**Подбор параметров для плохо разделимых данных (дополнение)**

Изображение выглядит как диаграмма

Автоматически созданное описание

Качество кластеризации для плохо разделимых данных, нас не устраивает, попробуем подобрать более подходящие параметры p и q.

* Сначала был произведен подбор параметра q в диапазоне [0.03, 0.1]
* Хорошие результаты давал q = 0.085, на нем я остановился
* Затем я подбирал параметр p, задачей было увеличить количество связанных опорных векторов, повысив точность кластеризации

По результатам подбора параметра p была получена таблица:

Изображение выглядит как стол

Автоматически созданное описание

Проиллюстрируем часть итераций подбора параметра:

Изображение выглядит как диаграмма

Автоматически созданное описание

Изображение выглядит как диаграмма

Автоматически созданное описаниеИзображение выглядит как диаграмма

Автоматически созданное описание

Таким образом, оптимальными параметрами для данного разбиения, исходя из критерия Калининского-Харабаша оказались p = 0.25, q = 0.085.

Однако если взглянуть на исходное разбиение по кластерам для моделируемой выборки, то можно заметить, что при таком наборе параметров мы теряем информацию о наличии 3го кластера. Это связано с тем, что критерий не учитывает настоящее разбиение исходных данных, не сравнивает исходное разбиение с получившимся.

Если нам важен 3й кластер, то лучше выбрать набор параметров (0.23, 0.085).

**Итог:** оптимальные параметры p и q зависят от конкретной исследуемой выборки. Их нужно подбирать для каждой выборки индивидуально, сначала следует взять маленький p и подбирать параметр q так, чтобы число кластеров было близко к правильному, а затем подобрать p так, чтобы снизить количество опорных векторов и увеличить число связанных опорных векторов.