Поиск аномалий

kNN ★ LOF

Обо мне

- Старший специалист по машинному обучению
- deep learning engineer
- NLP, CV, anomaly detection
- Open source contributor
- Выпускник и амбассадор Яндекс Практикума
- Выпускник DLS ФПМИ МФТИ

Аномалии


SCAMMER

попугаи	**удавы**
3.248357	-2.874754
2.930868	-2.826776
3.323844	-3.340012
3.761515	-2.883873
2.882923	-2.853464
2.882932	-3.357176
3.789606	-2.067113
…	…

попугаи	**удавы**
3.248357	-2.874754
2.930868	-2.826776
3.323844	-3.340012
3.761515	-2.883873
2.882923	-2.853464
2.882932	-3.357176
3.789606	-2.067113
…	…

индекс аномальности
-0.119150
-0.131275
-0.108670
-0.065472
-0.128972
-0.120056
-0.012170
…

Методы

$ pip install pyod

```
from pyod.models.knn import KNN

clf = KNN()

clf.fit(data)

scores = clf.decision_scores_
```

Методы

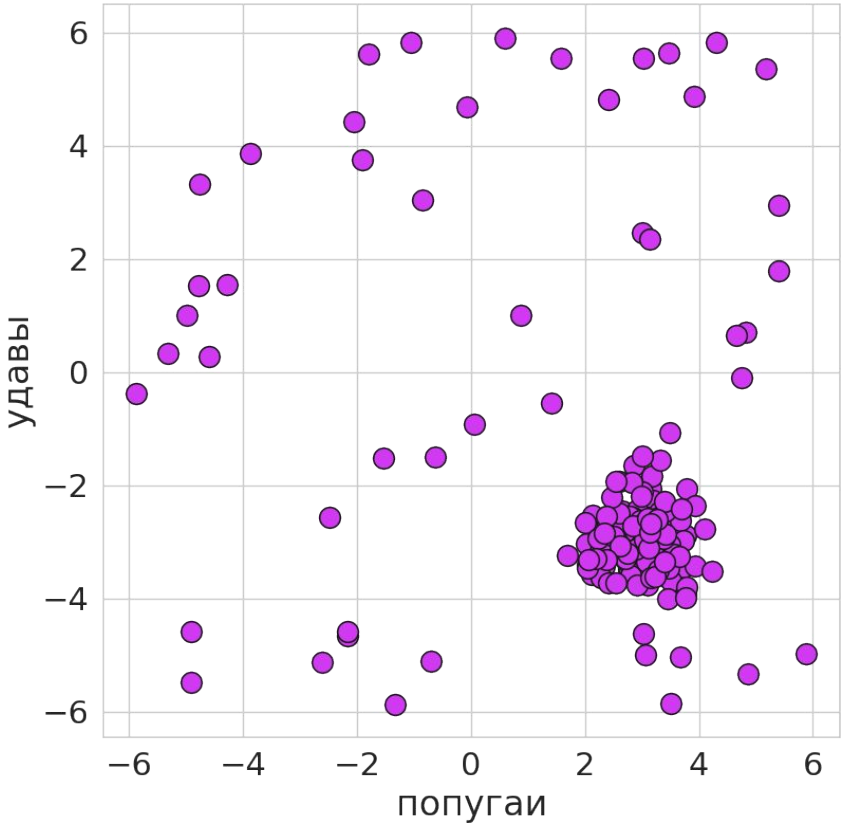
$ pip install pyod

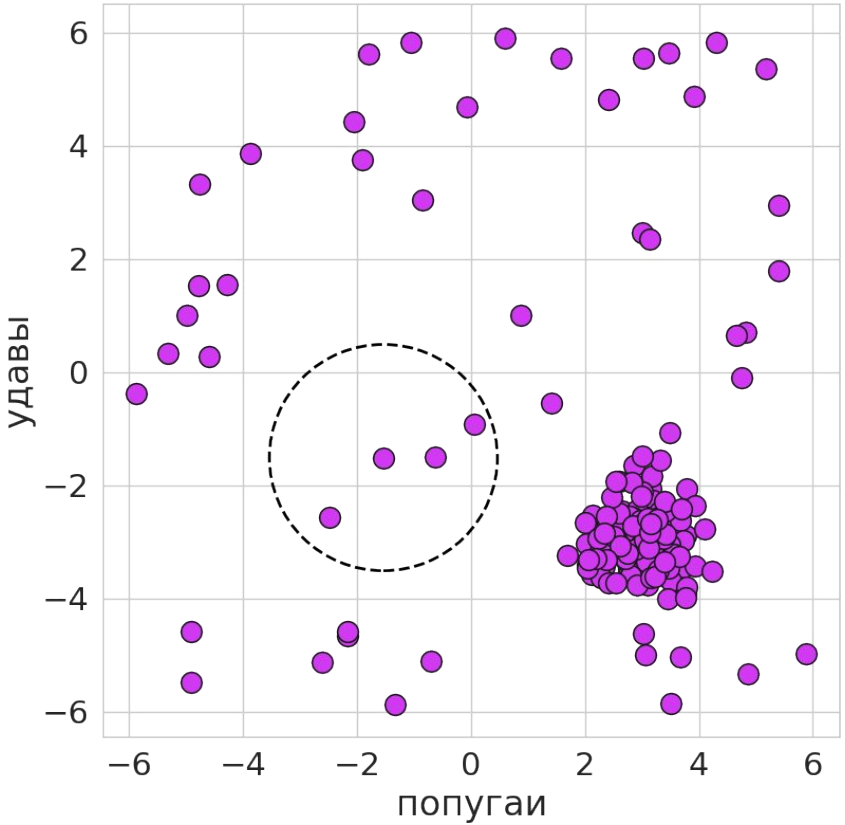
[1] Edwin Knorr and Raymond Ng. Algorithms for mining distance-based outliers in large datasets. In *Proc. of the VLDB Conference,* 392–403, New York, USA, September 1998.

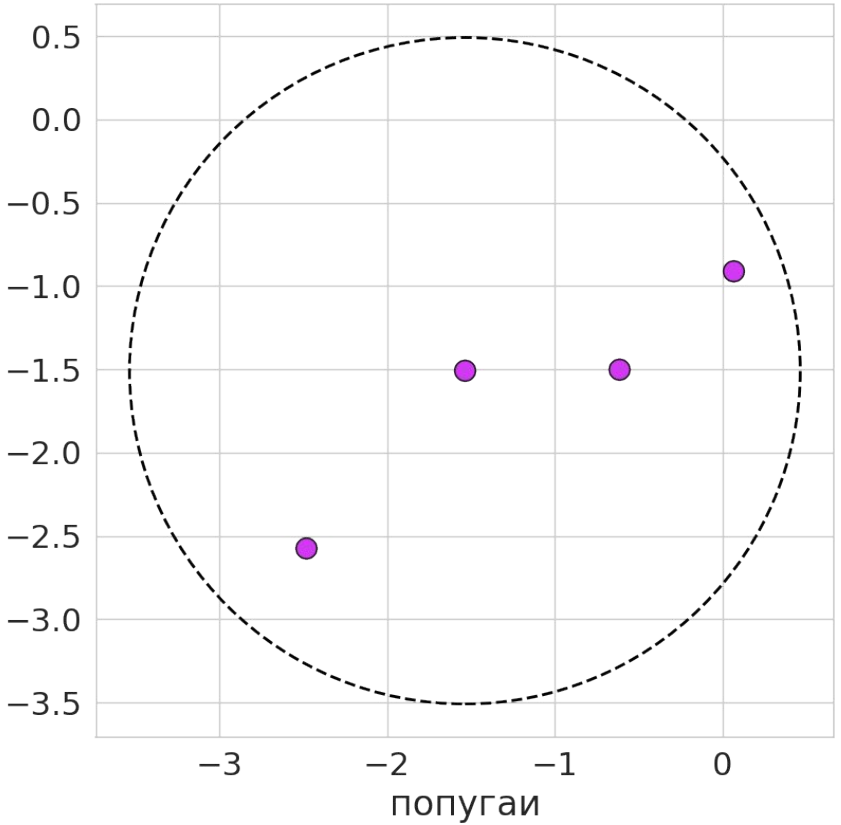
[2] Fabrizio Angiulli and Clara Pizzuti. Fast outlier detection in high dimensional spaces. In *European Conference on Principles of Data Mining and Knowledge Discovery,* 15–27. Springer, 2002.

