

Série TD 2

Classification – Les arbres de décision.

Exercice 1

Nous considérerons l'ensemble **S** d'exemples suivant ayant les attributs A, B, C, et D :

	A	B	C	D	Classe
E1	a1	b1	c1	d2	+
E2	a1	b2	c2	d2	+
E3	a1	b2	c3	d1	-
E4	a2	b1	c1	d1	-
E5	a2	b2	c1	d1	-
E6	a2	b2	c1	d2	+
E7	a1	b1	c1	d1	+
E8	a2	b1	c2	d2	-
E9	a3	b1	c3	d1	+
E10	a3	b2	c2	d2	+

1. Calculer l'entropie de l'ensemble **S** par rapport à la valeur de la classe.
2. Construire l'arbre de décision correspondant à l'ensemble **S** en utilisant l'algorithme ID3.
3. Etablir la matrice de confusion pour la table d'entraînement ci-dessus.
4. Donner la précision de l'arbre obtenu sur la table d'entraînement. Calculer la moyenne harmonique sur la même table.
5. Trouver la classe de l'exemple ayant les attributs (a2, b1, c3, d1) selon l'arbre construit.

NB : On donne le tableau suivant représentant les valeurs de l'entropie :

x\y	1	2	3	4	5	6
6	0,592	0,811	0,918	0,971	0,994	1
5	0,650	0,863	0,954	0,991	1	
4	0,722	0,918	0,985	1		
3	0,811	0,971	1			
2	0,918	1				
1	1					

Exercice 2

Dans un hôpital, on souhaite construire un arbre de décision pour la prédiction du risque des patients d'avoir une certaine maladie en fonction de leur âge et de deux symptômes booléens (vrai ou faux) appelés S1 et S2. Le risque est évalué selon trois valeurs F (faible), M (moyen) et E (élevé), l'âge est discrétisé selon trois valeurs (jeune, adulte et senior).

L'hôpital dispose de la table suivante :

N°	Age	S1	S2	Risque
1	Jeune	F	V	Faible
2	Jeune	V	V	Elevé
3	Adulte	F	F	Faible
4	Senior	V	F	Elevé
5	Senior	F	V	Moyen
6	Jeune	F	F	Faible
7	Adulte	V	F	Moyen
8	Adulte	V	V	Moyen
9	Senior	F	F	Faible
10	Senior	V	V	Elevé

1. Calculer l'entropie de l'ensemble d'exemples par rapport à la valeur de la classe *Risque*.
2. Construire l'arbre souhaité en utilisant l'algorithme ID3.
3. Donner le risque du patient ayant les attributs (Jeune, V, F) selon l'arbre construit.