****

**Московский авиационный институт**

**(Национальный исследовательский университет)**

**Институт № 3**

**Кафедра 311**

**Интеллектуальные системы**

**Лабораторная работа № 6**

**«Кластерный анализ»**

**Выполнил студент  
Веденеев Максим Кириллович**

**Группа М3З-501-БК**

**Дата 06.10.2023 г.**

**Принял преподаватель  
Кос Оксана Игоревна**

Оглавление

[Цель лабораторной работы 3](#__RefHeading___1)

[Глава 1. Кластеризация методом k-средних 4](#__RefHeading___2)

[Глава 2. Реализация 6](#__RefHeading___3)

[Итоги лабораторной работы 9](#__RefHeading___4)

[Список литературы 10](#__RefHeading___5)

# Цель лабораторной работы

Изучить:

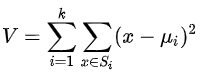
1. Кластерный анализ.
2. Реализовать его.

# Глава 1. Кластеризация методом k-средних

Кластерный анализ - многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя.

Метод k-средних— наиболее популярный метод кластеризации.

Действие алгоритма таково, что он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров:

(1)

где k — число кластеров, Si — полученные кластеры, i=1,2,k,а mu — центры масс всех векторов x из кластера S.

Алгоритм представляет собой версию EM-алгоритма, применяемого также для разделения смеси гауссиан. Он разбивает множество элементов векторного пространства на заранее известное число кластеров k.

Основная идея заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике.

Алгоритм завершается, когда на какой-то итерации не происходит изменения внутрикластерного расстояния. Это происходит за конечное число итераций, так как количество возможных разбиений конечного множества конечно, а на каждом шаге суммарное квадратичное отклонение V уменьшается, поэтому зацикливание невозможно.

На Рисунке 1 демонстрируется алгоритм разбития на кластеры:

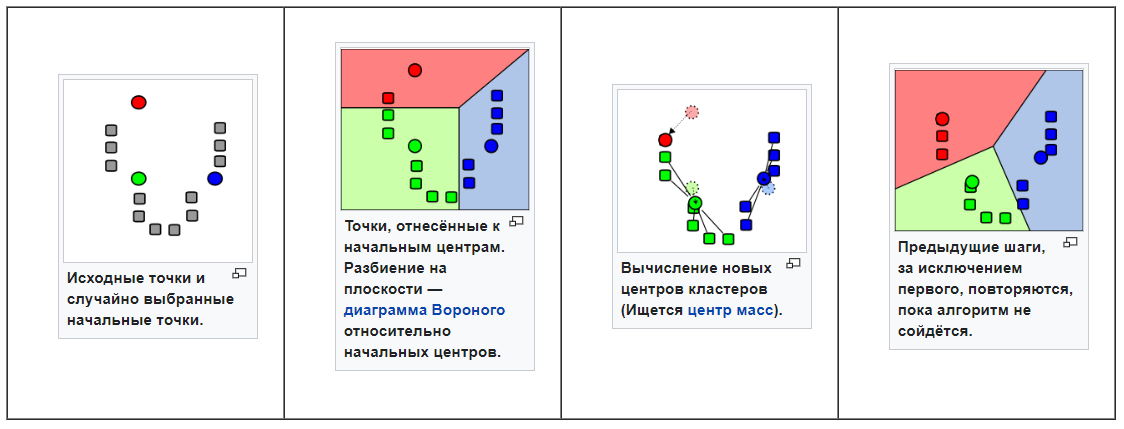


Рисунок 1 – схема работы алгоритма

# Глава 2. Реализация

import numpy as np

def main():

arr = np.array([2, 4, 10, 12, 3, 20, 30, 11, 25])

print("Input: " + str(arr))

m1, m2 = 0, 0 # Значение центроидов

m = 0 # Число итераций

flag = False # Флаг, который даст знать, что центры масс m1, m2 не меняются

a, b = arr[0], arr[1] # a, b сохраняют предыдущее состояние центра масс, чтобы сравнить с текущим значением

m1, m2 = a, b

cluster1, cluster2 = [], []

while not flag:

sum1, sum2 = 0, 0

cluster1, cluster2 = [], []

m += 1 # Подсчет числа итераций

k, j = 0, 0 # Элементы, которые будут записывать значения в первый и второй массив

for i in range(len(arr)):

if abs(arr[i] - m1) <= abs(arr[

i] - m2): # Принимаем за первую центроиду первое значение массива, за вторую - второе и сравниваем

cluster1.append(arr[i]) # Если проходит, то записываем в первый кластер

k += 1

else:

cluster2.append(arr[i])

j += 1

for i in range(k): # Вычисляем сумму для 1-го кластера

sum1 += cluster1[i]

for i in range(j): # Вычисляем сумму для 2-го кластера

sum2 += cluster2[i]

a, b = m1, m2 # Записываем старые центроиды

m1 = sum1 / k # Высчитываем новые центроиды

m2 = sum2 / j

flag = (m1 == a and m2 == b) # Если оба значения центроида = старому => выйти из цикла

print("\n\_\_\_\_\_\_\_\_\_\_\_\_ITERATION", m, "\_\_\_\_\_\_\_\_\_\_\_\_")

print("m1=", m1, "m2=", m2)

print("Cluster 1:\n", cluster1)

print("Cluster 2:\n", cluster2)

print("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_")

print("Final value of centorid: m1=", m1, "m2=", m2)

print()

print("Final cluster 1 is:\n", cluster1)

print("Final cluster 2 is:\n", cluster2)

if \_\_name\_\_ == "\_\_main\_\_":

main()

sssssss

# Итоги лабораторной работы

Мы освоили метод k-средних кластерного анализа.

# Список литературы

1. [[wikipedia.org](https://ru.wikipedia.org/wiki/Метод_k-средних)] – Метод k-средних