

**Enseignante : Dr. BERMAD**

**Durée: 1h30**

**TD 2 Data Mining:  
Clustering**

**A l'intention de: L3STID**

**Exercice 1 :**

Dans une étude industrielle, on a étudié 2 caractères:  $X_1$  et  $X_2$ , sur 6 individus  $W_1, W_2, \dots, W_6$ .  
Les données recueillies sont:

	$X_1$	$X_2$
$W_1$	-2	2
$W_2$	-2	-1
$W_3$	0	-1
$W_4$	2	2
$W_5$	-2	3
$W_6$	3	0

- On fait une classification par l'algorithme des centres mobiles avec, pour centres initiaux,  $c_1^0$  de coordonnées (1, 1) et  $c_2^0$  de coordonnées (2,3). Quels sont les clusters obtenus ?
- On considère la matrice de données  $X$  dans  $\mathbb{R}^2$  définie par:

$$X = \begin{pmatrix} 2 & 2 \\ 7.5 & 4 \\ 3 & 3 \\ 0.5 & 5 \\ 6 & 4 \end{pmatrix}$$

- Appliquer l'algorithme CAH (Clustering Ascendant Hiérarchique) en utilisant le saut minimal (single link).
- Tracer le dendrogramme correspondant

**Exercice2 :**

On interroge 6 individus en leur demandant leur sexe  $X_1$  (F : femme, H : homme), leur type de logement  $X_2$  (R : rural, U : urbain) et leur état civil  $X_3$  (C : célibataire, M : marié, A : autre).

On obtient :

	<b>X1</b>	<b>X2</b>	<b>X3</b>
<b>W1</b>	H	U	C
<b>W2</b>	F	U	C
<b>W3</b>	F	R	M
<b>W4</b>	F	U	A
<b>W5</b>	H	R	M
<b>W6</b>	H	R	A

1. En considérant l'indice de Jaccard, calculer  $d(w_1, w_2)$  et  $d(w_3, w_6)$ .
2. Est-ce que  $w_1$  est plus proche de  $w_2$ , que  $w_3$  de  $w_6$  ?