

Examen (documents autorisés)

Questions de cours (5 pts)

1. Proposer, avec une brève explication, trois applications de la fouille de données non vues dans le cours.
2. Expliquer le problème de sur-apprentissage et comment le résoudre.
3. Expliquer le problème de la non séparabilité linéaire dans les machines à vecteurs supports (SVM). Comment est-il résolu ? Quelle est l'utilité des noyaux ?

Exercice 1 Motifs fréquents (8 pts)

Soit la base de données transactionnelle suivante :

TID	Items	TID	Items
T01	A, B, C, D	T05	B, C, G
T02	A, C, D, F	T06	D, F, G
T03	C, D, E, G, A	T07	A, B, G
T04	A, D, F, B	T08	C, D, F, G

En utilisant un support minimum de 25% et un minimum de confiance de 60%, trouver :

1. Les motifs fréquents maximaux (de taille maximale)
2. Les motifs fréquents fermés (qui ne sont sous ensemble d'aucun autre motif)
3. Les règles solides d'association.

Exercice 2 Arbres de décision (7 pts)

Soit les exemples suivants ayant trois attributs et appartenant à deux classes :

N°	Att1	Att2	Att3	Classe	N°	Att1	Att2	Att3	Classe
1	A	L	T	C1	8	B	L	T	C1
2	A	H	T	C2	9	B	H	F	C1
3	A	H	F	C2	10	C	H	T	C2
4	A	H	F	C2	11	C	L	T	C2
5	A	L	F	C1	12	C	H	F	C1
6	B	H	T	C1	13	C	H	F	C1
7	B	H	F	C1	14	C	H	F	C1

En utilisant l'algorithme ID3 :

1. Calculer l'entropie de l'ensemble d'exemples par rapport à la valeur de la classe.
2. Construire l'arbre de décision appris de cette table.

★★★ Bonne chance ★★★

A.Djeffal