

Examen semestriel

Module "Datamining" - "Modèles pour le Datamining"

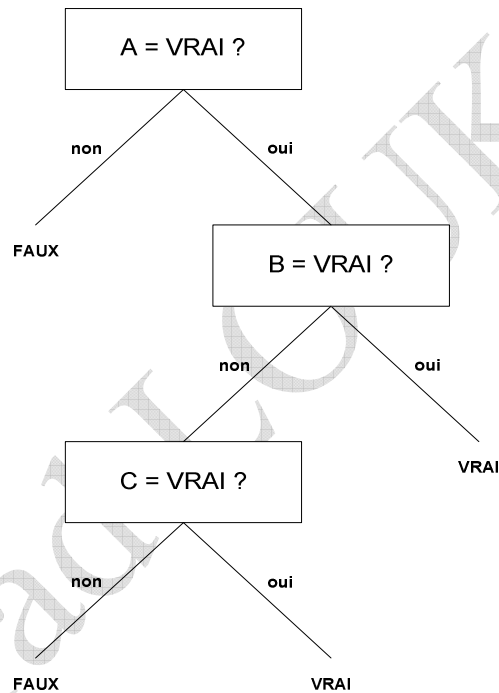
Durée : 01H30

**Exercice 1 (04 points) :**

Représentez sous forme d'un arbre de décision l'expression logique suivante : A et (B ou C).

Réponse :

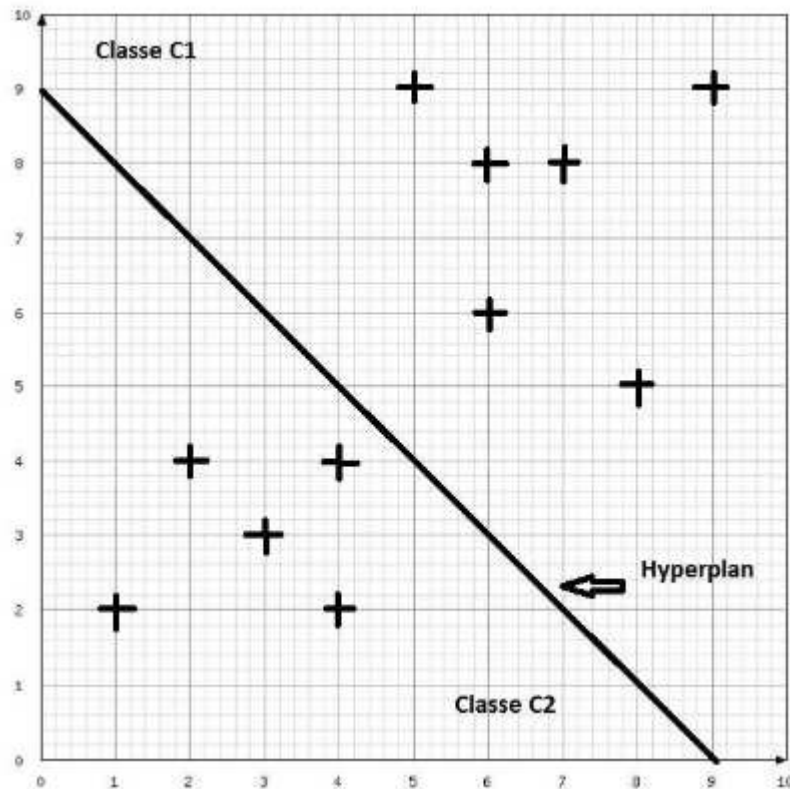
A et (B ou C).



(4 points)

**Exercice 2 (04 points) :**

La figure suivante représente la classification par SVM d'un ensemble de données. On y a représenté l'hyperplan séparateur entre les deux classes C1 et C2. Mais, deux points ont été volontairement masqués (cachés) : un point de la classe C1 et un point de la classe C2.



Question 1 : Donnez l'équation de l'hyperplan séparateur de la figure.

Réponse :

$$y = -x + 9$$

(1 point)

Question 2 : Pouvez-vous deviner quel est le point qui a été masqué de la classe C1 ? . Si oui, donnez ses coordonnées exactes, sinon situez-le approximativement .

Réponse :

Le point masqué de la classe C1 ne peut pas être situé exactement. Néanmoins , on peut donner l'équation qu'il doit vérifier :

Constatons que le point le plus proche de la classe C2 à l'hyperplan séparateur (support vecteur) est le point (4, 4). Logiquement, le point masqué de la classe C1 est celui qui doit jouer le rôle du "vecteur support" de cette classe. Il doit appartenir à la droite parallèle à l'hyperplan, son équation est :  $y = -x + 10$ . Il peut être n'importe où sur cette droite.

(1.5 point)

Question 3 : Même question pour le point masqué de la classe C2.

