

Corrigée type

Questions (9 points)

1. Est le processus de **découverte (l'extraction)** de connaissances dans les bases de données *en utilisant un* ensemble des technique et de méthodes du domaine des statistiques, des mathématiques et de l'informatique permettant **l'extraction**, à partir d'un important volume de données brutes. **(2 points)**

Domaines : **Marketing , Scoring, Médical, Prévention, Web (1 points)**

2. Dans les méthodes de clustering non hiérarchiques le nombre de cluster est défini avant le lancement de l'algorithme contrairement aux méthodes hiérarchique. Aussi les méthodes de clustering non hiérarchiques ne travaillent qu'avec des données numériques seulement. **(3 points)**

3. ID3 construit l'arbre de décision récursivement. A chaque étape de la récursion, il calcule parmi les attributs restant pour la branche en cours, celui qui maximisera le gain d'information. (l'attribut qui permettra le plus facilement de classer les exemples à ce niveau de cette branche de l'arbre). **(3 points)**

Exercice 1 : (6 points)

1.

Etape	w0	w1	w2	w3	w4	w5	w6	w7	x0	x1	x2	x3	x4	x5	x6	x7	Somme(xiwi)	o	c	w0	w1	w2	w3	w4	w5	w6	w7
1	2	1	0	1	0	-1	1	1	1	1	1	1	1	1	1	0	4	1	0	1	0	-1	0	-1	-2	0	1
2	1	0	-1	0	-1	-2	0	1	1	0	1	1	0	0	0	0	0	0	1	2	0	0	1	-1	-2	0	1
3	2	0	0	1	-1	-2	0	1	1	1	1	0	1	1	0	1	0	0	0	2	0	0	1	-1	-2	0	1
4	2	0	0	1	-1	-2	0	1	1	1	1	1	1	0	0	1	3	1	1	2	0	0	1	-1	-2	0	1
5	2	0	0	1	-1	-2	0	1	1	0	1	1	0	0	1	1	4	1	0	1	0	-1	0	-1	-2	-1	0
6	1	0	-1	0	-1	-2	-1	0	1	1	0	1	1	0	1	1	-1	0	1	2	1	-1	1	0	-2	0	1
7	2	1	-1	1	0	-2	0	1	1	1	0	1	1	1	1	1	3	1	0	1	0	-1	0	-1	-3	-1	0
8	1	0	-1	0	-1	-3	-1	0	1	1	1	1	0	0	0	0	0	0	1	2	1	0	1	-1	-3	-1	0
9	2	1	0	1	-1	-3	-1	0	1	1	1	1	1	1	1	1	-1	0	0	2	1	0	1	-1	-3	-1	0
10	2	1	0	1	-1	-3	-1	0	1	1	1	1	1	0	1	1	2	1	1	2	1	0	1	-1	-3	-1	0

2. Le perceptron obtenu ne donne pas toujours le bon résultat (se trompe par exemple sur le chiffre 4) car il n'a pas « suffisamment appris ». Le choix du critère d'arrêt de l'algorithme est la raison de ce résultat.