

## EMD SAD

**Exercice 1 : 12 pts**

On veut réaliser une étude sur un ensemble de ressources sur des cartes électroniques, on a 6 cartes, notés C1, C2, C3, C4, C5, C6, et 3 PCs à dispositions, P1, P2, P3. La matrice contient le nombre de ressources communes sur 2 éléments, par exemple 11 ressources sont présentes sur la carte C1 et 7 ressources sur la carte C1 et C2.

	C1	C2	C3	C4	C5	C6	P1	P2	P3
C1	11	7	4	6	6	7	4	4	3
C2	7	15	8	8	9	6	3	3	2
C3	4	8	13	7	7	4	2	3	2
C4	6	8	7	15	7	6	6	8	6
C5	6	9	7	7	12	4	3	5	4
C6	7	6	4	6	4	10	6	5	5
P1	4	3	2	6	3	6	13	10	9
P2	4	3	3	8	5	5	10	15	11
P3	3	2	2	6	4	5	9	11	12

On veut regrouper ces 9 composants en 3 composants, ou on a les mêmes composants. Pour cela on mesure la distance entre deux composants par la formule suivante :

$$d(A,B) = (nA + nB - 2nAB)/(nA + nB)$$

ou  $nA$  (resp.  $nB$ ) désigne le nombre de ressources présentes dans le composant  $A$  (resp. sur le composant  $B$ ) et  $nAB$  le nombre de ressources en commun entre les composants  $A$  et  $B$ . On obtient la matrice des distances suivante : a) On vous demande de donner la matrice des distances, b) De faire une Classification ascendante hiérarchique, c) De tracer le dendrogramme, et de décider combien de classes il est possible de retenir.

**Exercice 2 : 8 Pts**

On a une base de données contenant des étudiants la décision finale du jury, et leurs notes, 5 couples « entrée-sortie » : (Zaki, Admis), (Nada, Admis), (Ali, Ajourné), (Amir, Ajourné) et (Neila, Admis).

Les notes des étudiants sont pour les 4 matières étudiées :

(Zaki, 14, 12, 8 et 12), (Nada, 12, 12, 6 et 10), (Ali, 8, 9, 9 et 1), (Amir, 15, 11, 3 et 5), (Neila, 12, 9, 14 et 11).

L'étudiant Samy se présente avec ces notes qui sont : 9, 14, 15 et 6.

Avec l'algorithme  $k$  plus proches voisins ( $k=3$ ) et en choisissant la distance suivante : Distance = Racine (Somme (ABS( $X_i - X_j$ ))). Déterminez la classe de Samy.

## EMD SAD

**Exercice 1 : 12 pts**

On veut réaliser une étude sur un ensemble de ressources sur des cartes électroniques, on a 6 cartes, notés C1, C2, C3, C4, C5, C6, et 3 PCs à dispositions, P1, P2, P3. La matrice contient le nombre de ressources communes sur 2 éléments, par exemple 11 ressources sont présentes sur la carte C1 et 7 ressources sur la carte C1 et C2.

	C1	C2	C3	C4	C5	C6	P1	P2	P3
C1	11	7	4	6	6	7	4	4	3
C2	7	15	8	8	9	6	3	3	2
C3	4	8	13	7	7	4	2	3	2
C4	6	8	7	15	7	6	6	8	6
C5	6	9	7	7	12	4	3	5	4
C6	7	6	4	6	4	10	6	5	5
P1	4	3	2	6	3	6	13	10	9
P2	4	3	3	8	5	5	10	15	11
P3	3	2	2	6	4	5	9	11	12

On veut regrouper ces 9 composants en 3 composants, ou on a les mêmes composants. Pour cela on mesure la distance entre deux composants par la formule suivante :

$$d(A,B) = (nA + nB - 2nAB)/(nA + nB)$$

ou  $nA$  (resp.  $nB$ ) désigne le nombre de ressources présentes dans le composant  $A$  (resp. sur le composant  $B$ ) et  $nAB$  le nombre de ressources en commun entre les composants  $A$  et  $B$ . On obtient la matrice des distances suivante : a) On vous demande de donner la matrice des distances, b) De faire une Classification ascendante hiérarchique, c) De tracer le dendrogramme, et de décider combien de classes il est possible de retenir.

**Exercice 2 : 8 Pts**

On a une base de données contenant des étudiants la décision finale du jury, et leurs notes, 5 couples « entrée-sortie » : (Zaki, Admis), (Nada, Admis), (Ali, Ajourné), (Amir, Ajourné) et (Neila, Admis).

Les notes des étudiants sont pour les 4 matières étudiées :

(Zaki, 14, 12, 8 et 12), (Nada, 12, 12, 6 et 10), (Ali, 8, 9, 9 et 1), (Amir, 15, 11, 3 et 5), (Neila, 12, 9, 14 et 11).

