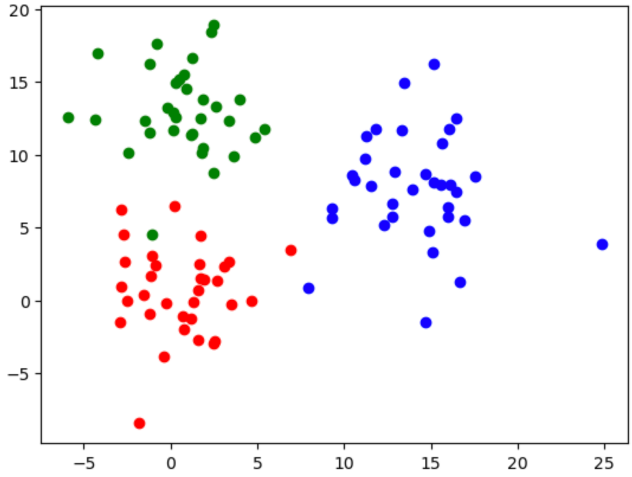
Подбор параметров для плохо разделимых данных (дополнение)



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

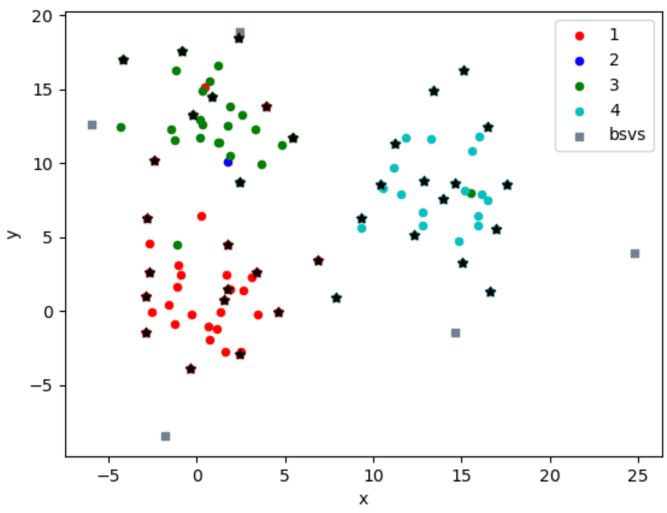
- Сначала был произведен подбор параметра q в диапазоне [0.03, 0.1]
- Хорошие результаты давал q = 0.085, на нем я остановился
- Затем я подбирал параметр p, задачей было увеличить количество связанных опорных векторов, повысив точность кластеризации

По результатам подбора параметра p была получена таблица:

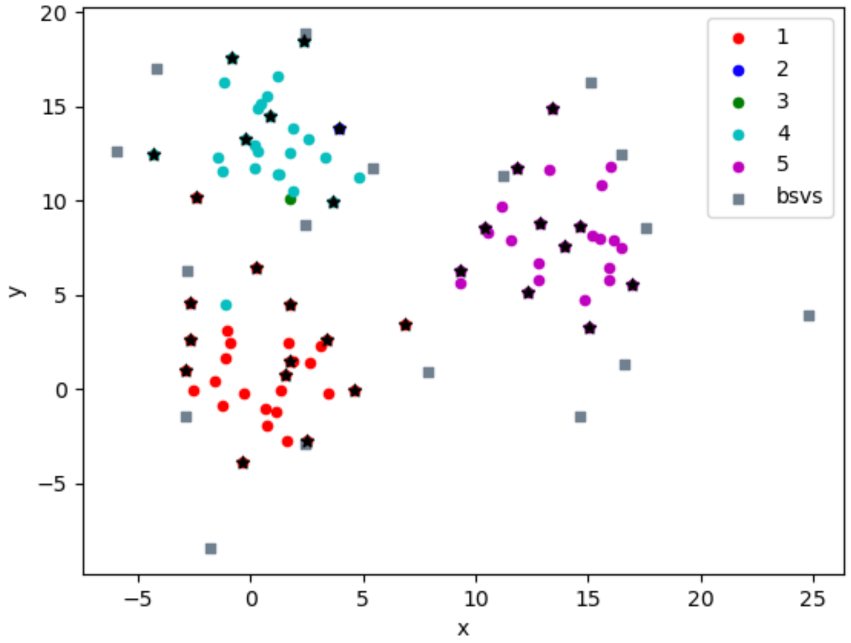
	p	q	clusters count	SVs	BSVs	CH
0	0.20	0.085	5.0	38.0	2.0	56.255193
1	0.23	0.085	4.0	36.0	5.0	76.227196
2	0.25	0.085	2.0	33.0	8.0	101.796659
3	0.30	0.085	5.0	30.0	17.0	65.285582
4	0.35	0.085	6.0	23.0	25.0	55.461256
5	0.40	0.085	7.0	23.0	30.0	53.805037
6	0.50	0.085	7.0	22.0	42.0	60.297453

Проиллюстрируем часть итераций подбора параметра:

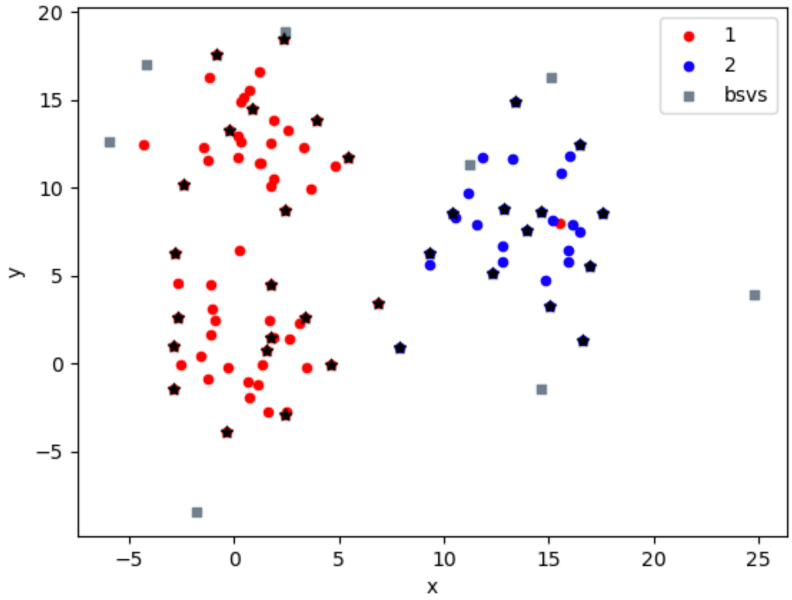
```
p, q, clusters count, SVs, BSVs, CH
[0.23, 0.085, 4, 36, 5, 76.22719602043819]
```



```
p, q, clusters count, SVs, BSVs, CH
[0.3, 0.085, 5, 30, 17, 65.28558212855133]
```




```
p, q, clusters count, SVs, BSVs, CH
[0.25, 0.085, 2, 33, 8, 101.7966586126122]
```



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.