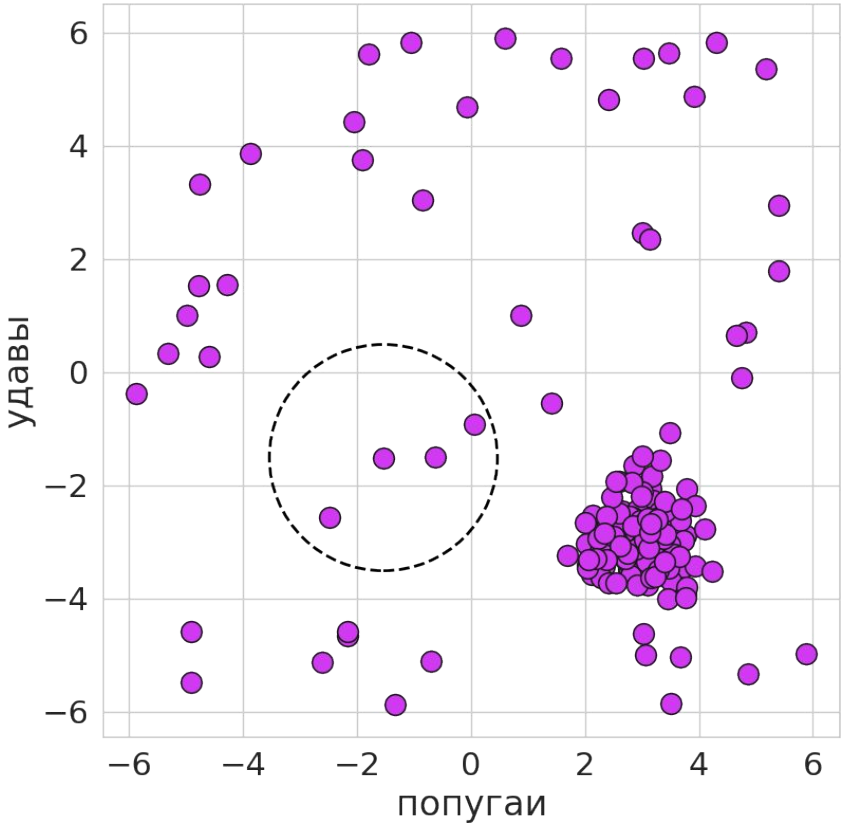
[3] Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *ACM Sigmod Record,* volume 29, 427–438. ACM, 2000.

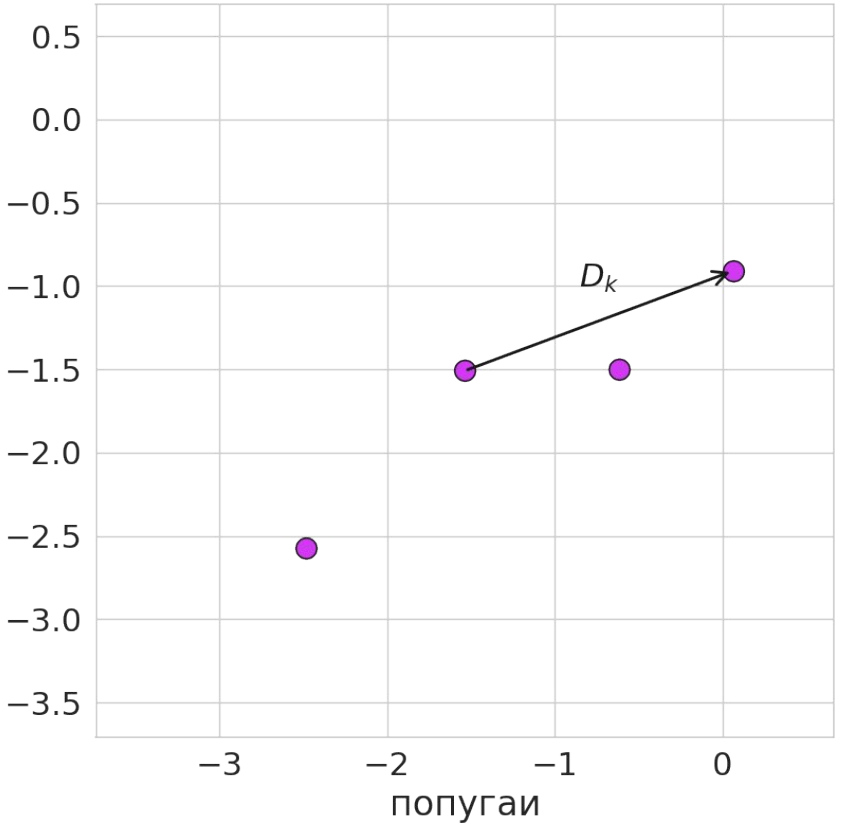
[4] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM sigmod record,* volume 29, 93–104. ACM, 2000.

