Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Ghardaïa
Faculté des Sciences et de la Technologie
Département des Mathématiques et d'Informatique

Concours d'accès à la formation de troisième cycle en vue de l'obtention du diplôme de doctorat au titre de l'année universitaire 2021-2022

Filière : Informatique

Spécialité : Systèmes Intelligents et Apprentissage Automatique

Module : Apprentissage Automatique **Durée : 02 heures** **Coefficient : 3**

Exercice 1. — Arbres de décision *(04 points)*

Considérons le tableau suivant, qui présente un ensemble d'apprentissage tiré aléatoirement d'une base de données d'achats des clients d'une entreprise.

Id	Âge	Salaire	Étudiant	A acheté
1	Petit	Haut	Non	Non
2	Petit	Haut	Non	Non
3	Moyen	Haut	Non	Oui
4	Sénior	Moyen	Non	Oui
5	Sénior	Bas	Oui	Oui
6	Sénior	Bas	Oui	Non
7	Moyen	Bas	Oui	Oui
8	Petit	Moyen	Non	Non
9	Petit	Bas	Oui	Oui
10	Sénior	Moyen	Oui	Oui
11	Petit	Moyen	Oui	Oui
12	Moyen	Moyen	Non	Oui
13	Moyen	Haut	Oui	Oui
14	Sénior	Moyen	Non	Non

On veut construire un arbre de décision pour prédire la classe (A acheté) d'un nouvel client, en se basant sur ce tableau. Utiliser la technique du *gain d'information* pour choisir l'attribut à la racine de l'arbre.
Où : $\boldsymbol{m}$ est le nombre de classes présentes dans D, $\boldsymbol{v}$ est le nombre de valeurs distinctes d'un attribut A.

Q.1.— En utilisant le gain d'information maximal, déterminer uniquement l'attribut racine de l'arbre (ayant le maximum de gain) puis en déduire l'arbre final sans faire de calculs supplémentaires.

Q.2.— Utiliser l'arbre de décision construit pour classifier et calculer la précision sur l'ensemble de test suivant :

Id	Âge	Salaire	Étudiant	A acheté
1	Petit	Haut	Non	Non
2	Sénior	Bas	Oui	Non
3	Moyen	Bas	Oui	Oui
4	Petit	Bas	Oui	Oui
5	Moyen	Moyen	Non	Oui
6	Sénior	Moyen	Non	Non

NB :

L'information avant le partitionnement d'un ensemble D	$\text{Info}(D) = -\sum_{i=1}^{m} p_i \times log_2(p_i)$
L'information après le partitionnement en utilisant un attribut un A	$\text{Info}_A(D) = \sum_{j=1}^{v} \frac{\lvert D_j \rvert}{\lvert D \rvert} \times \text{Info}(D_j)$
Gain d'information	$\text{Info (D)} - \text{Info}_A\text{(D)}$

Exercice 2. — Support Vector Machines *(07 points)*

On considère ici un problème de classification binaire vers $Y = \{-1, +1\}$ de données dans un espace de description $X \epsilon R^d$. On note $\{(x^i, y^i) \epsilon (X, Y), i \epsilon \{1, \ldots, n\}$ l'ensemble d'apprentissage considéré. La fonction de décision du classifieur considéré par $f_{w,b}(x) = sign(w^T x + b)$.

On considère dans un premier temps un ensemble de données linéairement séparable. Cet ensemble de données et la frontière sont représentés (en 2D) sur la Figure 1.

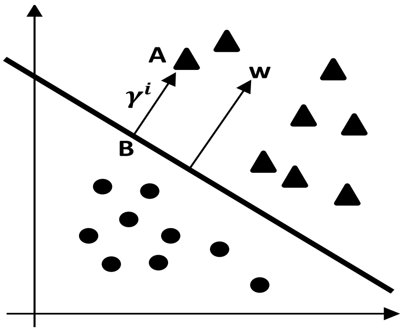


Figure 1 – Ensemble de données linéairement séparables

Sur cette figure, l'échantillon x^i et le label y^i est représenté par le point A. On s'intéresse à sa distance signée γ^i à la fonction de décision (dont le point le plus proche est représenté en B sur la figure).

