**Лабораторная № 3.**

**Кластеризация.**

**Решите задачу кластеризации на данных** pluton, votes.repub и swiss **с использованием стандартных пакетов в среде R.**

**ЧАСТЬ I Изучение метода**

Ответьте на вопросы: **Таблица 1**

|  |  |  |
| --- | --- | --- |
| **№** | **Вопрос** | **Ответ** |
|  | Как называются методы, изучаемые в лабораторной 3? |  |
|  | К какой категории методов ML они относятся? |  |
|  | Что может прогнозировать алгоритм, обученный данным методом? |  |
|  | В чем принципиальное отличие задачи классификации от задачи кластеризации? |  |
|  | Что значит «метод сошелся» для метода K-means? |  |
|  | Чем могут отличаться задачи кластеризации? |  |
|  | Приведите примеры «жесткой» и «мягкой» кластеризаций (не из указанных в лекции). |  |

**ЧАСТЬ II Выполнение задания**

1. Разбейте множество объектов из набора данных pluton на 3 кластера методом центров тяжести (К-means). Сравните качество разбиения в зависимости от макисмального числа итераций алгоритма.
2. Постройте дендрограмму для набора данных votes.repub (число голосов, поданных за республиканцев на выборах с 1856 по 1976 год). Проинтерпретируйте полученный результат.
3. Рассчитайте, какое минимальное число главных компонент объясняет больше 90% изменчивости в исходных данных swiss и добавлять значения этих компонент в исходный файл данных в виде новых переменных.

**ЧАСТЬ III Оформление работы**

Оформить работу надо в бумажном виде:

* 1. Титульный лист (тема лаб. 3!)
  2. Таблица 1 с ответами на вопросы.
  3. Для каждой из задач:
     1. Формулировка задачи.
     2. Описание признаков используемого датасета (массива данных).
     3. Программный код.
     4. Результат вычислений в виде графиков.
     5. Пояснения выведенных результатов.

**Литература**

1. Визуализация метода K-means

<https://www.naftaliharris.com/blog/visualizing-k-means-clustering/>