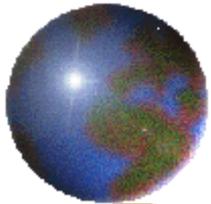


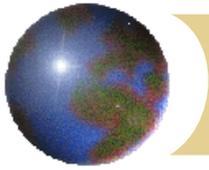


Université Badji Mokhtar Annaba  
Département de l'informatique  
Master 1 I.A.T.I



# *Reconnaissance de Formes* *(R.d.F)*

***Chargé de module :***  
***Dr Mohamed Amine YAKOUBI***

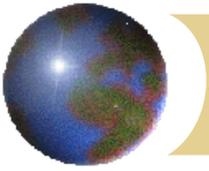


# *Objectifs*

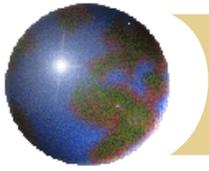
❖ *Rappel sur Reconnaissance des Formes ?*

❖ *Classification ?*

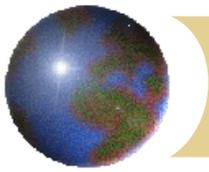
❖ *Algorithme K-MEANS*



*Rappel*  
*Reconnaissance des Formes*



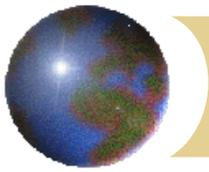
# *Classification*



# *Classification*

## *Définition?*

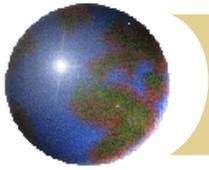
- ✚ La classification est une méthode mathématique **d'analyse de données**, il est appliqué sur des données numériques (**points, tableaux, images, sons, . . .etc.** ), pour faciliter l'étude d'une **population d'effectif important**.



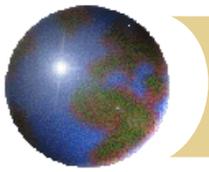
# *Classification*

## *Méthodes de classification?*

- ✚ Méthodes supervisées (classement)
- ✚ Méthodes non supervisées (classification / Clustering)



# *K Plus Proche Voisins (KPPV / KNN)*

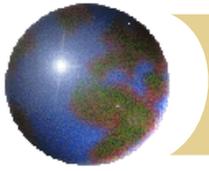


# *Classification*

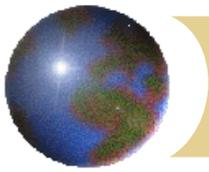
## *Méthodes de classification?*

- ✚ Méthodes supervisées (classement)

- ✚ Méthodes non supervisées (classification / Clustering)



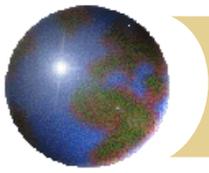
# *K moyennes (K-Means)*



# *K-Means*

## *Définition?*

- ✚ est une méthode de **partitionnement des données**. Étant donné des **points** et un entier  $k$ , le problème est de **diviser** les points en  $k$  groupes, souvent appelés **clusters**, de façon à **minimiser** une certaine fonction. On considère **la distance** d'un point à la **moyenne des points** de son cluster ; la fonction à minimiser est la **somme** des carrés de ces distances.

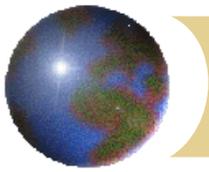


# *K-Means*

## *Algorithme Application?*



Exemple

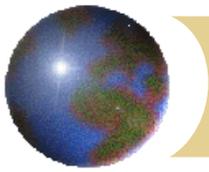


# *K-Means*

## *Exercise?*

- **Example**

Cluster the following eight points (with  $(x, y)$  representing locations) into three clusters A1(2, 10) A2(2, 5) A3(8, 4) A4(5, 8) A5(7, 5) A6(6, 4) A7(1, 2) A8(4, 9).



# *K-Means*

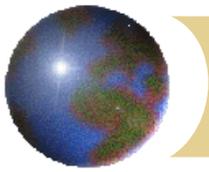
## *Exercise?*

- **Example**

Cluster the following eight points (with  $(x, y)$  representing locations) into three clusters A1(2, 10) A2(2, 5) A3(8, 4) A4(5, 8) A5(7, 5) A6(6, 4) A7(1, 2) A8(4, 9).

The distance function between two points  $a=(x1, y1)$  and  $b=(x2, y2)$  is defined as:  $\rho(a, b) = |x2 - x1| + |y2 - y1|$ .

- Use k-means algorithm to find the three cluster centers after the second iteration.



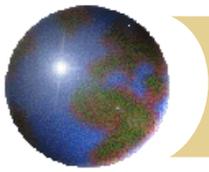
# *K-Means*

## *Exercise?*

Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

Iteration 1

		(2, 10)	(5, 8)	(1, 2)	
	Point	Dist Mean 1	Dist Mean 2	Dist Mean 3	Cluster
A1	(2, 10)				
A2	(2, 5)				
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				

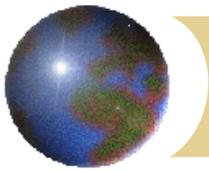


# *K-Means*

## *Exercise?*

Iteration 1

		(2, 10)	(5, 8)	(1, 2)	
	<b>Point</b>	<b>Dist Mean 1</b>	<b>Dist Mean 2</b>	<b>Dist Mean 3</b>	<b>Cluster</b>
A1	(2, 10)	0	5	9	1
A2	(2, 5)	5	6	4	3
A3	(8, 4)	12	7	9	2
A4	(5, 8)	5	0	10	2
A5	(7, 5)	10	5	9	2
A6	(6, 4)	10	5	7	2
A7	(1, 2)	9	10	0	3
A8	(4, 9)	3	2	10	2



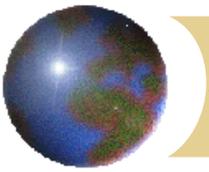
# *K-Means*

## *Exercise?*

Cluster 1  
(2, 10)

Cluster 2  
(8, 4)  
(5, 8)  
(7, 5)  
(6, 4)  
(4, 9)

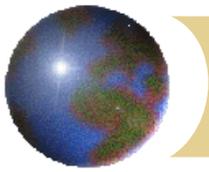
Cluster 3  
(2, 5)  
(1, 2)



# *K-Means*

## *Exercise?*

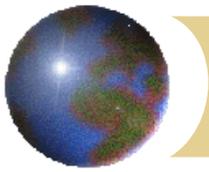
- Next, we need to re-compute the new cluster centers (means). We do so, by taking the mean of all points in each cluster.



# *K-Means*

## *Exercise?*

		(2, 10)	(6, 6)	(1.5, 3.5)	
	Point	Dist Mean 1	Dist Mean 2	Dist Mean 3	<b>Cluster</b>
A1	(2, 10)				
A2	(2, 5)				
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				



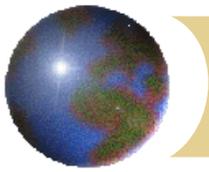
# *K-Means*

## *Exercise?*

- Next, we need to re-compute the new cluster centers (means). We do so, by taking the mean of all points in each cluster.

- 
- 
- 

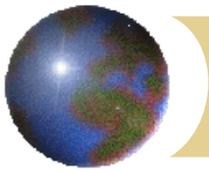




# *K-Means*

## *Exercise?*

		(3, 9.5)	(6.5, 5.25)	(1.5, 3.5)	
	Point	Dist Mean 1	Dist Mean 2	Dist Mean 3	<b>Cluster</b>
A1	(2, 10)				
A2	(2, 5)				
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				

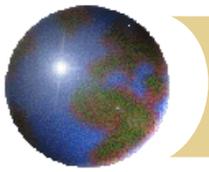


# *K-Means*

## *Exercise?*

- Next, we need to re-compute the new cluster centers (means). We do so, by taking the mean of all points in each cluster.

