Tariel
Rémi
GISSAG

Examen Classification Automatique (DM)

Exercice 1 - k-means

1) Le regroupement {(1,2), (9,12), (12,16,20)} ne représente pas une segmentation possible car 12 fait partie de 2 sous partition, la condition $\forall p \neq q, P_q \cap P_p = \emptyset$ n'est donc pas satisfaite

2). k = 2 avec 1 et 20 comme centre initiaux

étape 0
$C_0 = \{1, 20\}$

x	1	2	9	12	16	20
d(x,1)	0	1	8	11	15	19
d(x,20)	19	18	11	8	4	0

On obtient la partition suivante

$P^0 = \{(1,2,9), (12,16,20)\}$

étape 1

$C_0^1 = \frac{1+2+9}{3} = 4 \quad C_1^1 = \frac{12+16+20}{3} = 16$

$\Rightarrow C^1 = \{4, 16\}$

x	1	2	9	12	16	20
$d(x,4)$	3	2	5	8	12	16
$d(x,16)$	15	14	7	4	0	4

$\Rightarrow P^1 = \{(1,2,9), (12,16,20)\}$

On observe que $P^0 = P^1$ donc on a une stabilité des partitions.

$\Rightarrow P = \{(1,2,9), (12,16,20)\}$
pour $k = 2$

- $k = 3$ avec 1, 12 et 20 en centres initiaux

étape 0

$C^0 = \{1, 12, 20\}$

x	1	2	9	12	16	20
$d(x,1)$	0	1	8	11	15	19
$d(x,12)$	11	10	3	0	4	8
$d(x,20)$	19	18	11	8	4	0

$\Rightarrow P^0 = \{(1,2), (9,12,16), (20)\}$

étape 1

$C_0^1 = \frac{1+2}{2} = 1,5 \quad C_1 = \frac{9+12+16}{3} = 12,333 \quad C_2^1 = 20$

$\Rightarrow C^1 = \{1,5 ; 12,333 ; 20\}$

x	1	2	9	12	16	20
d(x, 1,5)	0,5	0,5	7,5	10,5	14,5	18,5
d(x, 12,333)	11,333	10,333	3,333	0,333	3,667	7,667
d(x, 20)	19	18	11	8	4	0

$\Rightarrow P^1 = \{(1,2) , (9,12,16) , (20)\}$

On observe que $P^0 = P^1$ donc on a une stabilité de partition

$\Rightarrow P = \{(1,2) , (9,12,16) , (20)\}$

3). <u>Calcul inertie totale</u>

On rappelle la formule de l'inertie totale d'un ensemble de points

$$I_T = \frac{1}{n} \sum_{i=1}^{n} d^2(w_i, g)$$

avec g le barycentre de notre ensemble de points.
Dans notre cas $n = 6$ et
$$g = \frac{1+2+9+12+16+20}{6} = 10$$

$d^2(1,g) = 81$ $\quad$ $d^2(12,g) = 4$
$d^2(2,g) = 64$ $\quad$ $d^2(16,g) = 36$
$d^2(9,g) = 1$ $\quad$ $d^2(20,g) = 100$

$$I_T = \frac{1}{6}(81+64+1+4+36+100) = 47,667$$

. Calcul inertie interclasse

Rappel de la formule

$I_{inter} = \frac{1}{n} \sum_{r=1}^{k} n_r d^2(g_r, g)$

avec g_r le barycentre de la partition r

$k = 2$	$k = 3$
$g_1 = 4 \quad g_2 = 16$	$g_1 = 1,5 \quad g_2 = 12,333 \quad g_3 = 20$
$n_1 = 3 \quad n_2 = 3$	$n_1 = 2 \quad n_2 = 3 \quad n_3 = 1$
$d^2(g_1, g) = d^2(4, 10) = 36$	$d^2(g_1, g) = d^2(1,5, 10) = 72,25$
$d^2(g_2, g) = d^2(16, 10) = 36$	$d^2(g_2, g) = d^2(12,333, 10) = 5,443$
	$d^2(g_3, g) = d^2(20, 10) = 100$
$I_{inter} = \frac{1}{6}(3 \times 36 + 3 \times 36) = 36$	$I_{inter} = \frac{1}{6}(2 \times 72,25 + 3 \times 5,443) = 43,471$

