

### Exercice 'Minimisation de l'entropie'

---

Soit l'échantillon suivant :

no	P1	P2	P3	Classe
1	0	V	N	A
2	1	V	I	A
3	0	F	O	B
4	1	V	N	A
5	1	V	O	B
6	1	F	N	A
7	0	F	O	B
8	0	V	I	A
9	0	F	N	B

Soit l'ensemble d'apprentissage constitué des exemples  $\{1, \dots, 9\}$ .

(NB: cet échantillon est une variation sur l'échantillon vu en TD.)

On veut construire un arbre de décision parfait en choisissant la propriété minimisant l'entropie.

1. Calculer  $E_0$  entropie de l'échantillon total.
2. a. Calculer  $E(P1)$ .  
2. b. Calculer  $E(P2)$ .  
2. c. Calculer  $E(P3)$ .  
2. d. Quelle propriété choisir en premier critère à la racine ?
3. Quels sont les nœuds terminaux à profondeur 1 ?
4. Pour les autres nœuds, continuer la construction.
5. Conclure.

### Exercice 'Entropie relatif'

---

Soit l'échantillon suivant :

no	x	y	z	Classe
1	0	1	0	+
2	1	0	0	-
3	0	1	1	-
4	0	0	1	+
5	1	0	1	-

1. Soit l'ensemble d'apprentissage constitué des exemples {1,...,5}.

Construire l'arbre de décision parfait  $t_1$  suivant le principe de minimisation de l'entropie.

Quel est le meilleur attribut à la racine de l'arbre ?

Quelle est sa profondeur ?

2. Construire un arbre de profondeur 2 utilisant l'attribut  $y$  à la racine de l'arbre.

Conclure.

## Exercice 'Foret'

---

Soit l'ensemble complet des exemples a 3 attributs et 2 valeurs (0 ou 1) avec leur classe:

no	x	y	z	Classe
1	0	0	0	+
2	0	0	1	-
3	0	1	0	+
4	0	1	1	+
5	1	0	0	-
6	1	0	1	-
7	1	1	0	-
8	1	1	1	+

1. Soit  $AD_i$  ( $1 \leq i \leq 5$ ) l'arbre de decision construit sur l'ensemble d'apprentissage  $E_i$  en suivant le principe de minimisation de l'entropie.

On a:

$E_1 = \{0, 1, 2, 3\}$        $E_2 = \{4, 5, 6, 7\}$        $E_3 = \{0, 2, 4, 6\}$   
 $E_4 = \{1, 3, 5, 7\}$        $E_5 = \{0, 1, 4, 5\}$

Construire  $AD_1, AD_2, AD_3, AD_4, AD_5$ .

2. Tester les  $AD_i$  sur l'ensemble total en comptant le nombre d'erreurs faites.

3. On appelle "foret" un ensemble d'arbres de decision decidant de la classe d'un exemple en faisant voter chaque arbre et en prenant le vote majoritaire.

Soit  $F_5$  la foret constituee des  $AD_i$  ( $1 \leq i \leq 5$ ).

Tester  $F_5$  sur l'ensemble total en comptant le nombre d'erreurs faites.

4. Soit  $F_3$  la foret constituee de  $AD_1, AD_3, AD_5$ .

Tester  $F_3$  sur l'ensemble total en comptant le nombre d'erreurs faites.

5. Conclure.