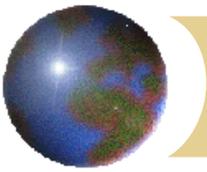




# *K-Means*

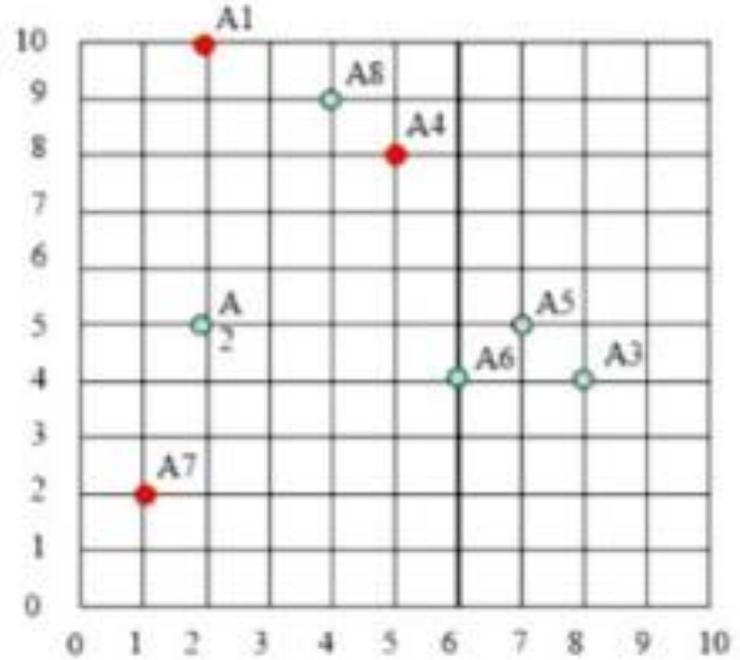
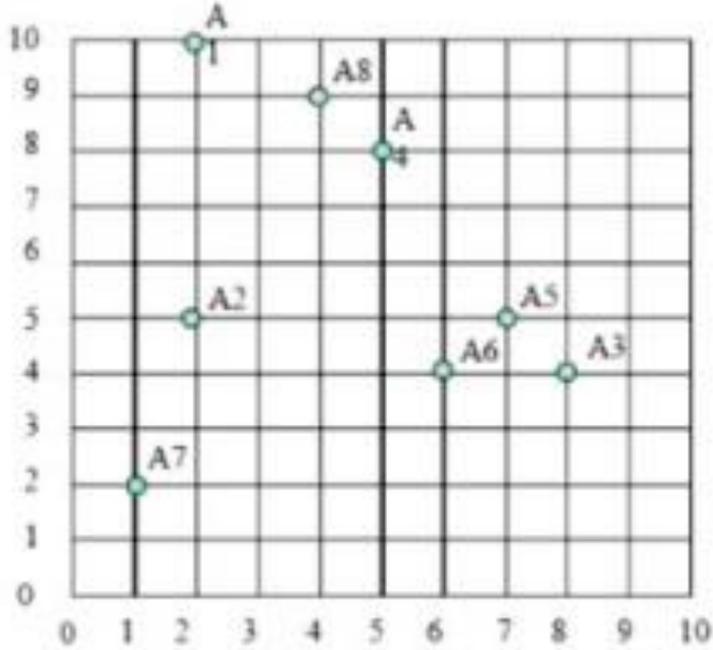
## *Exercise?*

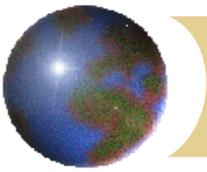
		(3.67, 9)	(7, 4.3)	(1.5, 3.5)	
	Point	Dist Mean 1	Dist Mean 2	Dist Mean 3	Cluster
A1	(2, 10)	2.67	10.7	7.5	1
A2	(2, 5)	5.67	10.7	7.5	1
A3	(8, 4)	9.67	10.7	7.5	1
A4	(5, 8)	2.67	10.7	7.5	1
A5	(7, 5)	7.67	10.7	7.5	1
A6	(6, 4)	7.67	10.7	7.5	1
A7	(1, 2)	9.67	10.7	7.5	1
A8	(4, 9)	0.67	10.7	7.5	1



