



Машинное обучение

НИЯУ МИФИ, Кафедра финансового мониторинга

Лабораторный практикум.

В.Ю. Радыгин, Д.Ю. Куприянов

Семестр 2. Лабораторная работа 4

Лабораторная работа 4 рассчитана на два занятия и работу дома. Её целью является изучение основ кластеризации данных с помощью методов Affinity Propagation, DBSCAN, Spectral Clustering.

Задание 1

1. Загрузите с сайта <https://sci2s.ugr.es/keel/datasets.php> набор статистических данных, указанный в вашем варианте. Разберитесь, какие данные приведены в наборе и какой атрибут является меткой класса.
2. На основе загруженного файла создайте Pandas DataFrame, подобрав правильные типы данных столбцов.
3. Выполните стандартизацию полученного дата фрейма.
4. Выполните расчёт главных компонент. Преобразуйте набор к новой системе координат и отсеките те координаты, которым соответствуют собственные значения, меньшие 1.
5. Выполните кластеризацию данных с помощью метода Affinity Propagation. Сравните число кластеров с реальным.
6. Используя метод силуэта, определите оптимальное число кластеров для кластеризации методами DBSCAN и Spectral Clustering. Сравните с реальным числом кластеров.
7. Выполните кластеризацию методами K-Means, DBSCAN и Spectral Clustering. Сравните с реальным числом кластеров с реальным числом кластеров
8. Постройте четыре раза проекцию по двум координатам с наибольшими собственными значениями, раскрасив точки в различные цвета в соответствии с реальными классами; классами, предсказанными k-means, и классами, предсказанными DBSCAN и Spectral Clustering.

Варианты

Задание 1

1. <https://sci2s.ugr.es/keel/dataset.php?cod=210>
2. <https://sci2s.ugr.es/keel/dataset.php?cod=209>
3. <https://sci2s.ugr.es/keel/dataset.php?cod=107>
4. <https://sci2s.ugr.es/keel/dataset.php?cod=72>