

# **Практическая работа №7 от 17.04.2020**

## **Методы кластеризации**

### ***Цель работы***

Практическое использование метода k-средних для задачи кластеризации.

### ***Задачи работы***

1. Научиться использовать метод k-средних.
2. Научиться анализировать качество кластеризации методов k-средних.

### ***Перечень обеспечивающих средств***

1. ПК.
2. Учебно-методическая литература.
3. Задания для самостоятельного выполнения.

### ***Общие теоретические сведения***

#### **Задача кластеризации**

Задача кластеризации – это пример задачи с обучением без учителя. В данной задаче рассматривается набор данных, о которых нет дополнительной информации о принадлежности к тем или иным классам.

Тем не менее, есть методы, позволяющие на основании анализа данных разделять их на кластеры по различным признакам.

#### **Метод k-средних и анализ качества кластеризации**

Метод k-средних используется для кластеризации на основе расстояния между отдельными элементами набора данных.

Можно использовать любое определение расстояния. В случае, если набор данных состоит из чисел или наборов (векторов) чисел, логично использовать простое Евклидово расстояние – квадратный корень из суммы квадратов разностей отдельных координат.

Метод k-средних оптимизирует (ищет минимум) величину инерции — сумму квадратов расстояний от всех элементов набора данных до центров построенных кластеров.

Более подробно метод рассматривается в лекции №3.

## **Задание**

- Сделайте форк репозитория [https://github.com/mosalov/k\\_mean\\_practice](https://github.com/mosalov/k_mean_practice).
- Откройте репозиторий в Binder(<https://mybinder.org/>) и запустите файл «k\_means.ipynb».
- Выполните представленный в файле код, чтобы понять, как применять метод k-средних с использованием библиотеки sklearn.
- Выполните инструкции, указанные в комментариях во второй части файла, начиная с «Загрузите файл "worldcities.csv" в DataFrame.»
- Сохраните файл Jupyter notebook с названием «Задание 7.ipynb» и загрузите его в репозиторий.

## **Требования к предоставлению результатов**

Требуется представить результаты в виде письма на адрес [mosalov.op@ut-mo.ru](mailto:mosalov.op@ut-mo.ru) с указанием ФИО и ссылки на репозиторий с сохраненным файлом Jupyter notebook.

## **Литература**

1. <http://datascientist.one/k-means-algorithm/>
2. <https://craftappmobile.com/кластеризация-методом-k-средних/>
3. <http://datareview.info/article/klasterizatsiya-s-pomoshhyu-metoda-k-srednih-na-python/>
4. <https://proglab.io/p/unsupervised-ml-with-python/>