# Лабораторна робота №5 – SOM та K-means кластеризація

## 1. Вступ

У цій лабораторній роботі розглядаються два методи кластеризації без учителя: самоорганізуючі карти (SOM) та алгоритм K-means. Обидва методи застосовано до класичного набору даних Iris. Результати кластеризації подано у вигляді графіків, таблиць та програмної реалізації.

## 2. Теоретичні відомості

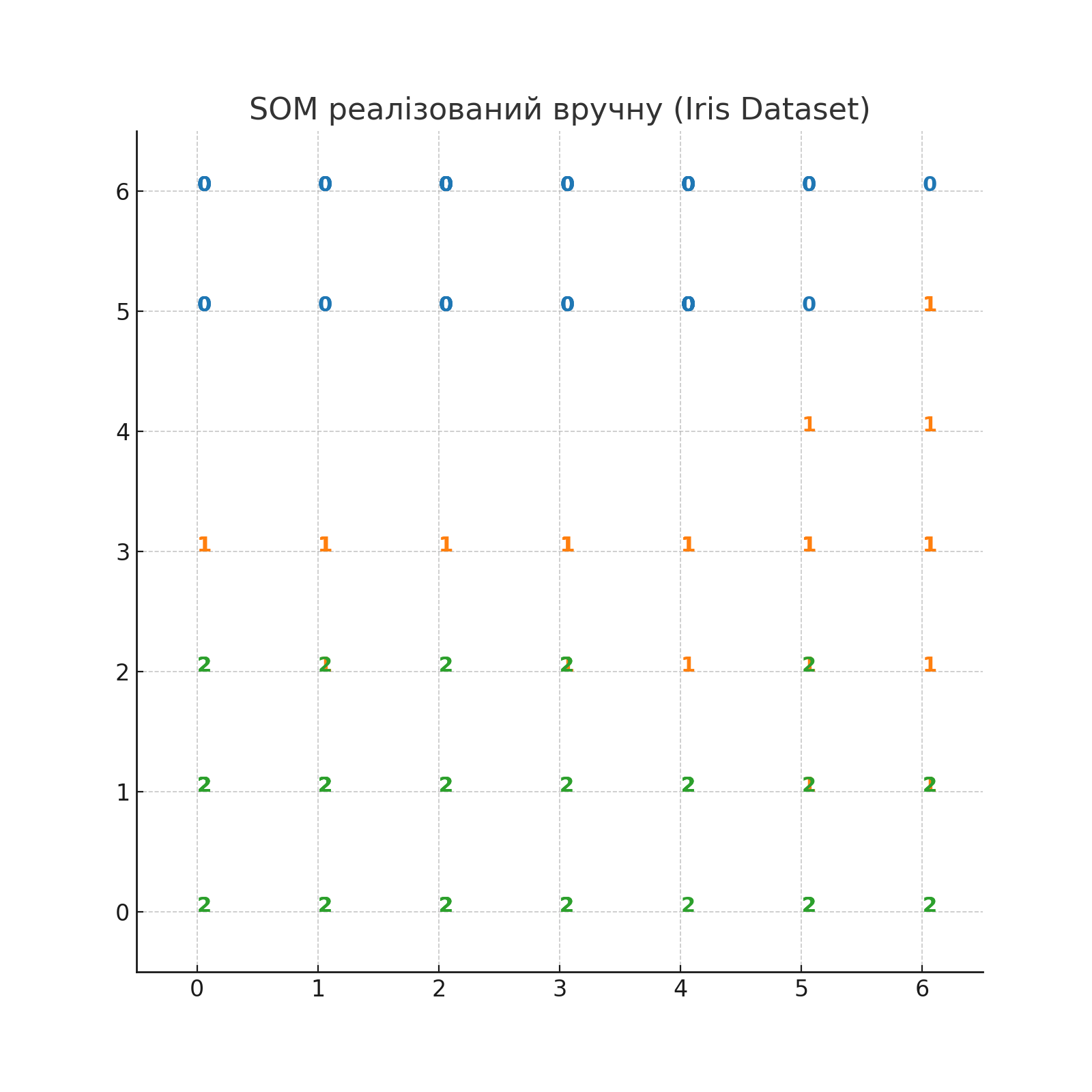
### 2.1. Self-Organizing Maps (SOM)

SOM (самоорганізуючі карти Кохонена) — це нейронна мережа, яка проектує дані з багатовимірного простору на двовимірну решітку зберігаючи топологію. Кожен вхідний вектор призначається до найбільш схожого вузла (нейрона), після чого оновлюються ваги цього вузла та його сусідів. SOM дозволяє візуалізувати та кластеризувати дані.

