Дніпровський національний університет імені Олеся Гончара Факультет прикладної математики Кафедра обчислювальної математики та математичної кібернетики

Лабораторна робота №3-4

Інтелектуальний аналіз даних

студента групи ПА-19-2

Ільяшенко Єгора

Варіант №7

Дніпро, 2022

Постановка задачі: «Реалізувати та застосувати алгоритм кластеризації методом k-середніх за випадковими та реальними даними.»

Хід виконання:

1. Власноруч реалізувати алгоритм кластеризації методом k-середніх.
2. Отримати та підготувати вибірку до застосування методом.
3. Застосувати написаний алгоритм до отриманої вибірки.
4. Підготувати звіт за виконаною роботою.

Коментарі щодо ходу виконання:

1. Алгоритм реалізується самостійно без використання готових рішень. Дозволяється використовувати готові рішення для перевірки правильності власної реалізації. Мова програмування - за власним вибором.
2. Вибірки даних дві: випадкові значення та "іриси Фішера" (дані є у файлах команди).
3. Знайти спосіб наглядно продемонструвати результати роботи алгоритму.

Кінцевий звіт має містити:

1. Постановка задачі
2. Теоретичні відомості
3. Хід роботи (реалізація алгоритму)
4. Вибірка даних (продемонструвати вибрані елементи вибірки, пояснити як отримували дані)
5. Результати

До завдання мають бути додані 3 файли:

1. Архів із кодом програми
2. Архів із вибіркою даних
3. Звіт (.doc, .docx, .odt, .pdf)

До оцінювання виконаного завдання також входить факт захисту роботи голосом.

Постановка задачі.

Розробити алгоритм кластеризації під назвою «K-Means» для n-векторної декартової системи координат.

Теоретичні відомості

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание



Програму розробив у Unity3D, візуалізація за допомогою ComputeShader.