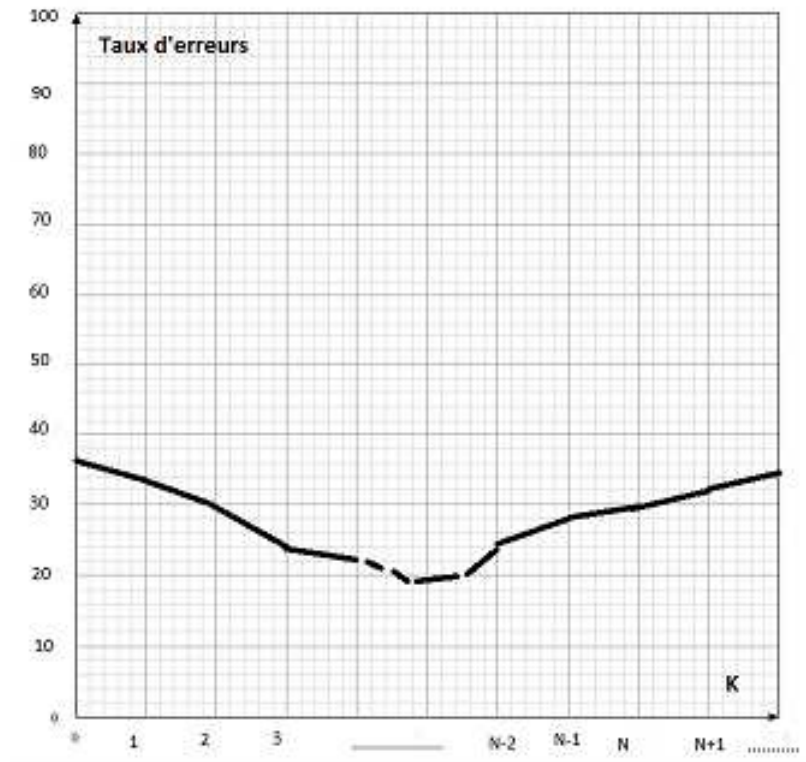
Réponse :

Le point masqué de la classe C2 ne peut pas être situé exactement. Il peut être n'importe où dans la région située en dessous de la droite  $y = -x + 8$  (droite qui passe par le support vecteur de la classe C2, et parallèle à l'hyperplan).

(1.5 point)

**Exercice 3 (04 points) :**

La figure suivante représente le taux d'erreurs de classification, en utilisant l'algorithme KNN, en fonction de K sur un ensemble de N données.



Question : La courbe vous semble-t-elle normale ? Si oui donnez une explication de ses variations, sinon expliquez ce qui est anormal.

Réponse :

Cette courbe comporte une anomalie dans sa dernière partie. Pour K supérieur à N , la courbe devrait normalement être constante : on doit avoir le même taux d'erreur pour toutes les valeurs de K avec  $K \geq N$ .

(4 points)

**Exercice 4 (08 points) :**

Dans un magasin de produits informatiques, on a enregistré les achats suivants faits par les clients (transactions) :

N°	Articles
01	CD, DVD, Jeu, Antivirus
02	CD, Jeu, Souris, Caméra, Antivirus
03	Antivirus, FlashDisk, CD, Jeu, Caméra
04	Jeu, DVD, Tablette
05	FlashDisk, Antivirus, Tablette, Souris

1/ Trouver tous les ensembles d'items fréquents vérifiant  $minsup \geq 60\%$ , en appliquant l'algorithme Apriori .

Réponse :

C1

Ensemble items	Support	%	Observation
{ CD }	03	60%	
{ DVD }	02	40%	Non retenu
{ Jeu }	04	80%	
{ Souris }	02	40%	Non retenu
{ Caméra }	02	40%	Non retenu
{ Tablette }	02	40%	Non retenu
{ Antivirus }	04	80%	
{ FlashDisk }	02	40%	Non retenu

F1 : Ensemble fréquent de taille 1

{ CD }
{ Jeu }
{ Antivirus }

(1 point)

C2

Ensemble items	Support	%	Observation
{ CD, Jeu }	03	60%	
{ CD, Antivirus }	03	60%	
{ Jeu, Antivirus }	03	60%	

F2 : Ensemble fréquent de taille 2

{ CD, Jeu }
{ CD, Antivirus }
{ Jeu, Antivirus }

(1 point)

C3

Ensemble items	Support	%	Observation
{ CD, Jeu, Antivirus }	03	60%	

F3 : Ensemble fréquent de taille 3

{ CD, Jeu, Antivirus }
------------------------

(1 point)

C4

Ensemble items	Support	%	Observation
$\emptyset$			

F4 : Ensemble fréquent de taille 4

$\emptyset$
-------------

Les ensembles fréquents sont :

{ CD }, { Jeu }, { Antivirus }, { CD, Jeu }, { CD, Antivirus }, { Jeu, Antivirus }, { CD, Jeu, Antivirus }

2/ Dédire toutes les règles d'associations .

Réponse :

Les ensemble fréquents	Règles	Confiance
{ CD, Jeu }	CD → Jeu	3/3, soit 100%
	Jeu → CD	3/4, soit 75%
{ CD, Antivirus }	CD → Antivirus	3/3, soit 100%
	Antivirus → CD	3/4, soit 75%
{ Jeu, Antivirus }	Jeu → Antivirus	3/4, soit 75%
	Antivirus → Jeu	3/4, soit 75%
{ CD, Jeu, Antivirus }	CD → Jeu, Antivirus	3/3, soit 100%
	Jeu → CD, Antivirus	3/4, soit 75%
	Antivirus → CD, Jeu	3/4, soit 75%
	CD, Jeu → Antivirus	3/3, soit 100%
	CD, Antivirus → jeu	3/3, soit 100%
	jeu, Antivirus → CD	3/3, soit 100%

(3 points)

3/ Quelles sont les règles ayant une confiance  $\geq 75\%$  ?

Réponse :

Toutes les règles construites ont une confiance  $\geq 75\%$

N°	Règles
01	CD → Jeu
02	Jeu → CD
03	CD → Antivirus
04	Antivirus → CD
05	Jeu → Antivirus
06	Antivirus → Jeu
07	CD → Jeu, Antivirus
08	Jeu → CD, Antivirus
09	Antivirus → CD, Jeu
10	CD, Jeu → Antivirus
11	CD, Antivirus → jeu
12	jeu, Antivirus → CD

(2 points)