Фрагмент реалізації SOM на Python:

weights = np.random.rand(width, height, input\_len)  
for iteration in range(num\_iterations):  
 for x in X\_scaled:  
 bmu\_i, bmu\_j = get\_bmu(x)  
 for i in range(width):  
 for j in range(height):  
 dist\_to\_bmu = (i - bmu\_i)\*\*2 + (j - bmu\_j)\*\*2  
 h = np.exp(-dist\_to\_bmu / (2 \* (sigma \*\* 2)))  
 weights[i, j] += lr \* h \* (x - weights[i, j])

Візуалізація результату:



Таблиця координат BMU для кожного прикладу:

|  |  |  |
| --- | --- | --- |
| BMU\_x | BMU\_y | Class |
| 2 | 6 | 0 |
| 4 | 5 | 0 |
| 4 | 6 | 0 |
| 4 | 6 | 0 |
| 2 | 6 | 0 |
| 0 | 5 | 0 |
| 3 | 6 | 0 |
| 2 | 6 | 0 |
| 5 | 6 | 0 |
| 4 | 5 | 0 |
| 1 | 5 | 0 |
| 3 | 6 | 0 |
| 4 | 5 | 0 |
| 5 | 6 | 0 |
| 0 | 6 | 0 |
| 0 | 6 | 0 |
| 0 | 5 | 0 |
| 2 | 6 | 0 |
| 0 | 5 | 0 |
| 1 | 6 | 0 |
| 2 | 5 | 0 |
| 1 | 6 | 0 |
| 3 | 6 | 0 |
| 3 | 5 | 0 |
| 3 | 6 | 0 |
| 4 | 5 | 0 |
| 3 | 5 | 0 |
| 2 | 6 | 0 |
| 2 | 5 | 0 |
| 4 | 6 | 0 |
| 4 | 5 | 0 |
| 2 | 5 | 0 |
| 0 | 6 | 0 |
| 0 | 6 | 0 |
| 4 | 5 | 0 |
| 4 | 5 | 0 |
| 2 | 5 | 0 |
| 2 | 6 | 0 |
| 5 | 6 | 0 |
| 2 | 6 | 0 |
| 2 | 6 | 0 |
| 6 | 6 | 0 |
| 4 | 6 | 0 |
| 3 | 5 | 0 |
| 1 | 6 | 0 |
| 5 | 5 | 0 |
| 1 | 6 | 0 |
| 4 | 6 | 0 |
| 1 | 5 | 0 |
| 4 | 5 | 0 |
| 0 | 3 | 1 |
| 2 | 3 | 1 |
| 0 | 3 | 1 |
| 6 | 3 | 1 |
| 1 | 3 | 1 |
| 4 | 3 | 1 |
