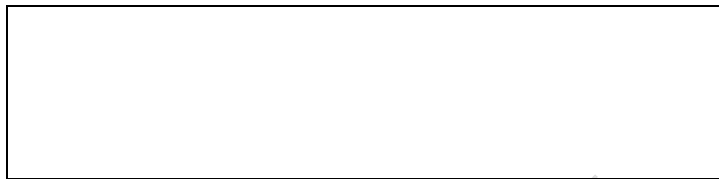


Examen semestriel

Module "Modèles pour le Datamining"

Durée : 01H30



Exercice 1 (03 points) :

a/ Expliquez le principe d'une classification supervisée et la différence avec une classification non supervisée.

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

b/ Parmi les algorithmes présentés en cours, précisez ceux qui font partie des méthodes supervisées et ceux qui appartiennent aux méthodes non supervisés.

Réponse :

Algorithmes de la classification supervisée	Algorithmes de la classification non supervisée
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

Exercice 2 (07 points) :

Le tableau suivant contient des données sur des individus d'une population décrits selon deux attributs : attribut 1 et attribut 2. On souhaite utiliser la méthode SVM pour classer ces données en deux classes : C1 et C2.

Tableau des données

N°	Attribut 1	Attribut 2	Classe
1	0	3	C1
2	3	3	C1
3	1	4	C1

4	2	5	C1
5	1	6	C2
6	2	7	C2
7	3	7	C2

8	3	8	C2
9	2	9	C2
10	4	9	C2
11	8	9	C2

Question 1 :

- Représentez sur le plan de la figure 1 les données du tableau précédent, en symbolisant la classe C1 par une croix (+) et la classe C2 par un cercle (o).
- Représentez l'hyperplan séparateur optimal en montrant clairement les supports de vecteur que vous utilisez.

Question 2 :

- On souhaite classer un douzième individu ayant Attribut1 égal à 4 et Attribut2 égal à 5.6. Redessinez tous les points sur la figure 2, en montrant la classe (croix ou cercle) de l'élément qui vient d'être ajouté.
- Représentez le nouvel hyperplan optimal.

