

Домашнее задание 13

Кластеризация и классификация.

Выполняя это задание, вы научитесь выявлять в наблюдениях скопления похожих друг на друга объектов – кластеры, а также относить вновь поступающие наблюдения к одному из ранее выявленных кластеров.

Задание

1. Найти в интернете многомерную выборку, т.е. результаты наблюдений за большим (не менее 100) числом объектов, для каждого из которых известно несколько количественных признаков. Наблюдаемые объекты должны быть разбиты на несколько (3 или более) классов, и для каждого объекта должно быть известно, к какому классу он относится.
2. С помощью метода k средних разбить массив наблюдений на кластеры. Сравнить получившиеся кластеры с реально наблюдаемыми классами.
3. Случайным образом разбить выборку на обучающую и контрольную. По обучающей выборке построить классификатор с помощью метода k ближайших соседей и наивный байесовский классификатор. Сравнить результаты их работы (процент правильных классификаций) на контрольной выборке.
4. По желанию можно также визуализировать результаты кластеризации и классификации, снизив с помощью метода главных компонент размерность наблюдений до двух и изобразив их в виде точек на плоскости.