Домашнее задание по машинному обучению №2

Дедлайн на полный балл – 9 марта

Дедлайн на половинный балл – 16 марта

Датасеты – blobs.csv (1-3), cancer.csv (4)

Результатом в задачах 1-3 должна быть визуализация кластеров и значение метрики Dunn index.

1. (2) Реализуйте алгоритм k-means и кластеризуйте точки датасета на [2,3,4,5] кластеров.

2. (2) Реализуйте алгоритм DBSCAN, найдите параметры для кластеризации на 4 кластера.

3. (3) Реализуйте алгоритм Agglomerative Clustering и выведите результаты кластеризации на [2,3,4,5] кластеров. Метрика на объединение - любая. Требование на связность тоже по желанию.

4. (2) Кластеризуйте данные из датасета cancer.csv (с нормировкой и без) на 2-10 кластеров методом k-means и посчитайте метрику purity по каждому разбиению.

Дополнительные задачи

1. (0) Реализуйте алгоритм Аffinity Propagation, найдите параметры для кластеризации на <4, 4, >4 кластеров.