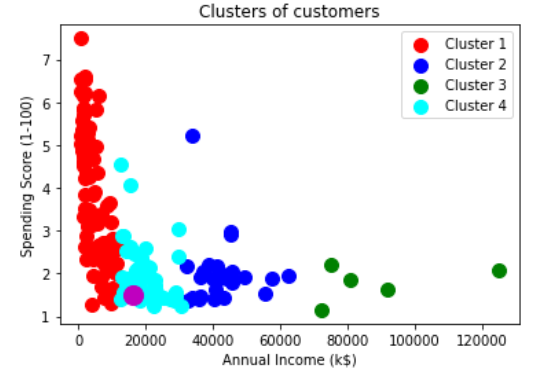
**Лабораторная работа № 10**

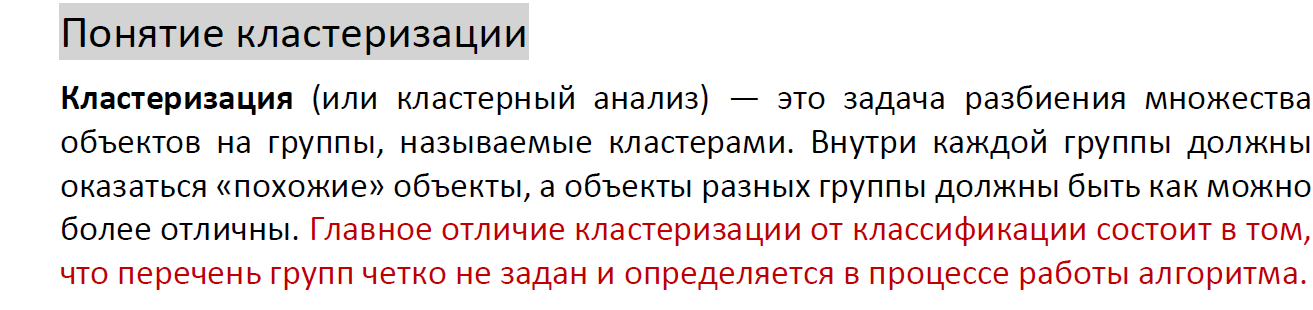
**Тема**: Алгоритмы кластеризации.

**Цель работы** *–* научитьсяприменять алгоритм K-means и алгоритм иерархической кластеризации.

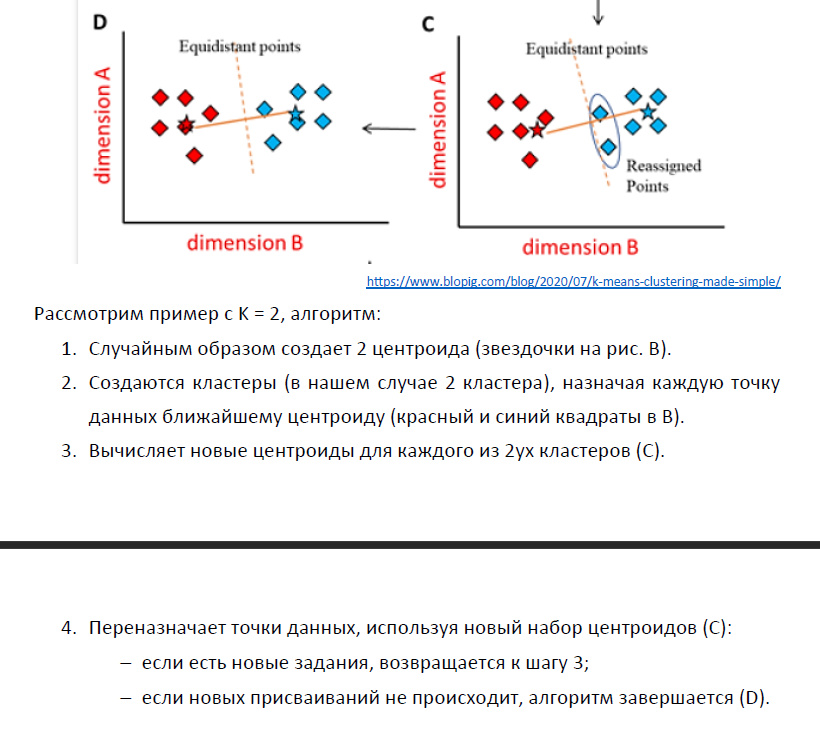
1. Для работы используйте данные из *Country-data.csv* (можно использовать и другие данные на свой выбор, подходящие для кластеризации)
2. Из датасета Country-data.csv выбрать 2 параметра, характеризующих вашу цель исследования (например, исследуем уровень медицины в стране, или благосостояние и т.д.).
3. Разделите данные на кластеры методом K-means, обоснуйте выбор количества кластеров.
4. Разделите данные на кластеры методом иерархической кластеризации, выберите с помощью дендрограммы оптимальное количество кластеров.
5. Визуализируйте результаты кластеризации двух методов и сравните их.
6. Из датасета выберите любую страну и визуализируйте ее в виде точки отличного цвета и размера на графике кластеров (*пример на рисунке, точка пурпурного цвета*).

**Вопросы:**

1. Что решают задачи кластеризации в машинном обучении?



1. Расскажите принцип работы метода K-means.



1. Как можно выбрать оптимальное количество кластеров в K-means?

При помощи графика зависимости wcss от количества кластеров

1. Расскажите принцип работы метода иерархической кластеризации.



1. Для чего можно использовать дендрограмму в методе иерархической кластеризации?

Для определения наиболее оптимального количества кластеров