| 2 | 3 | 1 |
| 6 | 5 | 1 |
| 1 | 3 | 1 |
| 5 | 2 | 1 |
| 6 | 4 | 1 |
| 3 | 3 | 1 |
| 6 | 3 | 1 |
| 3 | 3 | 1 |
| 4 | 3 | 1 |
| 0 | 3 | 1 |
| 4 | 2 | 1 |
| 5 | 3 | 1 |
| 6 | 2 | 1 |
| 6 | 3 | 1 |
| 3 | 2 | 1 |
| 3 | 3 | 1 |
| 6 | 1 | 1 |
| 3 | 3 | 1 |
| 1 | 3 | 1 |
| 1 | 3 | 1 |
| 1 | 3 | 1 |
| 1 | 2 | 1 |
| 3 | 3 | 1 |
| 5 | 4 | 1 |
| 6 | 3 | 1 |
| 6 | 3 | 1 |
| 5 | 3 | 1 |
| 5 | 1 | 1 |
| 4 | 2 | 1 |
| 2 | 3 | 1 |
| 0 | 3 | 1 |
| 6 | 2 | 1 |
| 4 | 3 | 1 |
| 5 | 3 | 1 |
| 5 | 3 | 1 |
| 3 | 3 | 1 |
| 5 | 3 | 1 |
| 6 | 5 | 1 |
| 5 | 3 | 1 |
| 4 | 3 | 1 |
| 4 | 3 | 1 |
| 3 | 3 | 1 |
| 6 | 5 | 1 |
| 4 | 3 | 1 |
| 2 | 0 | 2 |
| 4 | 1 | 2 |
| 1 | 1 | 2 |
| 3 | 1 | 2 |
| 3 | 0 | 2 |
| 0 | 1 | 2 |
| 5 | 2 | 2 |
| 0 | 2 | 2 |
| 6 | 0 | 2 |
| 1 | 0 | 2 |
| 2 | 2 | 2 |
| 5 | 0 | 2 |
| 1 | 1 | 2 |
| 4 | 1 | 2 |
| 4 | 0 | 2 |
| 2 | 1 | 2 |
| 1 | 2 | 2 |
| 0 | 0 | 2 |
| 0 | 1 | 2 |
| 6 | 1 | 2 |
| 1 | 0 | 2 |
| 4 | 1 | 2 |
| 0 | 1 | 2 |
| 5 | 0 | 2 |
| 2 | 1 | 2 |
| 0 | 2 | 2 |
| 5 | 0 | 2 |
| 3 | 2 | 2 |
| 3 | 0 | 2 |
| 0 | 2 | 2 |
| 0 | 1 | 2 |
| 0 | 0 | 2 |
| 3 | 0 | 2 |
| 5 | 1 | 2 |
| 6 | 1 | 2 |
| 0 | 1 | 2 |
| 2 | 0 | 2 |
| 2 | 2 | 2 |
| 3 | 2 | 2 |
| 1 | 1 | 2 |
| 2 | 0 | 2 |
| 1 | 1 | 2 |
| 4 | 1 | 2 |
| 1 | 0 | 2 |
| 2 | 0 | 2 |
| 1 | 1 | 2 |
| 6 | 0 | 2 |
| 3 | 1 | 2 |
| 2 | 1 | 2 |
| 3 | 2 | 2 |