L'étudiant Samy se présente avec ces notes qui sont : 9, 14, 15 et 6.

Avec l'algorithme  $k$  plus proches voisins ( $k=3$ ) et en choisissant la distance suivante : Distance = Racine (Somme (ABS( $X_i - X_j$ ))). Déterminez la classe de Samy.

Un laboratoire d'écologie étudie les espèces micro-animales (larves, ...) présentes dans les rivières et les étangs. Il réalise, dans 6 sites de rivière, notes  $R1, R2, R3, R4, R5$  et  $R6$ , et 3 sites d'étangs, notes  $E1, E2$  et  $E3$ , des prélèvements répétés qui lui permettent d'avancer une liste des espèces présentes dans chacun de ces sites et de repérer les espèces présentes dans plusieurs sites à la fois. La matrice suivante contient, pour chaque paire de sites  $A$  et  $B$ , le nombre d'espèces communes aux 2 sites. Ainsi on y lit par exemple que 11 espèces sont présentes au site  $R1$  et qu'il y a 7 espèces présentes à la fois au site  $R1$  et au site  $R2$ .

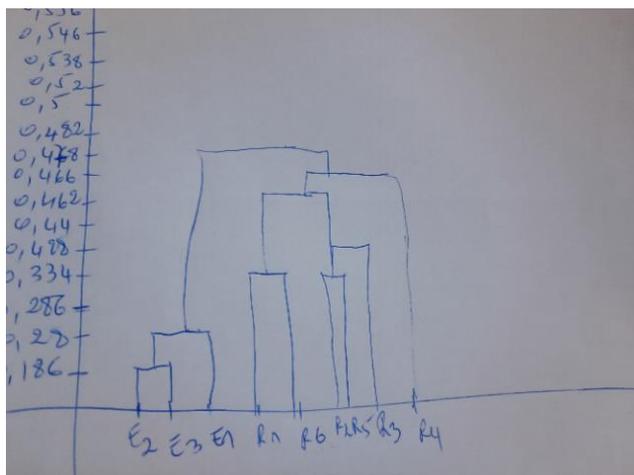
	$R1$	$R2$	$R3$	$R4$	$R5$	$R6$	$E1$	$E2$	$E3$
$R1$	11	7	4	6	6	7	4	4	3
$R2$	7	15	8	8	9	6	3	3	2
$R3$	4	8	13	7	7	4	2	3	2
$R4$	6	8	7	15	7	6	6	8	6
$R5$	6	9	7	7	12	4	3	5	4
$R6$	7	6	4	6	4	10	6	5	5
$E1$	4	3	2	6	3	6	13	10	9
$E2$	4	3	3	8	5	5	10	15	11
$E3$	3	2	2	6	4	5	9	11	12

On se propose de regrouper les 9 sites en trois ou quatre classes composées de sites où se sont pratiquement les mêmes espèces qui sont présentes. Pour réaliser cette classification, on propose de mesurer la distance entre deux sites  $A$  et  $B$  par la formule

$$d(A,B) = (nA + nB - 2nAB) / (nA + nB)$$

ou  $nA$  (resp.  $nB$ ) désigne le nombre d'espèces présentes au site  $A$  (resp. au site  $B$ ) et  $nAB$  le nombre d'espèces en commun entre les sites  $A$  et  $B$ . On obtient la matrice des distances suivante :

	$R1$	$R2$	$R3$	$R4$	$R5$	$R6$	$E1$	$E2$	$E3$
$R1$	0	0,462	0,666	0,538	0,478	0,334	0,666	0,692	0,74
$R2$	0,462	0	0,428	0,466	0,334	0,52	0,786	0,8	0,852
$R3$	0,666	0,428	0	0,5	0,44	0,652	0,846	0,786	0,84
$R4$	0,538	0,466	0,5	0	0,482	0,52	0,572	0,466	0,556
$R5$	0,478	0,334	0,44	0,482	0	0,636	0,76	0,63	0,666
$R6$	0,334	0,52	0,652	0,52	0,636	0	0,478	0,6	0,546
$E1$	0,666	0,786	0,846	0,572	0,76	0,478	0	0,286	0,28
$E2$	0,692	0,8	0,786	0,466	0,63	0,6	0,286	0	0,186
$E3$	0,74	0,852	0,84	0,556	0,666	0,546	0,28	0,186	0



Exercice 2 :

Les 3 distances les plus proches de Samy sont : 4,24 de Nada ; de 4,12 de Ali ; de 3,74 de Neila.

La classe la plus représentée par ces 3 voisins (en utilisant le Vote) est admis, donc on déclare Samy comme Admis.