

Листок 06. Задача кластеризации

Замечание обязательно проводим предварительную обработку данных: удаление пропущенных значений, нормировку, преобразование категориальных признаков

Упражнение 1. Для набора данных `countries` проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

Упражнение 2. Для набора данных `countries` найдите «оптимальное» число кластеров для метода

1. k-средних
2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin

Упражнение 3. Из набора данных `countries` возьмите переменные `sleep`, `totwrk`, `age`, `educ` и проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

Упражнение 4. Из набора данных `countries` возьмите переменные `sleep`, `totwrk`, `age`, `educ` и найдите «оптимальное» число кластеров для метода

1. k-средних
2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin

Упражнение 5. Для набора данных `Labour` проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

Упражнение 6. Для набора данных `Labour` найдите «оптимальное» число кластеров для метода

1. k-средних
2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin