
Examen

Questions de cours (4 pts)

1. En apprentissage supervisé, quelle est la différence fondamentale entre classification et régression ?
2. Expliquer en quelques mots ce qu'est le phénomène de sur-apprentissage.
3. Donner une stratégie pour éviter le sur-apprentissage lors de l'apprentissage d'arbres de décision.
4. Donner une méthode d'évaluation des techniques d'apprentissage et expliquer, en quelques lignes, son principe.

Exercice 1 Motifs fréquents (10.5 pts : 4 + 2.5 + 2.5 + 1.5)

Dans un supermarché, on dispose de la base de transactions suivante :

TID	Items	TID	Items
T1	Lait, Jus, Couches	T6	Lait, Couches, Pain, Beurre
T2	Pain, Beurre, Lait	T7	Pain, Beurre, Couches
T3	Lait, Couches, Sucre	T8	Jus, Couches
T4	Pain, Beurre, Sucre	T9	Lait, Couches, Pain, Beurre
T5	Jus, Sucre, Couches	T10	Jus, Sucre

En utilisant l'algorithme Apriori avec un support minimum de 20% et une confiance minimale de 75%, trouver :

1. Les motifs fréquents,
2. Les motifs fréquents fermés,
3. Les motifs fréquents maximaux,
4. Les règles solides d'association de type $A, B \Rightarrow C$.

Exercice 2 Classification (5.5 pts : 4 + 0.5 + 1)

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 age 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'age est discrétisé selon trois classes (jeune, adulte et senior).

L'hôpital dispose de la table suivante :

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

Questions :

1. Construire l'arbre souhaité en utilisant l'algorithme ID3.
2. Donner le risque du patient ayant les attributs (Jeune,V,F) selon l'arbre construit.
3. Trouver le risque du même patient en utilisant la classification bayésienne naïve.

★★★★ Bonne chance ★★★★★

Dr A.Djeffal