

Série de TD N°2

Clustering 1 (K-means et K-medoids)

Exercice1 :

On désire classifier l'ensemble des points suivants en trois classes:

$A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $B_1(5, 8)$, $B_2(7, 5)$, $B_3(6, 4)$, $C_1(1, 2)$, $C_2(4, 9)$

On suppose initialement que les points A_1 , B_1 et C_1 sont choisis comme centres. Utiliser l'algorithme K-means pour déterminer:

- (a) Les trois centres calculés après la première itération.
- (b) Les trois classes finales résultant de l'application de l'algorithme.

Exercice2 :

Soient 10 points (a,b,c,d,e,f,g,h,i,j) du plan dont les coordonnées sont les colonnes du tableau suivant:

a	b	c	d	e	f	g	h	i	j
1	2	3	5	4	6	8	10	12	10
2	5	2	4	7	7	3	4	1	0

Donner les partitions stabilisées ps_1 et ps_2 correspondants:

pour ps_1 aux centres de départ $\{a,b,i\}$

pour ps_2 aux centres de départ $\{a,d,i\}$

On fera fonctionner l'algorithme directement sur le graphique joint. Que peut-on conclure?

Exercice3:

On considère les 6 points $M_1 = (-2, 3)$, $M_2 = (-2, 1)$, $M_3 = (-2, -1)$, $M_4 = (2, -1)$, $M_5 = (2, 1)$ et $M_6 = (1, 0)$.

En supposant que les deux premiers points M_1 et M_2 sont les centres initiaux, décrire l'algorithme k-medoids pour la répartition de ces points.

Série de TD N°2- Corrigé Clustering 1 (K-means et K-medoids)

Exercice1 :

On désire classifier l'ensemble des points suivants en trois classes:

$A_1(2, 10)$, $A_2(2, 5)$, $A_3(8, 4)$, $B_1(5, 8)$, $B_2(7, 5)$, $B_3(6, 4)$, $C_1(1, 2)$, $C_2(4, 9)$

On suppose initialement que les points A_1 , B_1 et C_1 sont choisis comme centres. Utiliser l'algorithme K-means pour déterminer:

- (a) Les trois centres calculés après la première itération.
- (b) Les trois classes finales résultant de l'application de l'algorithme.

Réponse

- (a) Les trois centres calculés après la première itération.

Initialisation : $G_1 = \{A_1\}$, $G_2 = \{B_1\}$, $G_3 = \{C_1\}$ avec $c_1 = A_1$, $c_2 = B_1$ et $c_3 = C_1$

Itération 1 :

Pour A_2 : $d(A_2, c_1) = \sqrt{(|2-2|^2 + |5-10|^2)} = 5$, $d(A_2, c_2) = \sqrt{(|2-5|^2 + |5-8|^2)} = 4.24$, $d(A_2, c_3) = \sqrt{(|2-1|^2 + |5-2|^2)} = 3.16$ alors $G_3 = G_3 + A_2$

Pour A_3 : $d(A_3, c_1) = \sqrt{(|8-2|^2 + |4-10|^2)} = 8.48$, $d(A_3, c_2) = \sqrt{(|8-5|^2 + |4-8|^2)} = 5$, $d(A_3, c_3) = \sqrt{(|8-1|^2 + |4-2|^2)} = 7.28$ alors $G_2 = G_2 + A_3$

Pour B_2 : $d(B_2, c_1) = \sqrt{(|7-2|^2 + |5-10|^2)} = 5$, $d(B_2, c_2) = \sqrt{(|7-5|^2 + |5-8|^2)} = 3.60$, $d(B_2, c_3) = \sqrt{(|7-1|^2 + |5-2|^2)} = 6.70$ alors $G_2 = G_2 + B_2$

Pour B_3 : $d(B_3, c_1) = \sqrt{(|6-2|^2 + |4-10|^2)} = 7.21$, $d(B_3, c_2) = \sqrt{(|6-5|^2 + |4-8|^2)} = 4.12$, $d(B_3, c_3) = \sqrt{(|6-1|^2 + |4-2|^2)} = 5.38$ alors $G_2 = G_2 + B_3$

