|  |  |  |  |
| --- | --- | --- | --- |
| Министерство образования Республики Беларусь  Учреждение образования  Белорусский Государственный Университет Информатики и Радиоэлектроники | | | |
| Факультет | Компьютерных сетей и систем | | |
| Кафедра | Информатики  Дисциплина: Конструирование те технологии электронных вычислительных средств | | |
|  |  | | |
| ЛАБОРАТОРНАЯ РАБОТА №6  «Кластеризация»  БГУИР 1-40 81 04 | | | |
| Магистрант:  гр. 858642  Кукареко А.В. | |  | Проверил:  Стержанов М. В. |
| Минск, 2019 | | | |

# 

ХОД РАБОТЫ

**Задание.**

Набор данных ex6data1.mat представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит две переменные X1 и X2 - координаты точек, которые необходимо кластеризовать.

Набор данных bird\_small.mat представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит массив размером (16384, 3) - изображение 128x128 в формате RGB.

1. Загрузите данные ex6data1.mat из файла.
2. Реализуйте функцию случайной инициализации K центров кластеров.
3. Реализуйте функцию определения принадлежности к кластерам.
4. Реализуйте функцию пересчета центров кластеров.
5. Реализуйте алгоритм K-средних.
6. Постройте график, на котором данные разделены на K=3 кластеров (при помощи различных маркеров или цветов), а также траекторию движения центров кластеров в процессе работы алгоритма
7. Загрузите данные bird\_small.mat из файла.
8. С помощью алгоритма K-средних используйте 16 цветов для кодирования пикселей.
9. Насколько уменьшился размер изображения? Как это сказалось на качестве?
10. Реализуйте алгоритм K-средних на другом изображении.
11. Реализуйте алгоритм иерархической кластеризации на том же изображении. Сравните полученные результаты.
12. Ответы на вопросы представьте в виде отчета..

**Результат выполнения:**

1. Загрузите данные ex6data1.mat из файла.





Рисунок 1 – исходные данные файла ex6data1.mat.

2. Реализуйте функцию случайной инициализации K центров кластеров.

Функция инициализации K центров кластеров:



3. Реализуйте функцию определения принадлежности к кластерам.

Функция определения принадлежности к кластерам:



4. Реализуйте функцию пересчета центров кластеров.

Функция пересчета центров кластеров:



Вспомогательная функция подсчета принадлежности точек к кластерам:



5. Реализуйте алгоритм K-средних.

Функция реализующая алгоритм «K-средних»:



6. Постройте график, на котором данные разделены на K=3 кластеров (при помощи различных маркеров или цветов), а также траекторию движения центров кластеров в процессе работы алгоритма.



Рисунок 2 – график разделения данных на 3 кластера и траектории движения центроидов.

7. Загрузите данные bird\_small.mat из файла..





Рисунок 3 – исходные данные файла bird\_small.mat.

8. С помощью алгоритма K-средних используйте 16 цветов для кодирования пикселей.



9. Насколько уменьшился размер изображения? Как это сказалось на качестве?

Оригинальный размер bird\_small.mat файла составляет: 128 x 128 x 3 = 49152

Размер нового файла составляет: 16 x 3 + 16384 = 16432, что почти в 3 раза меньше оригинального размера.



Рисунок 4 – сравнение исходного изображения и сжатого.

Если посмотреть на рисунок 4, то к сожалению можно заметить разницу в качестве невооруженным взглядом.

10. Реализуйте алгоритм K-средних на другом изображении.



Рисунок 5 – сравнение другого исходного изображения и сжатого.

11. Реализуйте алгоритм иерархической кластеризации на том же изображении. Сравните полученные результаты.



Иерархическая кластеризация была осуществлена с помощью готового алгоритма «AgglomerativeClustering» из пакета «sklearn.cluster». В качестве метрики связанности использовалось евклидово расстояние, а в качестве оценки расстояния между кластерами был использован метод "ward" (метод Уорда).



Рисунок 6 – сравнение оригинального изображения, с сжатыми с помощью алгоритмов «иерархической кластеризации» и «к-средних».

Если взглянуть на рисунок 6, то можно заметить, что картинка сжатая с помощью «алгоритма иерархической кластеризации» выглядит лучше, чем картинка сжатая с помощью алгоритма «к-средних».

**Вывод.**

В ходе выполнения лабораторной работы я ознакомился с одним из направлений «обучения без учителя» - кластеризацией, реализовал алгоритм к-средних и применил его для разбиения набора точек на группы и для сжатия изображения. Так же ознакомился с алгоритмом «иерархической кластеризации» и применил его для сжатия изображения.