

Série TD 5

Exercice 1 : Clustering incrémental

Objectif

1. Effectuer un **clustering incrémental** sur les exemples, avec :
 - **Seuil de distance = 25**
 - Distance utilisée = **City Block** (distance de Manhattan)
 - Ordre des exemples = ordre dans la table
2. Construire la **table de correspondance classe-cluster**.
3. Donner l'affectation finale des clusters aux exemples.
4. Calculer le **taux de clustering** (qualité du clustering par rapport aux classes réelles).

a. Principe du clustering incrémental avec seuil

- On initialise sans cluster.
- Pour chaque exemple (dans l'ordre donné) :
 - Calculer la distance City Block entre cet exemple et chacun des centroïdes des clusters existants.
 - Si la distance minimale est inférieure au seuil (25), affecter l'exemple au cluster correspondant et mettre à jour le centroïde.
 - Sinon, créer un nouveau cluster avec cet exemple.

b. Distance City Block (Manhattan)

Pour deux points $x = (x_1, x_2, \dots, x_n)$ et $y = (y_1, y_2, \dots, y_n)$:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

c. Exemple d'application

Exemple	Attributs (x1, x2, ...)	Infection (Classe)
1	(10, 5)	Oui
2	(12, 7)	Oui
3	(30, 40)	Non
4	(28, 38)	Non
5	(11, 6)	Oui

Initialisation

- Premier exemple (1) crée le premier cluster avec centroïde = (10, 5).

Exemple 2 : (12, 7)

- Distance au cluster 1 : $|12-10| + |7-5| = 2 + 2 = 4 < 25 \rightarrow$ appartient au cluster 1
- Mise à jour centroïde cluster 1 :

$$\frac{(10, 5) + (12, 7)}{2} = (11, 6)$$

Exemple 3 : (30, 40)

- Distance au cluster 1 : $|30-11| + |40-6| = 19 + 34 = 53 > 25 \rightarrow$ nouveau cluster créé (cluster 2)

Exemple 4 : (28, 38)

- Distance au cluster 1 : $|28-11| + |38-6| = 17 + 32 = 49 > 25$
- Distance au cluster 2 : $|28-30| + |38-40| = 2 + 2 = 4 < 25 \rightarrow$ appartient au cluster 2
- Mise à jour centroïde cluster 2 :

$$\frac{(30, 40) + (28, 38)}{2} = (29, 39)$$

Exemple 5 : (11, 6)

- Distance au cluster 1 : $|11-11| + |6-6| = 0 < 25 \rightarrow$ appartient au cluster 1
- Mise à jour centroïde cluster 1 :

$$\frac{(11, 6) \times 2 + (11, 6)}{3} = (11, 6)$$

(Centroïde ne change pas ici.)

d. : Table de correspondance classe-cluster

Cluster	Exemples associés	Classe majoritaire	Nb d'exemples dans cluster	Nb d'exemples de la classe majoritaire
1	1, 2, 5	Oui	3	3
2	3, 4	Non	2	2

e. : Calcul du taux de clustering

$$Taux = \frac{\text{nombre d'exemples bien classés}}{\text{nombre total d'exemples}} = \frac{3 + 2}{5} = 1 = 100\%$$

Exercice 2 :

Données

On vous fournit 8 exemples avec 2 attributs, ainsi que leur classe réelle (non utilisée directement dans le clustering) :

Exemple	Attribut 1	Attribut 2	Classe
1	5	5	A
2	6	6	A
3	15	15	B
4	16	16	B
5	5	6	A
6	14	14	B
7	25	25	C
8	26	26	C

Questions :

- Effectuer un **clustering incrémental** sur ces exemples, dans l'ordre donné, avec les paramètres suivants :
 - Distance : City Block (Manhattan)
 - Seuil de distance : 5
- Pour chaque nouvel exemple, affecter au cluster existant le plus proche si la distance au centroïde est ≤ 5 , sinon créer un nouveau cluster.
- Mettre à jour le centroïde du cluster à chaque affectation.
- À la fin, établir la table de correspondance **classe-cluster** (pour chaque cluster, indiquer la classe majoritaire).
- Calculer le taux de clustering (nombre d'exemples correctement regroupés selon la classe majoritaire de leur cluster).