Partie 1 : Formulation du SVM *(partie 1 sur 04 points)*

Q.1.1.— Sachant que $w/\|w\|$ est un vecteur unitaire orthogonal à la frontière de décision. Donner l'expression de la fonction de minimisation primale en fonction de x^i, y^i, w et b.

Q.1.2.— Proposer, selon les conditions de Karush-Kuhn-Tucker (KKT), une formulation duale à base de Lagrangien afin de résoudre ce problème d'optimisation quadratique de la question **Q.1.1**

Q.1.3.— Donner la solution analytique de la minimisation de ce Lagrangien selon les paramètres w et b.

Q.1.4.— Donner la fonction de décision (classification) optimale obtenue après optimisation de cette formulation duale du SVM qu'on notera $f(x)$.

Q.1.5.— Interpréter les résultats dans les cas suivants : $f(x) = 0$, $f(x) > 0$, $f(x) < 0$, $f(x) = +1$ et $f(x) = -1$.

Q.1.6.— Que cela implique-t-il si l'on souhaite éloigner au maximum (au sens géométrique) les points de la frontière de décision ?

Q.1.7.— D'après les conditions de Karush-Kuhn-Tucker (KKT) concernant les propriétés de la solution optimale d'un Lagrangien, on a : $\alpha_i\left(y^i\left(w^T x^i + b\right) - 1\right) \geq 0, \forall i \epsilon \{1..N\}$ Qu'en déduire pour les paramètres α_i obtenus à l'optimum ?

Q.1.8.— Quel rapport y a-t-il entre les α_i et les vecteurs supports de vecteurs.

Partie 2 : Application numérique (cas linéaire) *(partie 2 sur 03 points)*

Nous supposons que nous ayant les points de données étiquetés positivement suivants,

$$\left\{\begin{pmatrix}3\\1\end{pmatrix},\begin{pmatrix}3\\-1\end{pmatrix},\begin{pmatrix}6\\1\end{pmatrix},\begin{pmatrix}6\\-1\end{pmatrix}\right\}$$

et les points de données étiquetés négativement suivants,

$$\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\-1\end{pmatrix},\begin{pmatrix}-1\\0\end{pmatrix}\right\}$$

Q.2.1.— Représenter graphiquement les données d'apprentissage en indiquant les points positifs par des triangles (▲) et les points négatifs par des ronds (●).

Q.2.2.— Donner et signaler sur le graphe les vecteurs supports. Quel est leur nombre ?

Q.2.3.— On sait que ces vecteurs supports sont augmentés d'une dimension égale à 1 représentant le biais. En déduire les vecteurs supports S_i^* selon cette nouvelle dimension.

Q.2.4.— Écrire les systèmes linéaires paramétriques représentant la formulation du problème de classification en fonctions des paramètres α_i (multiplicateurs de Lagrange) des vecteurs supports S_i^* puis résoudre le système linéaire.

Q.2.5.— En déduire les valeurs du vecteurs $\hat{\omega}$ et du biais $\hat{b}$.

Q.2.6.— Tracer l'hyperplan séparateur.

Exercice 3. — Clustering par K-means et HCA *(05 points)*

Soit l'ensemble D des points suivants : D= { 2, 5, 8, 10, 11, 18, 20 }. On veut répartir les données de D en trois (3) clusters, en utilisant l'algorithme **K-means**. La distance **d1** entre un point **p** et un centre **c** est calculée ainsi : **d1(p , c) = |p – c|** (la valeur absolue de la différence entre le point **p** moins le centre **c**).

Q.1.— Appliquez **K-means** en choisissant comme centres initiaux des 3 clusters respectivement : 8, 10 et 11.

Q.2.— Refaire la même question **Q.1.** avec une autre distance **d2**, telle que :

d2(p, c) = max (0, p - c) (le maximum entre **0** et **(p – c)**).

Q.3.— Toujours avec la distance **d2**, appliquer, cette fois-ci, l'algorithme de Classification Hiérarchique Ascendante (**HCA**) sur l'ensemble D.

Q.4.— Tracer le dendrogramme correspondant au clustering obtenu par **HCA**.

NB : Montrez les différentes étapes de calcul sous forme de tableaux. La distance **d2** minimale doit être non nulle. Si le nombre est composé, on prendra le saut minimal.

Exercice 4. — Astuce Noyau *(04 points)*

Q.1.— Expliquer le principe général des méthodes à noyau.

Q.2.— Expliquer le concept de l'astuce noyau par le biais de l'application :

$$\Phi: X \to H$$

$$(x_1, x_2) \mapsto \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right)$$

Bonne Chance

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Ghardaïa
Faculté des Sciences et de la Technologie
Département des Mathématiques et d'Informatique

Concours d'accès à la formation de troisième cycle en vue de l'obtention du diplôme de doctorat au titre de l'année universitaire 2021-2022

Filière : Informatique

Spécialité : Systèmes Intelligents et Apprentissage Automatique

Module : Apprentissage Automatique **Durée : 02 heures** **Coefficient : 3**

Corrigé type

Exercice 1.— Arbres de Décision *(04 points)*

1- Construction de l'arbre

A- Attribut au nœud racine :

Les attributs candidats : 'Age', 'Salaire', 'Etudiant ?'
Utilisation de la technique du gain d'information pour choisir l'attribut :
Commençant par calculer Info(D), tel que D est l'ensemble d'apprentissage, ayant 14 instances et (m = 2) classes :

$$\text{Info}(D) = -\sum_{i=1}^{m} p_i \times log_2(p_i)$$

= - (9/14) * $\log_2$(9/14) – (5/14)*$\log_2$(5/14) = **0.940 bits** **(0,5 pts)**

On calcule maintenant : **Info$_{Age}$(D), Info$_{Salaire}$(D), Info$_{Etudiant\ ?}$(D).**

$$Info_{Age} = \sum_{j=1}^{v=3} \frac{|D_j|}{|D|} Info(D_j)$$

= (5/14) Info(D_1) + (4/14) Info(D_2) + (5/14) Info(D_3)
= (5/14) Info([2,3]) + (4/14) Info([4,0]) + (5/14) Info([3,2])
= (5/14) * 0.971 + (4/14) * 0 + (5/14) 0.971
= 0.693 bits **(0,5 pts)**

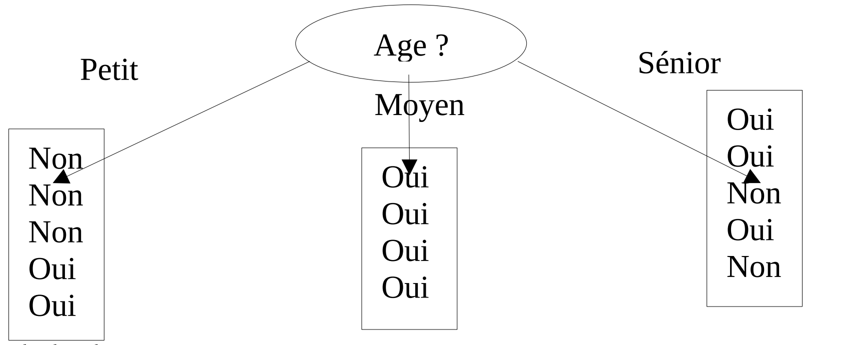
$$Info_{Salaire} = \sum_{j=1}^{v=3} \frac{|D_j|}{|D|} Info(D_j)$$

= (4/14) Info(D_1) + (6/14) Info(D_2) + (4/14) Info(D_3)
= (4/14) Info([2,2]) + (6/14) Info([4,2]) + (4/14) Info([3,1])
= (4/14) * 1 + (6/14) * 0.918 + (4/14) * 0.811
= 0.910 bits **(0,5 pts)**

$$Info_{Etudiant?} = \sum_{j=1}^{v=2} \frac{|D_j|}{|D|} Info(D_j)$$

= (7/14) Info(D_1) + (7/14) Info(D_2)
= (7/14) Info([6,1]) + (7/14) Info([3,4])
= (7/14) * 0.59 + (7/14) * 0.984
= **0.787 bits** **(0,5 pts)**
On peut alors calculer les gains pour tous les attributs :
Gain(Age) = 0.940 – 0.693 = **0.247**

