

EXAMEN S2

Exercice 1

Soit les transactions suivantes :

Tid	Item
T1	Pain, Lait, Beurre, chocolat, Fromage
T2	Pain, Beurre, Eau, Confiture, Fromage
T3	Fromage, chocolat, Pain, Beurre, Confiture
T4	Beurre, Lait, Jus
T5	chocolat, Fromage, Jus, Eau
T6	Fromage, chocolat, Pain, Beurre, Confiture

- 1) Avec un support minimum de 60%, trouver l'ensemble des motifs fréquents.
- 2) Trouver les motifs fréquents fermés et les motifs fréquents maximaux.
- 3) Avec un minimum de confiance de 70%, trouver les règles d'association de la forme $A \rightarrow B, C$.

Exercice 2

Soient les données suivantes :

$A1=(3,11)$, $A2=(3,6)$, $A3=(9,5)$, $A4=(6,9)$, $A5=(8,6)$, $A6=(7,5)$, $A7=(2,3)$, $A8=(5,10)$.

- 1) Faire une classification de ces données en utilisant l'algorithme k-means pour $k=3$. On considère les classes de départ $C1=(A1,A2,A3)$, $C2=(A4,A5)$, $C3=(A6,A7,A8)$, On utilisera city-blocs pour mesurer la distance entre les données.

Exercice 3

- 1) La distance choisie (euclidienne, city block,...) influence t'elle ou non le résultat d'une approche de clustering?
- 2) Quel est l'avantage et l'inconvénient majeur du clustering incrémental?
- 3) Si on considère, dans une approche d'extraction de règles d'association, que chaque item possède un poids représentant son importance, donnez la valeur de calcul du support?