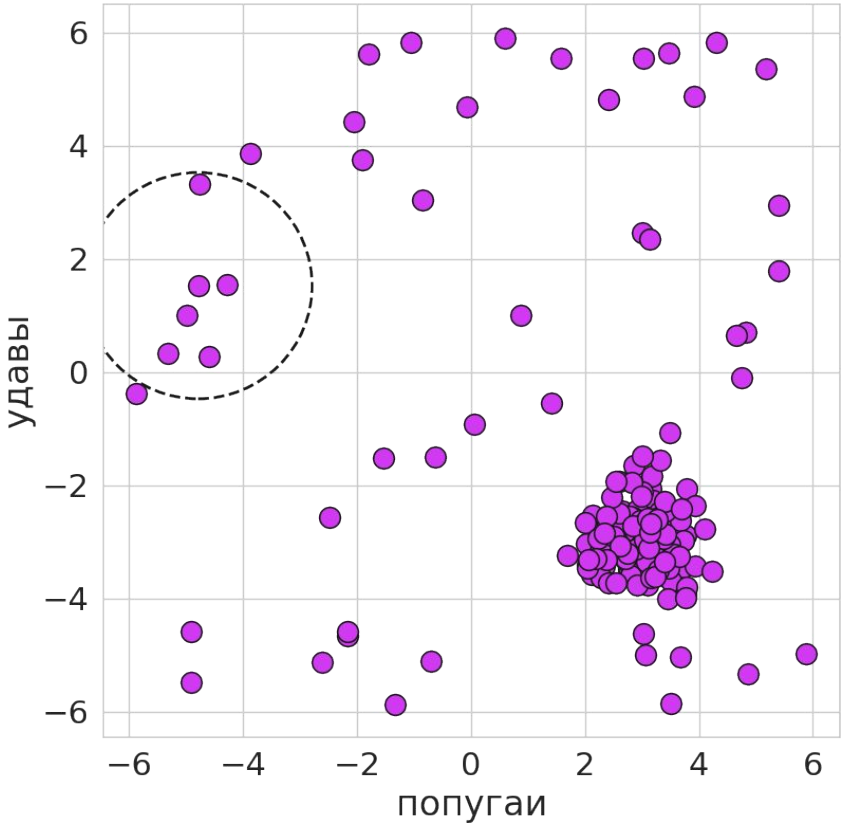

6
4
2
0
−2
−4
−6
удавы
−6
−4
−2
0
2
4
6
попугаи

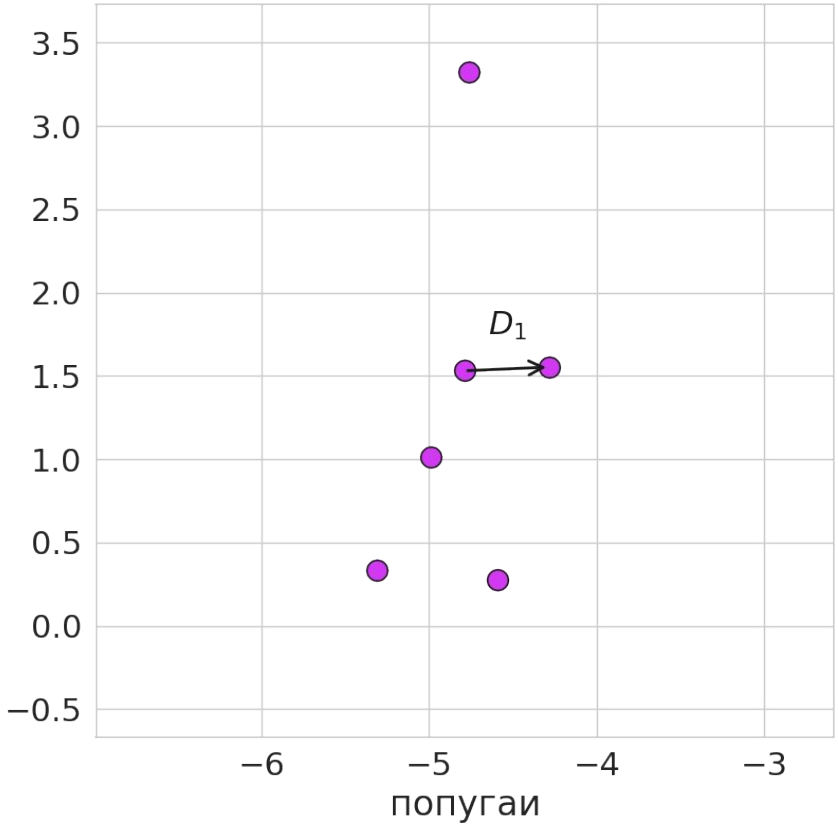

удавы
6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6
попугаи

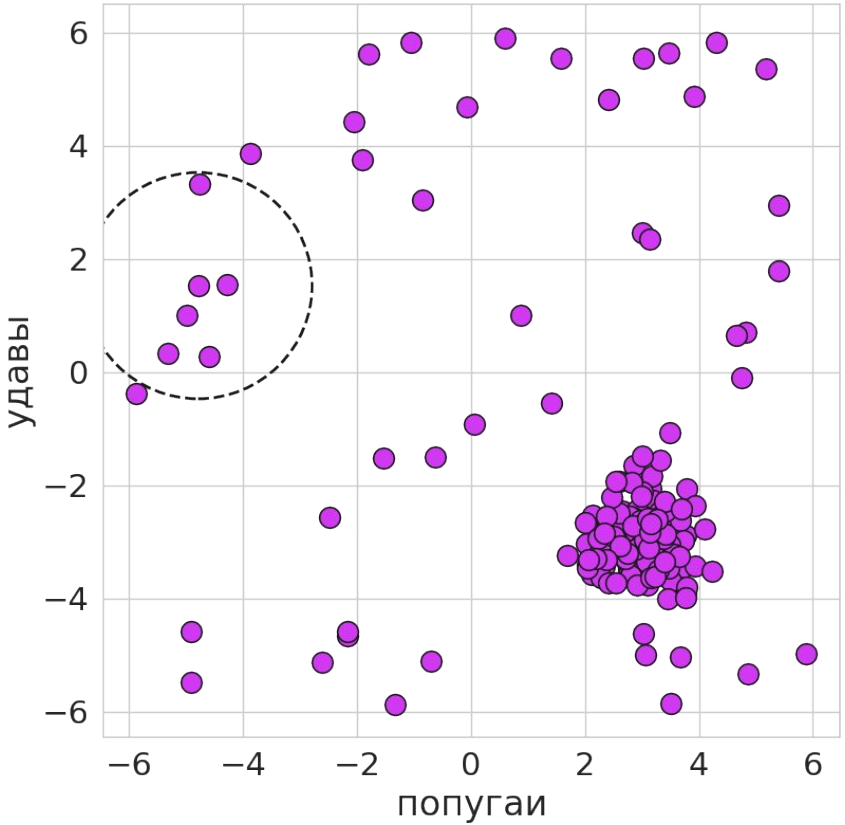

0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−3
−2
−1
0
попугаи

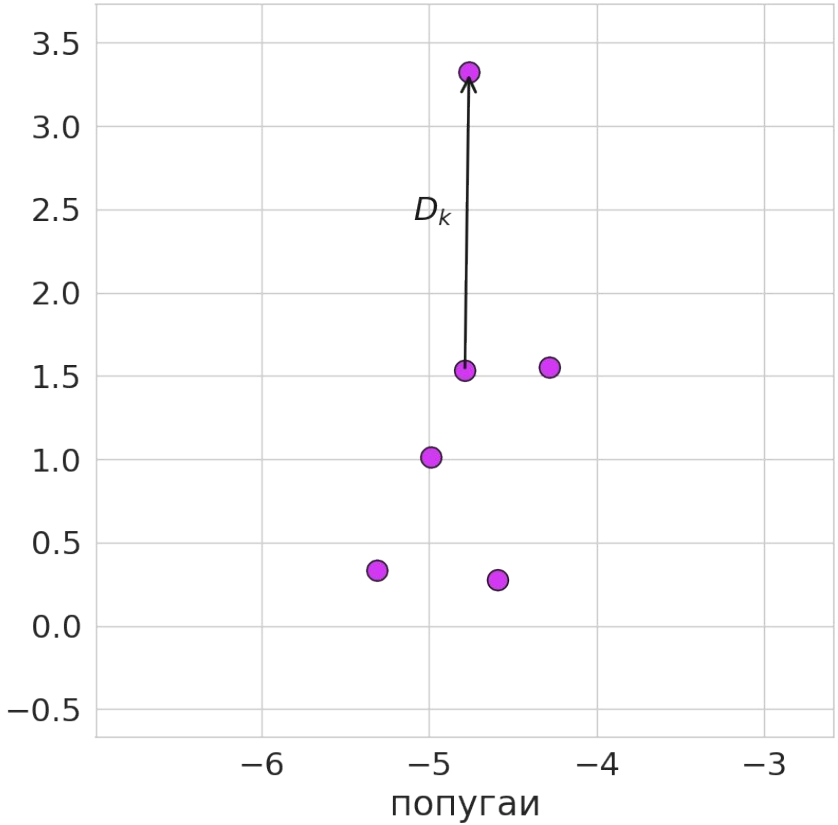

6
4
2
0
−2
−4
−6
удавы
−6
−4
−2
0
2
4
6
попугаи


0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
D_k
−3
−2
−1
0
попугаи

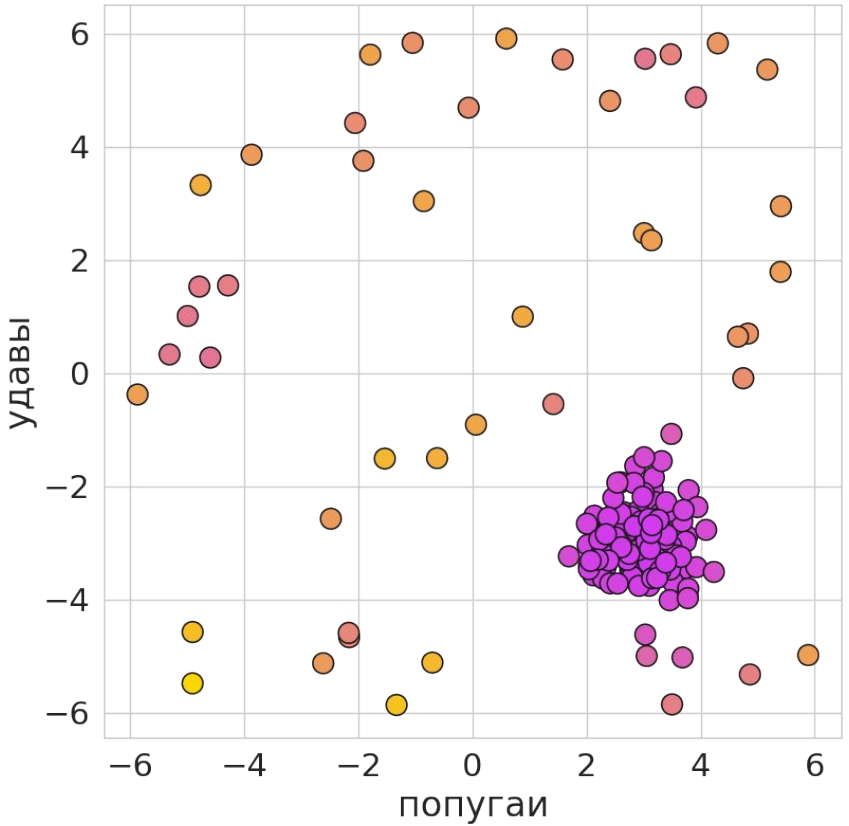

удавы
