 (examen 2015 session principale)

1. Donner un pseudocode du fonctionnement de l'algorithme des arbres de décision. [2pts]

Procédure construire-arbre(X)

Si tous les individus I appartiennent à la même modalité de la variable décisionnelle **ALORS** créer un nœud feuille portant le nom de cette classe : Décision

SINON

choisir le meilleur attribut pour créer un nœud // l'attribut qui sépare le mieux, le test associé à ce nœud sépare X en des branches
construire-arbre(Xd), ..., construire-arbre(Xg)

FIN

2. En déduire la variable la **plus décisive** par rapport à l'appartenance d'un individu à l'origine orientale. (Donner le calcul complet et la formule utilisée) [3pts=2pts(formule)+1pt(calcul)]

Indice de Gini :

1-somme(fréquence de la classe décisionnelle dans le nœud)

11 individus : Oriental = Oui : 6, Oriental = Non : 5 donc

▪ **Indice de Gini avant séparation au NIVEAU DE LA RACINE :** $IG(\text{Oriental}) = 1 - ((5/11)^2 + (6/11)^2) = 0.4958678$

▪ **Indice de Gini de la variable Yeux :**

4 Noir : 4 Oui, 0 Non $IG(Y=\text{Noir}) = 1 - ((4/4)^2 + (0/4)^2) = 0$

3 Brun : 2 Oui, 1 Non $IG(Y=\text{Brun}) = 1 - ((2/3)^2 + (1/3)^2) = 0$

4 Bleu : 0 Oui, 4 Non $IG(Y=\text{Bleu}) = 1 - ((0/4)^2 + (4/4)^2) = 0$

▪ **Indice de Gini de la variable Cheveux :**

4 Noir : 3 Oui, 1 Non $IG(Ch=\text{Noir}) = 1 - ((3/4)^2 + (1/4)^2) = 0$

4 Blanc : 3 Oui, 1 Non $IG(Ch=\text{Blanc}) = 1 - ((3/4)^2 + (1/4)^2) = 0$

3 Blond : 0 Oui, 3 Non $IG(Ch=\text{Blond}) = 1 - ((0/3)^2 + (3/3)^2) = 0$

▪ **Indice de Gini de la variable Taille :**

6 Petit : 3 Oui, 3 Non $IG(T=\text{Petit}) = 1 - ((3/6)^2 + (3/6)^2) = 0.5$

5 Grand : 3 Oui, 2 Non $IG(T=\text{Grand}) = 1 - ((3/5)^2 + (2/5)^2) = 0$

La variable la plus décisive est celle qui maximise $[IG(\text{avant séparation}) - IG(\text{fils1}) + \dots + IG(\text{filsn})]$, donc la couleur des **Yeux** est la variable la plus décisive par rapport à

l'appartenance d'un individu à l'origine orientale

	Yeux	Cheveux	Taille	Oriental
1	Noir	Noir	Petit	Oui
2	Noir	Blanc	Grand	Oui
3	Noir	Blanc	Petit	Oui
4	Noir	Noir	Grand	Oui
5	Brun	Noir	Grand	Oui
6	Brun	Blanc	Petit	Oui
7	Bleu	Blond	Grand	Non
8	Bleu	Blond	Petit	Non
9	Bleu	Blanc	Grand	Non
10	Bleu	Noir	Petit	Non
11	Brun	Blond	Petit	Non

Exercice2 (examen 2016 session principale)

$IG[\text{age}, 18 ; 35] = 1 - ((2/4)^2 + (2/4)^2) = 0,5$ $IG[\text{age}, 36 ; +] = 1 - ((2/6)^2 + (4/6)^2) = 0,44$ $IG[\text{age}] = 0,94$	$IG[\text{sexe}, H] = 1 - ((4/6)^2 + (2/6)^2) = 0,44$ $IG[\text{sexe}, F] = 1 - ((0/4)^2 + (4/4)^2) = 0$ $IG[\text{sexe}] = 0,44$
$IG[\text{Propriétaire}, N] = 1 - ((1/3)^2 + (2/3)^2) = 0,44$ $IG[\text{Propriétaire}, O] = 1 - ((3/7)^2 + (4/7)^2) = 0,49$ $IG[\text{Propriétaire}] = 0,93$	->La racine de l'arbre est la variable SEXE
$IG[\text{age}, 18 ; 35, \text{sexe}=H] = 1 - ((2/3)^2 + (1/3)^2) = 0,44$ $IG[\text{age}, 36 ; +, \text{sexe}=H] = 1 - ((2/3)^2 + (1/3)^2) = 0,44$ $IG[\text{age}, \text{sexe}=H] = 0,888$	$IG[\text{Propriétaire}, N, \text{sexe}=H] = 1 - ((1/2)^2 + (1/2)^2) = 0,5$ $IG[\text{Propriétaire}, O, \text{sexe}=H] = 1 - ((3/4)^2 + (1/4)^2) = 0,375$ $IG[\text{Propriétaire}, \text{sexe}=H] = 0,875$

