Лабораторная работа №2. Кластерный анализ.

Задание. Реализовать алгоритм кластеризации k-means (k-средних, алгоритм Ллойда).
Вход: Dataset (обучающая выборка), k (количество кластеров)
Выход: Метки кластеров для каждого объекта; центры кластеров.
Дополнительные условия.
1. Dataset: Iris. Можно выбрать другой датасет, но с небольшим количеством количественных признаков.
2. Библиотеки: все стандартные библиотеки для Data Science, кроме sklearn.
3. Параметр k (количество кластеров) задается пользователем при запуске.
4. Метрика расстояния: евклидовое
5. Метрика качества кластеризации при фиксированном k: сумма квадратов внутрикластерных расстояний (в каждом кластере рассчитывается сумма квадратов расстояний до центра и суммируются).

Требования к реализации.
1. Датасет в коде должен под переменной X (двумерный массив numpy).
2. Модульная структура программы. Рекомендация: Distance (функция расстояния), InitializationCentre (начальная инициализация центров кластеров), Expectation (кластеризация с текущими центрами), Maximization (перерасчет координат центров), Quality (функционал качества кластеризации). Можно реализовать в виде класса Kmeans (ООП вариант приветствуется).
3. Программа должна быть применима к любому датасету.
4. Так как алгоритм k-means зависит от начальной инициализации центров кластеров нужно в программе несколько раз (например, m=10) получить кластеризацию из различных случайных инициализаций и выбрать среди них наилучшую по метрике качества. Можно количество инициализаций задать вначале программы как параметр.
5. Построить двумерный график (выбрать два признака) с цветными метками реальных классов ирисов и на ней обозначить «крестиками» центры кластеров – нужно для визуальной проверки адекватности кластеризации.
6. Приветствуется (и даже необходимо) применение удобных библиотечных методов и функций.
7. Комментарии в коде.

Материалы.
- Лекция Воронцова К.В. «Курс Машинное обучение» 2019: https://www.youtube.com/watch?v=SZkrxWhI5qM&list=PLJOzdkh8T5krxc4HsHbB8g8f0hu7973fK – Машинное обучение. Кластеризация и частичное обучение. К.В. Воронцов, Школа анализа данных, Яндекс.
- http://www.machinelearning.ru/wiki/images/5/52/Voron-ML-Clustering-SSL-slides.pdf – презентация «Кластеризация и частичное обучение», алгоритм Ллойда