6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6
попугаи


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
D_1
−6
−5
−4
−3
попугаи


6
4
2
0
−2
−4
−6
удавы
−6
−4
−2
0
2
4
6
попугаи


3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
D_k
−6
−5
−4
−3
попугаи

k-Nearest Neighbors

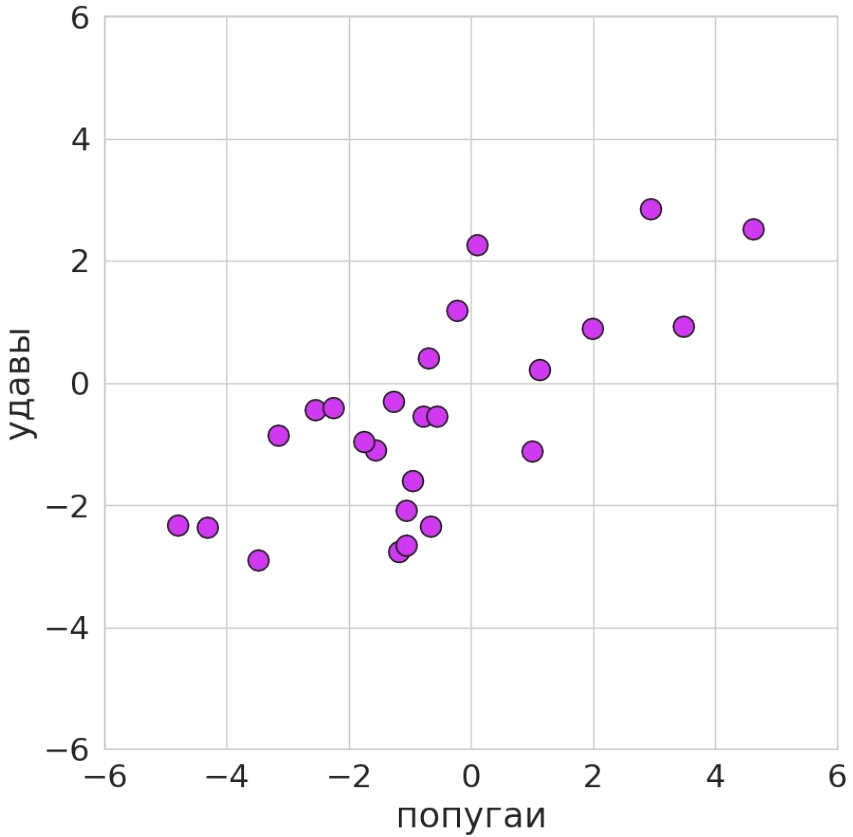


Algorithm

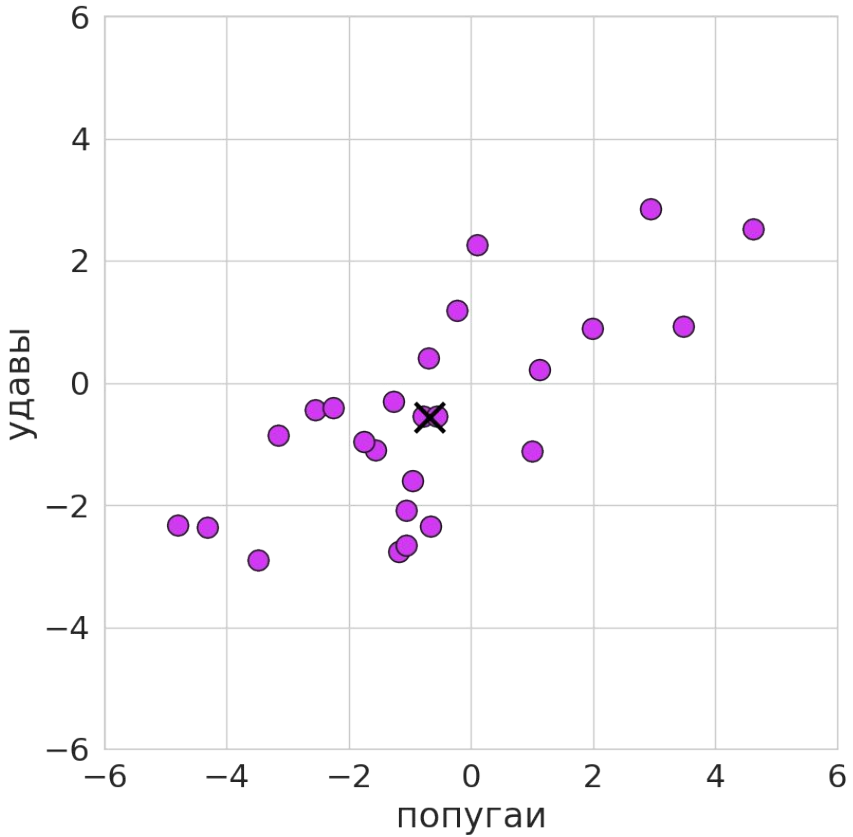
"""…
algorithm : {'auto', **'ball_tree'**, 'kd_tree', 'brute'}, default='auto'
Algorithm used to compute the nearest neighbors:

- **'ball_tree'** will use :class:`**BallTree**`
…"""