# *K-Means*

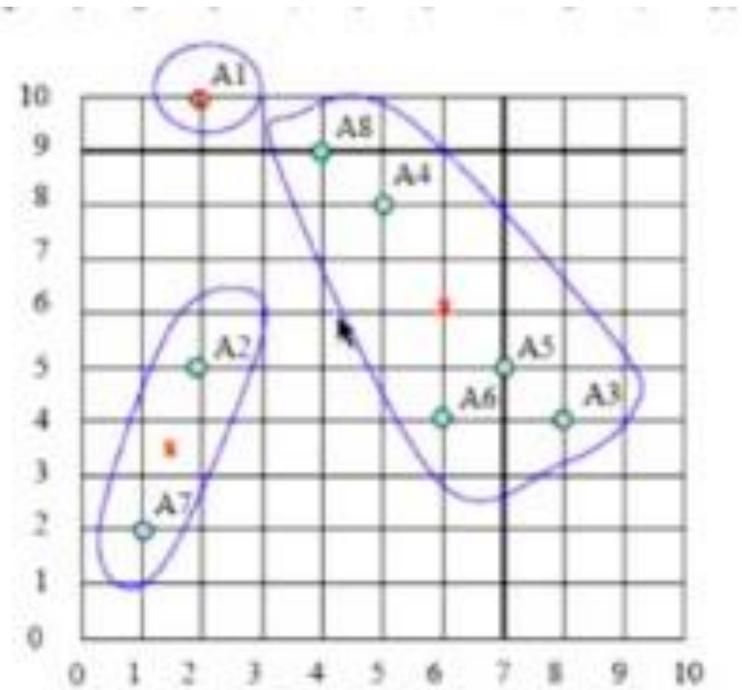
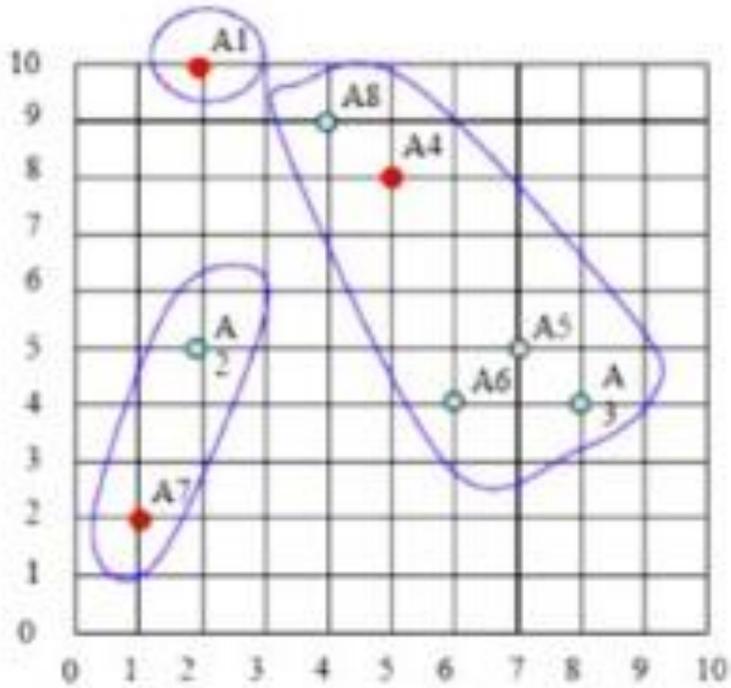
## *Exercice?*