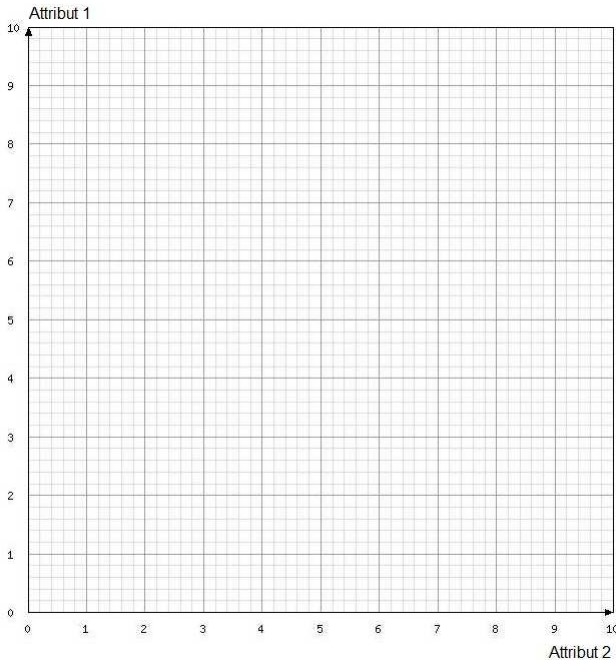


Figure 1

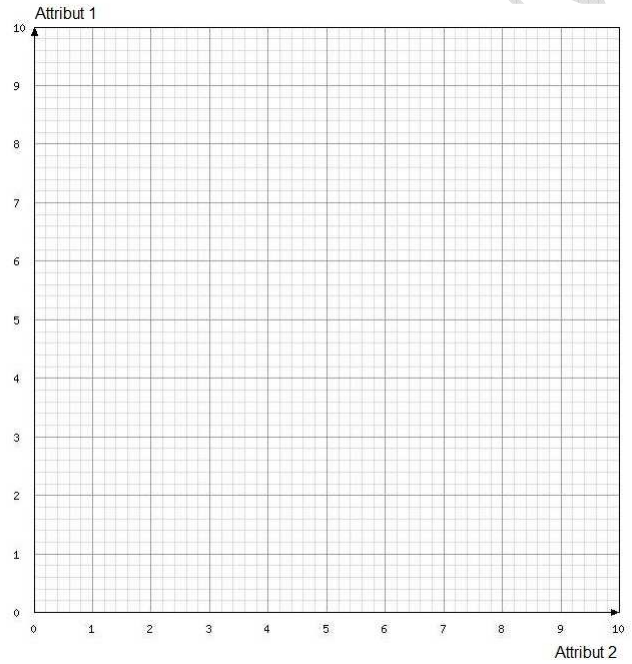


Figure 2

Question 3 : On ajoute un autre point (2, 3) dont on connaît la classe, qui est C1.

Les données sont-elles toujours linéairement séparables ? Si oui donnez l'équation du nouvel hyperplan optimal. Sinon, dites ce que prévoit SVM dans ce cas.

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

Exercice 3 (07 points) :

Le tableau suivant contient des données sur des individus d'une population décrits selon deux attributs : attribut 1 et attribut 2. La classe d'un individu peut être : C1, ou C2, ... ou C6.

Tableau des données

N°	Attribut 1	Attribut 2	Classe
1	1	2	C1
2	2	6	C1

3	2	5	C2
4	2	1	C3
5	4	2	C5

6	5	6	C4
7	6	5	C3
8	6	1	C6

Question 1 :

- Représentez sur le plan de la figure 3 suivante les données du tableau précédent.

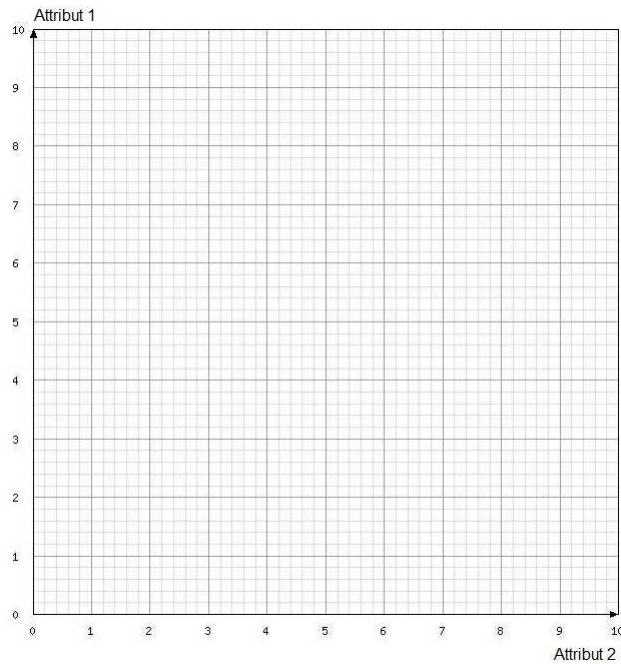


Figure 3

Question 2 :

- On veut classer un nouvel individu U ayant comme attributs (1, 4) en utilisant la méthode KNN. Quelle sera la classe de U si on choisit $k=3$. Justifiez.

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 3 :

- On utilise maintenant la variante de KNN qui utilise la distance $1/d^2$ (inverse de la distance au carré) pour calculer les voisins. Quelle sera la classe de U avec $k=3$? Justifiez .

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Exercice 4 (03 points) :

Question 1 : Quelles sont les qualités d'un bon clustering ?.

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 2: Quelles sont les recommandations générales pour le choix du paramètre k dans la méthode k plus proches voisins ?

Réponse :

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....
.....
.....

Dr Mourad LOUKAM