

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:

- 1) Проверить различные алгоритмы кластеризации (иерархические, на основе строгого разбиения, параметрические вероятностные, непараметрические вероятностные и т.д.)
- 2) Поискать подходящее преобразование признаков, включая поворот/растяжение, добавление полиномиальных признаков или другие элементарные преобразования.

Построить графики всех попарных проекций переменных (всего 6), где цветом указать принадлежность кластеру, а ошибочные наблюдения выделить отдельным цветом или другой формой маркера.