. Calcul inertie intraclasse

Rappel de la formule

$I_{intra} = \frac{1}{n} \sum_{r=1}^{k} n_r I(N_r)$

avec $I(N_r)$ l'inertie total de la partition r

$k=2$

$I(N_1) = \frac{1}{3}(d^2(w_i, g)$

$d^2(1,4) = 9$
$d^2(2,4) = 4$
$d^2(9,4) = 25$

$I(N_2) = \frac{1}{3} \sum_{i=1}^{3} d^2(w_i, g)$

$d^2(12,16) = 16$
$d^2(16,16) = 0$
$d^2(20,16) = 16$

$I(N_1) = 12,667$
$I(N_2) = 10,667$

$I_{intra} = 11,667$

$k=3$

$I(N_1) = \frac{1}{2}(d^2(1, 1,5) + d^2(2, 1,5))$

$d^2(1, 1,5) = 0,25$
$d^2(2, 2,5) = 0,25$

$I(N_2) = \frac{1}{3}(d^2(9, 12,333) + d^2(12, 12,333) + d^2(16, 12,333)$

$d^2(9, 12,333) = 11,109$
$d^2(12, 12,333) = 0,111$
$d^2(16, 12,333) = 13,447$

$I(N_3) = \frac{1}{1} d^2(20,20)$

$d^2(20,20) = 0$

$I(N_1) = 0,25$
$I(N_2) = 8,222$
$I(N_3) = 0$

$I_{intra} = 4,194$

. Calcul du R^2

Rappel de la formule :
$R^2 = \frac{I_{intra}}{I_{tot}}$

$k=2$

$R^2 = \frac{11,667}{47,667} = 0,245$

$k=3$

$R^2 = \frac{4,194}{47,667} = 0,088$

4)

	I_{intra}	I_{inter}	R^2
$k=2$	11,667	36	0,245
$k=3$	4,194	43,471	0,088

Le R^2 étant meilleur pour la classification avec $k=2$, ~~tout comme l'~~, on décide donc de choisir la segmentation avec $k=2$
$\{(1,2,9),(12,16,20)\}$

Exercice 2 - CAH

1) Quand 2 sites ont exactement les mêmes espèces on a l'égalité suivante :

$$n_A = n_B = n_{AB}$$

donc
$$d(A,B) = \frac{n_A + n_B - 2n_{AB}}{n_A + n_B}$$
$$= \frac{n_A + n_A - 2_{AA}}{n_A + n_A}$$
$$d(A,B) = 0$$

2) Quand 2 sites n'ont aucune espèce en commun on a l'égalité suivante :

$$n_{AB} = 0$$

donc
$$d(A,B) = \frac{n_A + n_B - 2n_{AB}}{n_A + n_B}$$
$$= \frac{n_A + n_B}{n_A + n_B}$$
$$d(A,B) = 1$$

3) $d(S_1, S_2) = \dfrac{n_{S_1} + n_{S_2} - 2\, n_{S_1S_2}}{n_{S_1} + n_{S_2}}$

$n_{S_1} = 10$
$n_{S_2} = 15$
$n_{S_1S_2} = 5$

$d(S_1, S_2) = \dfrac{10 + 15 - 2 \times 5}{10 + 15} = \dfrac{15}{25} = \dfrac{3}{5} = 0{,}6$

$d(S_1, S_3) = \dfrac{n_{S_1} + n_{S_3} - 2\, n_{S_1S_3}}{n_{S_1} + n_{S_3}}$

$n_{S_1} = 10$
$n_{S_3} = 11$
$n_{S_1S_3} = 8$

$d(S_1, S_3) = \dfrac{10 + 11 - 2 \times 8}{10 + 11} = \dfrac{21 - 16}{21} = \dfrac{5}{21} = 0{,}238$

4) Matrice de distance

	S_1	S_2	S_3	S_4
S_1	0	0,6	0,238	0,545
S_2	0,6	0	0,692	0,185
S_3	0,238	0,692	0	0,739
S_4	0,545	0,185	0,739	0

5)

	S_1	$S_2 S_4$	S_3
S_1	0	0,545	0,238
$S_2 S_4$	0,545	0	0,692
S_3	0,238	0,692	0

2e regroupement

	$S_1 S_3$	$S_2 S_4$
$S_1 S_3$	0	0,545
$S_2 S_4$	0,545	0

3e regroupement

Dendogramme

Le nombre de classes à retenir est 2 : $\{S_2, S_4\}$ et $\{S_1, S_3\}$