

Sciences des données - TP2

Alexandre Theisse Louis-Vincent Capelli Tom Sartori

November 15, 2023

Question 1

Question 2

Question 3

Mesure de séparation des clusters

Voici les résultats obtenus pour les 2 mesures de séparation des clusters pour les différents nombres de clusters.

On constate que le score silhouette est maximal pour 2 clusters et diminue constamment par la suite. Même le score maximal (0.43) est plus proche de 0 que de 1, ce qui indique que les clusters sont assez mal séparés. Les scores correspondants aux grands nombres de clusters sont encore plus faibles (proches de 0.2), ce qui indique un fort chevauchement entre les clusters.

On constate également cette tendance à l'augmentation du chevauchement en observant les matrices d'overlaps. On remarque que les valeurs augmentent avec le nombre de clusters, ce qui indique que les clusters sont de plus en plus chevauchants. En effet, les meilleures valeurs d'overlaps obtenues entre 2 clusters sont supérieures à 1, ce qui indique que les clusters ne sont pas séparés et les moins bonnes valeurs sont de l'ordre de 3 ce qui indique un chevauchement important.

2 clusters

Silhouette score : 0.43

Overlaps 2 à 2 :

	C1	C2
C1	-	1.53
C2	-	-

3 clusters

Silhouette score : 0.34

Overlaps 2 à 2 :

	C1	C2	C3
C1	-	1.40	1.74
C2	-	-	1.19
C3	-	-	-

4 clusters

Silhouette score : 0.30

Overlaps 2 à 2 :

	C1	C2	C3	C4
C1	-	2.68	1.70	1.43
C2	-	-	1.41	1.67
C3	-	-	-	1.14
C4	-	-	-	-

5 clusters

Silhouette score : 0.26

Overlaps 2 à 2 :

	C1	C2	C3	C4	C5
C1	-	1.39	2.50	2.56	1.13
C2	-	-	1.62	1.66	1.21
C3	-	-	-	1.02	2.09
C4	-	-	-	-	2.14
C5	-	-	-	-	-

6 clusters

Silhouette score : 0.22

Overlaps 2 à 2 :

	C1	C2	C3	C4	C5	C6
C1	-	1.25	2.32	2.97	1.04	1.30
C2	-	-	1.71	2.14	1.31	1.03
C3	-	-	-	1.18	2.45	1.64
C4	-	-	-	-	3.15	2.04
C5	-	-	-	-	-	1.36
C6	-	-	-	-	-	-

7 clusters

Silhouette score : 0.20

8 clusters

Silhouette score : 0.20

9 clusters

Silhouette score : 0.19

10 clusters

Silhouette score : 0.19

Cas particulier des 3 clusters