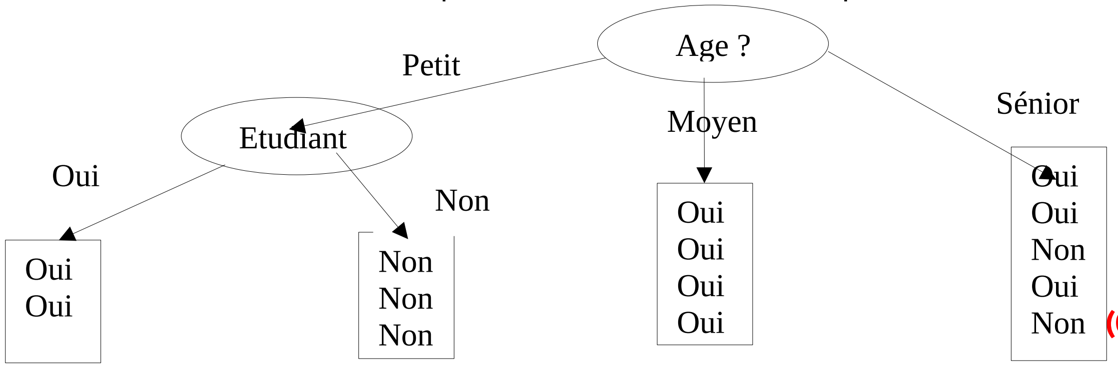
Gain(Salaire) = 0.940 – 0.910 = 0.030
Gain(Etudiant?) = 0.940 – 0.787 = 0.153
Le gain d'information est maximal en utilisant l'attribut '**Age**' pour le partitionnement à la racine de l'arbre de décision. L'attribut 'Age' possède trois valeurs/modalités (Petit, Moyen, Sénior), donc on aura trois branches à partir de la racine, une par valeur.



(0,5 pts)

B- Déduire le reste de l'arbre

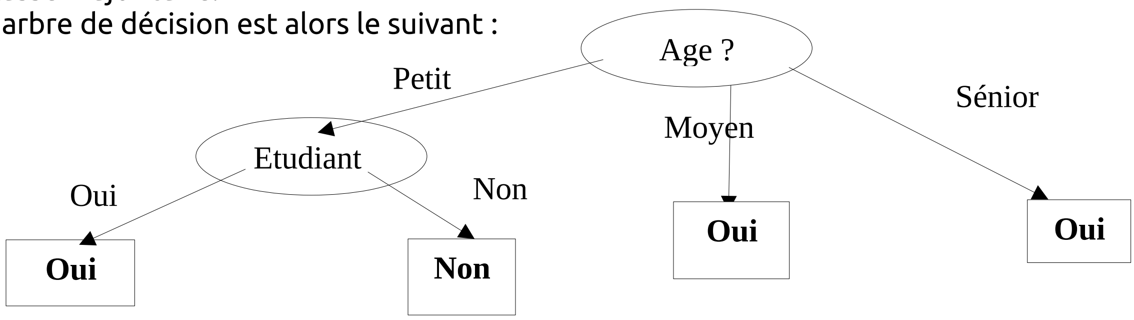
On procède à l'exploration, en excluant cette fois l'attribut 'Age'. La deuxième branche est pure, donc pas besoin de l'explorer. Pour la première branche (Age=Petit), on peut remarquer que la séparation en se basant sur l'attribut '**Etudiant**' permet d'avoir deux branches pures.



(0,5 pts)

Pour la troisième branche (Age=Sénior), on peut remarquer que la séparation en se basant sur l'attribut '**Salaire'** ou sur l'attribut '**Etudiant**' ne permettront pas donner des ensembles pures. On peut décider d'arrêter l'exploration. La classe choisie dans cette branche est 'Oui' puisque c'est la classe majoritaire.

L'arbre de décision est alors le suivant :



2- Précision de classification de l'ensemble de test **(0,5 pts)**

Id	Age	Salaire	Etudiant ?	A acheté ?	Décision classifieur
1	Petit	Haut	Non	Non	Non
2	Sénior	Bas	Oui	Non	Oui
3	Moyen	Bas	Oui	Oui	Oui
4	Petit	Bas	Oui	Oui	Oui
5	Moyen	Moyen	Non	Oui	Oui
6	Sénior	Moyen	Non	Non	Oui

Précision = # décision correctes / taille ensemble de test
= 4/6 = 0.66 **(0,5 pts)**

Exercice 2. — SVM *(07 points)*

Q.1.1— Expression de la fonction de minimisation primale en fonction de x^i, y^i, w et

