**Інститут прикладної математики та фундаментальних наук**

**Кафедра прикладної математики**

**Звіт**

**до лабораторної роботи №3**

**з дисципліни**

**“Математичні основи штучного інтелекту”**

**студента групи ПМ-31**

**Подубінського Назара**

**Прийняв:**

**Пабирівський В.В.**

**Львів 2023**

**Тема:** Кластеризація даних.

**Мета:** Навчитись реалізовувати програму алгоритму кластеризації та перевірити роботу алгоритму на тестовій множині даних.

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

**Етапи виконання завдання**

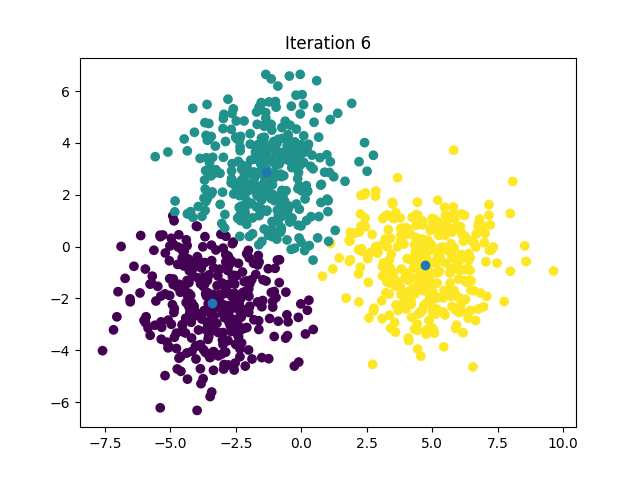
1. Вивчити із використанням запропонованих літературних джерел зміст задачі кластеризації та методи кластеризації даних.
2. Згенерувати тестову послідовність з N значень (для визначеності, можна покласти N≥1000), що є парами дійсних чисел на одиничному квадраті.
3. Реалізувати допоміжну функцію для обчислення 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, C-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).

***Посилання на GitHub:*** [***https://github.com/nazarpodubinskyi/MOSHI\_LAB/tree/master/Lab\_3\_***](https://github.com/nazarpodubinskyi/MOSHI_LAB/tree/master/Lab_3_)

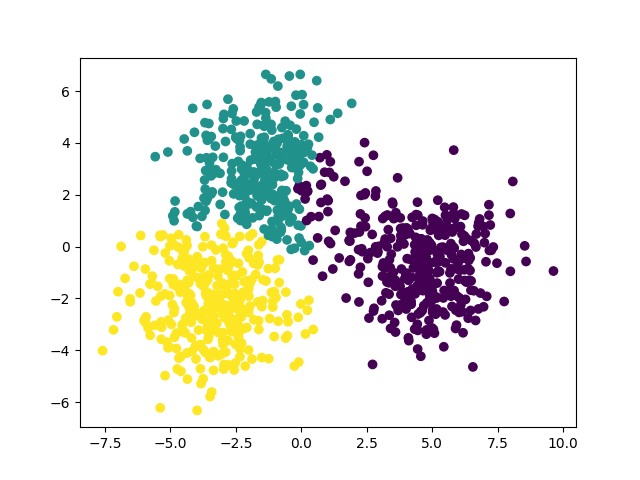
**Порівняння результатів двох методів**

**Кількість кластерів = 3, кількість точок = 1000.**

**Метод К-середніх**

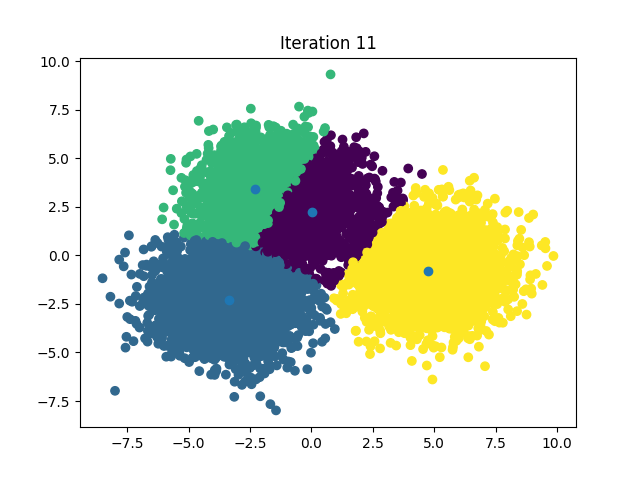
****

**Метод ієрархії**

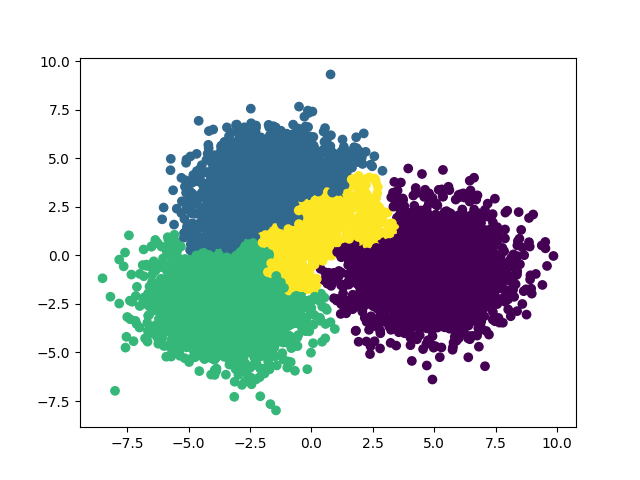
****

**Кількість кластерів = 4, кількість точок = 10000.**

**Метод К-середніх**

****

**Метод ієрархії**

****

**Висновок.** Якщо порівнювати два методи для одинакових даних можна дійти висновку, що їхні результати схожі. Метод ієрархії працює швидше оскільки йому не потрібно як в методі К-середніх використовувати ітерації.