Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра информатики

Дисциплина: Машинное обучение (МО)

Отчет о лабораторной работе №6

«Кластеризация»

|  |
| --- |
| **Выполнил:** |
| **Студент гр. 858341** |
| **Немкович А. В.** |

|  |
| --- |
| **Проверил:** |
| **Стержанов М. В.** |

**Минск 2019**

Содержание

[Постановка задачи 3](#_Toc25275643)

[Ход выполнения 4](#_Toc25275644)

[Выводы 8](#_Toc25275645)

# Постановка задачи

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.

**Задание.**

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.

# Ход выполнения

**Задание 1**: Загрузка данных ex6data1.mat из файла.



Рис 1. – код загрузки данных (ex6data1.mat)

**Задание 2**: Реализация функции случайной инициализации K центров кластеров.

Для выполнения этой задачи использовалась функция kMeansInitCentroids:

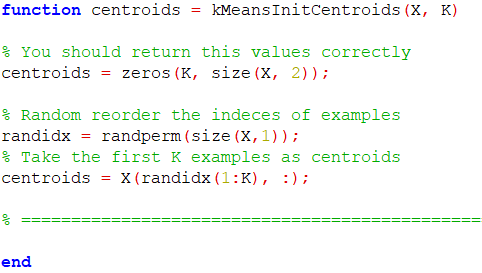


Рис 2. – Код функции kMeansInitCentroids

**Задание 3**: Реализация функции определения принадлежности к кластерам.

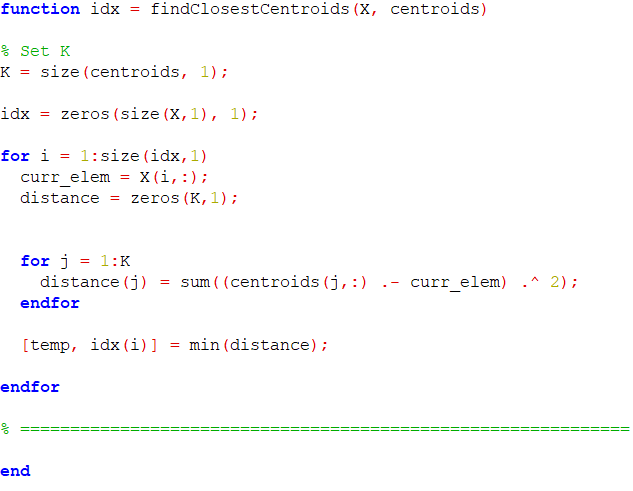
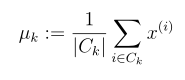


Рис 3. – Код функции определения принадлежности к кластерам

**Задание 4**: Реализация функции пересчета центров кластеров.

Формула пересчета центра *k*-го кластера:



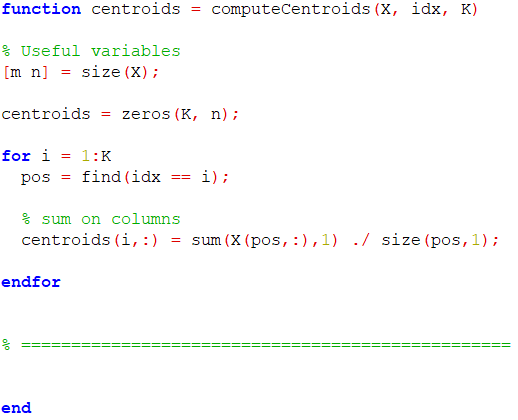


Рис 4. – Код функции пересчета центров кластеров

**Задание 5**: Реализация алгоритма K-средних.

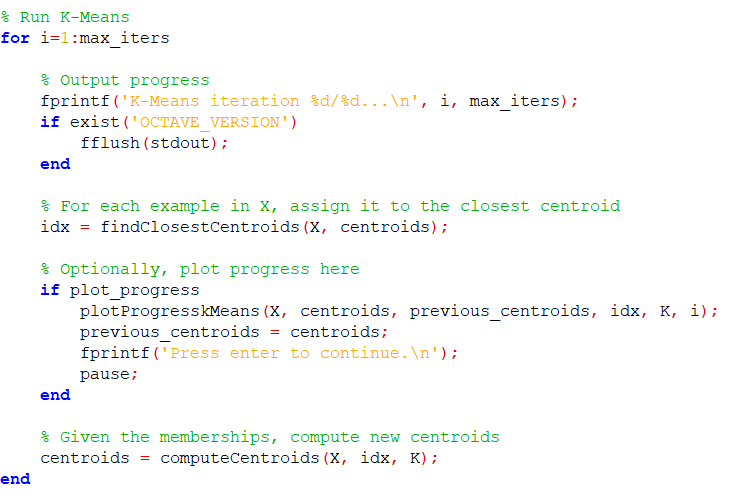


Рис 5. – Код алгоритма K-средних

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

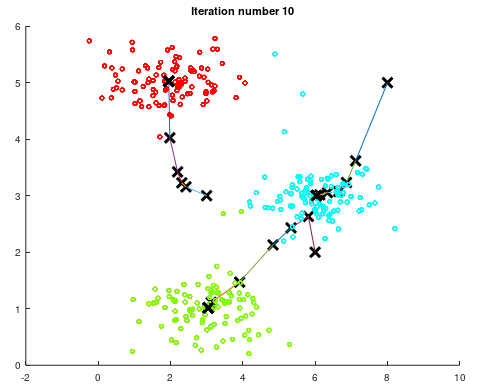


Рис 6. – График разделения данных на 3 кластера

**Задание 7**: Загрузка данных bird\_small.mat из файла.



Рис 7. – Исходное изображение 128х128

**Задание 8**: Использование 16 цветов для кодирования пикселей с помощью алгоритма K-средних.

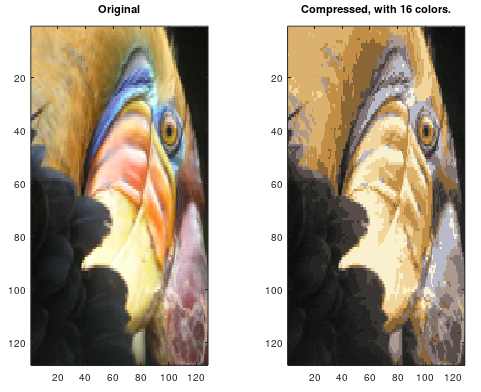


Рис 9. – Сравнение исходного изображения и сжатого (16 цветов)

# Выводы

Мы научились использовать алгоритм K-средних для решения задачи классификации, а также для сжатия изображения, используя для кодировки его пикселей только *n*-ное количество цветов.

Алгоритм K-средних:

* Инициализации K центров кластеров;
* Определение принадлежности к кластерам объектов выборки;
* Пересчет центров кластеров.