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2$$

0,25 point

$$s.t. \qquad y^i(\mathbf{w}^T\mathbf{x}^i + b) \geq 1, \forall i \in \{1, \ldots, n\}$$

Q.1.2— Formulation duale

0,25 point

$$\begin{cases} L(\omega, b, \alpha) = \frac{1}{2}\|\omega\|^2 - \sum_i \alpha_i[y_i[\langle x_i, \omega\rangle + b] - 1] \\ \forall\ i\ \alpha_i \geq 0 \end{cases}$$

Q.1.3— Solution analytique

$$\nabla_\omega L = \omega - \sum_i \alpha_i y_i x_i = 0 \implies \omega = \sum_i \alpha_i y_i x_i$$

0,5 point

0,5 point

$$\nabla_b L = \sum \alpha_i y^i = 0$$

Q.1.4— Fonction de décision optimale

$$\hat{f}(x) = \langle\hat{\omega}, x\rangle + \hat{b}$$

0,25 point

où $\hat{\omega}$ et $\hat{b}$ les valeurs estimés de ω et b dont la solution est :

$$\sum_i \alpha_i - \frac{1}{2} = \sum_i \sum_j \alpha_i\, \alpha_j y_i y_j \langle x_i, x_j\rangle$$

$$avec \sum_i \alpha_i y_i = 0$$

Q.1.5— Valeurs de f*(x)

1,25 point

$f^*(x) = 0$: on sur la frontière
$f^*(x) > 0$: Classe «+ »
$f^*(x) < 0$: Classe « - »
$f^*(x) = +1$: on est sur la droite délimitant les vecteurs support de la classe « + »
$f^*(x) = -1$: on est sur la droite délimitant les vecteurs support de la classe « - »

Q.1.6—

0,5 point

Si l'on s'éloigne au maximum des points de la frontières de décision la marge de la décision séparant les points « + » et « - » augmente.

Q.1.7—

0,5 point

$\alpha_i = 0 \implies y_i(\langle \omega, x_i\rangle + b) > 1$
$\alpha_i > 0 \implies y_i(\langle \omega, x_i\rangle + b) = 1$

Q.1.8—

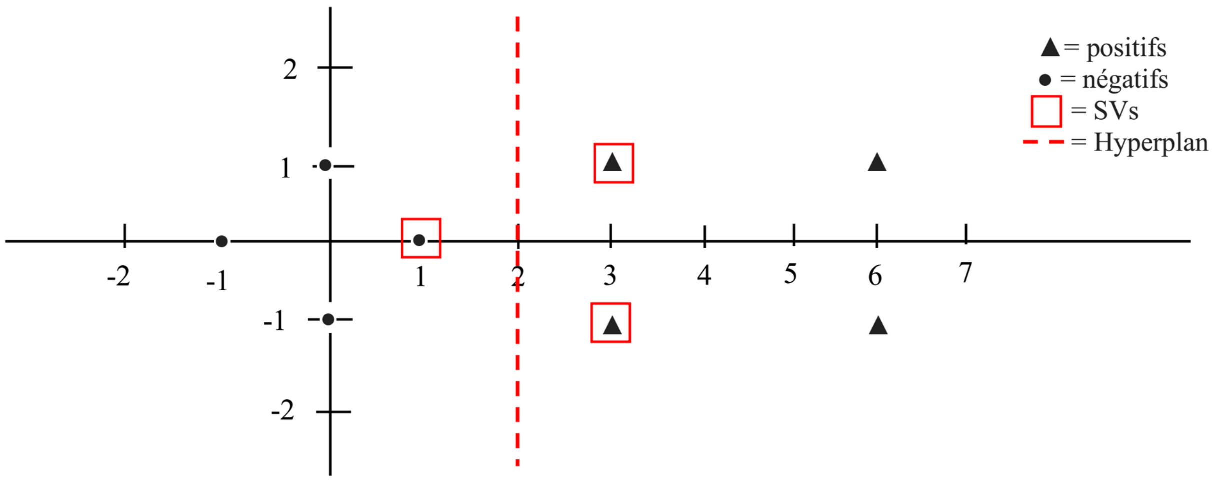
0,25 point

Les $\alpha_i > 0$ pour les données vecteurs supports et nulles en dehors des vecteurs supports.