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

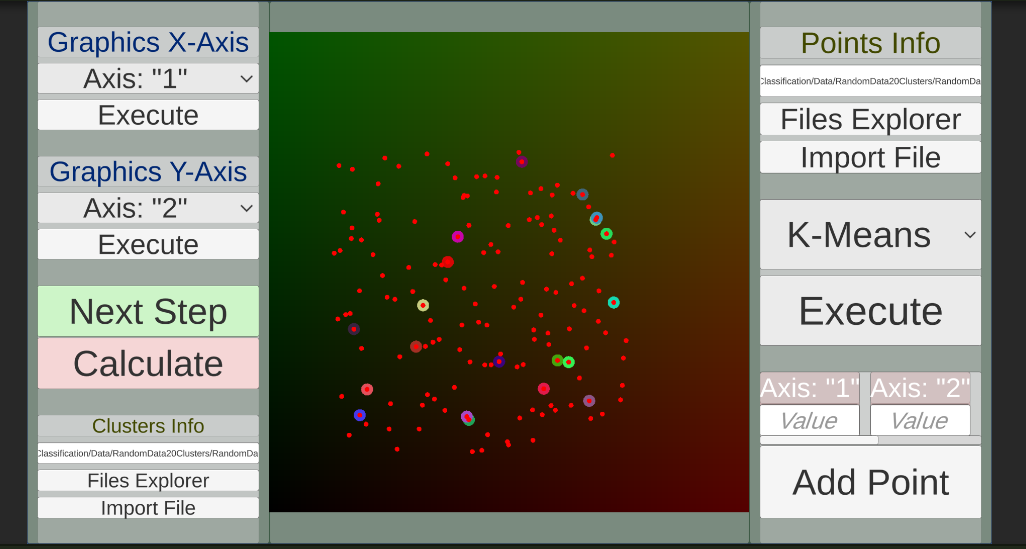
Изображение выглядит как текст

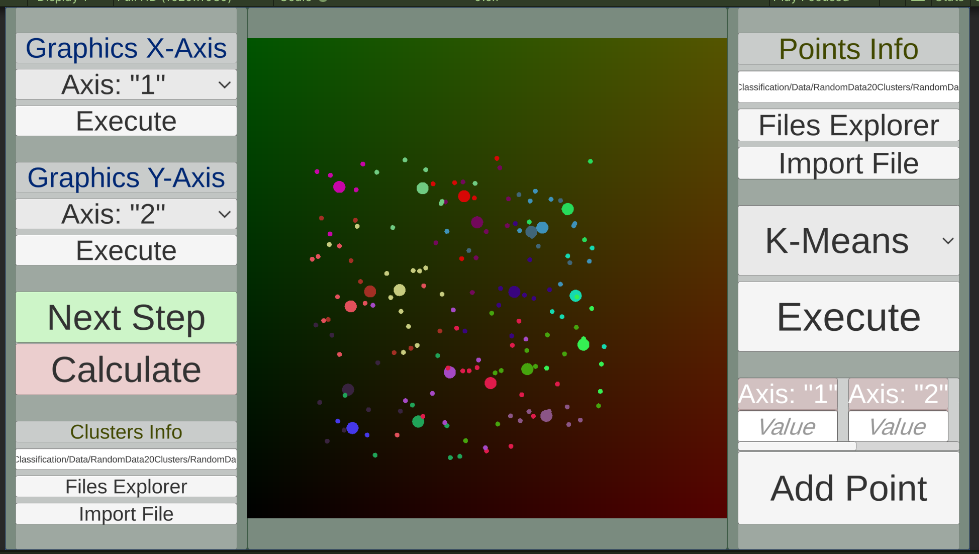
Автоматически созданное описание

Також у програмі є генератор пар (точки і кластери) (EditorOnly)

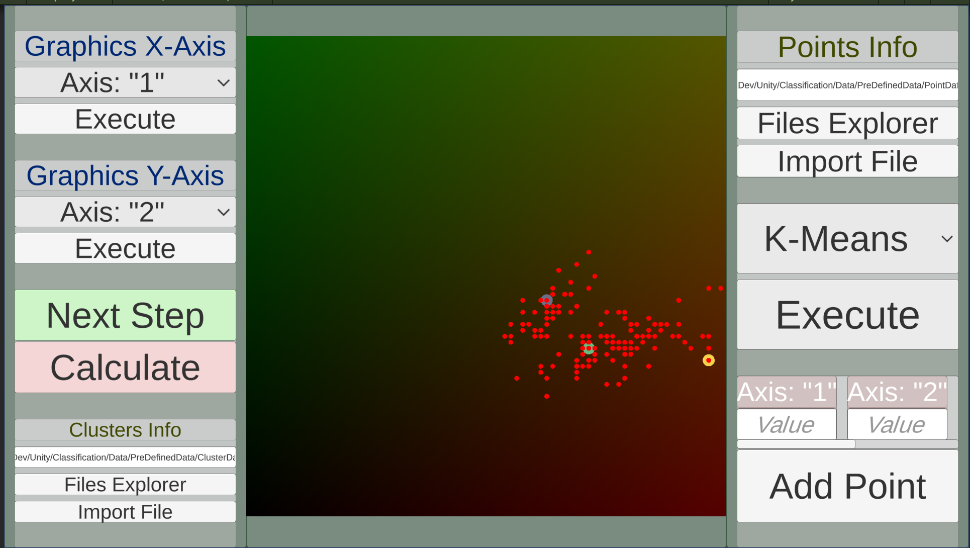
Вибірка точок двох видів: Іриси Фішера та Випадкова вибірка.