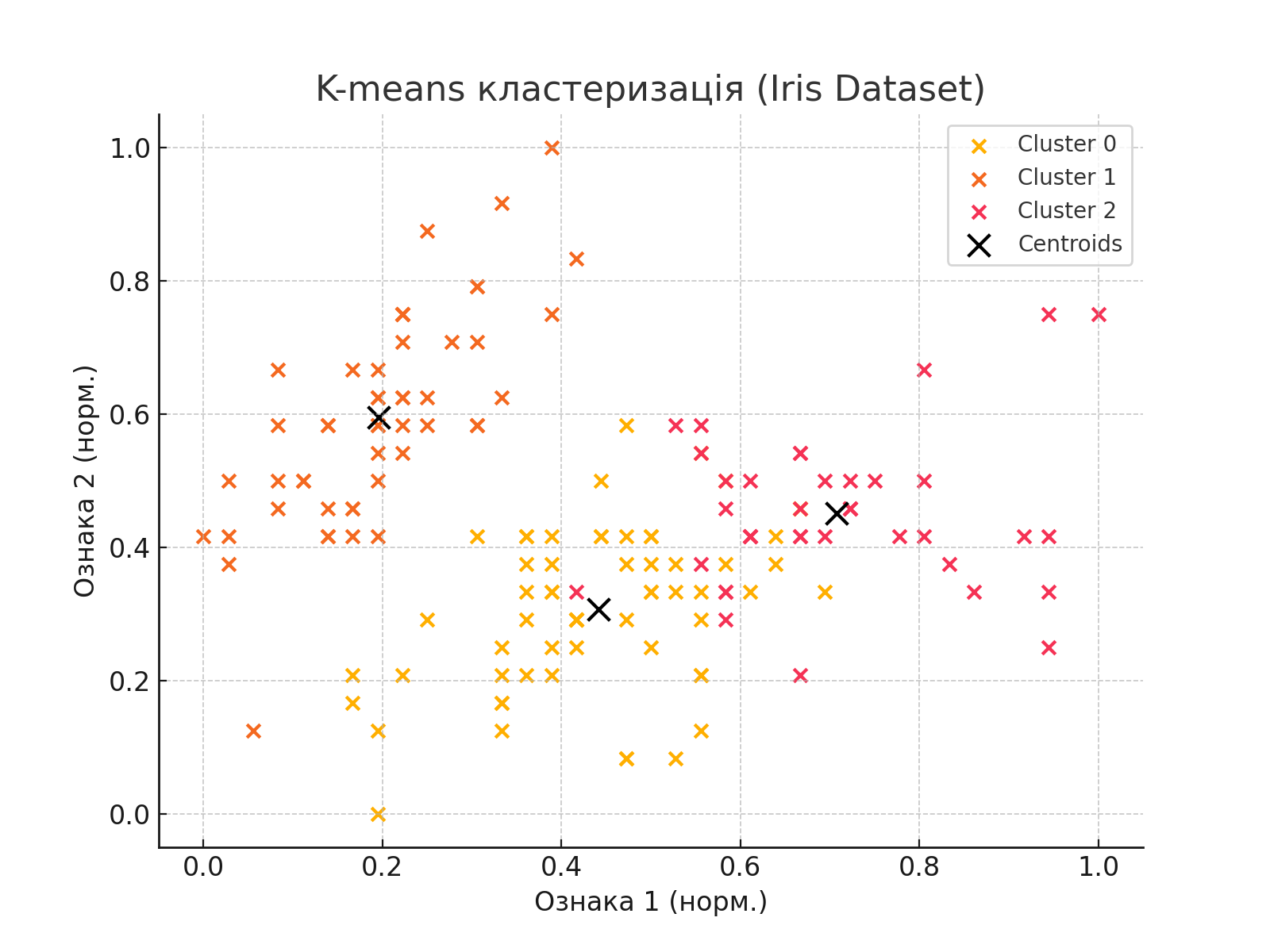
### 2.2. Алгоритм K-means

Алгоритм K-means кластеризує дані, поділяючи їх на k груп, мінімізуючи суму квадратів відстаней до центрів кластерів. На кожній ітерації точки призначаються до найближчого центру, після чого центри оновлюються.

Фрагмент реалізації K-means на Python:

for iteration in range(num\_iterations):  
 distances = np.array([[np.linalg.norm(x - c) for c in centroids] for x in X\_scaled])  
 cluster\_labels = np.argmin(distances, axis=1)  
 new\_centroids = np.array([X\_scaled[cluster\_labels == i].mean(axis=0) for i in range(k)])  
 if np.allclose(centroids, new\_centroids, atol=1e-4):  
 break  
 centroids = new\_centroids

Графік результатів кластеризації методом K-means:



Таблиця порівняння кластерів з класами:

|  |  |  |  |
| --- | --- | --- | --- |
| Ознака 1 | Ознака 2 | Кластер | Клас |
| 0.22 | 0.62 | 1.0 | 0.0 |
| 0.17 | 0.42 | 1.0 | 0.0 |
| 0.11 | 0.50 | 1.0 | 0.0 |
| 0.08 | 0.46 | 1.0 | 0.0 |
| 0.19 | 0.67 | 1.0 | 0.0 |
| 0.31 | 0.79 | 1.0 | 0.0 |
| 0.08 | 0.58 | 1.0 | 0.0 |
| 0.19 | 0.58 | 1.0 | 0.0 |
| 0.03 | 0.38 | 1.0 | 0.0 |
| 0.17 | 0.46 | 1.0 | 0.0 |
| 0.31 | 0.71 | 1.0 | 0.0 |
| 0.14 | 0.58 | 1.0 | 0.0 |
| 0.14 | 0.42 | 1.0 | 0.0 |
| 0.00 | 0.42 | 1.0 | 0.0 |
| 0.42 | 0.83 | 1.0 | 0.0 |
| 0.39 | 1.00 | 1.0 | 0.0 |
| 0.31 | 0.79 | 1.0 | 0.0 |
| 0.22 | 0.62 | 1.0 | 0.0 |
| 0.39 | 0.75 | 1.0 | 0.0 |
| 0.22 | 0.75 | 1.0 | 0.0 |
| 0.31 | 0.58 | 1.0 | 0.0 |
| 0.22 | 0.71 | 1.0 | 0.0 |
| 0.08 | 0.67 | 1.0 | 0.0 |
| 0.22 | 0.54 | 1.0 | 0.0 |
| 0.14 | 0.58 | 1.0 | 0.0 |
| 0.19 | 0.42 | 1.0 | 0.0 |
| 0.19 | 0.58 | 1.0 | 0.0 |
| 0.25 | 0.62 | 1.0 | 0.0 |
| 0.25 | 0.58 | 1.0 | 0.0 |
| 0.11 | 0.50 | 1.0 | 0.0 |
| 0.14 | 0.46 | 1.0 | 0.0 |
| 0.31 | 0.58 | 1.0 | 0.0 |
| 0.25 | 0.87 | 1.0 | 0.0 |
| 0.33 | 0.92 | 1.0 | 0.0 |
| 0.17 | 0.46 | 1.0 | 0.0 |
| 0.19 | 0.50 | 1.0 | 0.0 |
| 0.33 | 0.62 | 1.0 | 0.0 |
| 0.17 | 0.67 | 1.0 | 0.0 |
| 0.03 | 0.42 | 1.0 | 0.0 |
| 0.22 | 0.58 | 1.0 | 0.0 |
| 0.19 | 0.62 | 1.0 | 0.0 |
| 0.06 | 0.12 | 1.0 | 0.0 |
| 0.03 | 0.50 | 1.0 | 0.0 |
| 0.19 | 0.62 | 1.0 | 0.0 |
| 0.22 | 0.75 | 1.0 | 0.0 |
| 0.14 | 0.42 | 1.0 | 0.0 |
| 0.22 | 0.75 | 1.0 | 0.0 |
| 0.08 | 0.50 | 1.0 | 0.0 |
| 0.28 | 0.71 | 1.0 | 0.0 |
| 0.19 | 0.54 | 1.0 | 0.0 |
| 0.75 | 0.50 | 2.0 | 1.0 |
| 0.58 | 0.50 | 0.0 | 1.0 |
| 0.72 | 0.46 | 2.0 | 1.0 |
| 0.33 | 0.12 | 0.0 | 1.0 |
| 0.61 | 0.33 | 0.0 | 1.0 |
| 0.39 | 0.33 | 0.0 | 1.0 |
| 0.56 | 0.54 | 0.0 | 1.0 |
| 0.17 | 0.17 | 0.0 | 1.0 |
| 0.64 | 0.38 | 0.0 | 1.0 |
| 0.25 | 0.29 | 0.0 | 1.0 |
| 0.19 | 0.00 | 0.0 | 1.0 |
| 0.44 | 0.42 | 0.0 | 1.0 |
| 0.47 | 0.08 | 0.0 | 1.0 |
| 0.50 | 0.38 | 0.0 | 1.0 |
| 0.36 | 0.38 | 0.0 | 1.0 |
| 0.67 | 0.46 | 0.0 | 1.0 |
| 0.36 | 0.42 | 0.0 | 1.0 |
| 0.42 | 0.29 | 0.0 | 1.0 |
| 0.53 | 0.08 | 0.0 | 1.0 |
| 0.36 | 0.21 | 0.0 | 1.0 |
| 0.44 | 0.50 | 0.0 | 1.0 |
| 0.50 | 0.33 | 0.0 | 1.0 |
| 0.56 | 0.21 | 0.0 | 1.0 |
| 0.50 | 0.33 | 0.0 | 1.0 |
| 0.58 | 0.38 | 0.0 | 1.0 |
| 0.64 | 0.42 | 0.0 | 1.0 |
| 0.69 | 0.33 | 0.0 | 1.0 |
| 0.67 | 0.42 | 2.0 | 1.0 |
| 0.47 | 0.38 | 0.0 | 1.0 |