Partie 2 — Application

Q.2.1— Représentation graphique + Nombre de VS

0,5 point



Q.2.2— Repérage sur le graphe + Nombre de VS

0,5 point

Nombre de vecteur support = 3

Les SVs :

$$\left\{S_1 = \begin{pmatrix}1\\0\end{pmatrix},\ S_2 = \begin{pmatrix}3\\1\end{pmatrix},\ S_3 = \begin{pmatrix}3\\-1\end{pmatrix}\right\}$$

Q.2.3— Vecteurs support augmentés d'une dimension

Donc après l'application des SVs, on obtient une unité supplémentaire :

0,25 point

$$\left\{S_1^* = \begin{pmatrix}1\\0\\1\end{pmatrix}, S_2^* = \begin{pmatrix}3\\1\\1\end{pmatrix}, S_3^* = \begin{pmatrix}3\\-1\\1\end{pmatrix}\right\}$$ Données positives en 3 d

Q.2.4— Vecteurs support augmentés d'une dimension

0,75 point

$$\begin{cases}\alpha_1 S_1^* . S_1^* + \alpha_2 S_2^* . S_1^* + \alpha_3 S_3^* . S_1^* = -1\\ \alpha_1 S_1^* . S_2^* + \alpha_2 S_2^* . S_2^* + \alpha_3 S_3^* . S_2^* = +1\\ \alpha_1 S_1^* . S_3^* + \alpha_2 S_2^* . S_3^* + \alpha_3 S_3^* . S_3^* = +1\end{cases}$$

$$\alpha_1\begin{pmatrix}1\\0\\1\end{pmatrix}.\begin{pmatrix}1\\0\\1\end{pmatrix}+\alpha_2\begin{pmatrix}3\\1\\1\end{pmatrix}.\begin{pmatrix}1\\0\\1\end{pmatrix}+\alpha_3\begin{pmatrix}3\\-1\\1\end{pmatrix}.\begin{pmatrix}1\\0\\1\end{pmatrix}=-1$$

$$\alpha_1\begin{pmatrix}1\\0\\1\end{pmatrix}.\begin{pmatrix}3\\1\\1\end{pmatrix}+\alpha_2\begin{pmatrix}3\\1\\1\end{pmatrix}.\begin{pmatrix}3\\1\\1\end{pmatrix}+\alpha_3\begin{pmatrix}3\\-1\\1\end{pmatrix}.\begin{pmatrix}3\\1\\1\end{pmatrix}=+1$$

$$\alpha_1\begin{pmatrix}1\\0\\1\end{pmatrix}.\begin{pmatrix}3\\-1\\1\end{pmatrix}+\alpha_2\begin{pmatrix}3\\1\\1\end{pmatrix}.\begin{pmatrix}3\\-1\\1\end{pmatrix}+\alpha_3\begin{pmatrix}3\\-1\\1\end{pmatrix}.\begin{pmatrix}3\\-1\\1\end{pmatrix}=+1$$

$$\alpha_1(1+0+1)+\alpha_2(3+0+1)+\alpha_3(3+0+1)=-1$$
$$\alpha_1(3+0+1)+\alpha_2(9+1+1)+\alpha_3(9-1+1)=+1$$
$$\alpha_1(3+0+1)+\alpha_2(9-1+1)+\alpha_3(9+1+1)=+1$$

$$2\alpha_1+4\alpha_2+4\alpha_3=-1$$
$$4\alpha_1+11\alpha_2+9\alpha_3=+1$$
$$4\alpha_1+9\alpha_2+11\alpha_3=+1$$

$$\boxed{\Rightarrow \begin{cases}\alpha_1=-3{,}5\\ \alpha_2=0{,}75\\ \alpha_3=0{,}75\end{cases}}$$

Q.2.5— Valeurs du vecteurs $\widehat{\boldsymbol{\omega}}$ et du biais $\widehat{\boldsymbol{b}}$ 0,**5** point

On sait que $\omega=\sum_i \alpha_i y_i x_i$ et que par conséquent pour les vecteurs de support $\widehat{\omega}=\sum_i \alpha_i S_i^*$

Donc $\widehat{\omega}=-3{,}5\begin{pmatrix}1\\0\\1\end{pmatrix}+0{,}75\begin{pmatrix}3\\1\\1\end{pmatrix}+0{,}75\begin{pmatrix}3\\-1\\1\end{pmatrix}=\begin{pmatrix}1\\0\\2\end{pmatrix}$

On sait que pour l'hyperplan qui constitue la meilleure fonction de décision $\hat{y}=\hat{f}(x)=\langle\widehat{\omega},x\rangle+\hat{b}$ et que le support de vecteur qui le porte est d'augmenté d'une dimension qui constitue le biais.