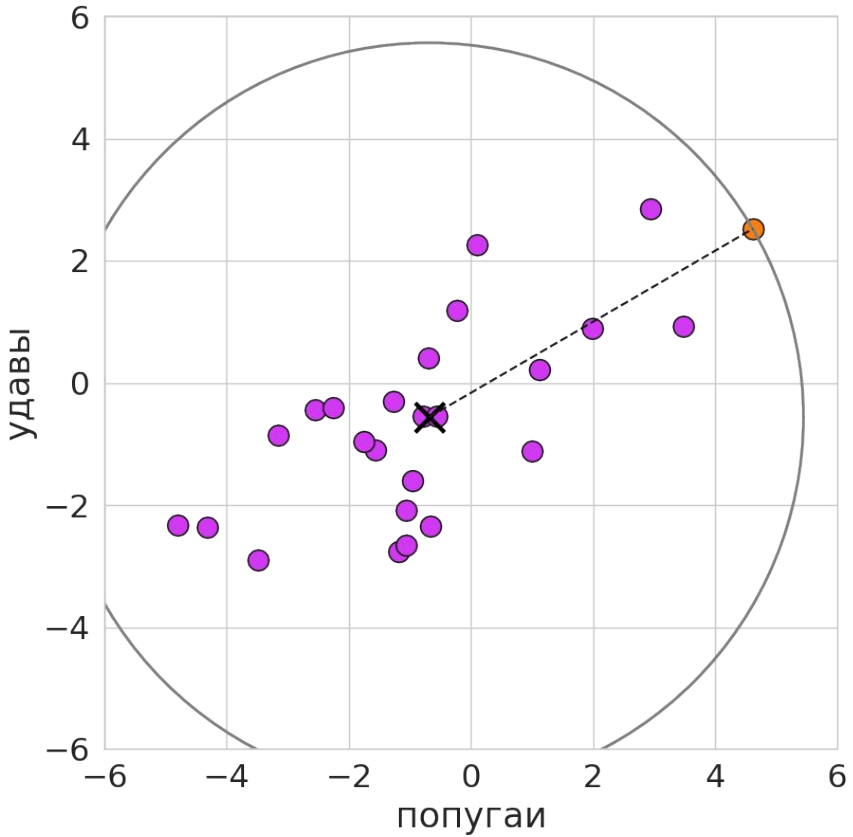
Ball Tree



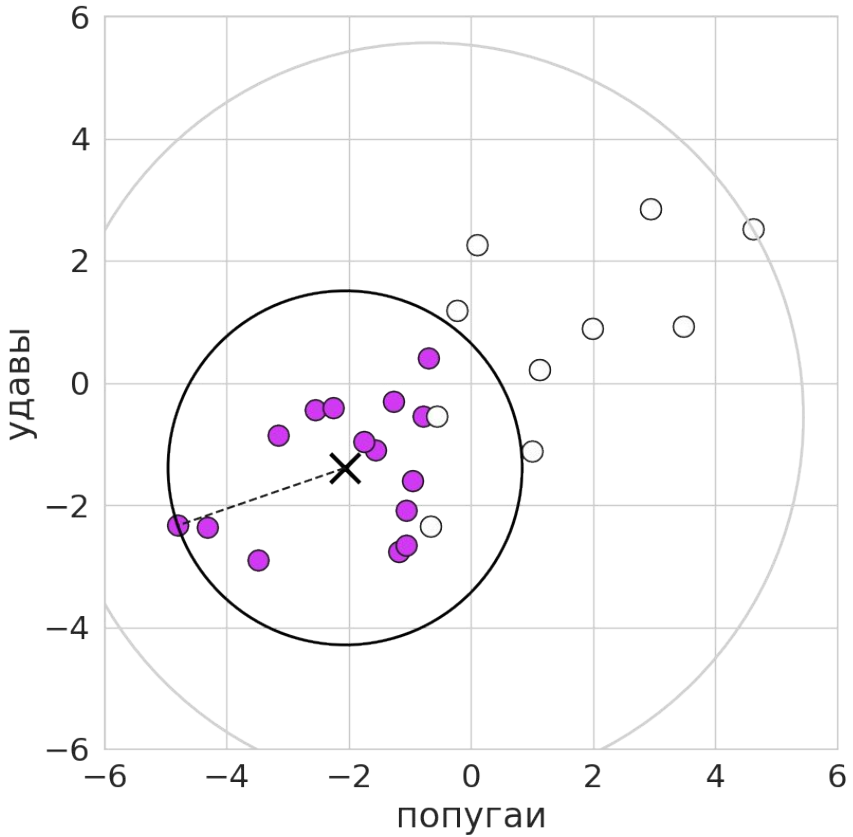
Ball Tree



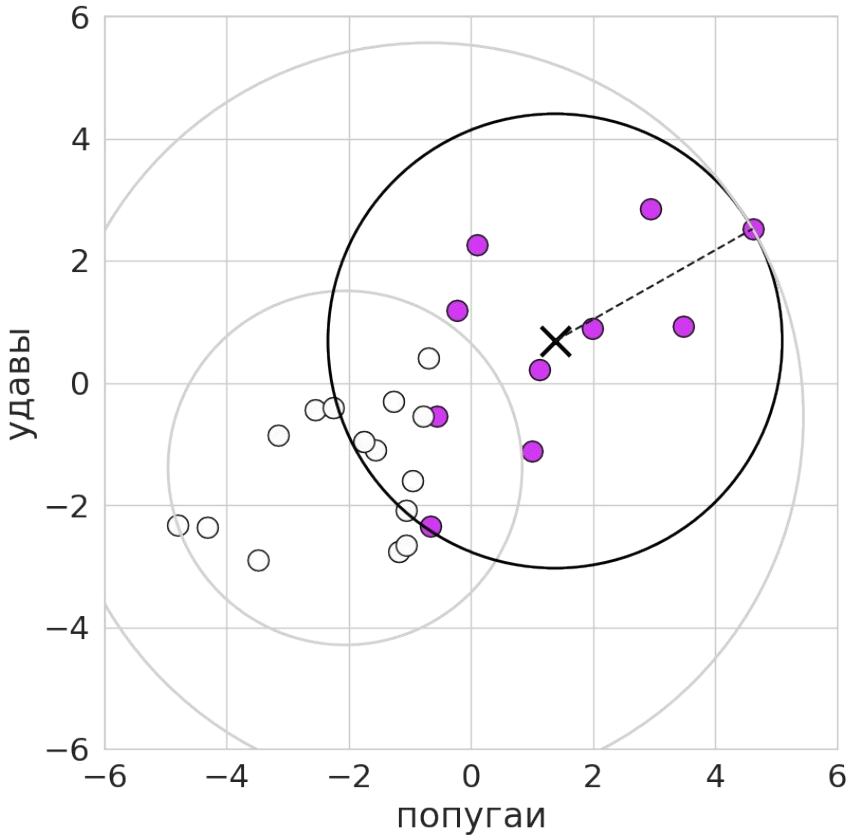
Ball Tree



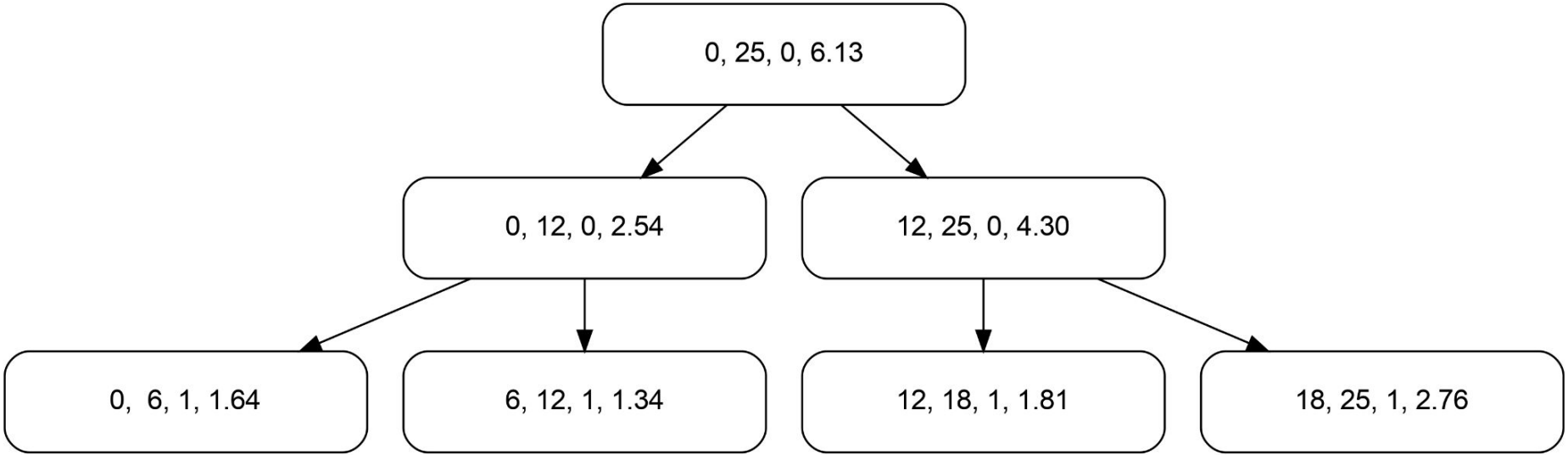
Ball Tree

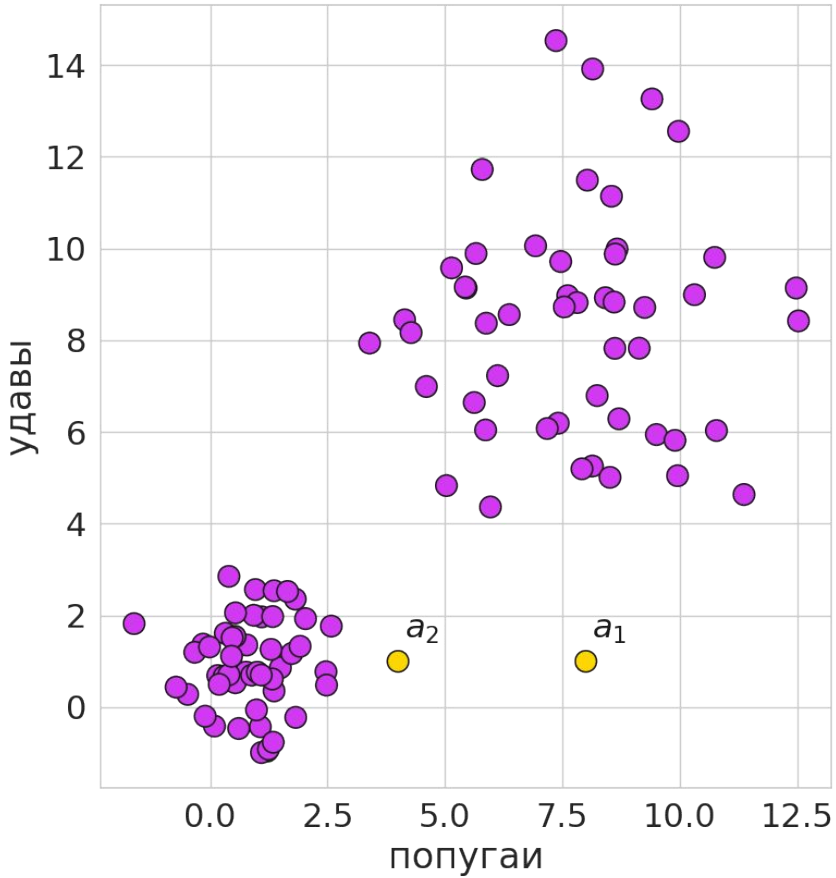


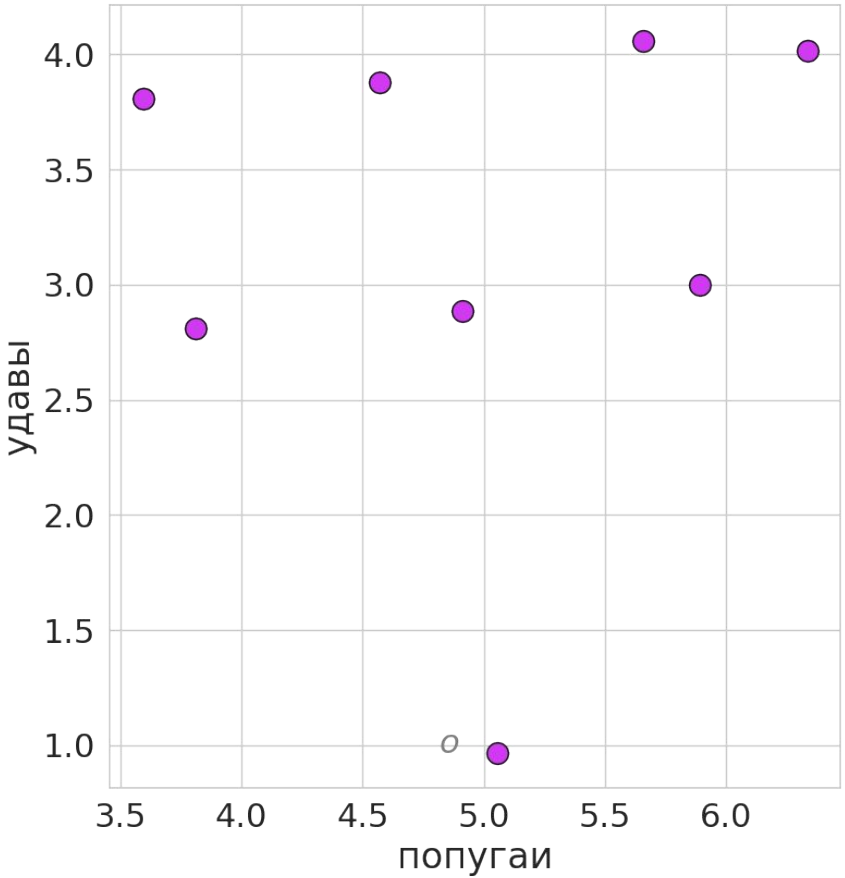
Ball Tree

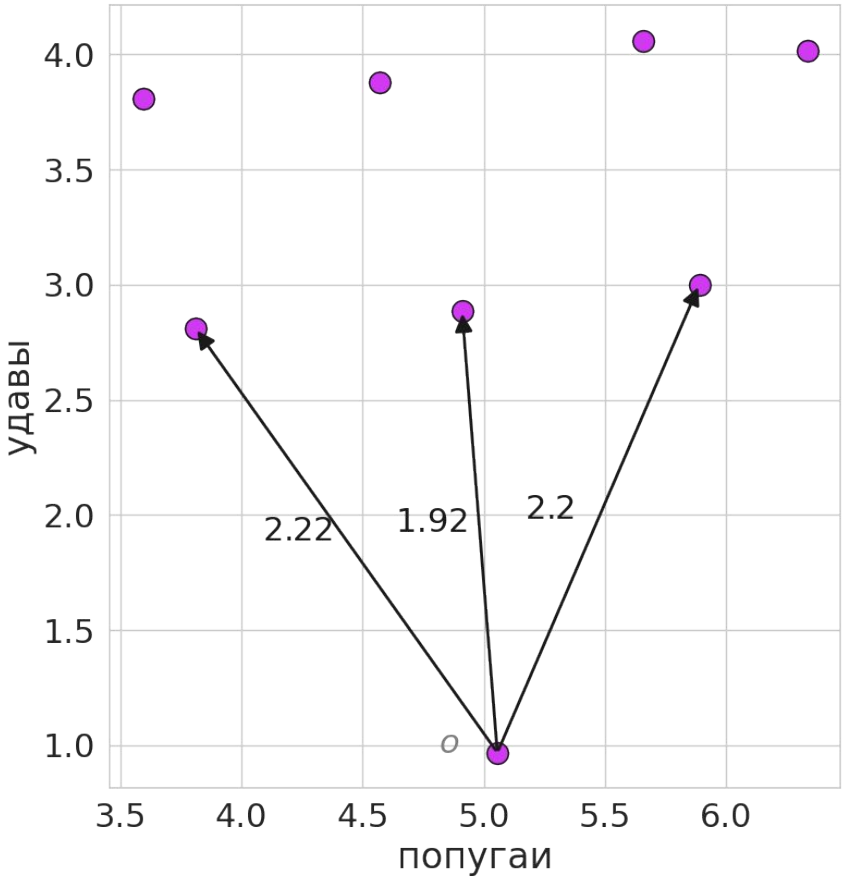


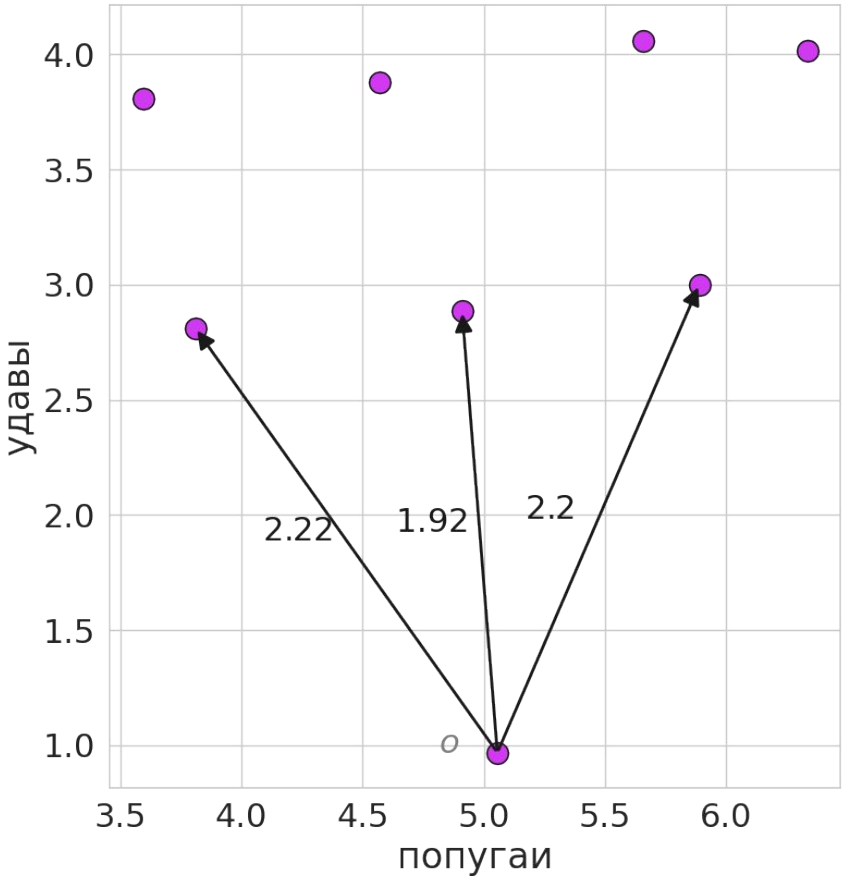
