

Лабораторна робота 3

МОШІ

Угрина Максим ПМ-31

«Кластеризація даних»

Завдання

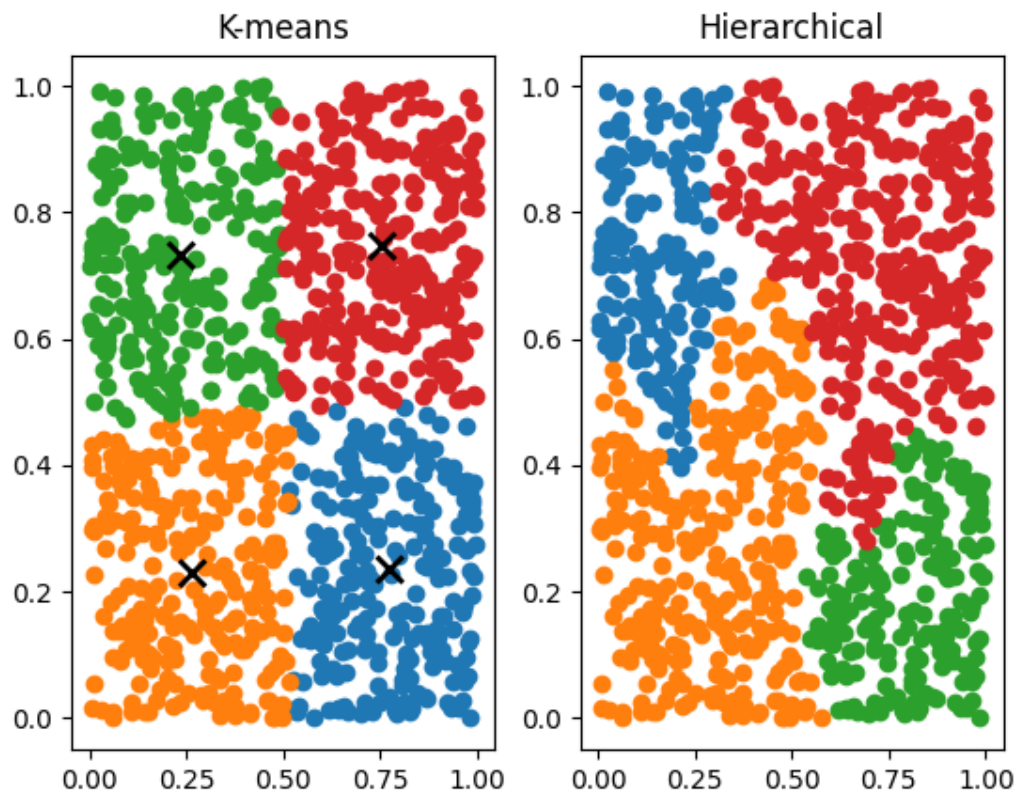
Розробити програмну реалізацію алгоритму кластеризації та перевірити роботу алгоритму на тестовій множині даних.

Етапи виконання завдання

1. Вивчити із використанням запропонованих літературних джерел зміст задачі кластеризації та методи кластеризації даних.
2. Згенерувати тестову послідовність з N значень (для визначеності, можна покласти $N \geq 1000$), що є парами дійсних чисел на одиничному квадраті.
3. Реалізувати допоміжну функцію для обчислення міри віддалі.
4. Реалізувати допоміжну функцію для виконання алгоритму кластеризації за методом К-середніх (англ. *K-means*).
5. Реалізувати допоміжну функцію для виконання алгоритму кластеризації за будь-яким іншим (за вибором) методом (ієрархічний, С-середніх, мінімального покриваючого дерева, по-шарової кластеризації, ART1, тощо).
6. Безпосередньо реалізувати кластеризацію даних двома методами та порівняти результати кластеризації.
7. Порівняйте кількість кластерів та якість кластеризації (можна просто оцінити середньо-зважене розмірів утворених кластерів згідно заданої міри віддалі для кожного із методів).

Результат

Figure 1



K-means:

Cluster 1: 242 points, center at (0.7714793548486061, 0.23485985088963457)

Cluster 2: 245 points, center at (0.26122177365294036, 0.22918370606885766)

Cluster 3: 235 points, center at (0.2321149527295248, 0.7333808485212716)

Cluster 4: 278 points, center at (0.7547473887135258, 0.7473727229741652)

Hierarchical:

Cluster 1: 160 points, center at (0.15442836458358142, 0.7312524330265692)

Cluster 2: 302 points, center at (0.2945959974075594, 0.2812856662100375)

Cluster 3: 189 points, center at (0.8006696172591731, 0.19942210762094828)

Cluster 4: 349 points, center at (0.7165072755292037, 0.7262486231273266)