Donc nous pouvons le décomposer en $\widehat{\omega}=\begin{pmatrix}1\\0\end{pmatrix}$ *et* $\hat{b}=-2$

Q.2.6— Tracer l'hyperplan séparateur 0,25 point

Voir graphe la ligne en pointillé

Exercice 3. — Clustering par K-means et HCA (05 points)

Q.1.— K-means pour d1(p , c) = |p − c|

µ1 = 8, **µ2** = 10, **µ3** = 11

Ptx	**2**	**5**	**8**	10	11	18	20
d(µ1)	6	3	0	2	3	10	12
d(µ2)	8	5	2	0	1	8	10
d(µ3)	9	6	3	1	0	7	9
Classe	C1	C1	C1	C2	C3	C3	C3

C1 = {2 ; 5 ; 8}, C2={10}, C3={11 ; 18 ; 20}

0,**5** point

µ1 = 5, **µ2** = 10, **µ3** = 16,3

Ptx	**2**	**5**	**8**	10	11	18	20
d(µ1)	3	0	3	5	6	13	15
d(µ2)	8	5	2	0	1	8	10
d(µ3)	14,33	11,33	8,33	6,33	5,33	2,33	4,33
Classe	C1	C1	C2	C2	C2	C3	C3

C1 = {2 ; 5 }, C2={8, 10, 11}, C3={18 ; 20}

0,**5** point

µ1 = 3,5, **µ2** = 9,66, **µ3** = 19

Ptx	**2**	**5**	**8**	10	11	18	20
d(µ1)	1,5	1,5	4,5	6,5	7,5	14,5	16,5
d(µ2)	7,66	4,66	1,66	0,33	1,33	8,34	10,34
d(µ3)	17	14	11	9	8	1	1
Classe	C1	C1	C2	C2	C2	C3	C3

C1 = {2 ; 5 }, C2={8, 10, 11}, C3={18 ; 20}

0,**5** point

Q.2.— K-means pour d2(p , c) = max (0, p − c)

µ1 = 8, **µ2** = 10, **µ3** = 11

Ptx	**2**	**5**	**8**	10	11	18	20
d(µ1)	0	0	0	2	2	10	12
d(µ2)	0	0	0	0	1	8	10
d(µ3)	0	0	0	0	0	7	9
Classe	C1	C2	C3	C3	C3	C3	C3

C1 = {2}, C2={5}, C3={8; 10; 11; 18; 20}

0,**5** point

µ1 = **2**, **µ2** = **5**, **µ3** = 13,4

Ptx	**2**	**5**	**8**	10	11	18	20
d(µ1)	0	3	6	8	9	16	18
d(µ2)	0	0	3	5	6	13	15
d(µ3)	0	0	0	0	0	4,6	6,6
Classe	C1	C2	C3	C3	C3	C3	C3

C1 = {2}, C2={5}, C3={8; 10; 11; 18; 20}

0,**5** point

Q.3.— HCA pour d2(p , c) = max (0, p − c)

	2	**5**	**8**	10	11	18	20
2	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0
8	6	3	0	0	0	0	0
10	8	5	2	0	0	0	0
11	9	6	3	**1**	0	0	0
18	16	13	10	8	7	0	0
20	18	15	12	10	9	2	0

0,**25** point

Joindre 10 et 11 ➔ {10, 11}

	2	**5**	**8**	{10, 11}	18	20
2	0	0	0	0	0	0
5	3	0	0	0	0	0
8	6	3	0	0	0	0
{10, 11}	8	5	2	0	0	0
18	16	13	10	8	0	0
20	18	15	12	10	**2**	0

0,**25** point

Joindre 18 et 20 ➔ {18, 20}

	2	**5**	**8**	{10, 11}	{18, 20}
2	0	0	0	0	0
5	3	0	0	0	0
8	6	3	0	0	0
{10, 11}	8	5	**2**	0	0
{18, 20}	16	13	10	7	0