Étape 1 : Initialisation

- Pas de cluster au départ.

Étape 2 : Traitement exemple par exemple

Exemple 1 : (5, 5)

- Aucun cluster, créer **Cluster 1** avec centroïde = (5,5).

Exemple 2 : (6, 6)

- Distance à cluster 1 :

$$|6 - 5| + |6 - 5| = 1 + 1 = 2 \leq 5$$

- Affecter à **Cluster 1**.
- Mise à jour du centroïde :

$$\frac{(5, 5) + (6, 6)}{2} = (5.5, 5.5)$$

Exemple 3 : (15, 15)

- Distance à cluster 1 :

$$|15 - 5.5| + |15 - 5.5| = 9.5 + 9.5 = 19 > 5$$

Pas de cluster à distance $\leq 5 \rightarrow$ créer **Cluster 2** avec centroïde = (15,15)

Exemple 4 : (16, 16)

- Distance à cluster 1 :

$$|16 - 5.5| + |16 - 5.5| = 10.5 + 10.5 = 21 > 5$$

Distance à cluster 2 :

$$|16 - 15| + |16 - 15| = 1 + 1 = 2 \leq 5$$

- Affecter à **Cluster 2**.
- Mise à jour centroïde cluster 2 :

$$\frac{(15, 15) + (16, 16)}{2} = (15.5, 15.5)$$

Exemple 5 : (5, 6)

- Distance à cluster 1 :

$$|5 - 5.5| + |6 - 5.5| = 0.5 + 0.5 = 1 \leq 5$$

Distance à cluster 2 :

$$|5 - 15.5| + |6 - 15.5| = 10.5 + 9.5 = 20 > 5$$

- Affecter à **Cluster 1**.
- Mise à jour centroïde cluster 1 (maintenant 3 points) :

$$\text{Nouveau centroïde} = \frac{(5, 5) + (6, 6) + (5, 6)}{3} = \frac{16}{3}, \frac{17}{3} \approx (5.33, 5.67)$$

Exemple 6 : (14, 14)

- Distance à cluster 1 :

$$|14 - 5.33| + |14 - 5.67| = 8.67 + 8.33 = 17 > 5$$

- Distance à cluster 2 :

$$|14 - 15.5| + |14 - 15.5| = 1.5 + 1.5 = 3 \leq 5$$

- Affecter à **Cluster 2**.
- Mise à jour centroïde cluster 2 (maintenant 3 points) :

$$\frac{(15, 15) + (16, 16) + (14, 14)}{3} = (15, 15)$$

Exemple 7 : (25, 25)

- Distance à cluster 1 :

$$|25 - 5.33| + |25 - 5.67| = 19.67 + 19.33 = 39 > 5$$

Distance à cluster 2 :

$$|25 - 15| + |25 - 15| = 10 + 10 = 20 > 5$$

Pas de cluster proche → créer **Cluster 3** avec centroïde = (25, 25)

Exemple 8 : (26, 26)

- Distance à cluster 1 :

$$|26 - 5.33| + |26 - 5.67| = 20.67 + 20.33 = 41 > 5$$

Distance à cluster 2 :

$$|26 - 15| + |26 - 15| = 11 + 11 = 22 > 5$$

Distance à cluster 3 :

$$|26 - 25| + |26 - 25| = 1 + 1 = 2 \leq 5$$

- Affecter à **Cluster 3**
- Mise à jour centroïde cluster 3 :

$$\frac{(25, 25) + (26, 26)}{2} = (25.5, 25.5)$$

Étape 3 : Résumé des clusters

Cluster	Exemples	Centroïde approximatif	Classe majoritaire	Nb Exemples
1	1, 2, 5	(5.33, 5.67)	A	3
2	3, 4, 6	(15, 15)	B	3
3	7, 8	(25.5, 25.5)	C	2

Étape 4 : Taux de clustering

- Tous les points du cluster 1 ont la classe A → 3 corrects
- Tous les points du cluster 2 ont la classe B → 3 corrects
- Tous les points du cluster 3 ont la classe C → 2 corrects

Total = 3 + 3 + 2 = 8

Nombre total d'exemples = 8

$$\text{Taux de clustering} = \frac{8}{8} = 100\%$$