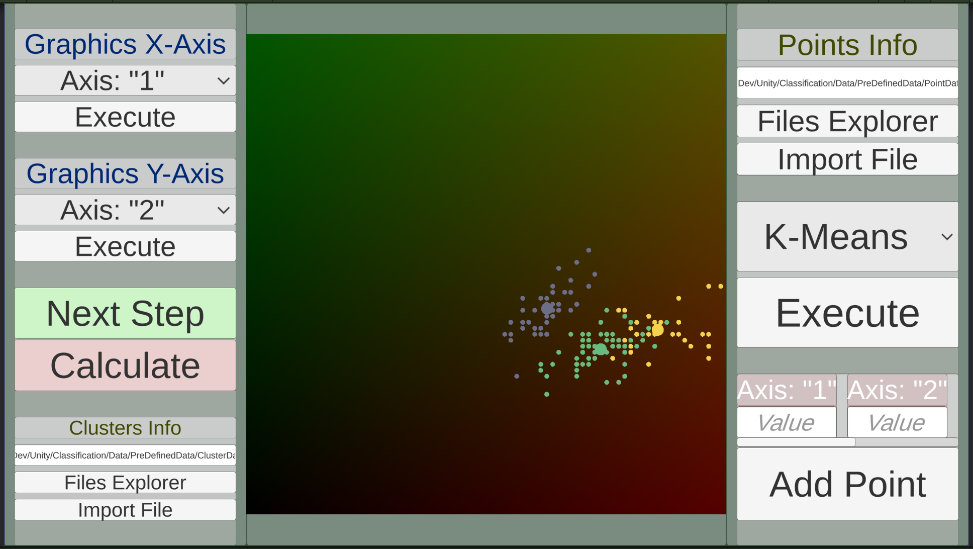
Приклад виконання рандомізованої вибірки





Приклад виконання Ірисів Фішера





Також у програму можна додати додаткову точку.

Керування камерою:

WASD – camera movement

QE – camera scale

Изображение выглядит как текст

Автоматически созданное описание

Дані записуються у такому вигляді

Кожна строка – окрема точка

Строка розділяється за допомогою символа ‘|’ на декілька значень.

Перше значення – векторна розмірність.

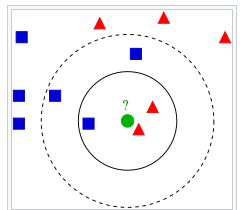
Далі ми перечислюємо координати вектора

Изображение выглядит как текст

Автоматически созданное описание

Кластери записуються схожим чином. Але ми маємо одне додаткове значення. Це ім’я кластера.

K-Nearest



При кластеризації методом K-найближчих, ми повинні задавати параметр k – кількість найближчих точок, які ми будемо брати до уваги. Метод застосовується на вже кластерізованих вибірках.

Изображение выглядит как текст

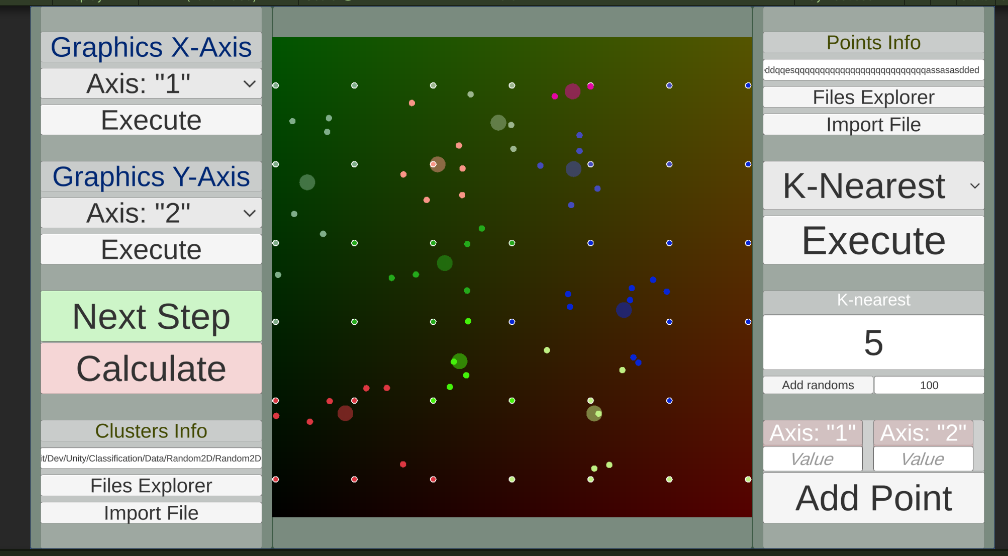
Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание



Висновки: Виконуючи лабораторну роботу №3-4, я розібрався у методах кластеризації, а саме K-Means, K-Nearests. Розробив програму на платформі Unity3D, використовуючи мову програмування C#.

Візуалізація за допомогою шейдерної мови програмування HLSL. У програмі можна імпортувати файли з точками та кластерами, додавати точки під час виконання програми, а також генерувати нові вибірки (за допомогою редактора Unity3D).