Ball Tree


0, 25, 0, 6.13
0, 12, 0, 2.54
12, 25, 0, 4.30
0, 6, 1, 1.64
6, 12, 1, 1.34
12, 18, 1, 1.81
18, 25, 1, 2.76

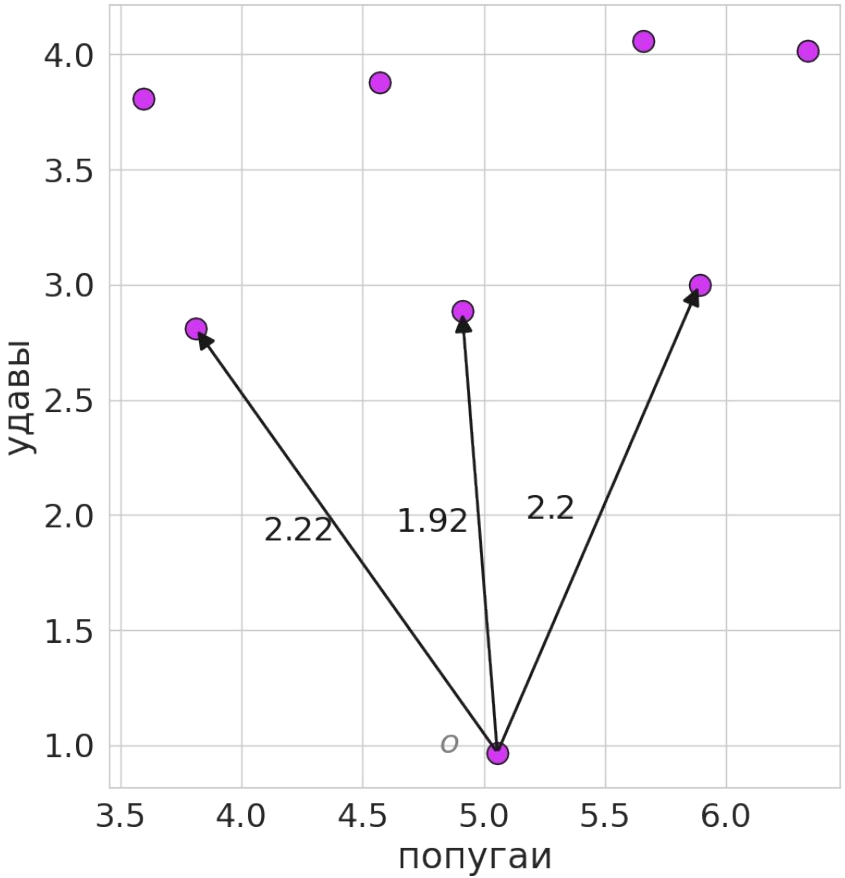

14
12
10
8
6
4
2
0
удавы
a_2
a_1
0.0
2.5
5.0
7.5
10.0
12.5
попугаи


4.0
3.5
3.0
2.5
2.0
1.5
1.0
удавы
3.5
4.0
4.5
5.0
5.5
6.0
попугаи


4.0
3.5
3.0
2.5
2.0
1.5
1.0
удавы
2.22
1.92
2.2
o
3.5
4.0
4.5
5.0
5.5
6.0
попугаи