0,**25** point

Joindre **8** et {10, 11} ➔ {8, 10, 11}

	2	**5**	{8, 10, 11}	{18, 20}
2	0	0	0	0
5	**3**	0	0	0
{8, 10, 11}	8	5	0	0
{18, 20}	16	13	7	0

0,**25** point

Joindre 2 et 5 ➔ {2, 5}

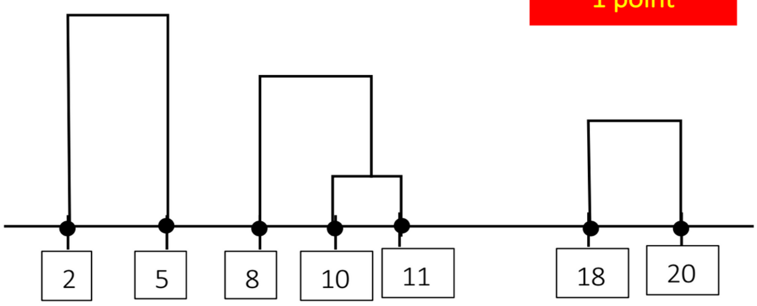
	{2, 5}	{8, 10, 11}	{18, 20}
{2, 5}	0	0	0
{8, 10, 11}	5	0	0
{18, 20}	13	7	0

0,**25** point

3 clusters : C1 = {2,5} **;** C2 ={8,10,11} **;** C3={18, 20} 0,**25** point

Q.4.— Dendogramme

1 point



Exercice 4. — Astuce noyau (04.00 points)

1. Le principe des méthodes à noyaux peut être décrit comme suit : (3 * 0.50 = 1.50 pts)
   - Utiliser une application (projection) Φ pour incorporer les exemples depuis un espace d'entrée X à un espace (vectoriel) de redescription H (feature space) de haute dimension.
   - Dans H, les motifs (patterns) peuvent être découvertes en tant que relations linéaires.
   - Nous pouvons calculer le produit scalaire des exemples dans H directement à partir des données de l'espace d'entrée X en utilisant une fonction Noyau.
2. L'astuce noyau (kernel trick) (2 (calcul) + 0.5 (conclusion) = 2.50 pts)

Soient $\mathrm{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\mathrm{X}' = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$

$$\begin{aligned}
\langle \boldsymbol{\Phi}(\boldsymbol{X}), \boldsymbol{\Phi}(\boldsymbol{X}') \rangle = {} & \langle \boldsymbol{\Phi}(\boldsymbol{x_1}, \boldsymbol{x_2}), \boldsymbol{\Phi}(\boldsymbol{x_1'}, \boldsymbol{x_2'}) \rangle \\
& = \langle (\mathbf{1}, \sqrt{\mathbf{2}}\boldsymbol{x_1}, \sqrt{\mathbf{2}}\boldsymbol{x_2}, \boldsymbol{x_1^2}, \boldsymbol{x_2^2}, \sqrt{\mathbf{2}}\boldsymbol{x_1}\boldsymbol{x_2}), (\mathbf{1}, \sqrt{\mathbf{2}}\boldsymbol{x_1'}, \sqrt{\mathbf{2}}\boldsymbol{x_2'}, \boldsymbol{x_1'^2}, \boldsymbol{x_2'^2}, \sqrt{\mathbf{2}}\boldsymbol{x_1'}\boldsymbol{x_2'}) \rangle \\
& = \mathbf{1} + \mathbf{2}\boldsymbol{x_1}\boldsymbol{x_1'} + \mathbf{2}\boldsymbol{x_2}\boldsymbol{x_2'} + \boldsymbol{x_1^2}\boldsymbol{x_1'^2} + \boldsymbol{x_2^2}\boldsymbol{x_2'^2} + \mathbf{2}\,\boldsymbol{x_1}\boldsymbol{x_2}\boldsymbol{x_1'}\boldsymbol{x_2'} = (\mathbf{1} + \langle \boldsymbol{X}, \boldsymbol{X}' \rangle)^{\mathbf{2}} \\
& = \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}').
\end{aligned}$$

Donc le calcul de la similarité entre $\Phi(X)\,et\ \Phi(X')$ se traduit par un produit scalaire entre X et X' (un noyau polynomial).