Pour C_2 : $d(C_2, c_1) = \sqrt{(|4-2|^2 + |9-10|^2)} = 2.23$, $d(C_2, c_2) = \sqrt{(|4-5|^2 + |9-8|^2)} = 1.41$, $d(C_2, c_3) = \sqrt{(|4-1|^2 + |9-2|^2)} = 7.61$ alors $G_2 = G_2 + C_2$

Les trois clusters après la première itération sont :

$G_1 = \{A_1\}$, $G_2 = \{B_1, A_3, B_2, B_3, C_2\}$, $G_3 = \{C_1, A_2\}$

Les trois centres après la première itération sont :

$c_1 = A_1$, $c_2 = ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6,6)$, $c_3 = ((1+2)/2, (5+2)/2) = (1.5, 3.5)$

- (b) Les trois classes finales résultant de l'application de l'algorithme.

Itération 2 :

$G_1 = \{A_1\}$, $G_2 = \{A_2, A_3, B_1, B_2, B_3, C_2\}$, $G_3 = \{C_1\}$ avec $c_1 = A_1$, $c_2 = (6,6)$ et $c_3 = (1.5, 3.5)$

Pour A2 : $d(A2,c1) = \sqrt{(|2-2|^2 + |5-10|^2)} = 5$, $d(A2,c2) = \sqrt{(|2-6|^2 + |5-6|^2)} = 4.12$, $d(A2,c3) = \sqrt{(|2-1.5|^2 + |5-3.5|^2)} = 1.58$
alors $G3 = G3 + A2$

Pour A3 : $d(A3,c1) = \sqrt{(|8-2|^2 + |4-10|^2)} = 8.48$, $d(A3,c2) = \sqrt{(|8-6|^2 + |4-6|^2)} = 2.82$, $d(A3,c3) = \sqrt{(|8-1.5|^2 + |4-3.5|^2)} = 6.51$
alors $G2 = G2 + A3$

Pour B1 : $d(B1,c1) = \sqrt{(|5-2|^2 + |8-10|^2)} = 3.60$, $d(B1,c2) = \sqrt{(|5-6|^2 + |8-6|^2)} = 2.23$, $d(B1,c3) = \sqrt{(|5-1.5|^2 + |8-3.5|^2)} = 5.70$
alors $G2 = G2 + B1$

Pour B2 : $d(B2,c1) = \sqrt{(|7-2|^2 + |5-10|^2)} = 5$, $d(B2,c2) = \sqrt{(|7-6|^2 + |5-6|^2)} = 1.41$, $d(B2,c3) = \sqrt{(|7-1.5|^2 + |5-3.5|^2)} = 5.70$
alors $G2 = G2 + B2$

Pour B3 : $d(B3,c1) = \sqrt{(|6-2|^2 + |4-10|^2)} = 7.21$, $d(B3,c2) = \sqrt{(|6-6|^2 + |4-6|^2)} = 2$, $d(B3,c3) = \sqrt{(|6-1.5|^2 + |4-3.5|^2)} = 4.52$
alors $G2 = G2 + B3$

Pour C1 : $d(C1,c1) = \sqrt{(|1-2|^2 + |2-10|^2)} = 8.06$, $d(C1,c2) = \sqrt{(|1-6|^2 + |2-6|^2)} = 6.40$, $d(C1,c3) = \sqrt{(|1-1.5|^2 + |2-3.5|^2)} = 1.58$
alors $G3 = G3 + C1$

Pour C2 : $d(C2,c1) = \sqrt{(|4-2|^2 + |9-10|^2)} = 2.23$, $d(C2,c2) = \sqrt{(|4-6|^2 + |9-6|^2)} = 3.6$, $d(C2,c3) = \sqrt{(|4-1.5|^2 + |9-3.5|^2)} = 6.04$
alors $G1 = G1 + C2$

Les trois clusters après la première itération sont :

$G1 = \{A1,C2\}$, $G2 = \{A3,B1,B2,B3\}$, $G3 = \{A2,C1\}$

Les trois centres après la première itération sont :

$c1 = ((2+4)/2, (10+9)/2) = (3, 9.5)$, $c2 = ((8+5+7+6)/4, (4+8+5+4)/4) = (6.5, 5.25)$, $c3 = ((1+2)/2, (5+2)/2) = (1.5, 3.5)$