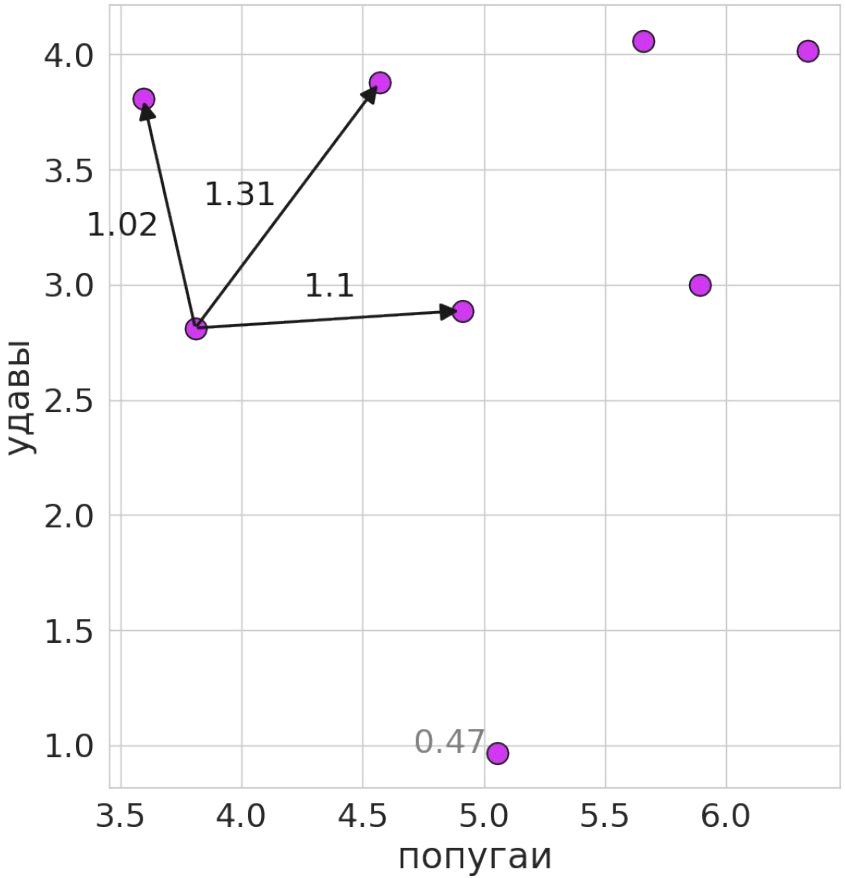


Среднее = (2.22 + 1.92 + 2.20) / 3 = 2.11



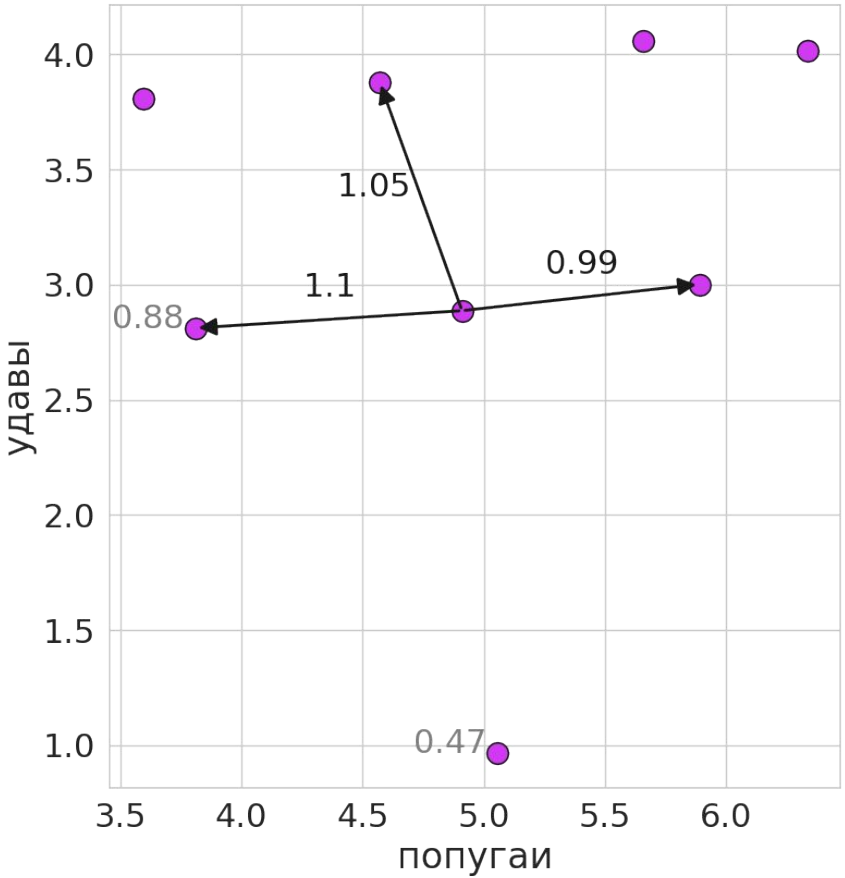
Среднее = (2.22 + 1.92 + 2.20) / 3 = 2.11

Плотность = 1 / 2.11 = 0.47



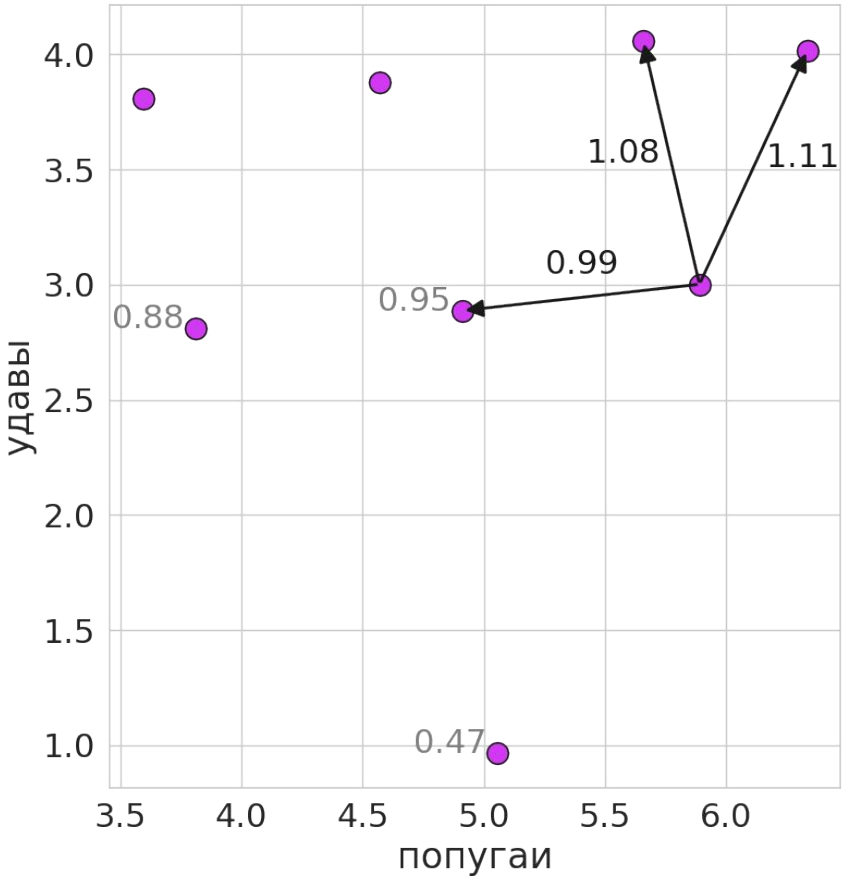
Среднее = (1.02 + 1.31 + 1.10) / 3 = 1.14

Плотность = 1 / 1.14 = 0.88



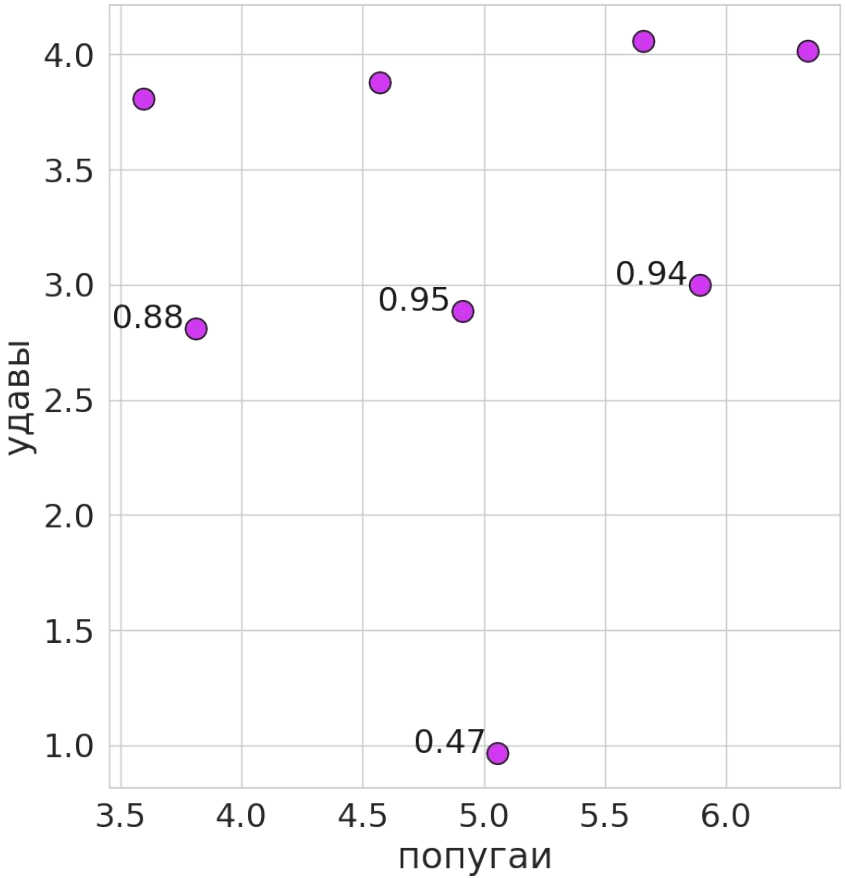
Среднее = (1.10 + 1.05 + 0.99) / 3 = 1.05