| 0.39 | 0.25 | 0.0 | 1.0 |
| 0.33 | 0.17 | 0.0 | 1.0 |
| 0.33 | 0.17 | 0.0 | 1.0 |
| 0.42 | 0.29 | 0.0 | 1.0 |
| 0.47 | 0.29 | 0.0 | 1.0 |
| 0.31 | 0.42 | 0.0 | 1.0 |
| 0.47 | 0.58 | 0.0 | 1.0 |
| 0.67 | 0.46 | 0.0 | 1.0 |
| 0.56 | 0.12 | 0.0 | 1.0 |
| 0.36 | 0.42 | 0.0 | 1.0 |
| 0.33 | 0.21 | 0.0 | 1.0 |
| 0.33 | 0.25 | 0.0 | 1.0 |
| 0.50 | 0.42 | 0.0 | 1.0 |
| 0.42 | 0.25 | 0.0 | 1.0 |
| 0.19 | 0.12 | 0.0 | 1.0 |
| 0.36 | 0.29 | 0.0 | 1.0 |
| 0.39 | 0.42 | 0.0 | 1.0 |
| 0.39 | 0.38 | 0.0 | 1.0 |
| 0.53 | 0.38 | 0.0 | 1.0 |
| 0.22 | 0.21 | 0.0 | 1.0 |
| 0.39 | 0.33 | 0.0 | 1.0 |
| 0.56 | 0.54 | 2.0 | 2.0 |
| 0.42 | 0.29 | 0.0 | 2.0 |
| 0.78 | 0.42 | 2.0 | 2.0 |
| 0.56 | 0.38 | 2.0 | 2.0 |
| 0.61 | 0.42 | 2.0 | 2.0 |
| 0.92 | 0.42 | 2.0 | 2.0 |
| 0.17 | 0.21 | 0.0 | 2.0 |
| 0.83 | 0.38 | 2.0 | 2.0 |
| 0.67 | 0.21 | 2.0 | 2.0 |
| 0.81 | 0.67 | 2.0 | 2.0 |
| 0.61 | 0.50 | 2.0 | 2.0 |
| 0.58 | 0.29 | 2.0 | 2.0 |
| 0.69 | 0.42 | 2.0 | 2.0 |
| 0.39 | 0.21 | 0.0 | 2.0 |
| 0.42 | 0.33 | 2.0 | 2.0 |
| 0.58 | 0.50 | 2.0 | 2.0 |
| 0.61 | 0.42 | 2.0 | 2.0 |
| 0.94 | 0.75 | 2.0 | 2.0 |
| 0.94 | 0.25 | 2.0 | 2.0 |
| 0.47 | 0.08 | 0.0 | 2.0 |
| 0.72 | 0.50 | 2.0 | 2.0 |
| 0.36 | 0.33 | 0.0 | 2.0 |
| 0.94 | 0.33 | 2.0 | 2.0 |
| 0.56 | 0.29 | 0.0 | 2.0 |
| 0.67 | 0.54 | 2.0 | 2.0 |
| 0.81 | 0.50 | 2.0 | 2.0 |
| 0.53 | 0.33 | 0.0 | 2.0 |
| 0.50 | 0.42 | 0.0 | 2.0 |
| 0.58 | 0.33 | 2.0 | 2.0 |
| 0.81 | 0.42 | 2.0 | 2.0 |
| 0.86 | 0.33 | 2.0 | 2.0 |
| 1.00 | 0.75 | 2.0 | 2.0 |
| 0.58 | 0.33 | 2.0 | 2.0 |
| 0.56 | 0.33 | 0.0 | 2.0 |
| 0.50 | 0.25 | 0.0 | 2.0 |
| 0.94 | 0.42 | 2.0 | 2.0 |
| 0.56 | 0.58 | 2.0 | 2.0 |
| 0.58 | 0.46 | 2.0 | 2.0 |
| 0.47 | 0.42 | 0.0 | 2.0 |
| 0.72 | 0.46 | 2.0 | 2.0 |
| 0.67 | 0.46 | 2.0 | 2.0 |
| 0.72 | 0.46 | 2.0 | 2.0 |
| 0.42 | 0.29 | 0.0 | 2.0 |
| 0.69 | 0.50 | 2.0 | 2.0 |
| 0.67 | 0.54 | 2.0 | 2.0 |
| 0.67 | 0.42 | 2.0 | 2.0 |
| 0.56 | 0.21 | 0.0 | 2.0 |
| 0.61 | 0.42 | 2.0 | 2.0 |
| 0.53 | 0.58 | 2.0 | 2.0 |
| 0.44 | 0.42 | 0.0 | 2.0 |

## 3. Висновки

У цій лабораторній роботі реалізовано два популярних алгоритми кластеризації без вчителя: SOM та K-means. Обидва методи продемонстрували здатність виявляти структуру у даних. SOM дозволив візуалізувати просторове розміщення об'єктів, а K-means — розділити вибірку на кластери, близькі до реальних класів.