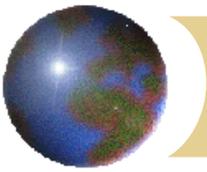




# *K-Means*

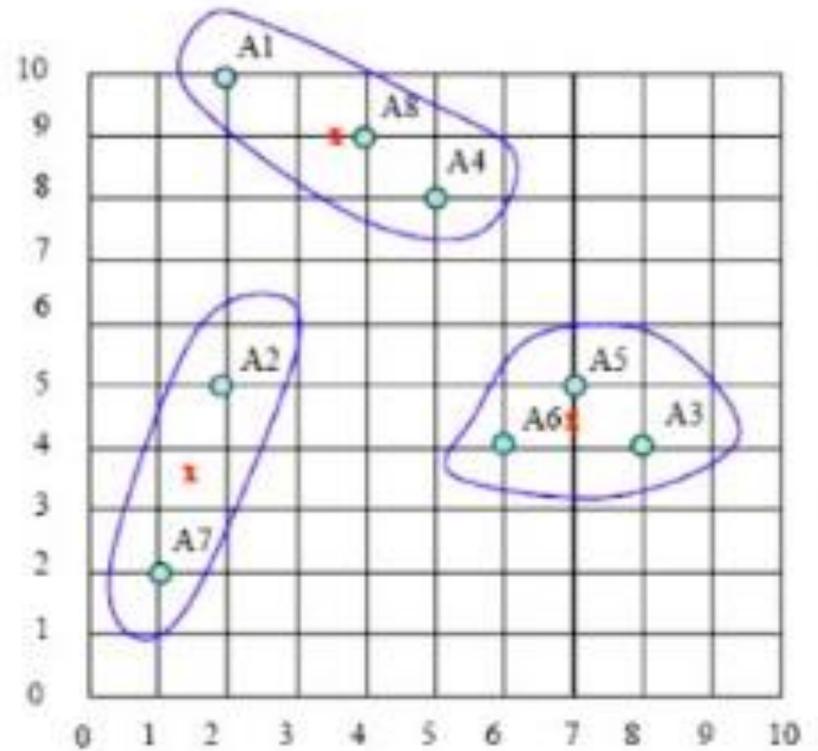
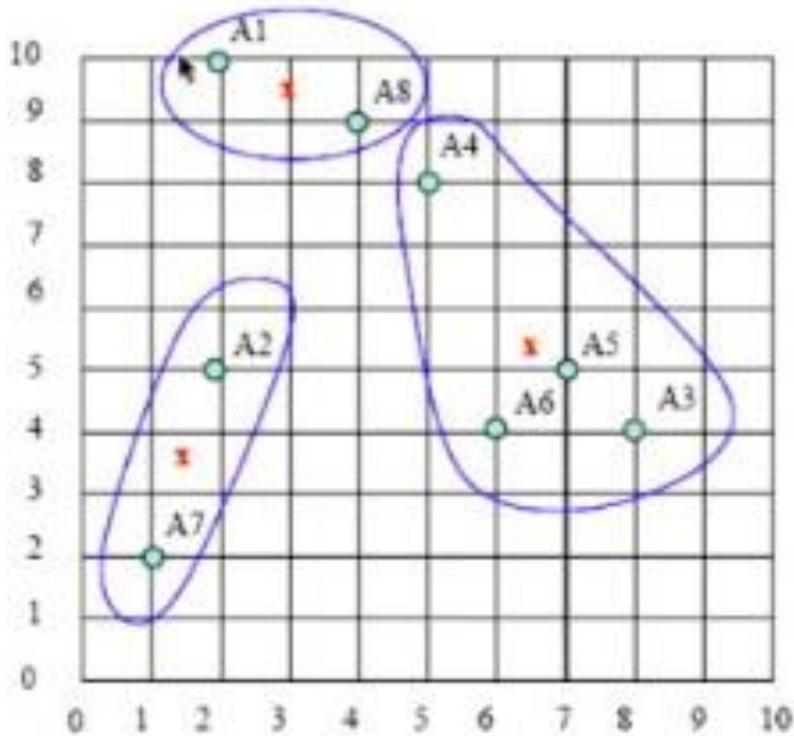
## *Exercice?*

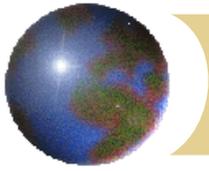




# *K-Means*

## *Exercice?*





***Merci Pour Votre Attention***