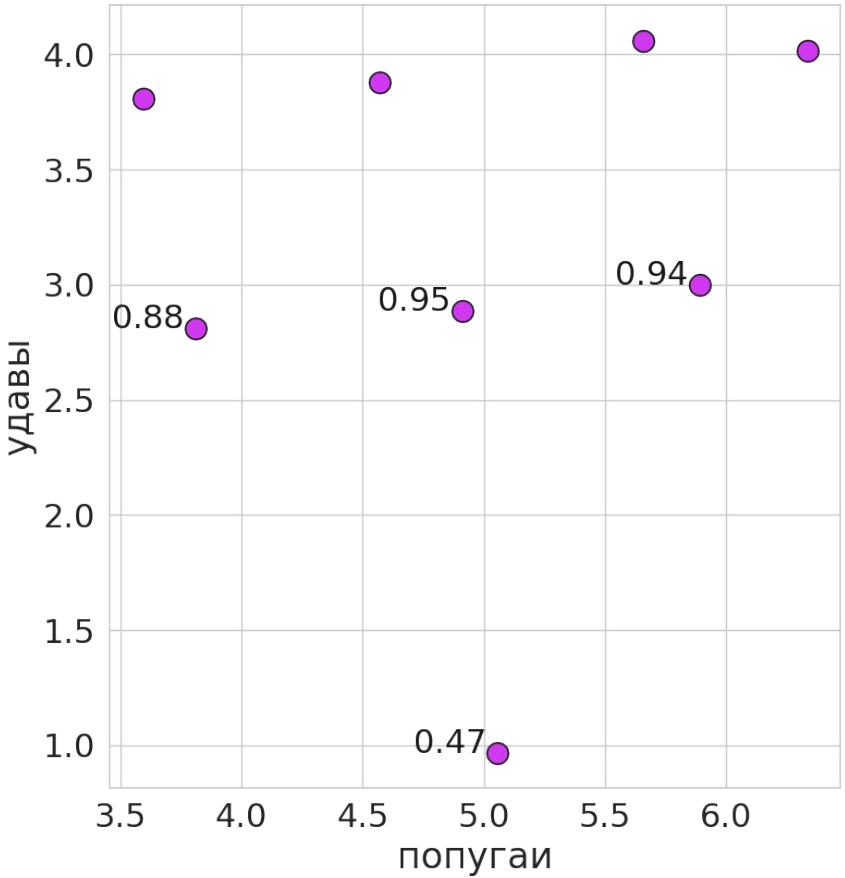
Плотность = 1 / 1.05 = 0.95



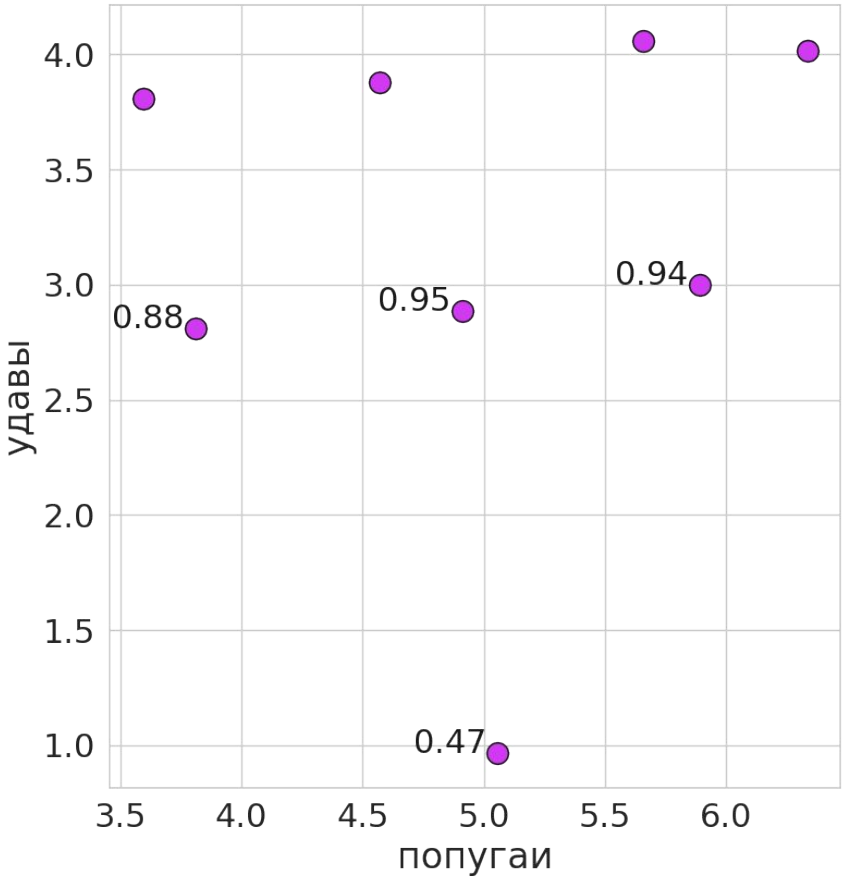
Среднее = (0.99 + 1.08 + 1.11) / 3 = 1.06

Плотность = 1 / 1.06 = 0.94


4.0
3.5
3.0
2.5
2.0
1.5
1.0
удавы
0.88
0.95
0.94
0.47
3.5
4.0
4.5
5.0
5.5
6.0
попугаи

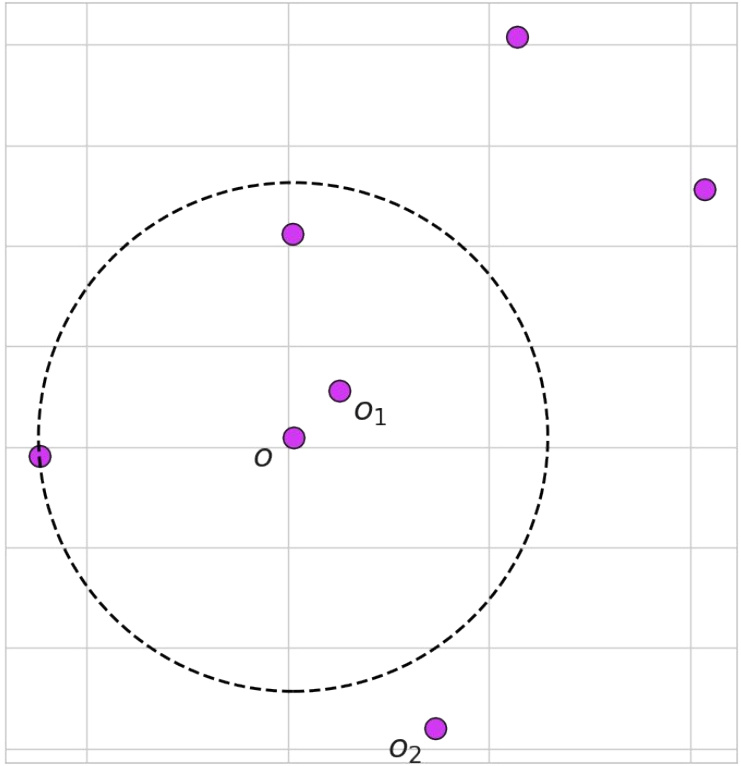


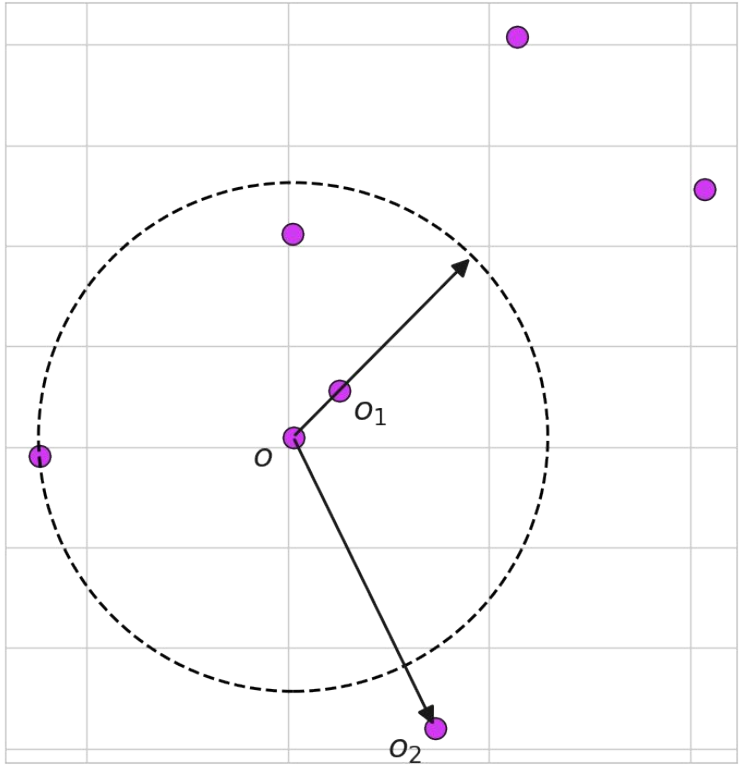
(0.88 + 0.95 + 0.94) / 3 = 0.92



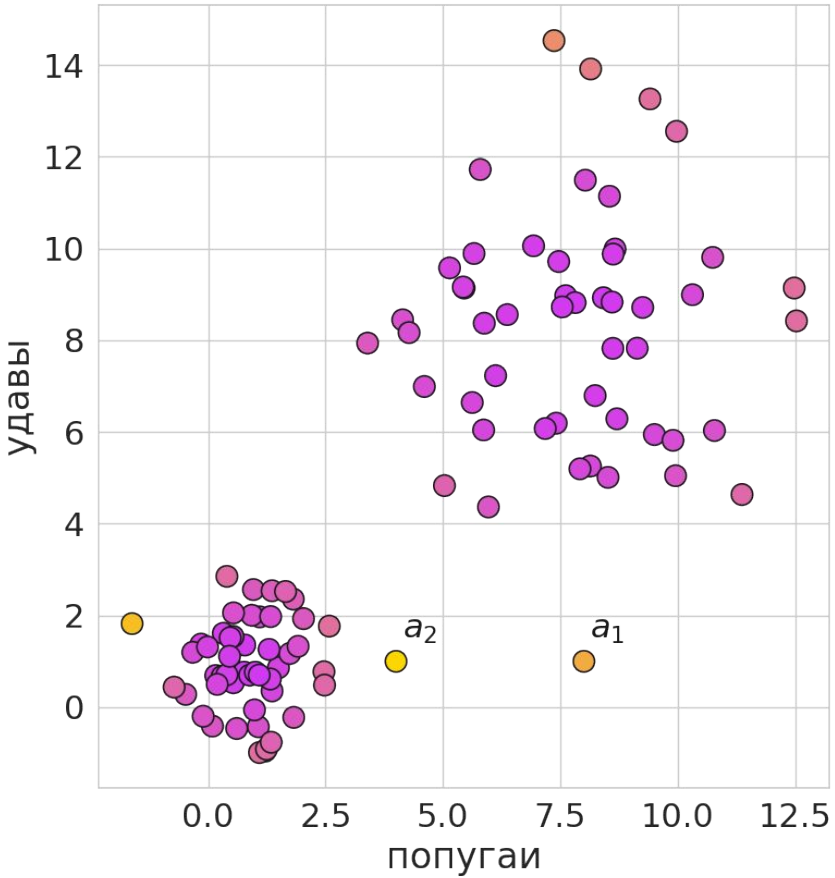
(0.88 + 0.95 + 0.94) / 3 = 0.92

LOF = 0.92 / 0.47 = 1.96


o_1
o
o_2


o_1
o
o_2

Local Outlier Factor



Резюме

Вопросы?