

Зеленым обозначено условие для первого варианта, желтым для второго

1. Реализовать кластеризацию набора данных IRIS по признакам PetalWidth, SepalWidth, PetalLength, SepalLength так, чтобы максимально минимизировать ошибку принадлежности к классу (переменная Species). Саму переменную Species в качестве признаковой не использовать и ни при каких преобразованиях входного пространства или настройки параметров кластеризации не учитывать. Кластеров в результате должно быть три. Желательно получить ошибку меньше 10 наблюдений, более 20 – задача не засчитывается. Рекомендуется:
  - Проверить различные алгоритмы кластеризации (иерархические или k-means, параметрические или непараметрические вероятностные)
  - Поискать подходящее преобразование признаков, включая поворот/растяжение, добавление полиномиальных признаков или другие элементарные преобразования.
2. Рассчитать и визуализировать индекс силуэта, найти наиболее типичного и наиболее нетипичного представителя каждого кластера.
3. Построить графики всех попарных проекций переменных (всего 6) и проекцию на плоскость с помощью LDA (MDS), где цветом указать полученную принадлежность кластеру, маркером истинный класс, а ошибочные наблюдения выделить отдельным цветом.