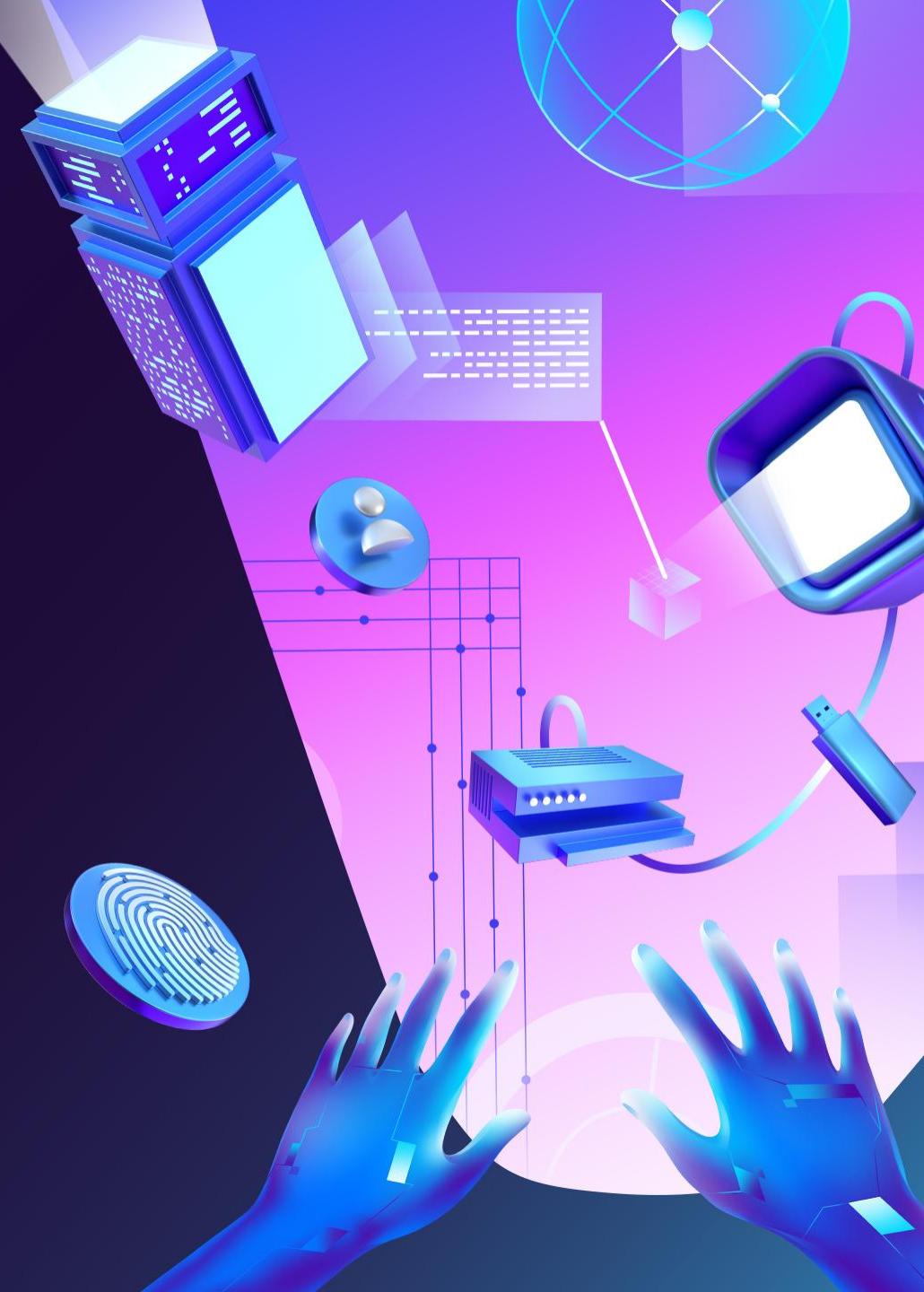


Михаил Васильев

Поиск аномалий с Python

от теории к практике





Обо мне

- Старший специалист по машинному обучению
- Makves (входит в группу компаний «Гарда»)
- deep learning engineer
- NLP, CV, anomaly detection
- Open source contributor

Михаил Васильев





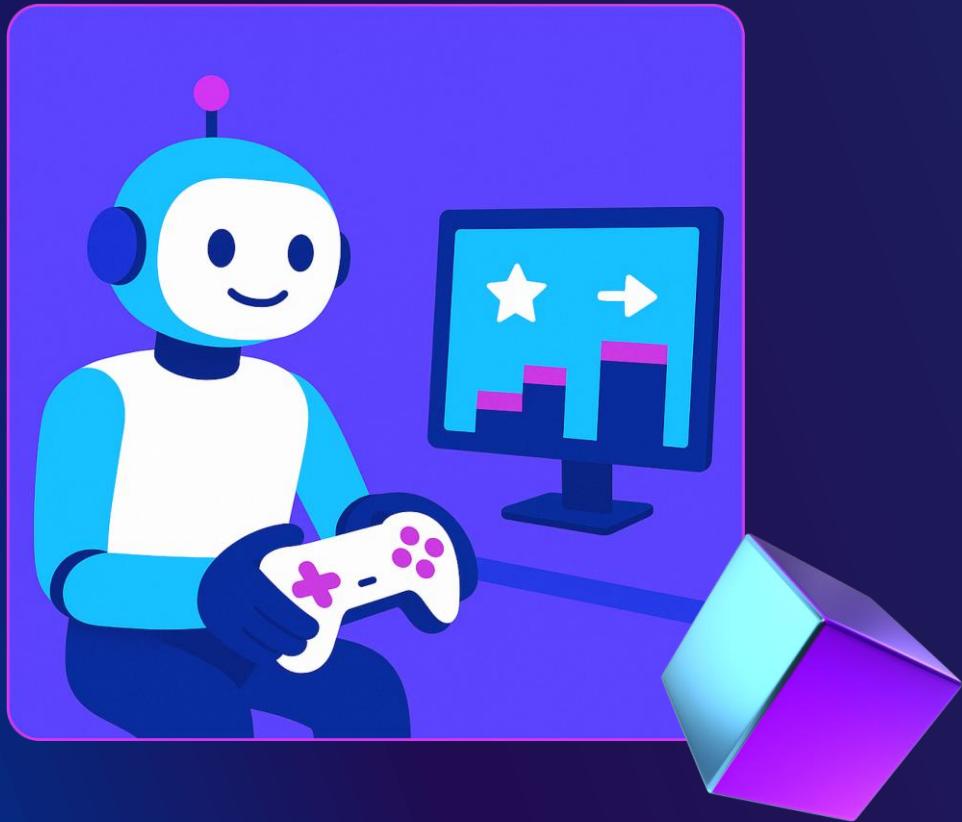
Поиск аномалий

Применение



phd X pt

Применение



phd X pt

Применение

phd X pt

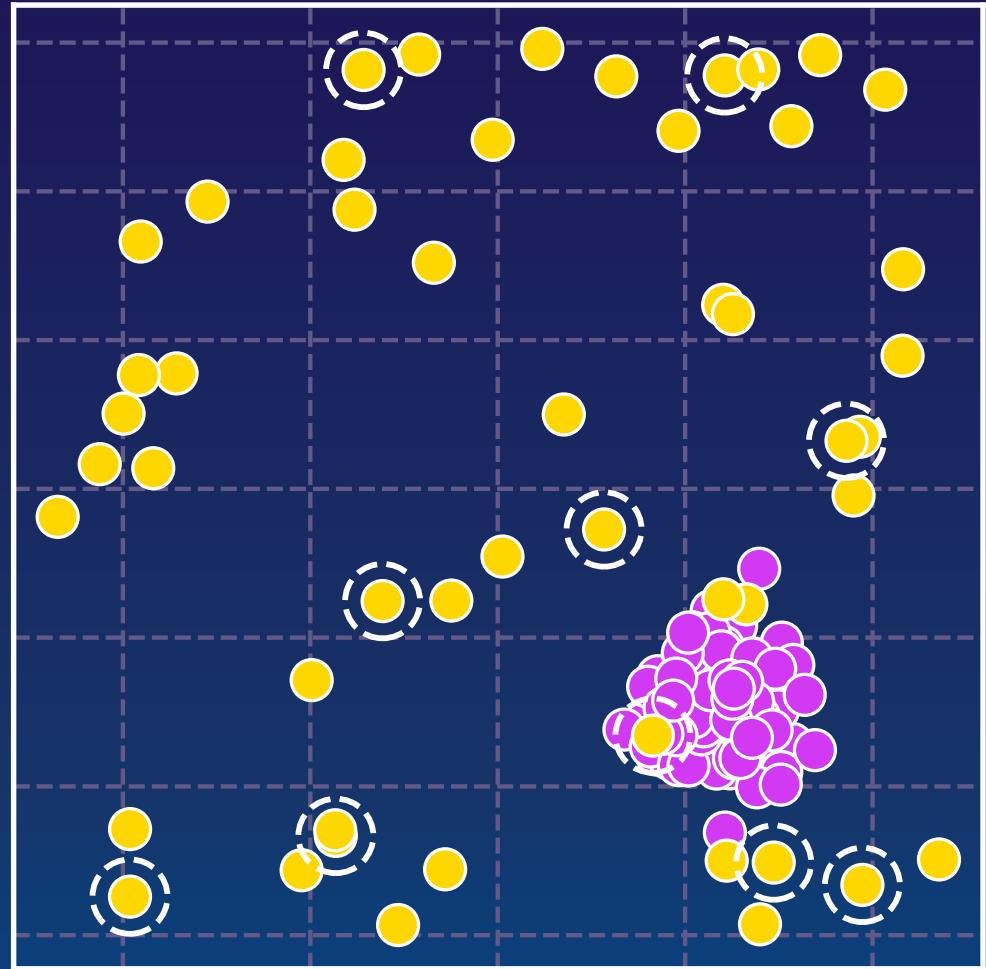


Применение

phd X pt

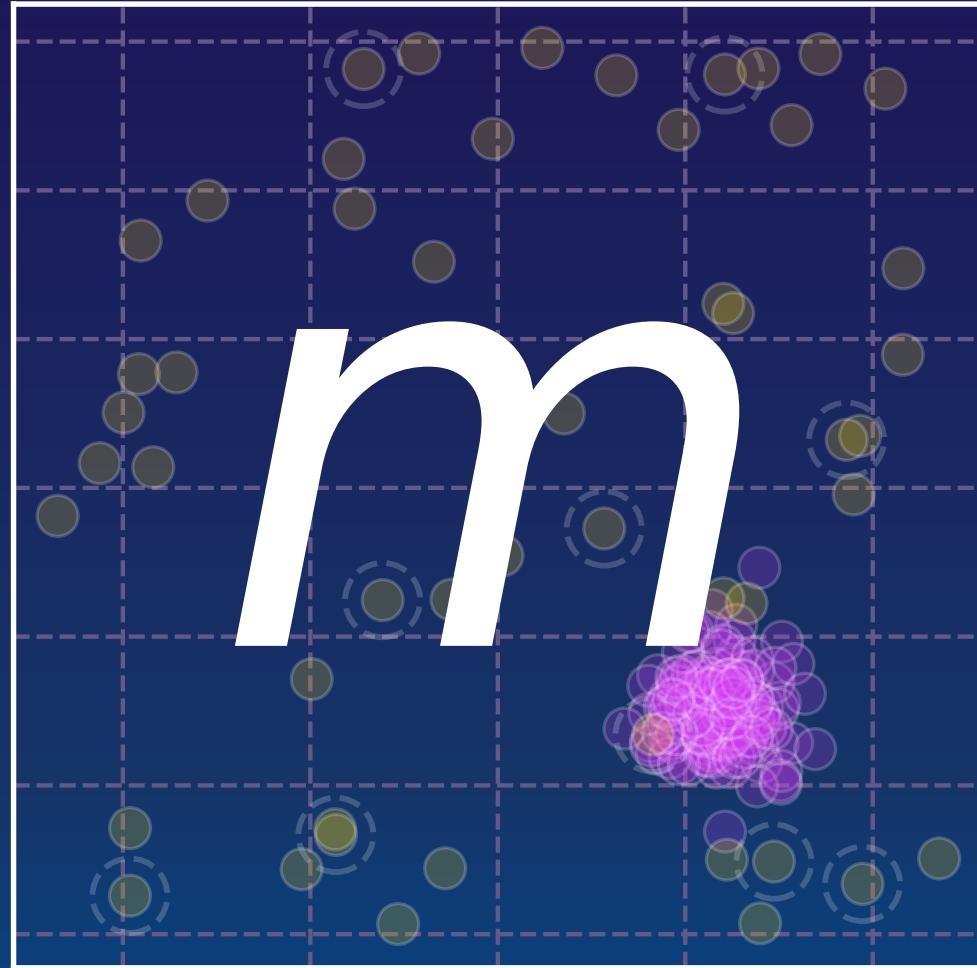
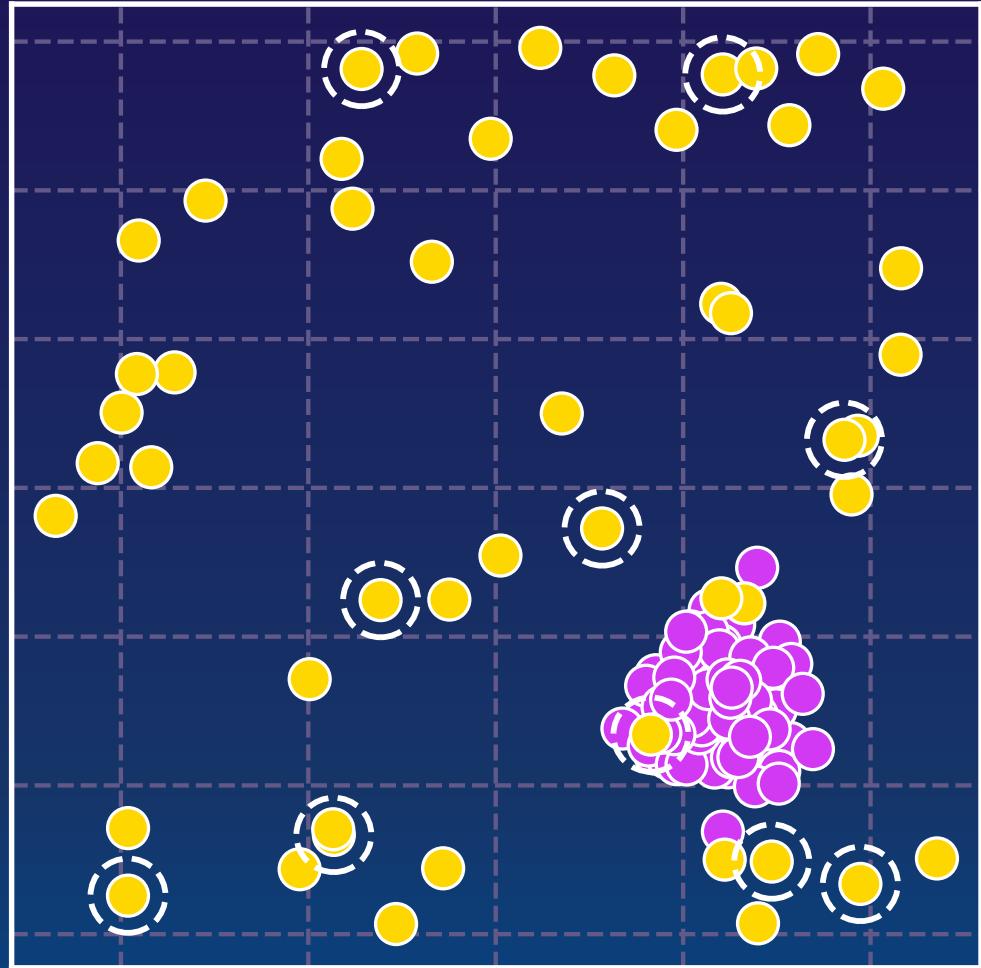


Этапы



Этапы

$\bar{p}hd$ X pt



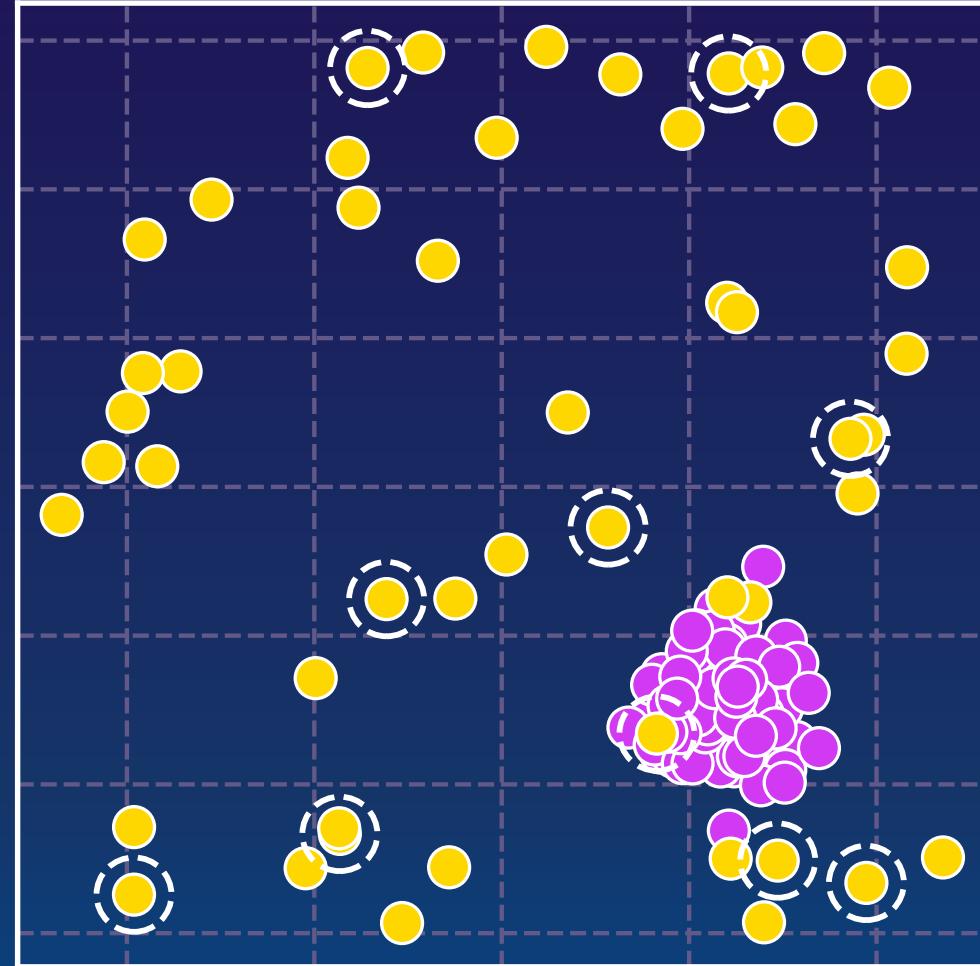
Применение

phd X pt

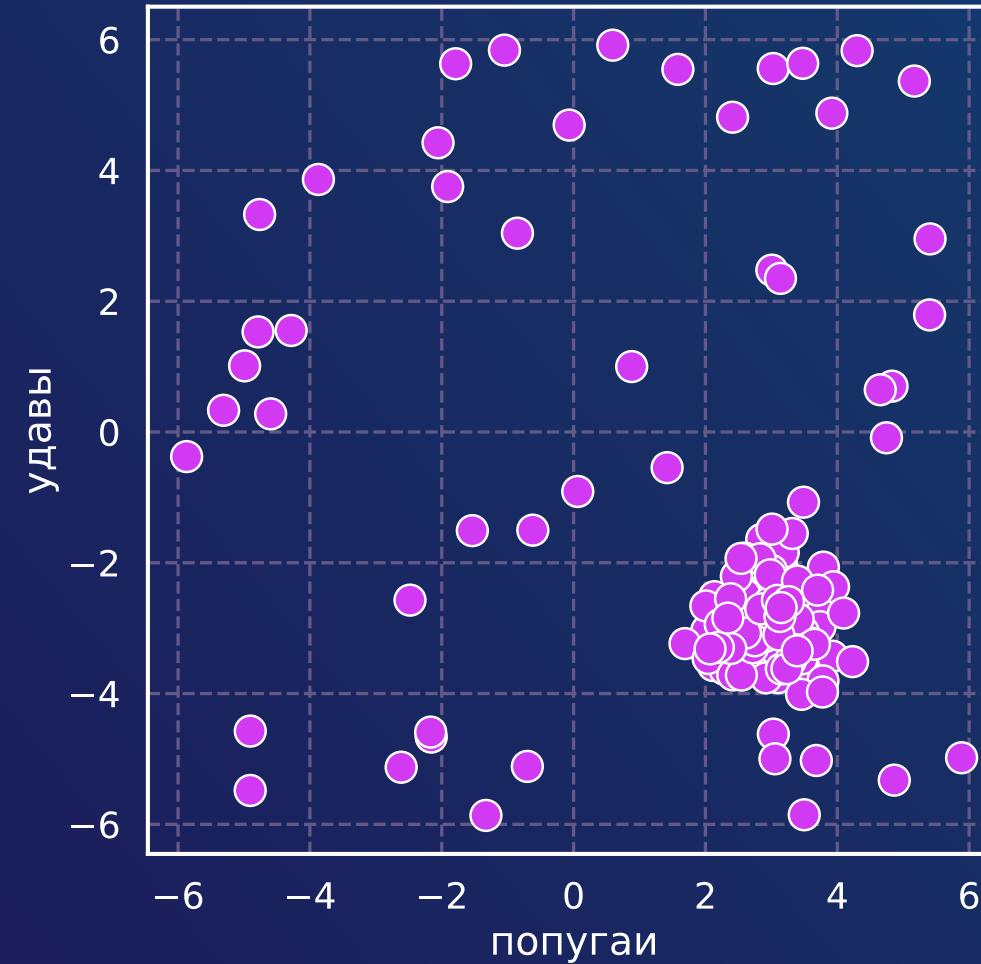


Сбор данных

phd X pt



Попугай	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
...	...

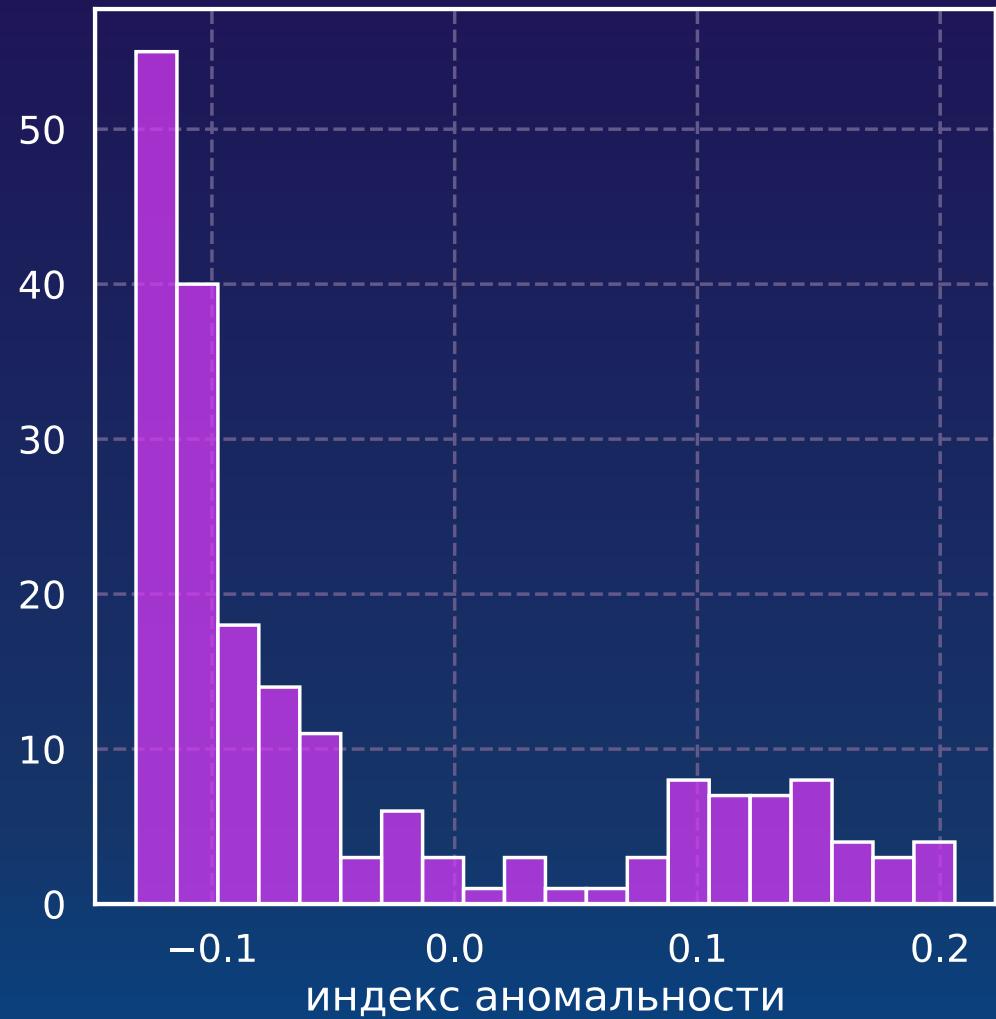


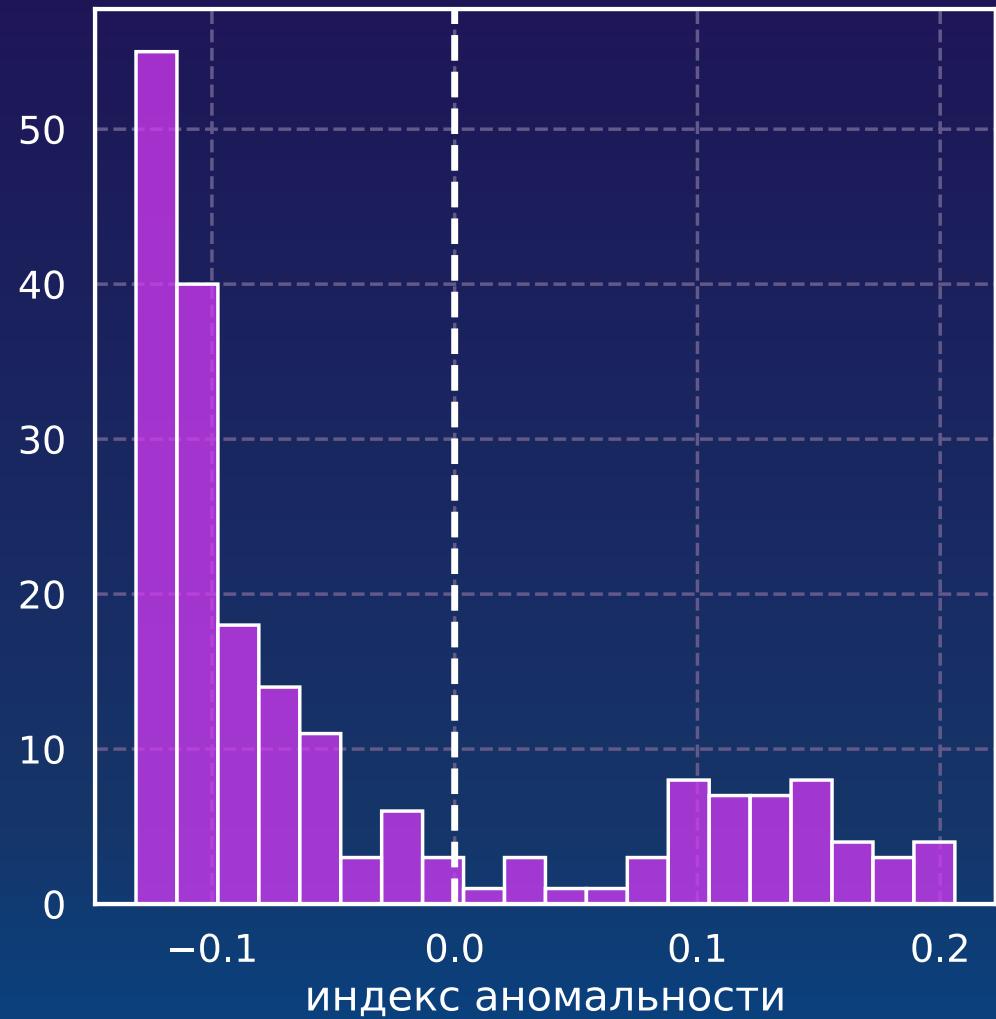
Попугай	Удавы
3.248357	-2.874754
2.930868	-2.826776
3.323844	-3.340012
3.761515	-2.883873
2.882923	-2.853464
2.882932	-3.357176
3.789606	-2.067113
...	...

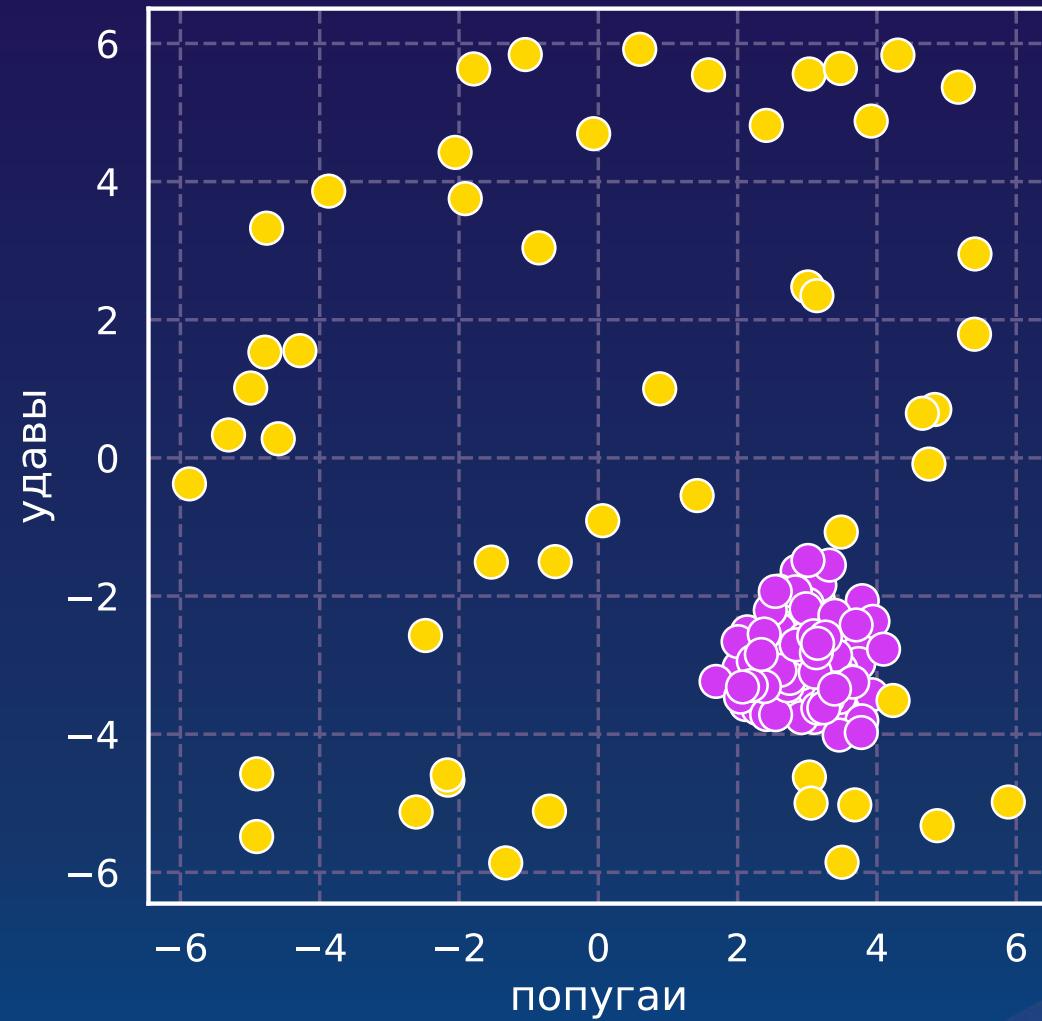
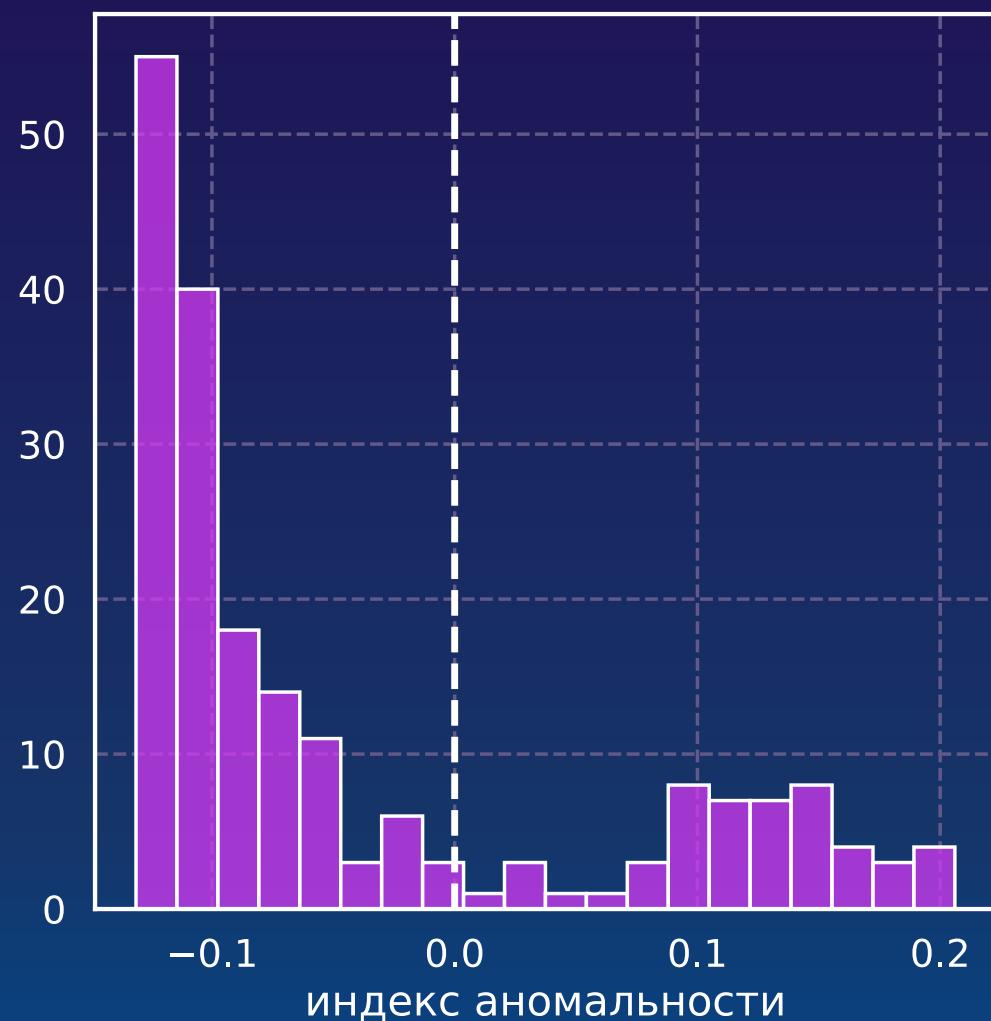
Индекс аномальности
-0.119150
-0.131275
-0.108670
-0.065472
-0.128972
-0.120056
-0.012170
...

```
data['scores'] = scores  
data.sort_values(by='scores', ascending=False)
```

Попугай	Удавы	Индекс аномальности
-5.869948	-0.376072	0.206049
5.886062	-4.981947	0.204401
5.169088	5.366983	0.195944
-4.905527	-5.482081	0.190407
-3.870725	3.862321	0.185113
-1.329580	-5.863756	0.183137
...







Банки

phd X pt



- ❖ Фрода
- ❖ Мошенники



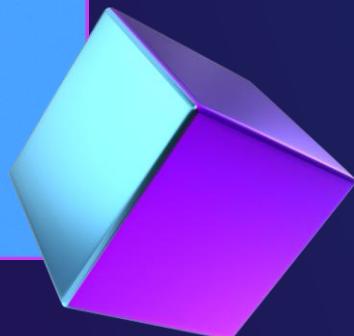
Кибербезопасность



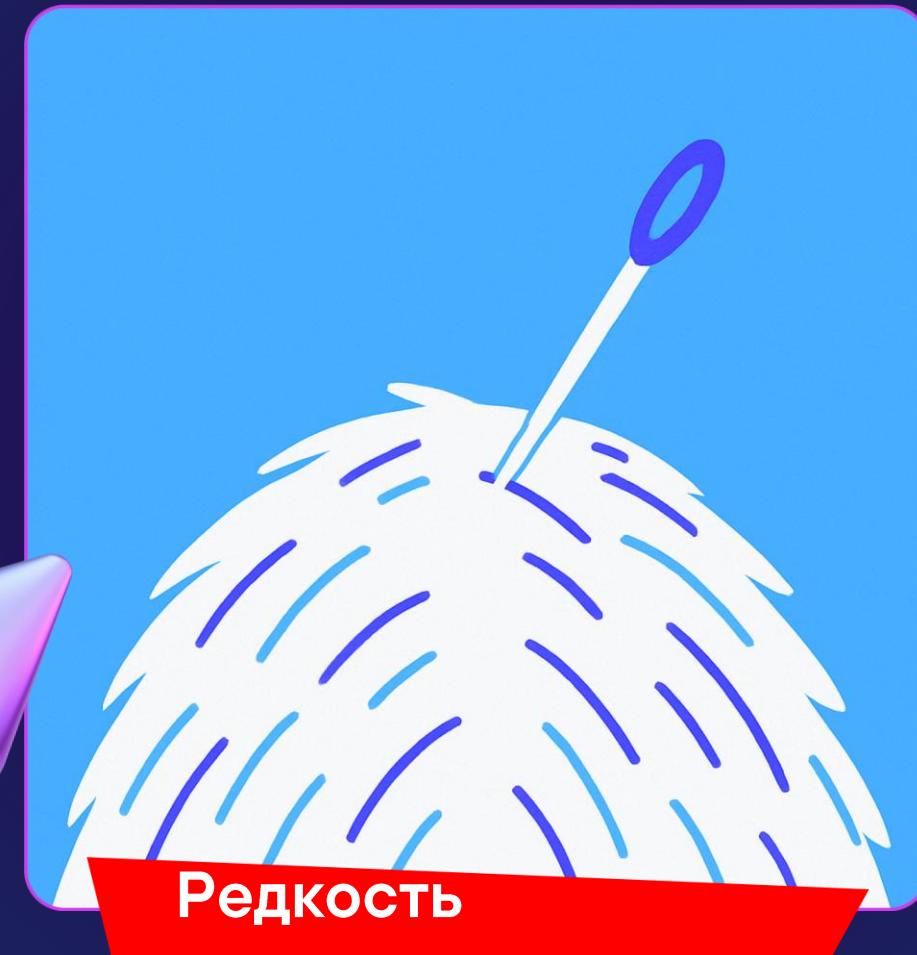
- ❖ Хакеры
- ❖ Корпоративный шпионаж



Таблицы



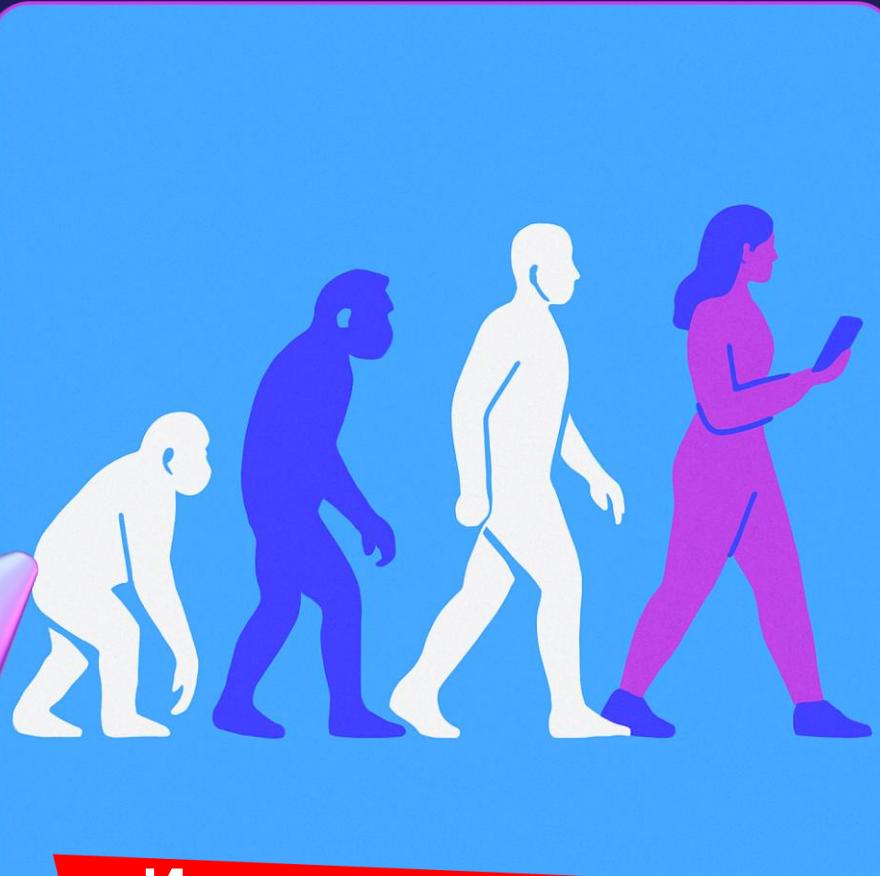
Свойства



Свойства

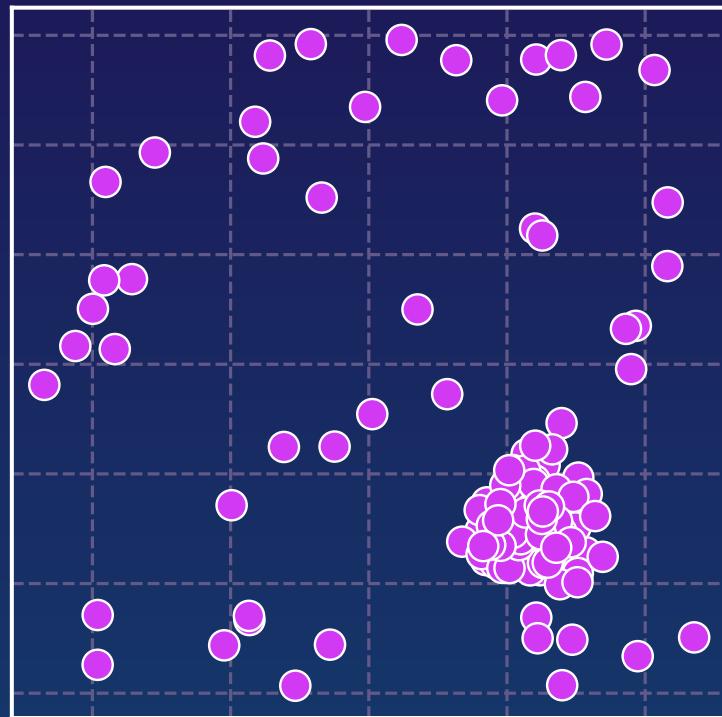


Свойства

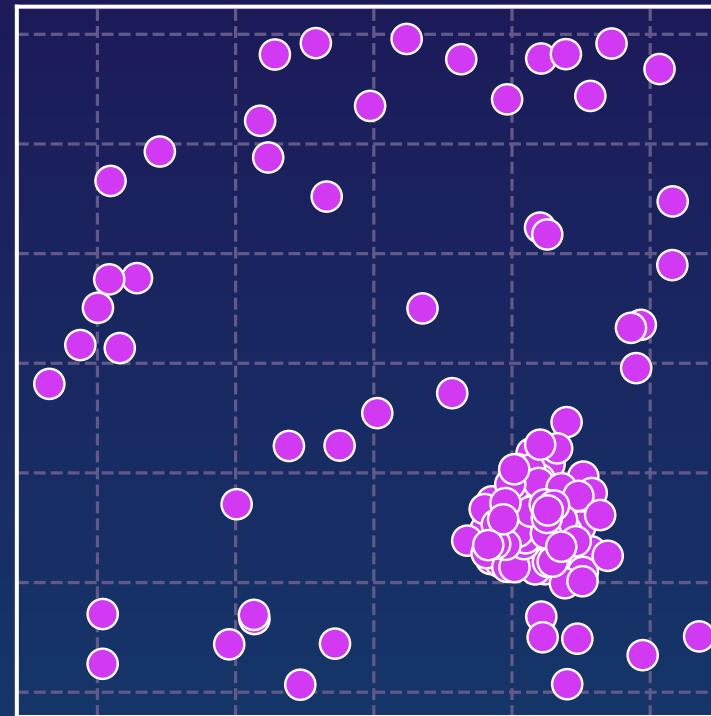
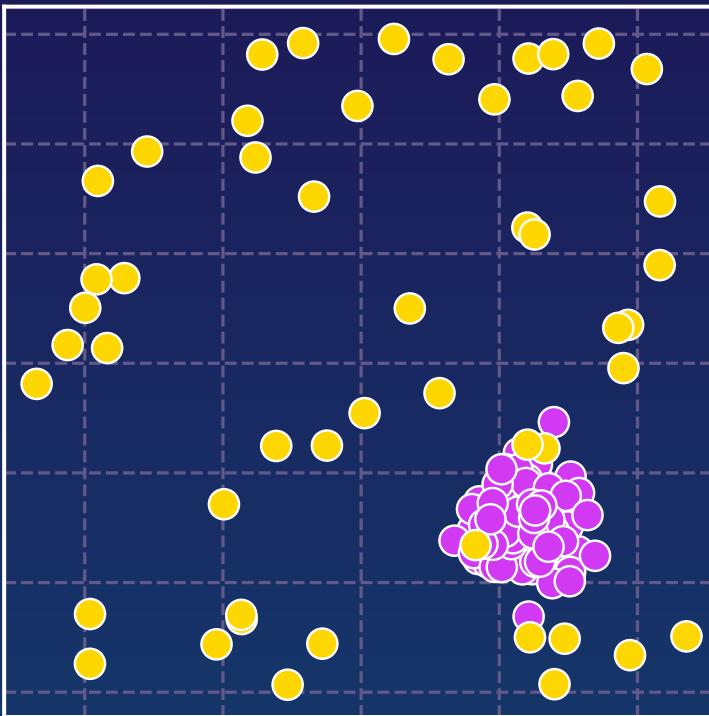


Изменчивость

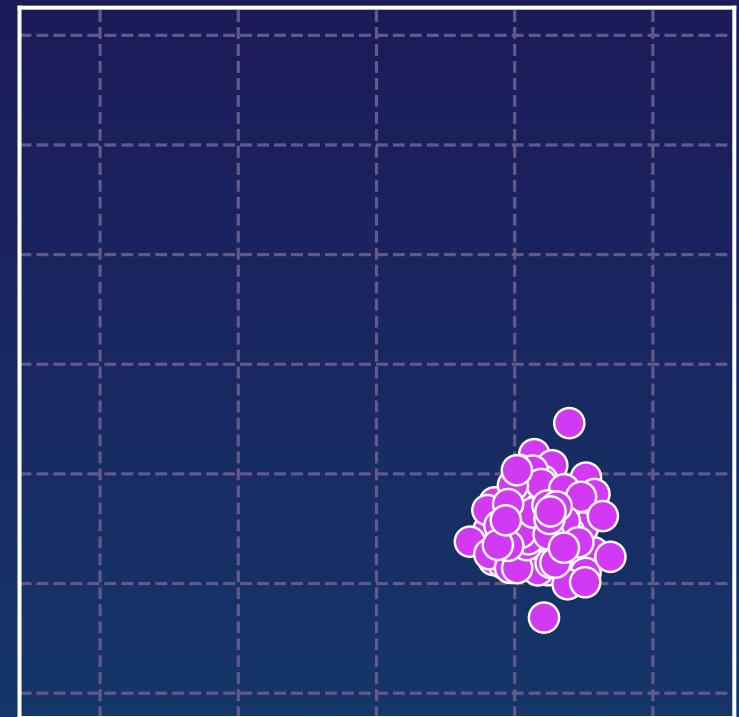
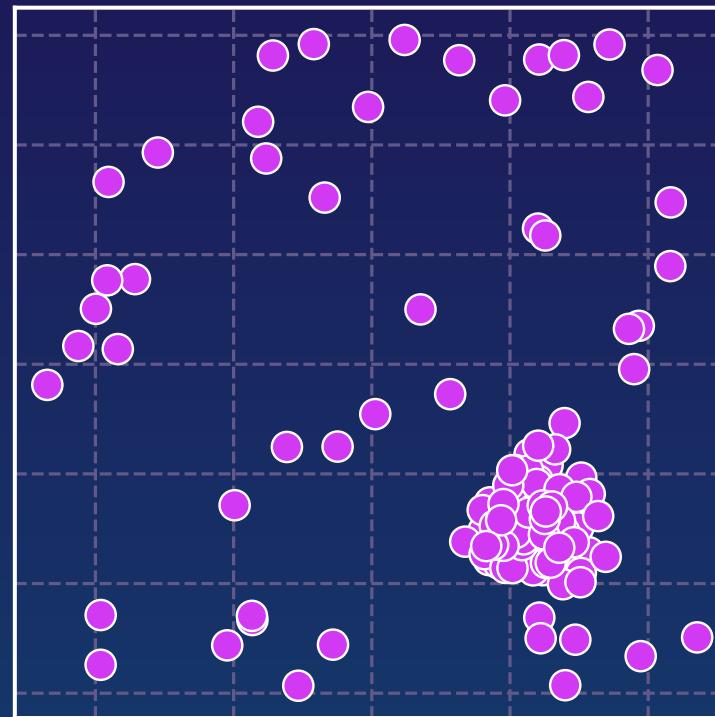
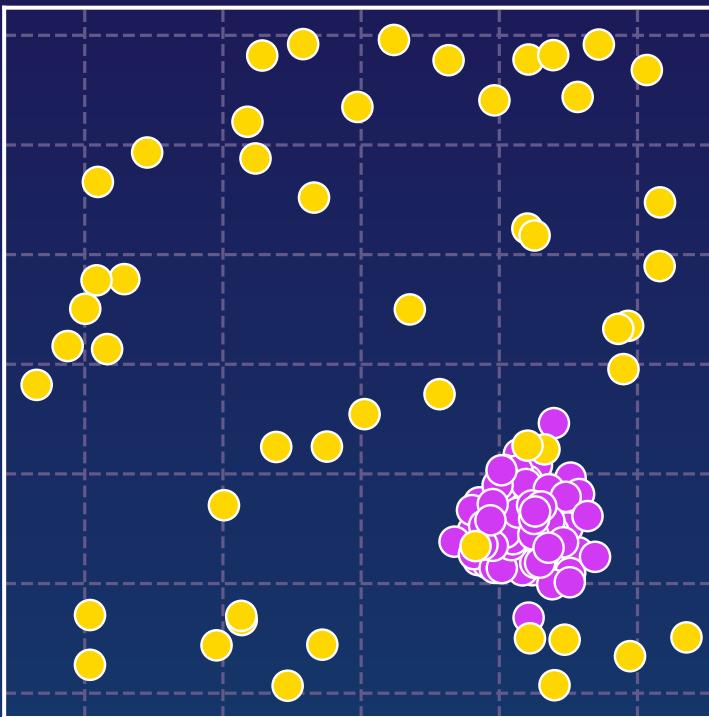
Виды задач



Виды задач

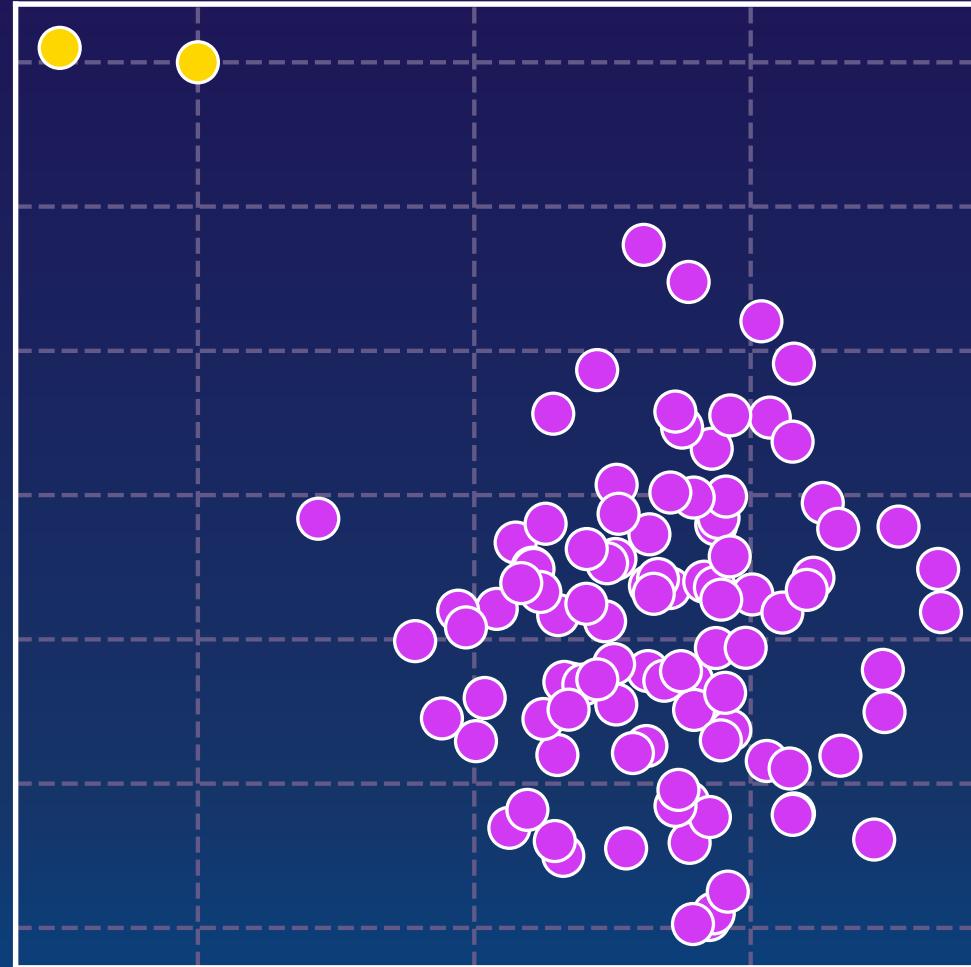


Виды задач



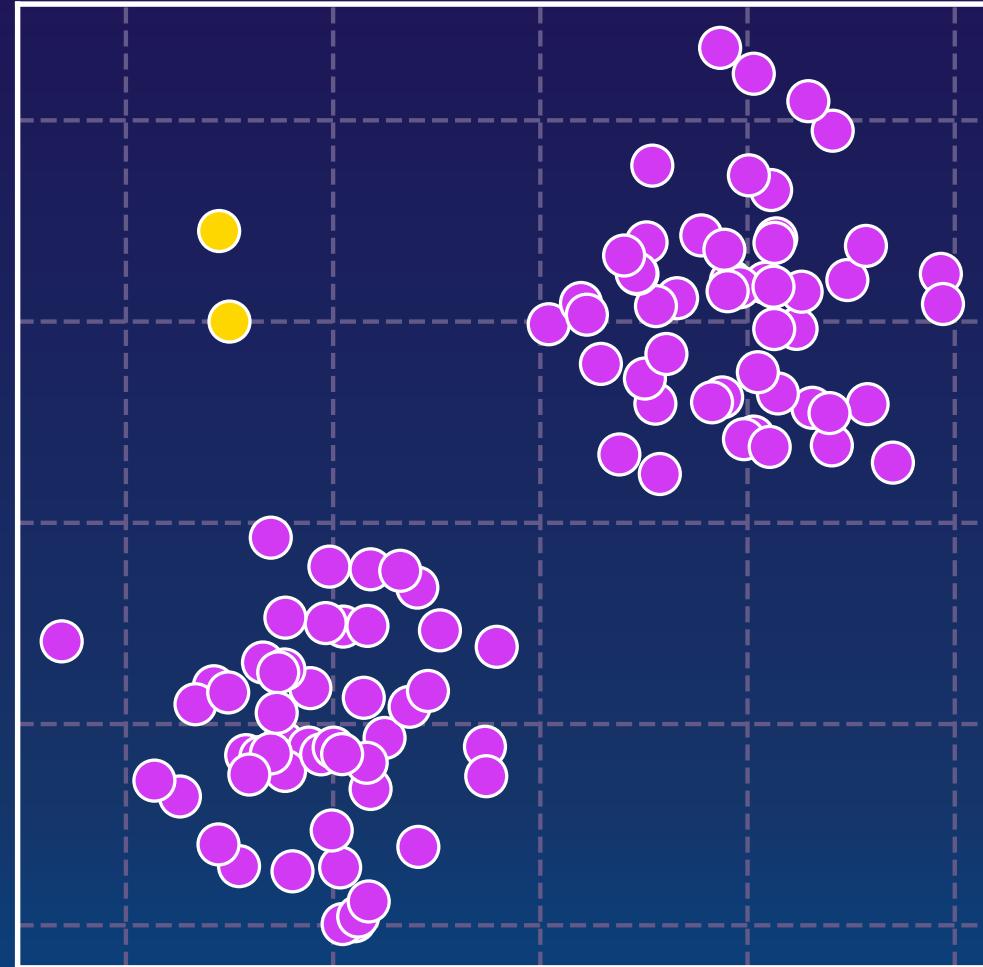
Глобальные

phd X pt



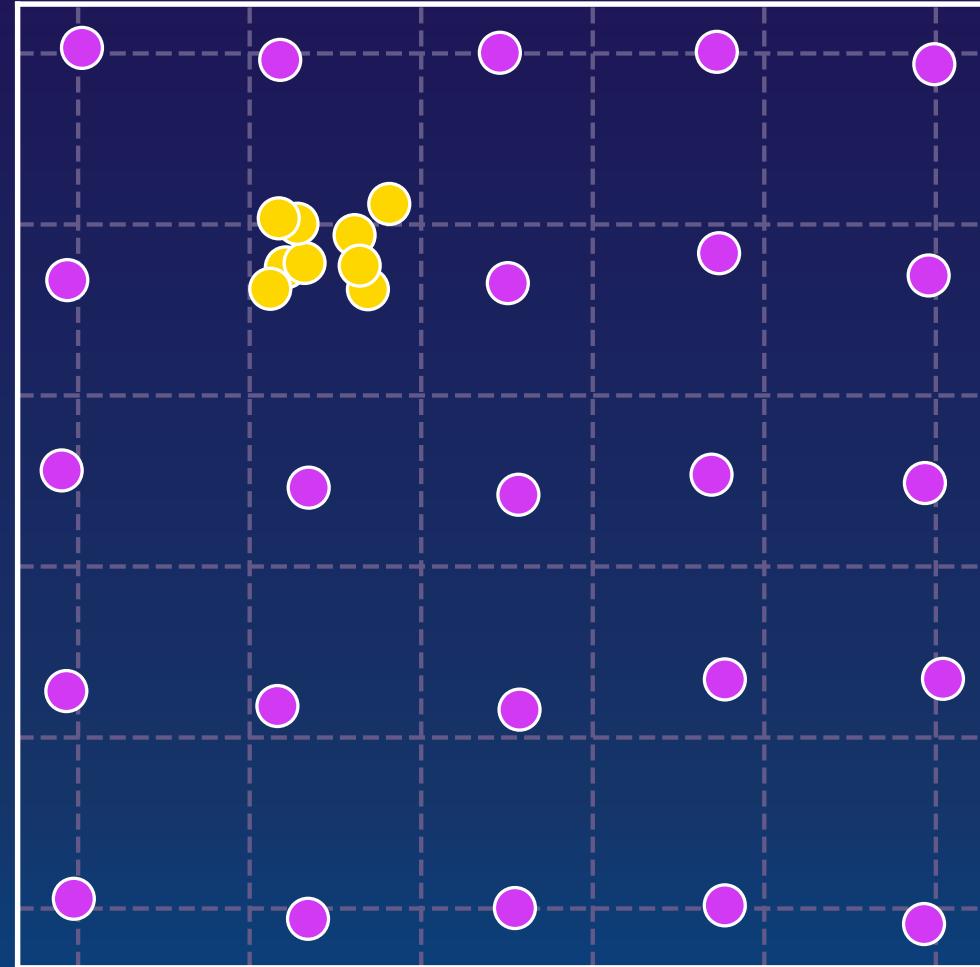
Контекстные

phd X pt



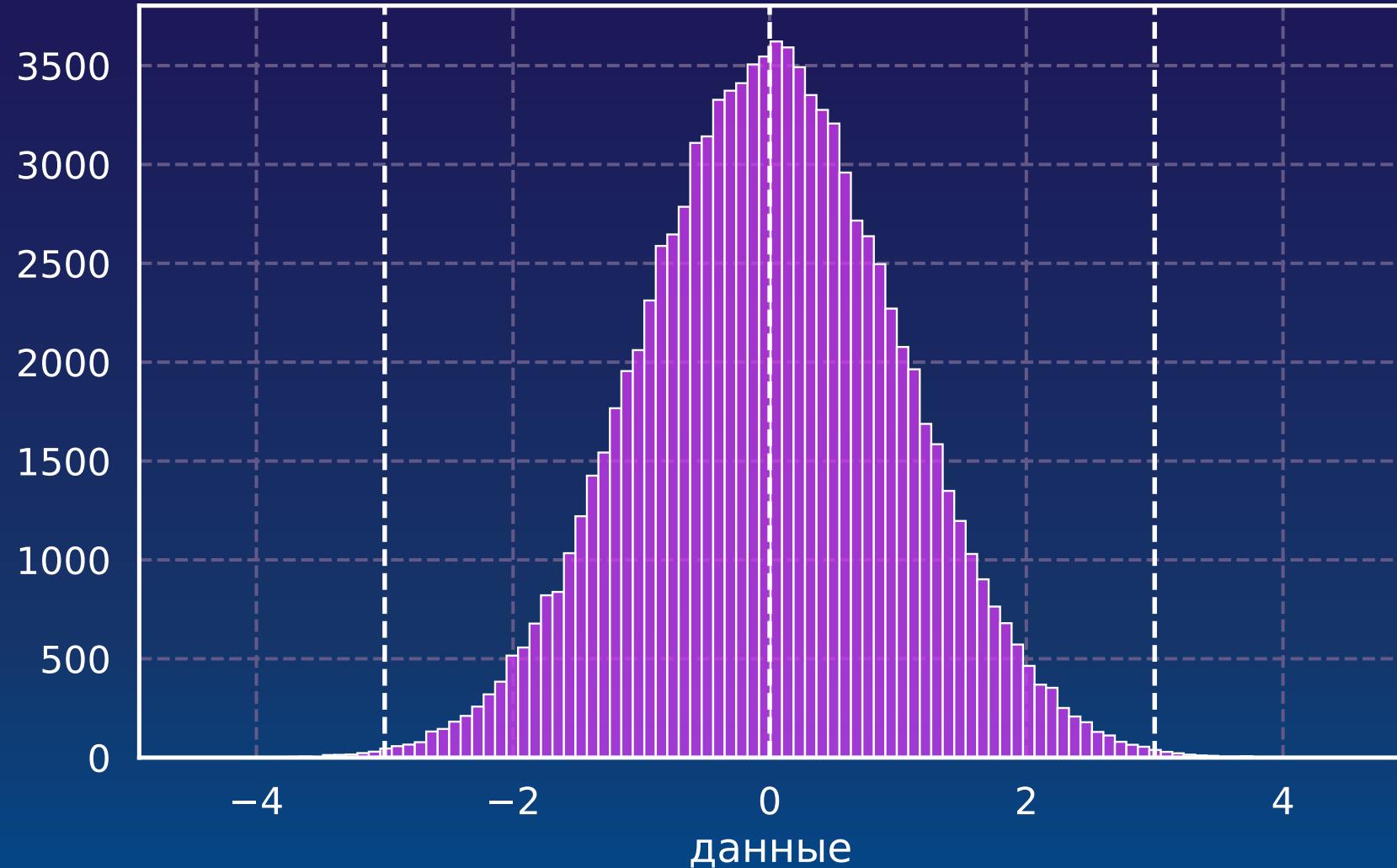
Коллективные

phd X pt



Три сигмы

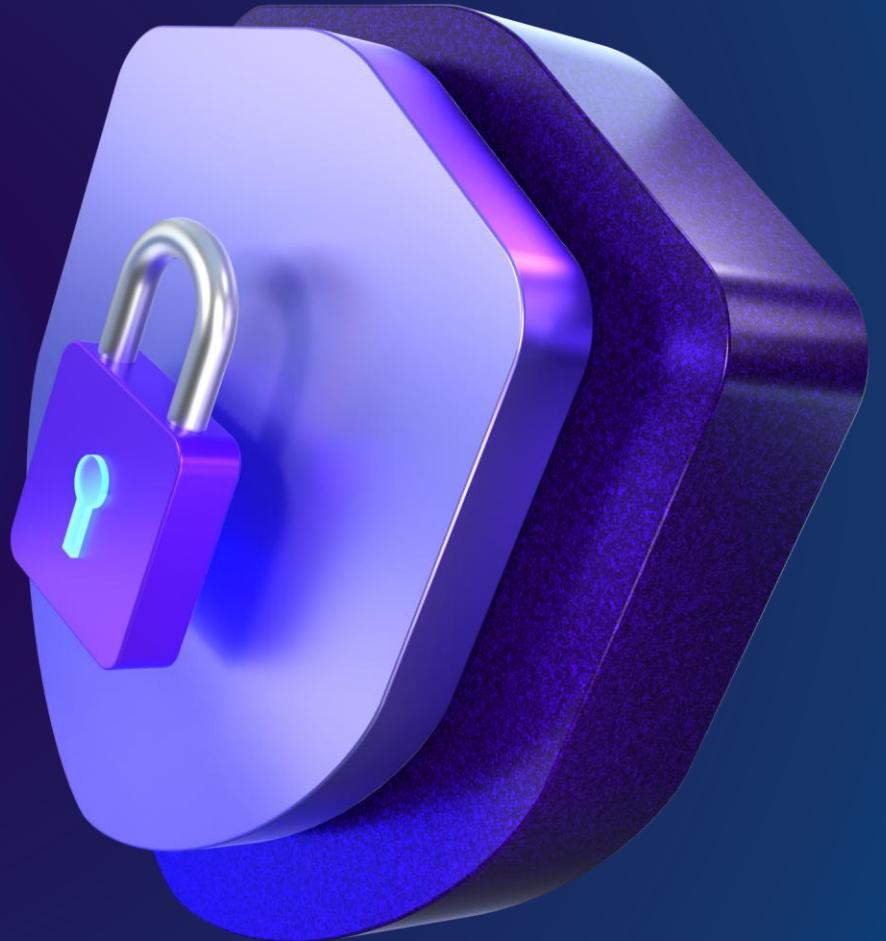
$\bar{\rho} \text{hd}$ X pt



Методы

\$ pip install pyod

```
from pyod.models.iforest import IForest  
clf = IForest()  
clf.fit(data)  
scores = clf.decision_scores_
```

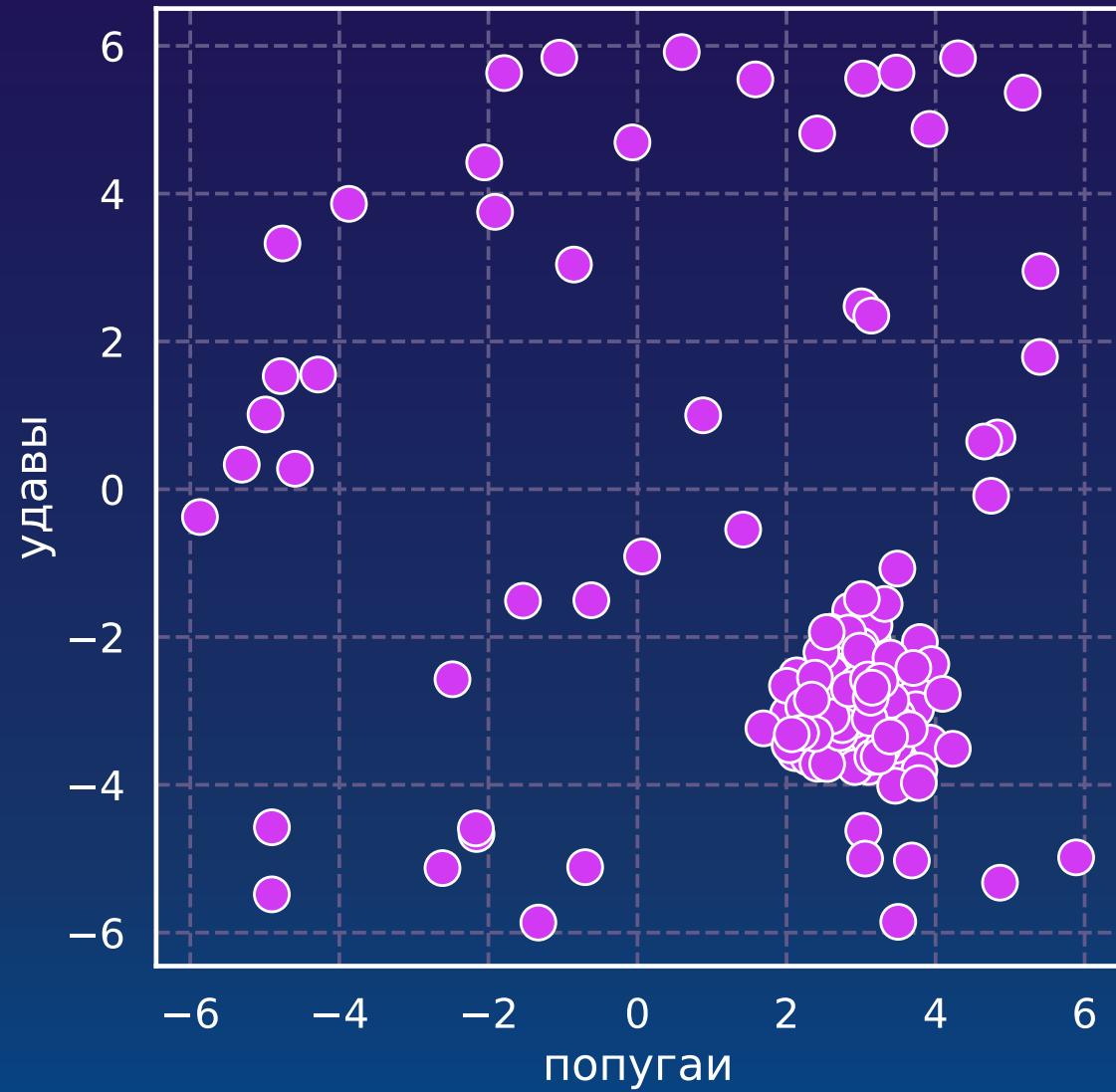


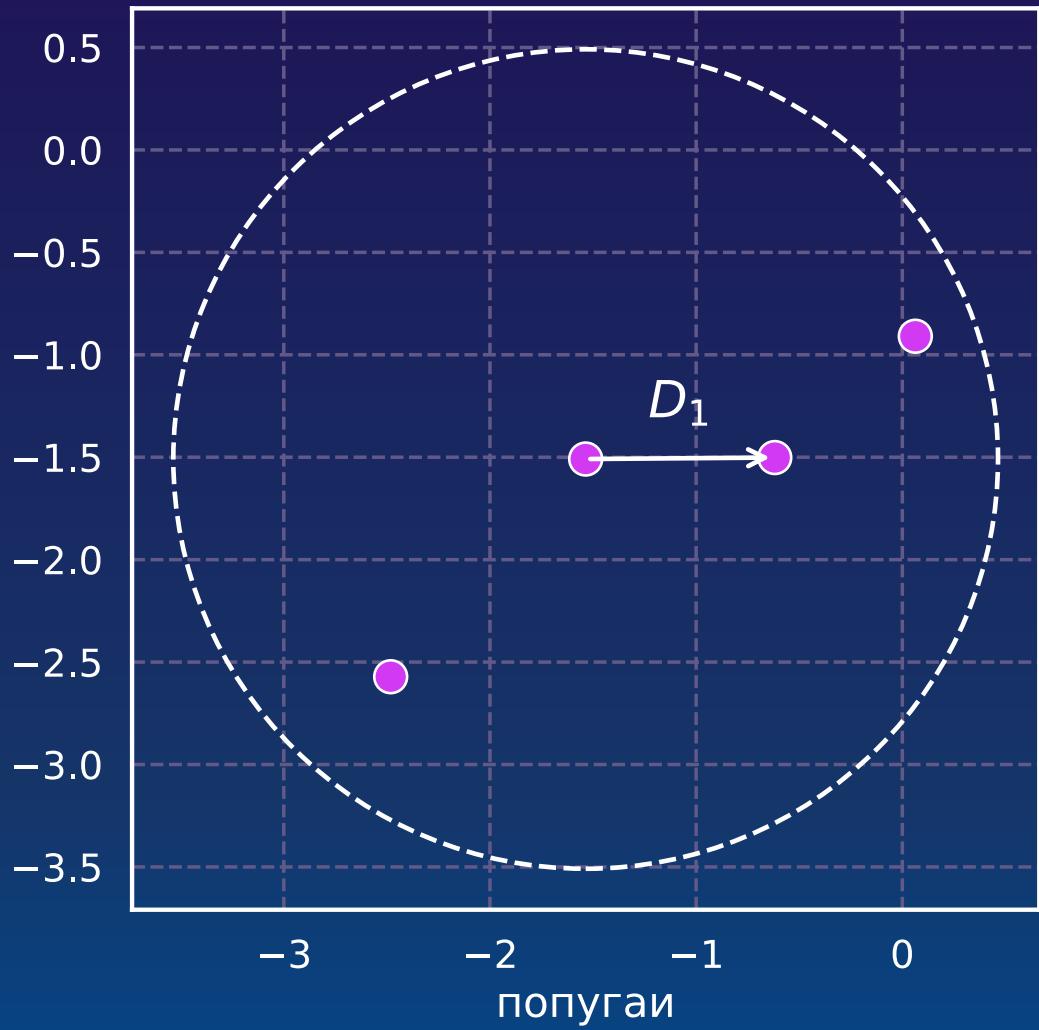
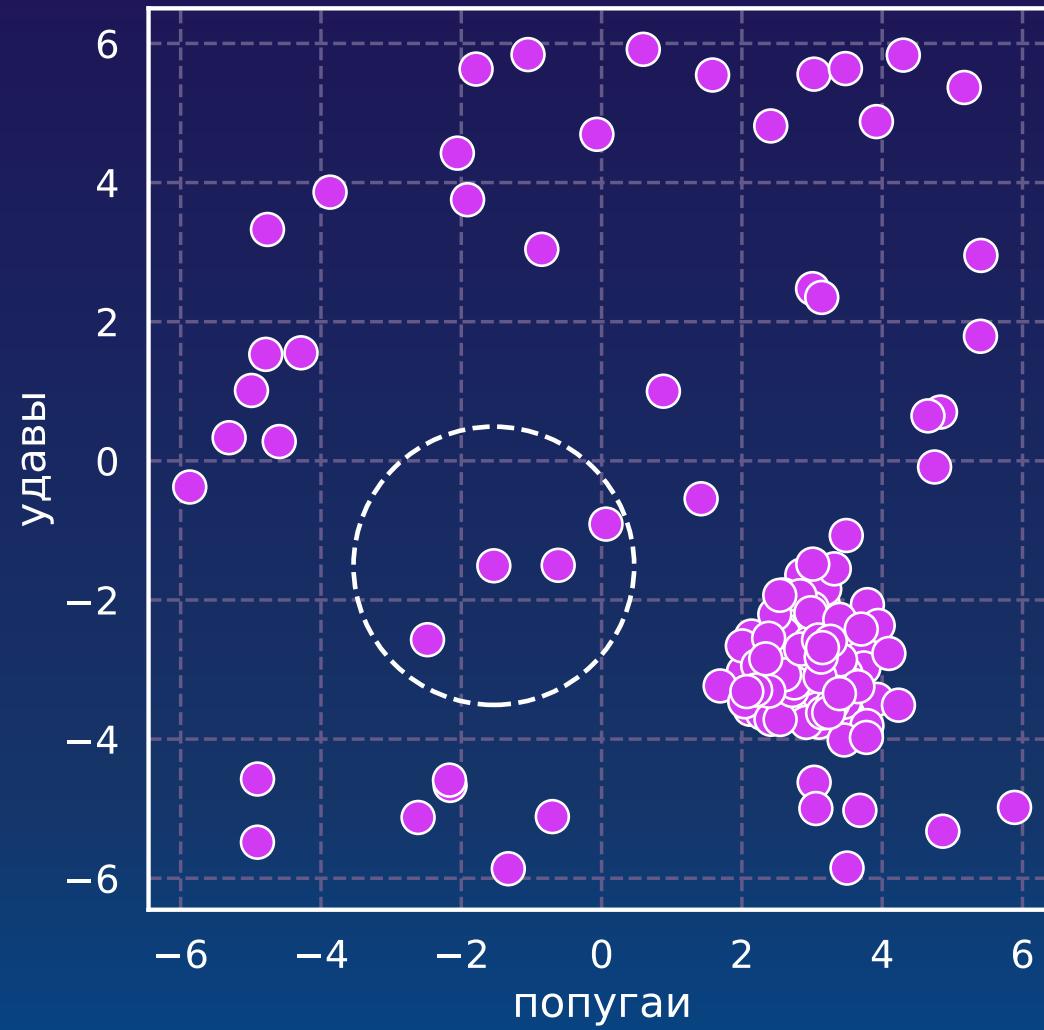


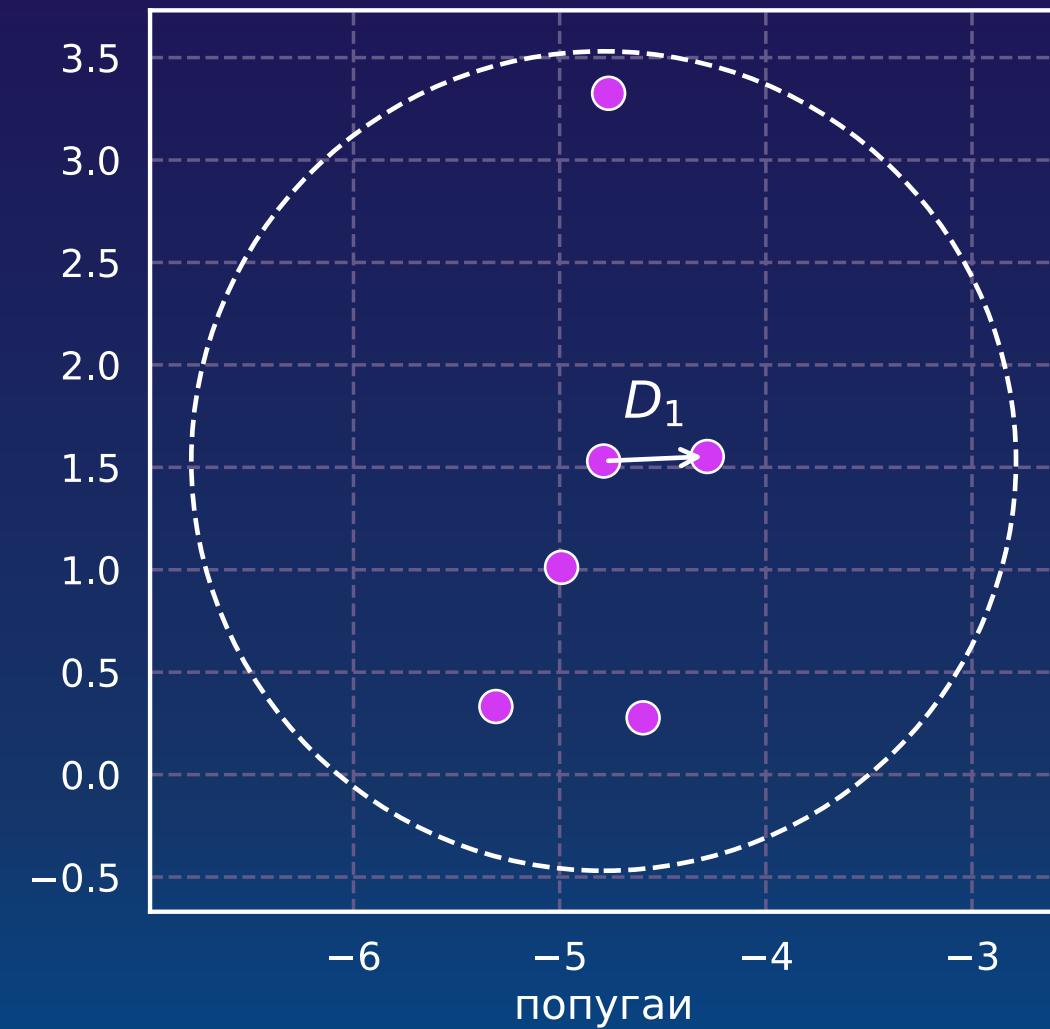
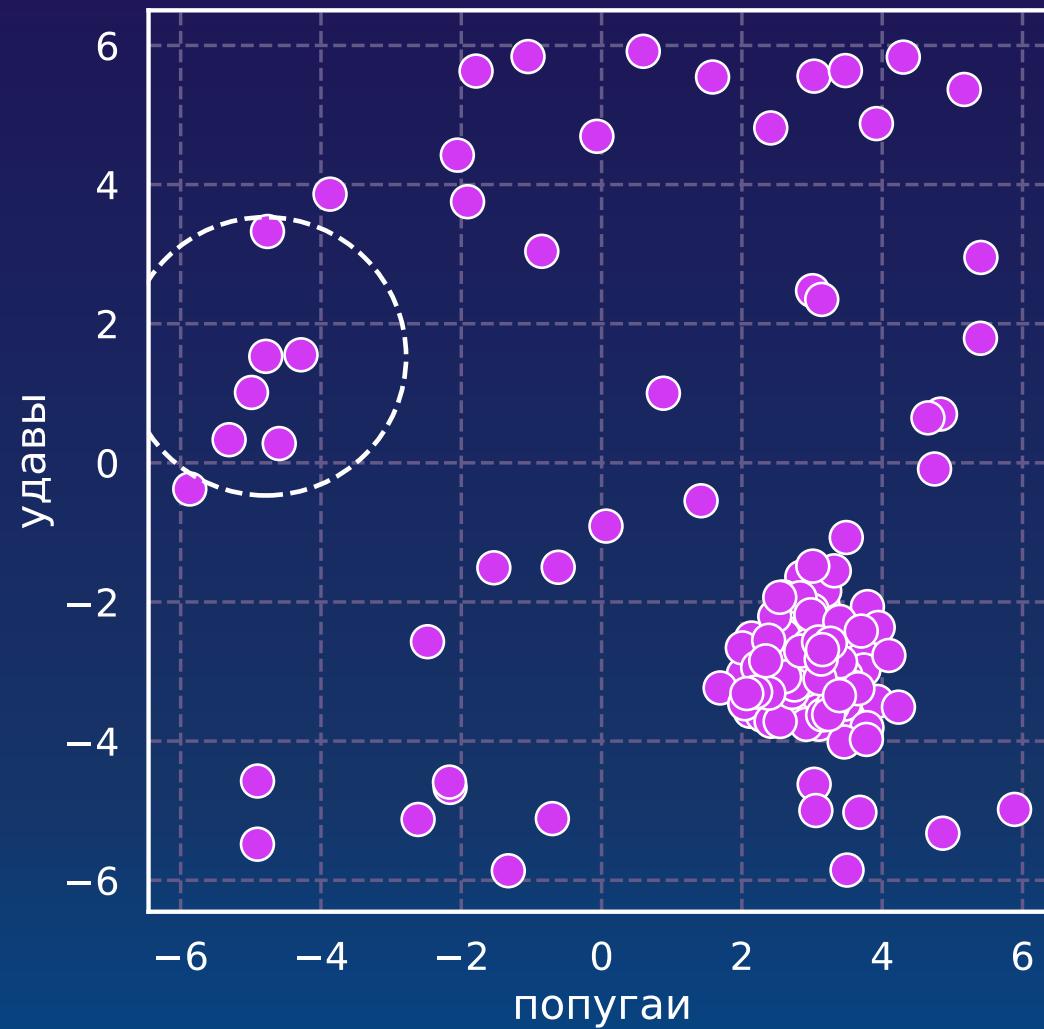
A man is known
by the company
he keeps

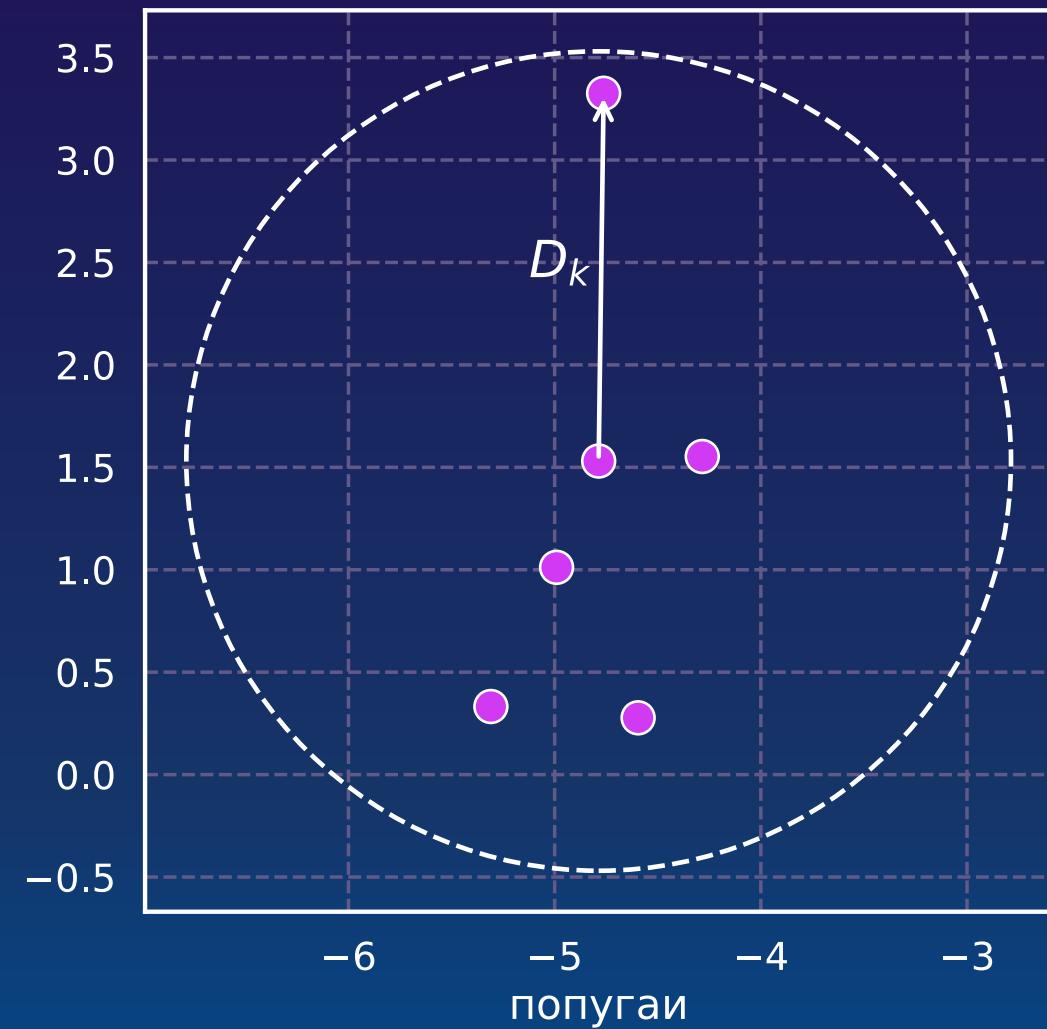
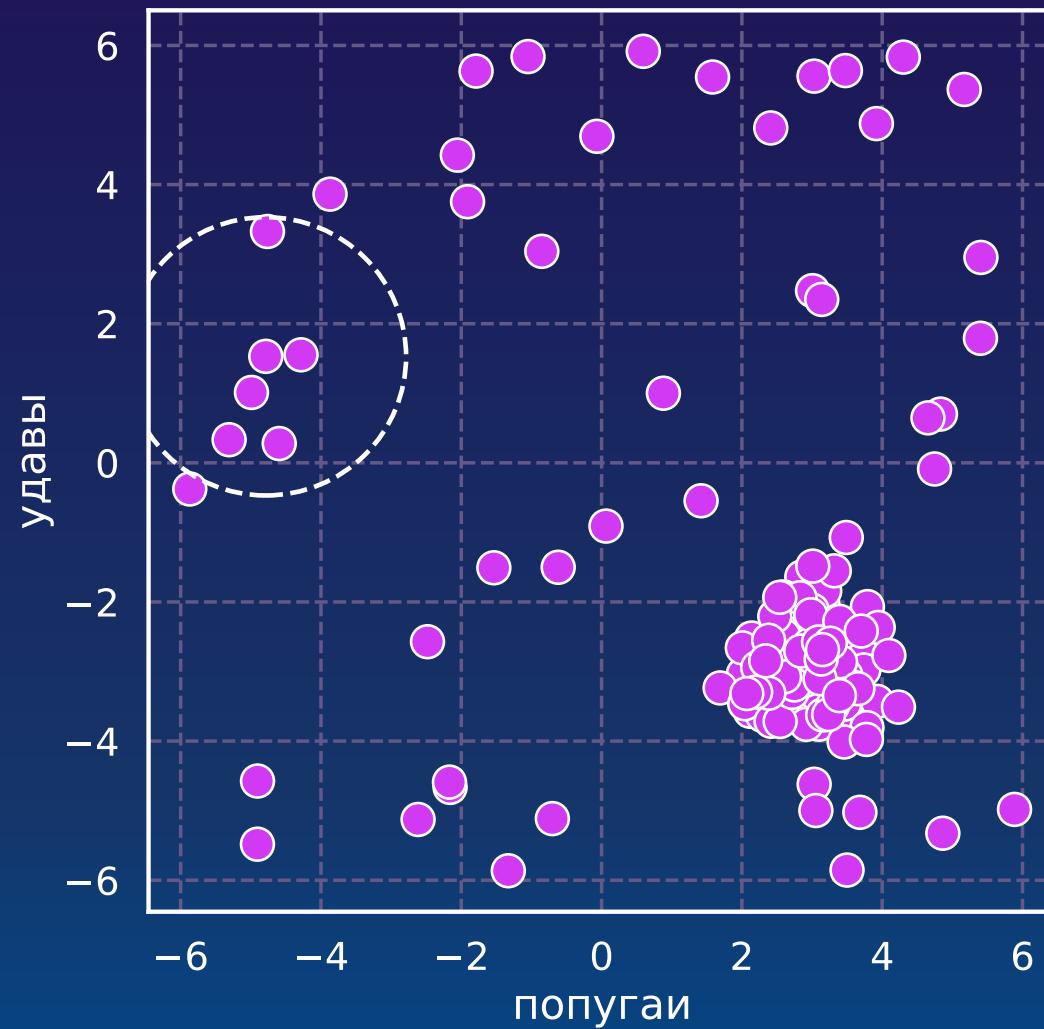
The Canterbury Tales by Geoffrey Chaucer

Proximity



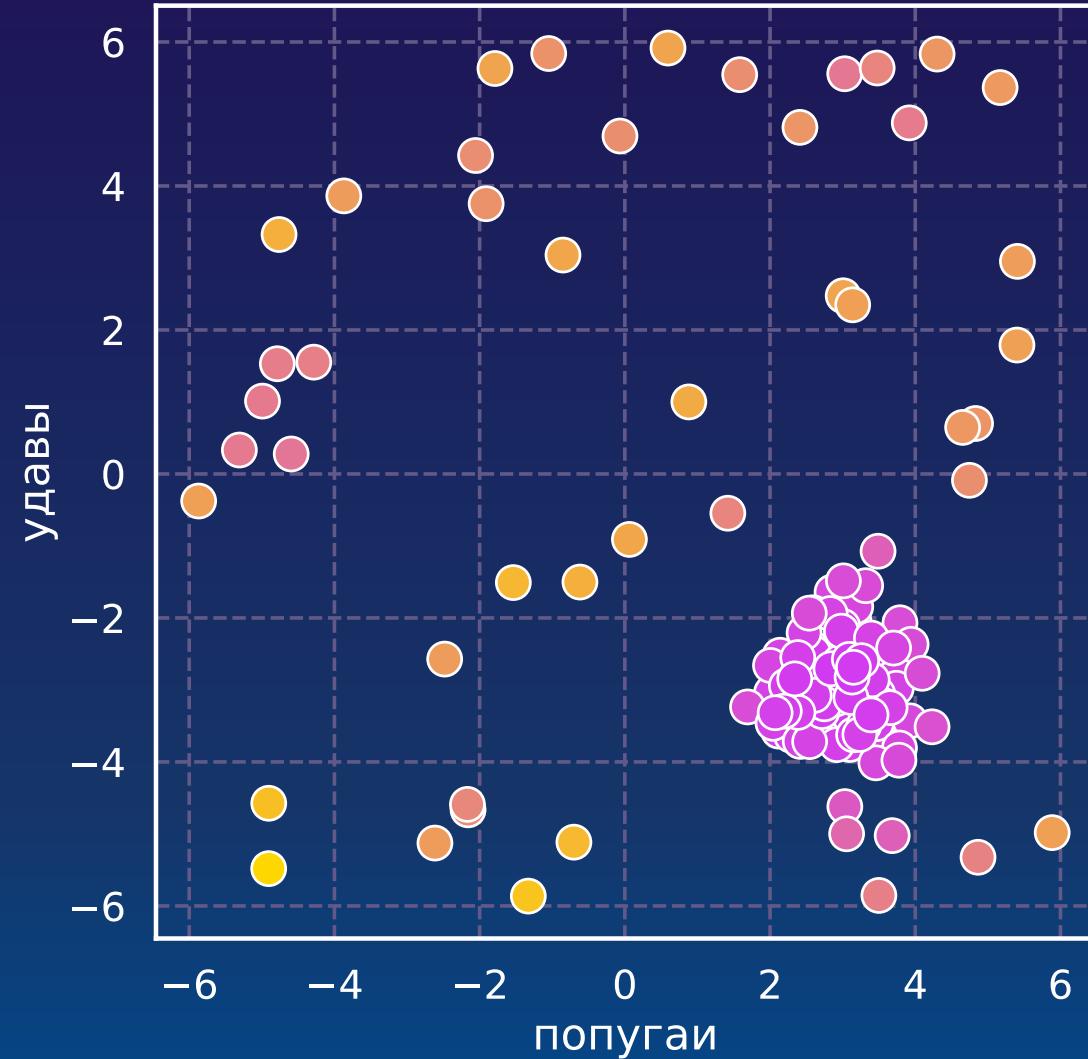






k Nearest Neighbors

phd X pt

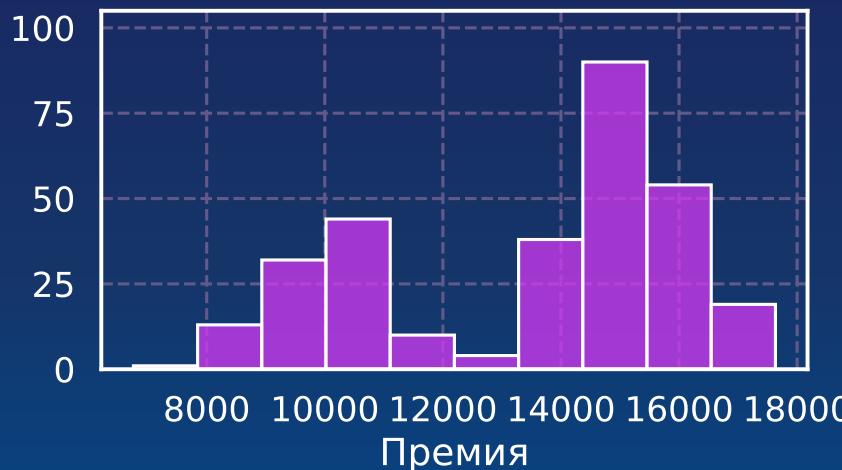
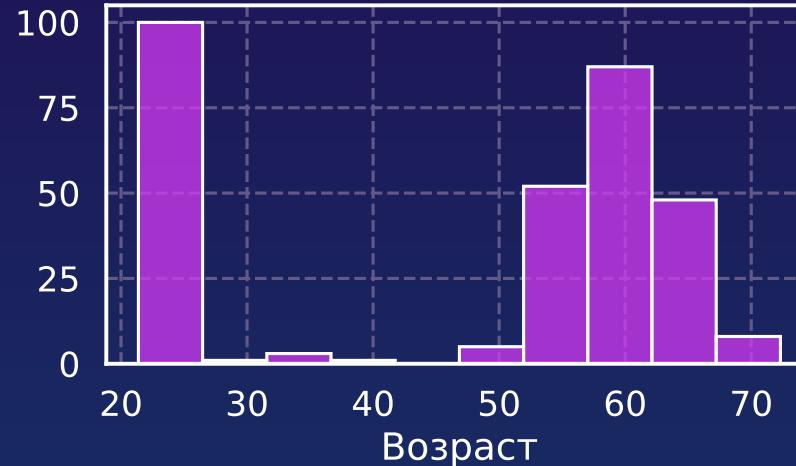


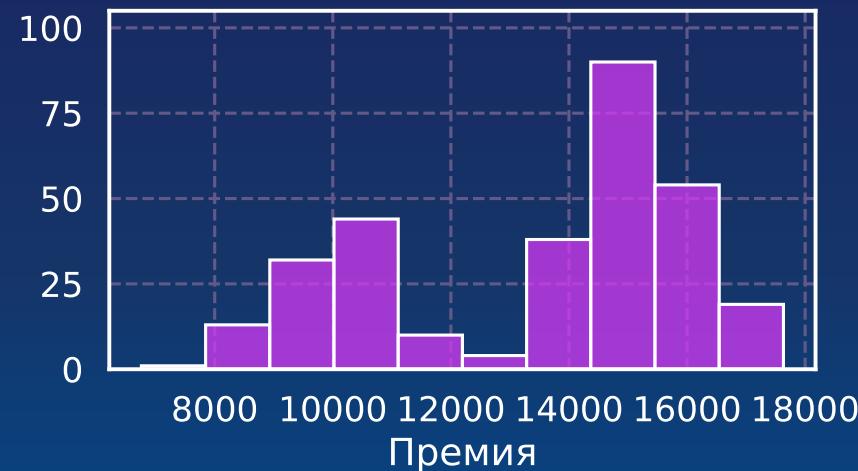
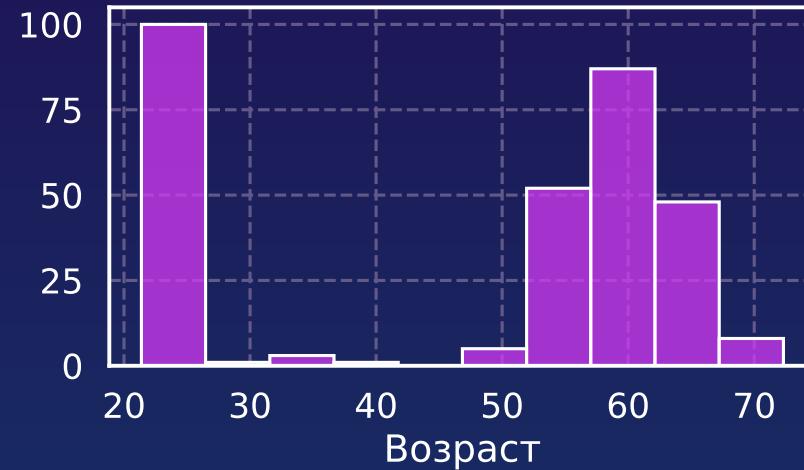


Il faut savoir
se faire petit parfois

Joe Dassin

Probabilistic



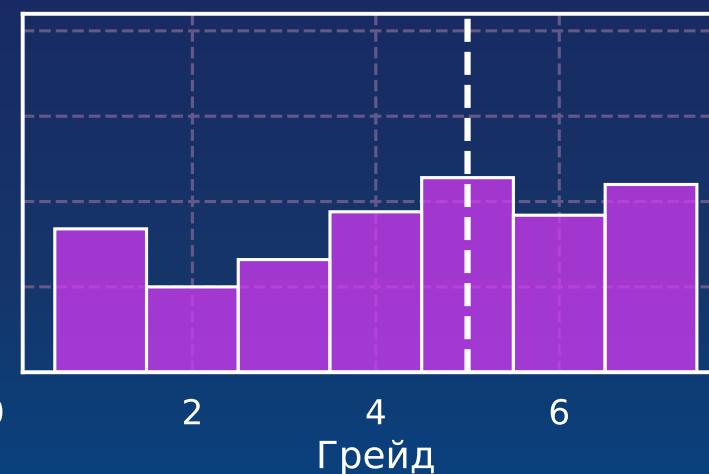
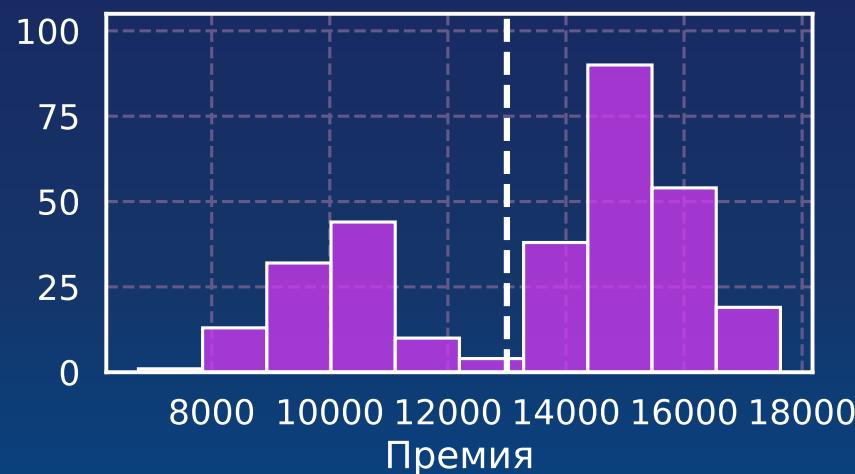
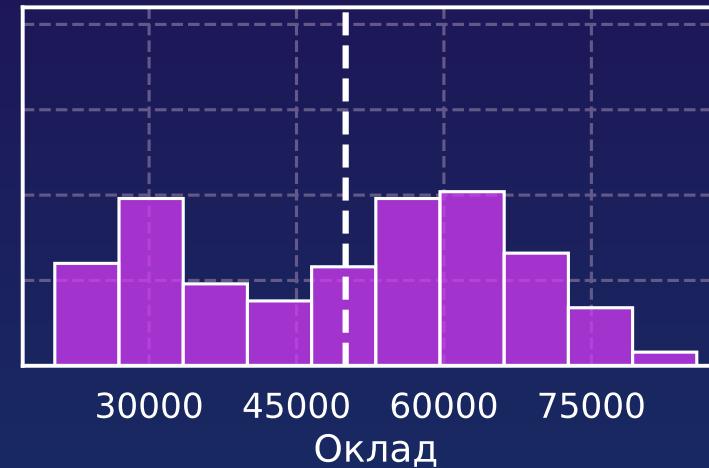
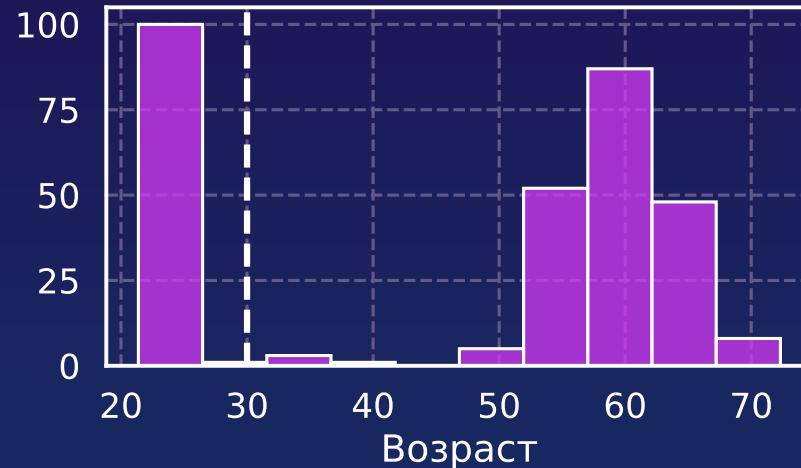


age = 30

salary = 50000

bonus = 13000

grade = 5

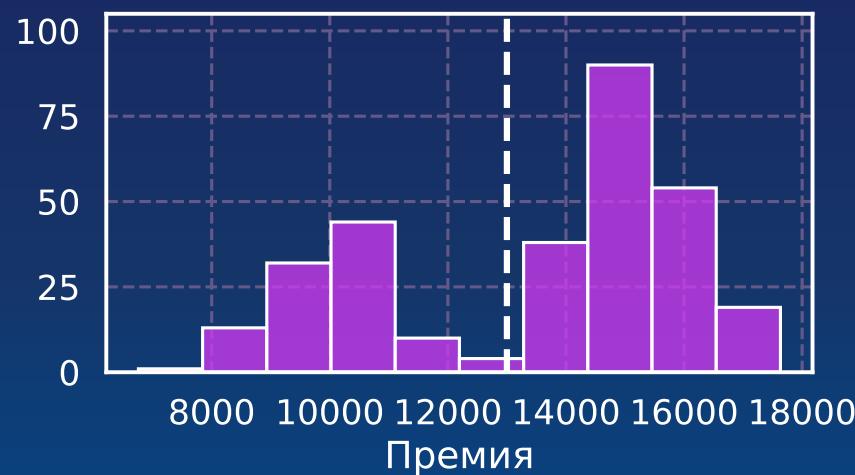
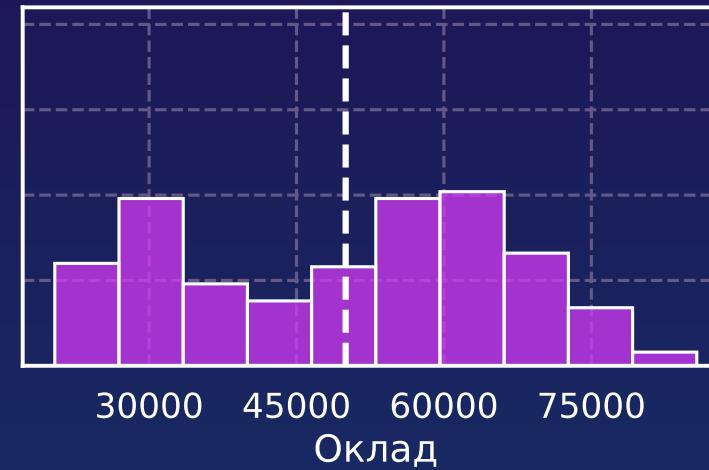
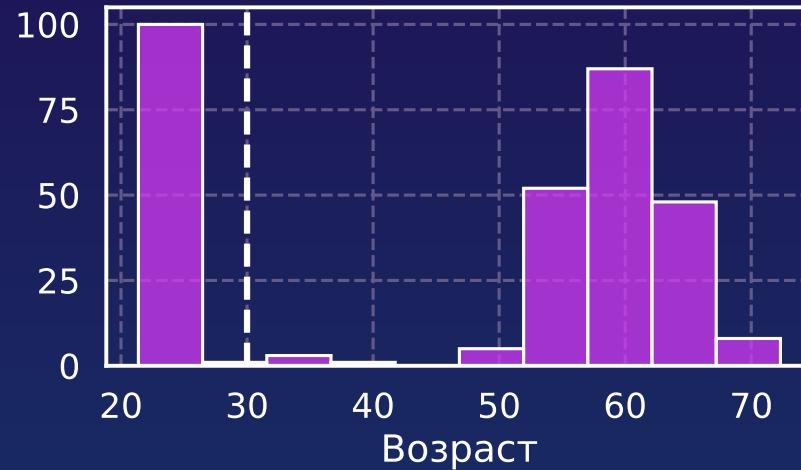


age = 30

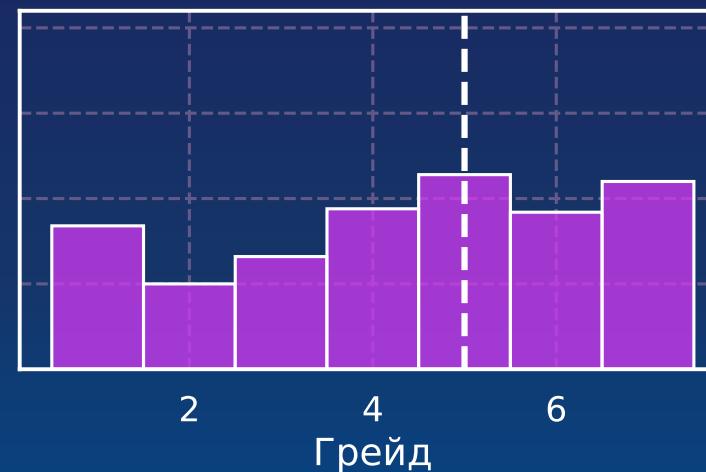
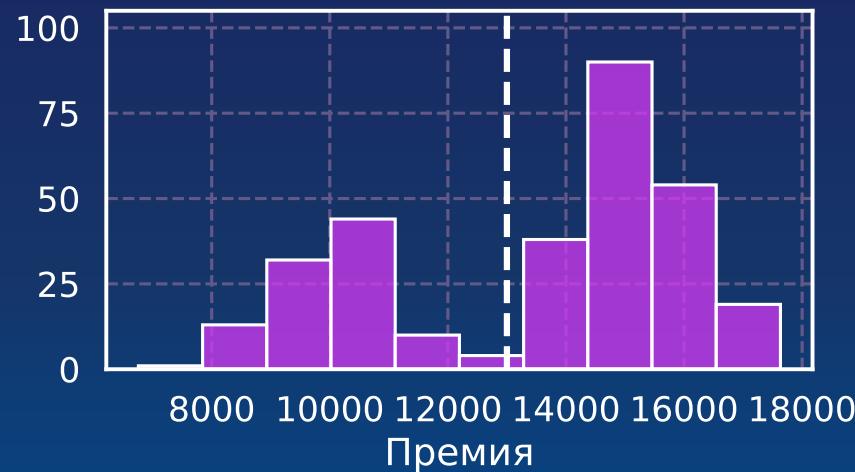
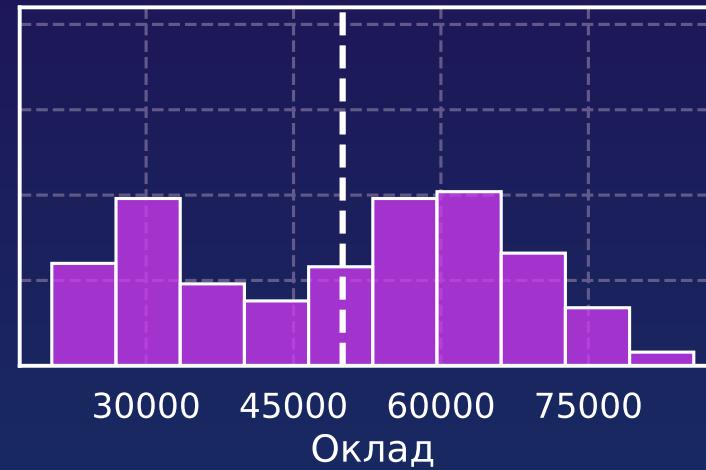
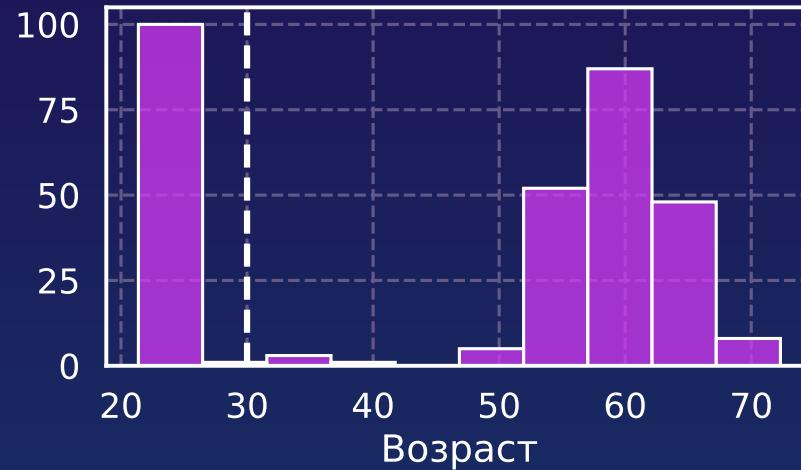
salary = 50000

bonus = 13000

grade = 5



age = 30 → 1
salary = 50000 → 29
bonus = 13000 → 4
grade = 5 → 57



age = 30

→ 1/100

salary = 50000

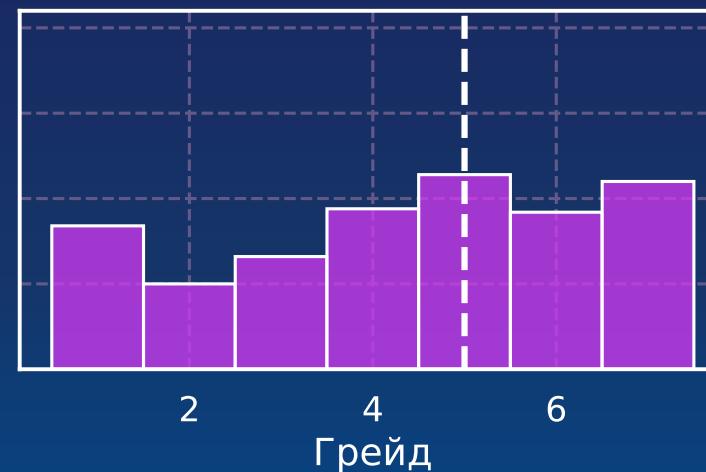
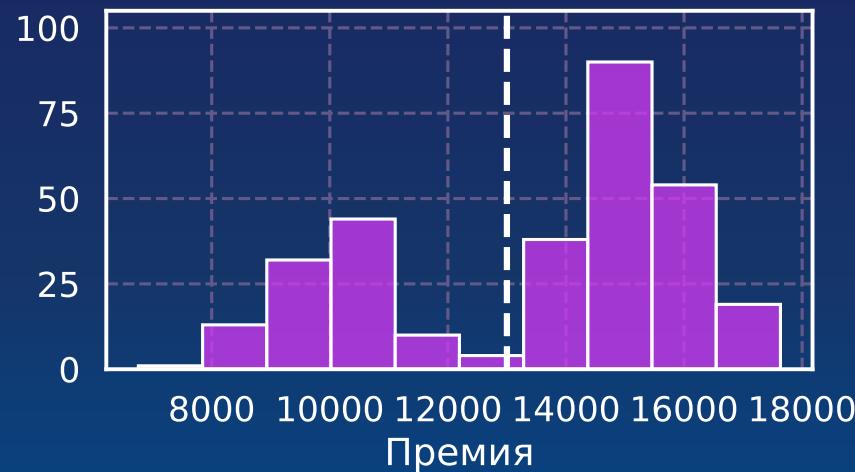
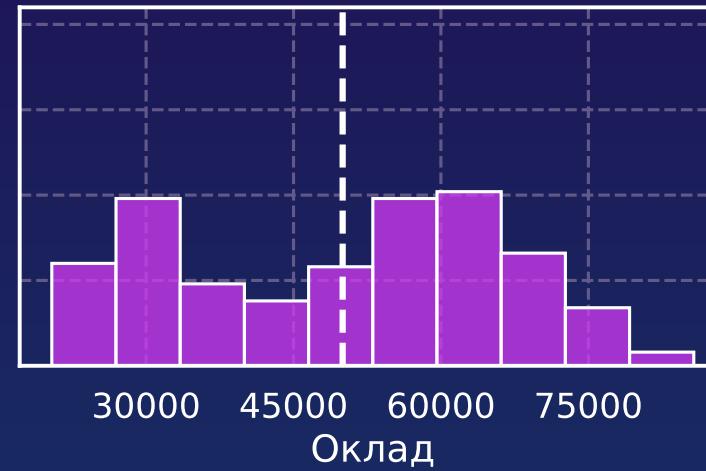
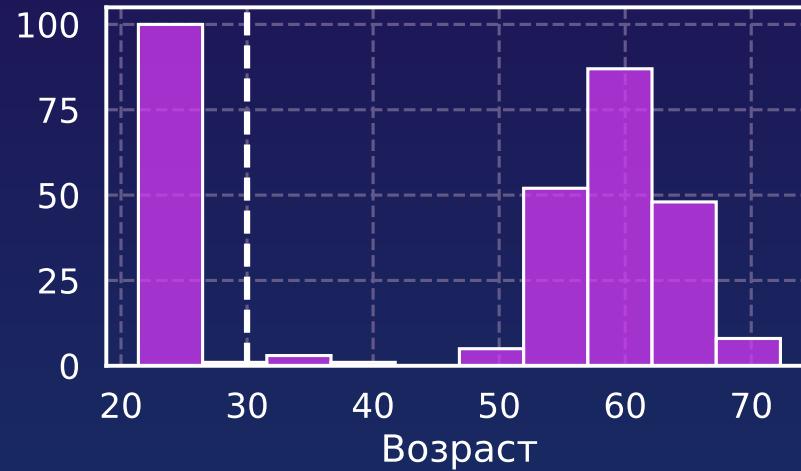
→ 29/51

bonus = 13000

→ 4/90

grade = 5

→ 57/57



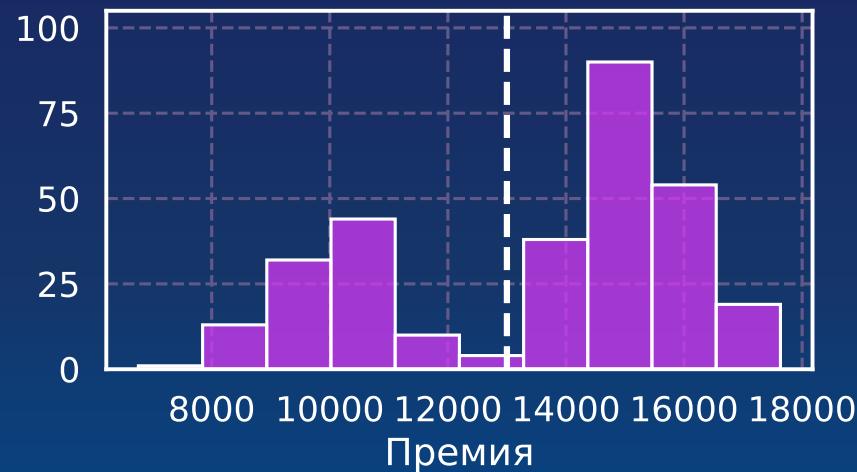
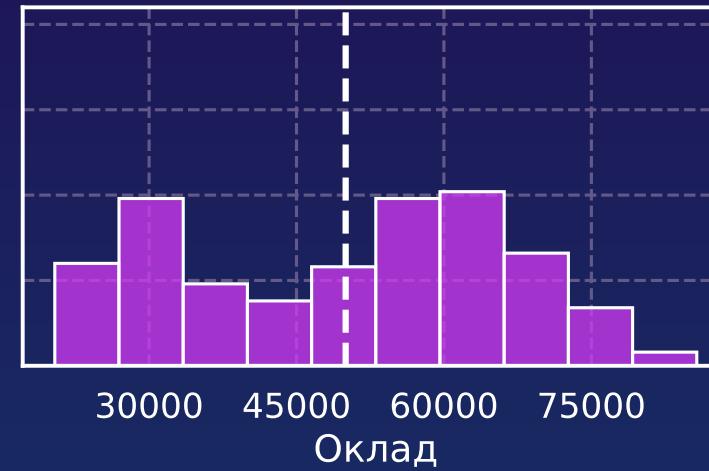
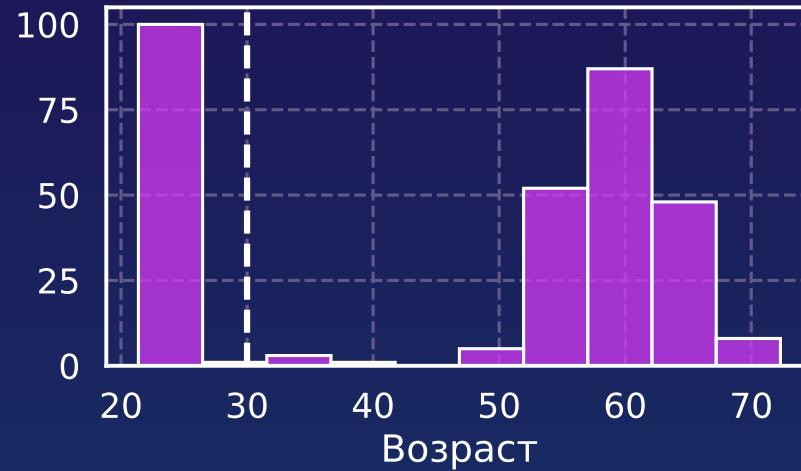
age = 30 → 0.01

salary = 50000 → 0.57

bonus = 13000 → 0.04

grade = 5 → 1.00

mean = 0.4



age = 30 → 0.01

salary = 50000 → 0.57

bonus = 13000 → 0.04

grade = 5 → 1.00

mean = 0.4

score = 1 / 0.4 = 2.46

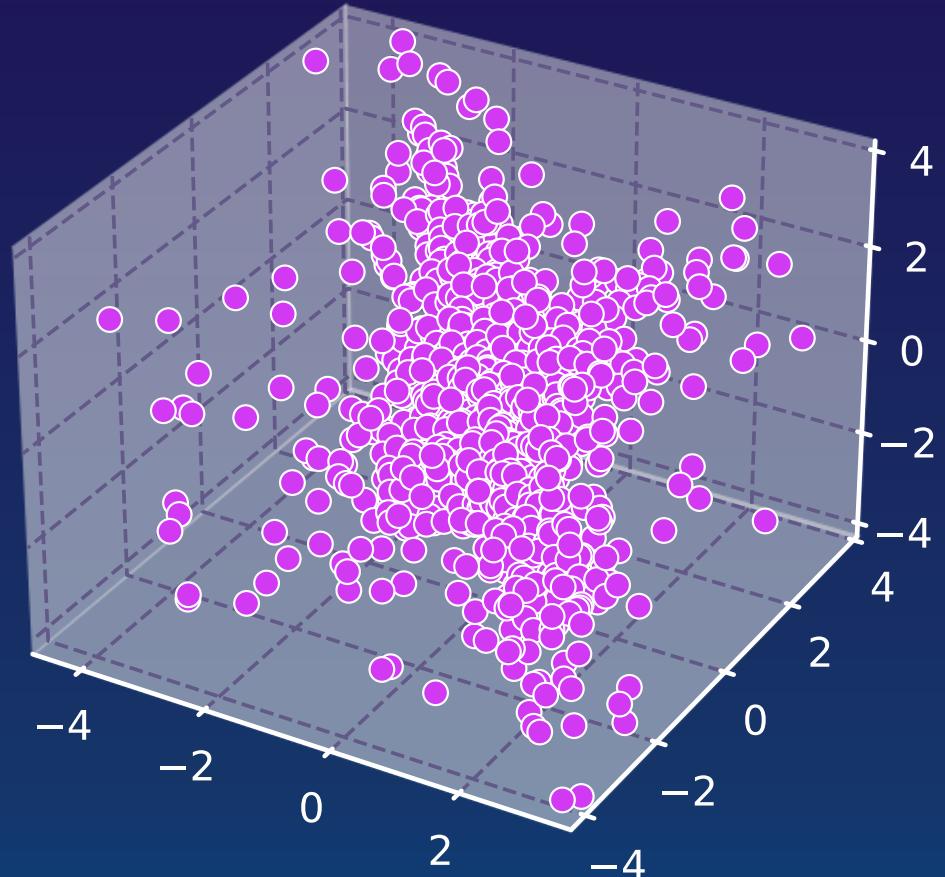


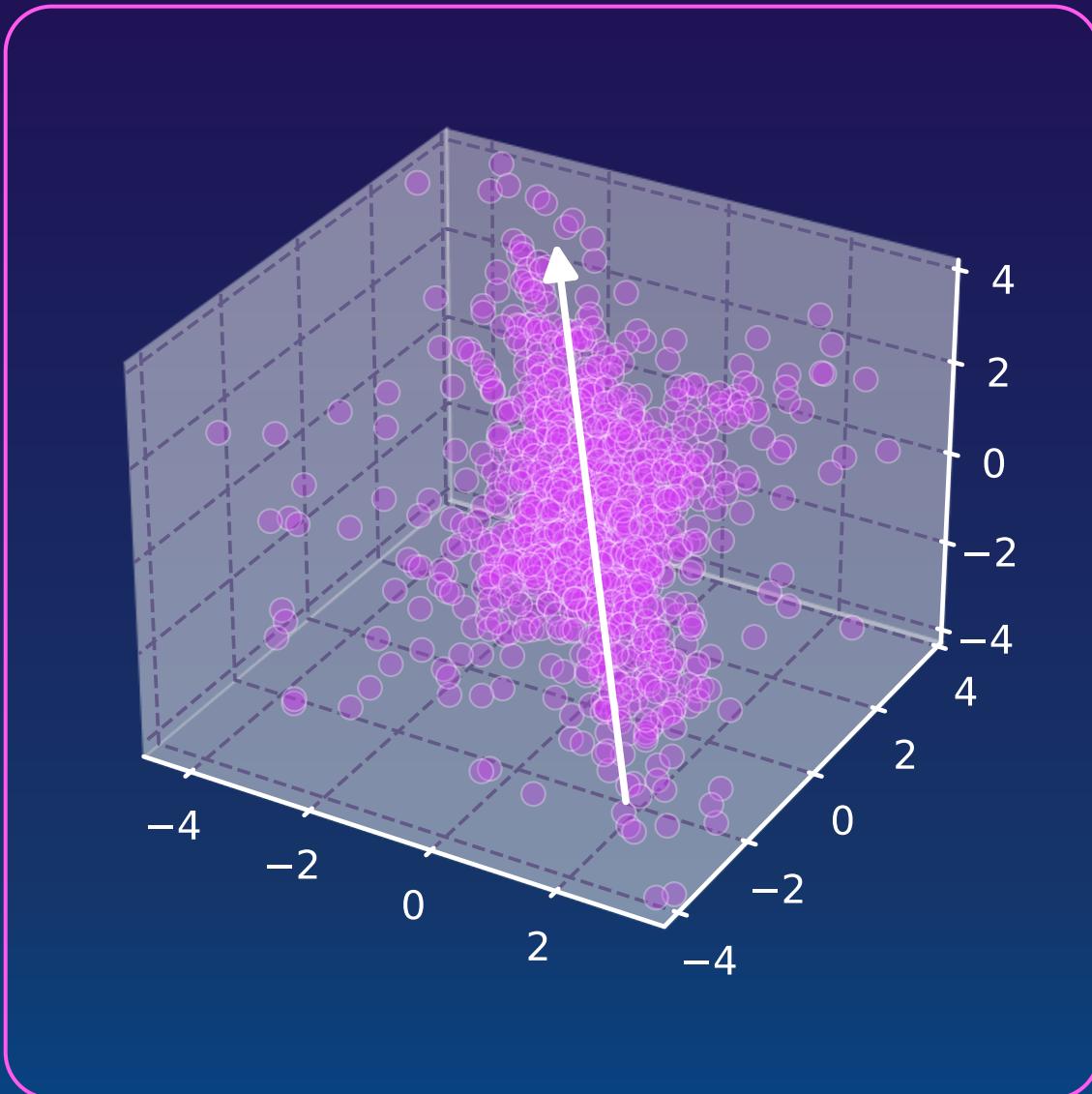
In der Beschränkung
zeigt sich erst der Meister

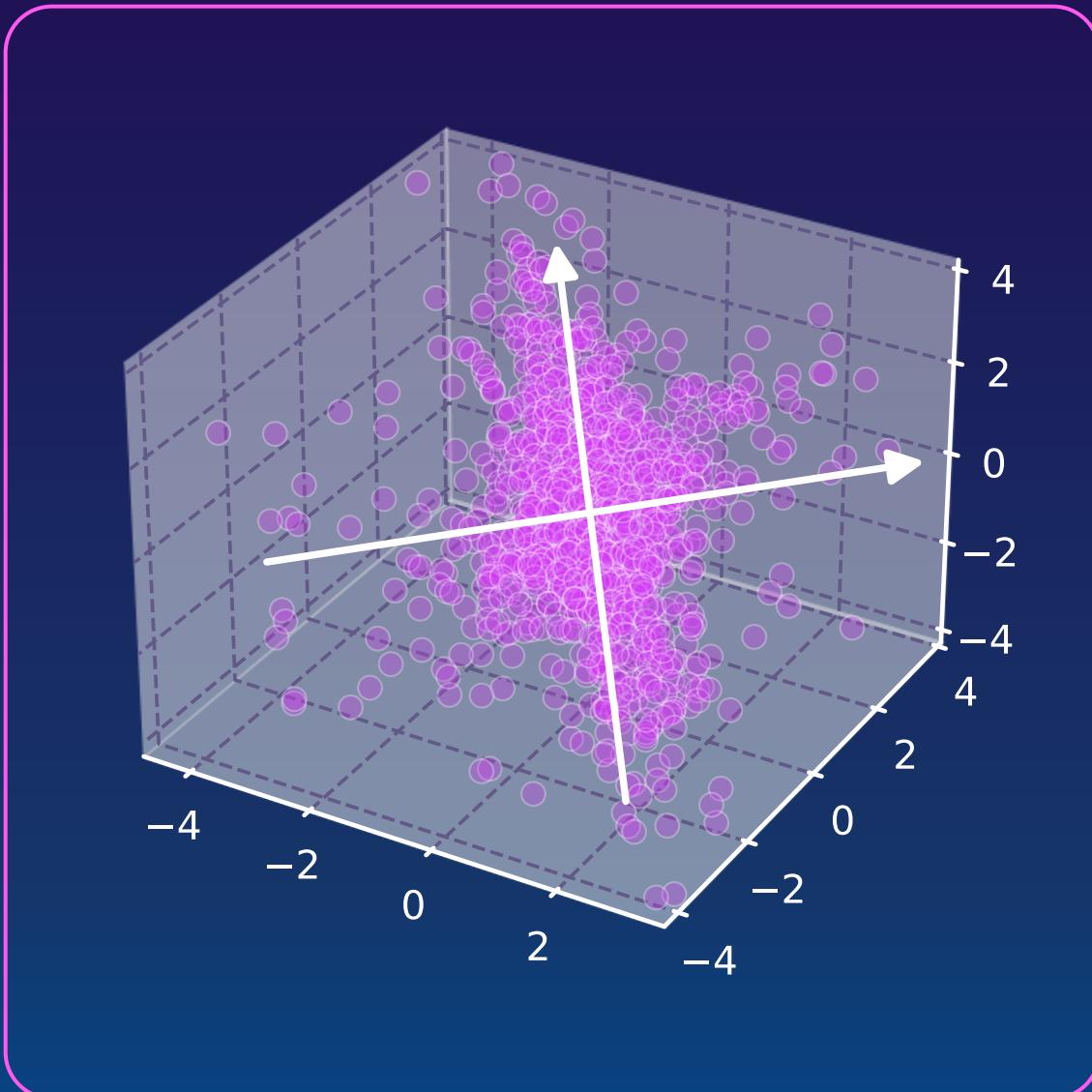
«Sonette» von Johann Wolfgang von Goethe

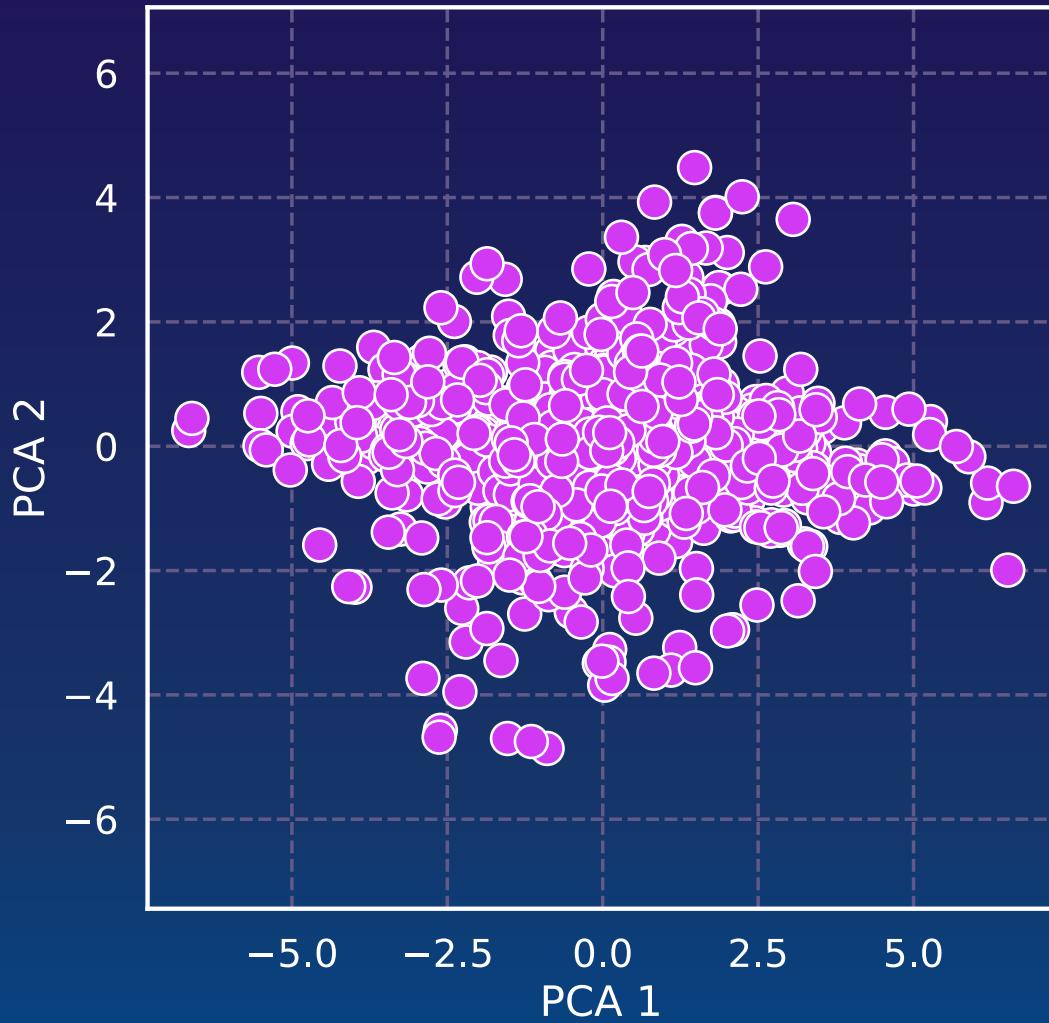
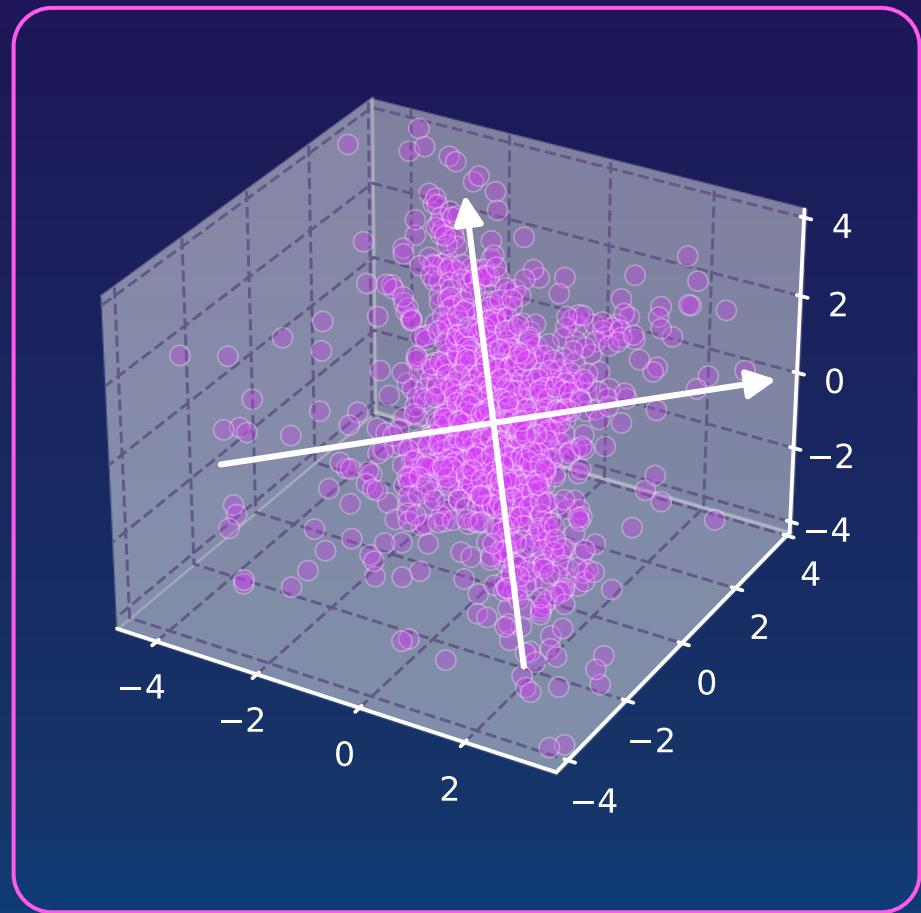
Linear

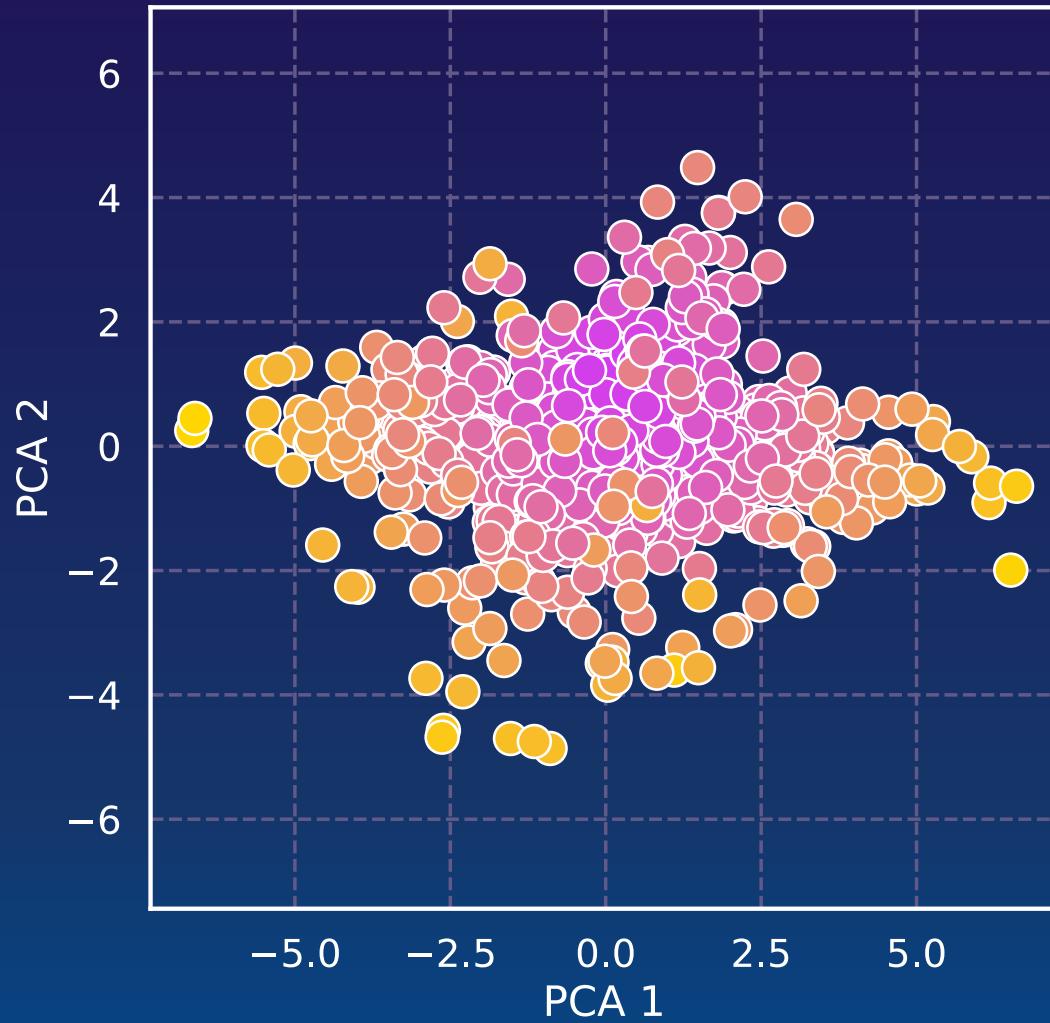
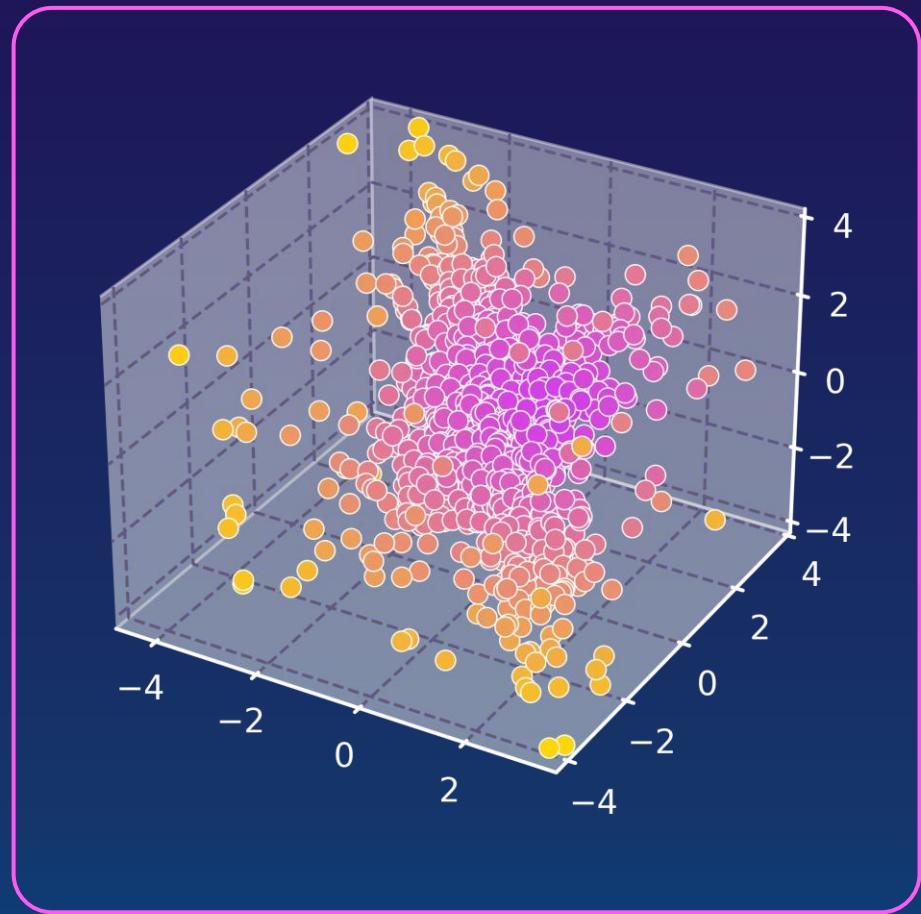
$\bar{\text{phd}}$ X pt

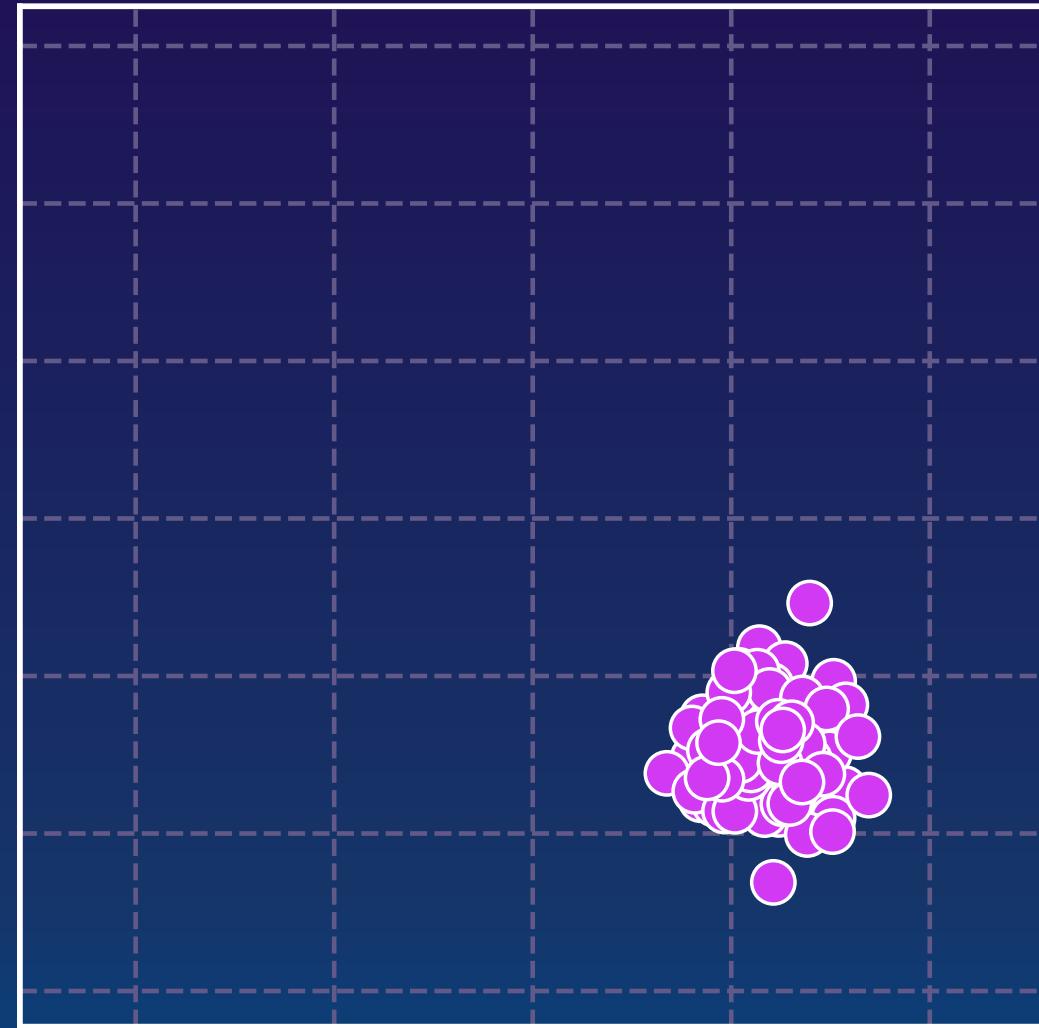






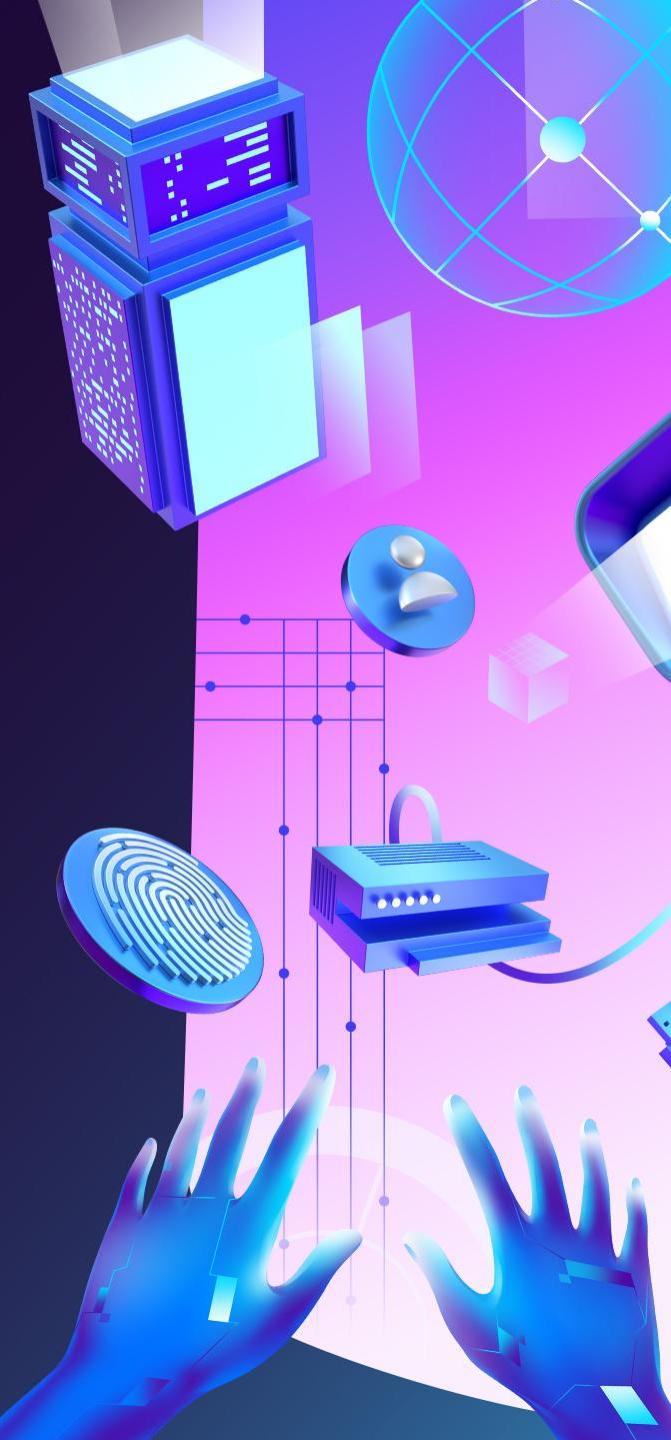




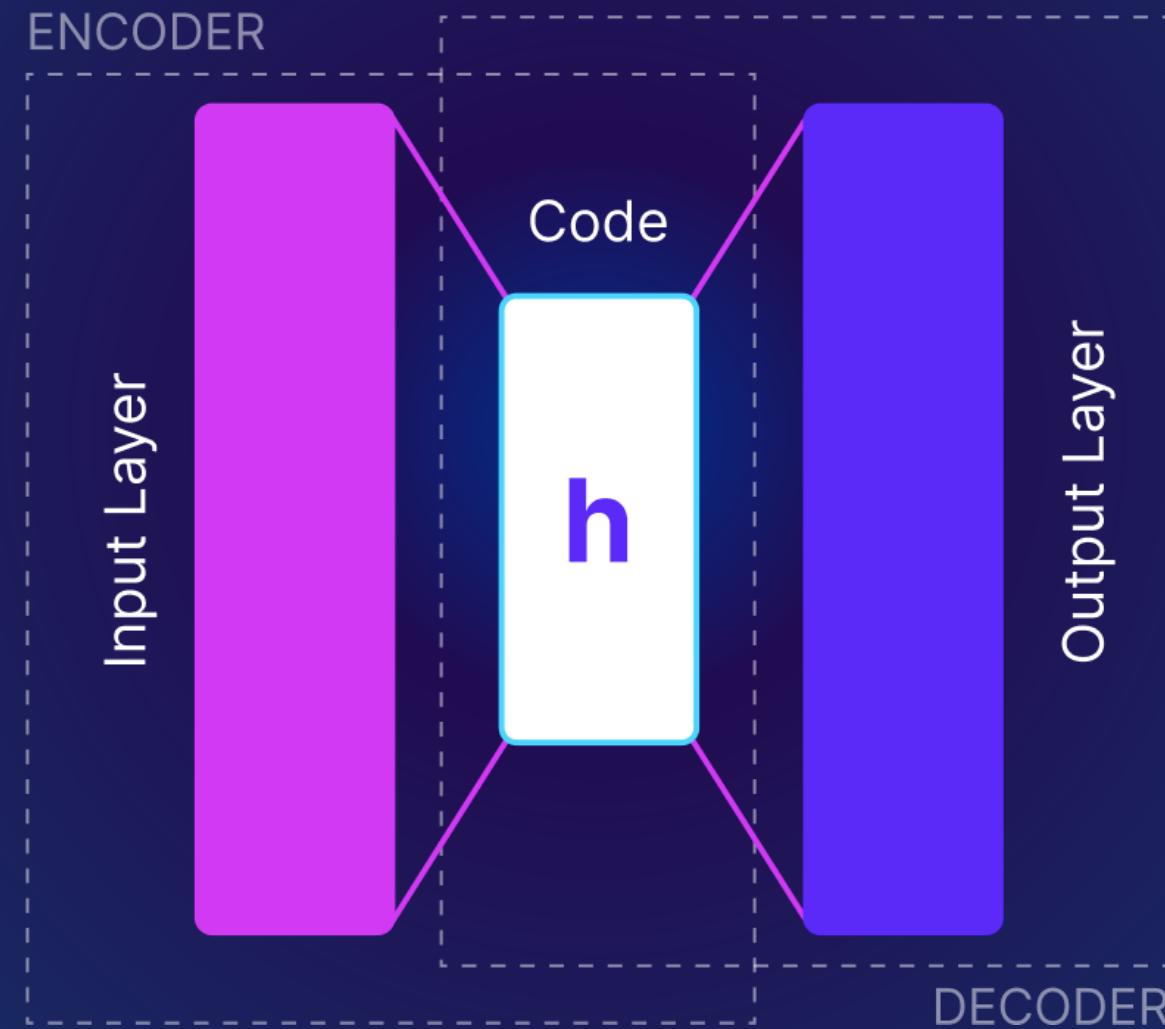


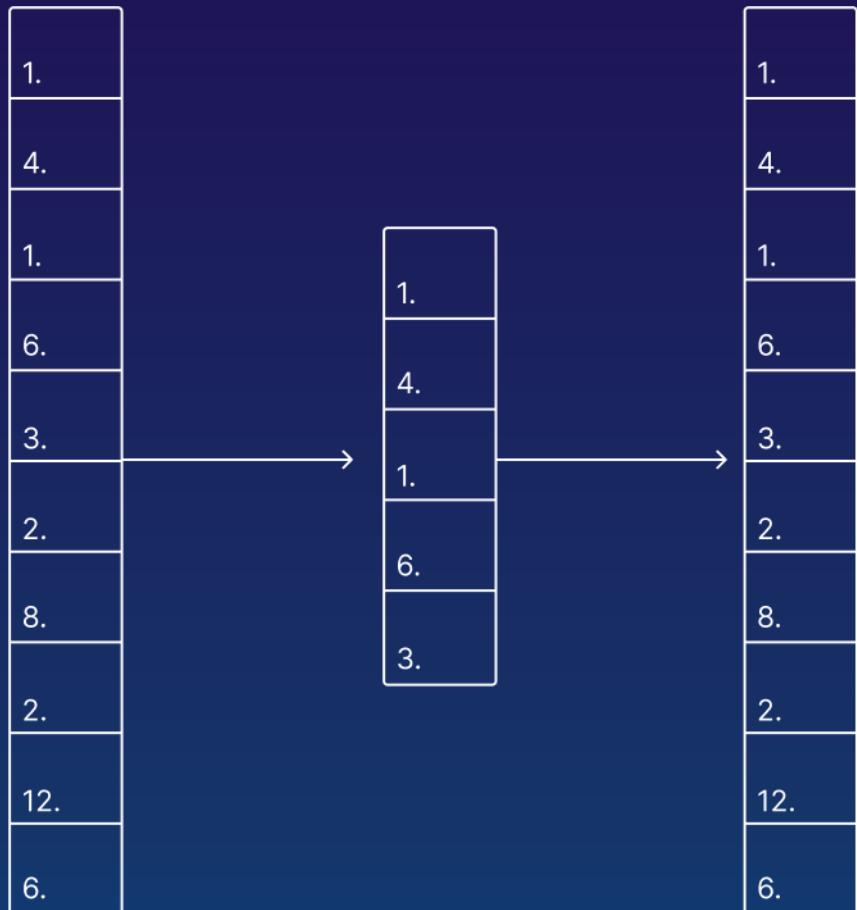
**«Все счастливые семьи
похожи друг на друга, каждая
несчастливая семья
несчастлива по-своему»**

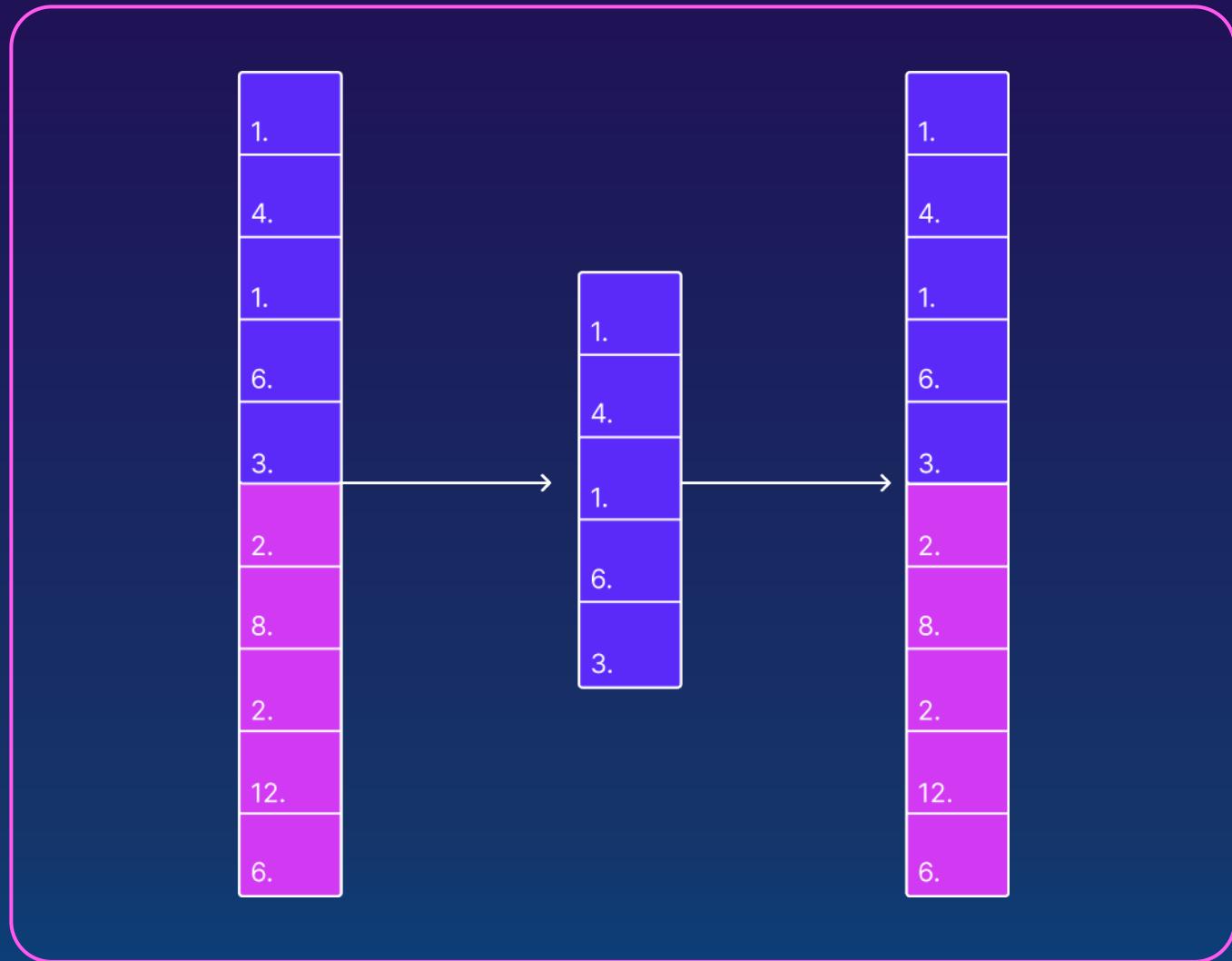
«Анна Каренина» Лев Толстой

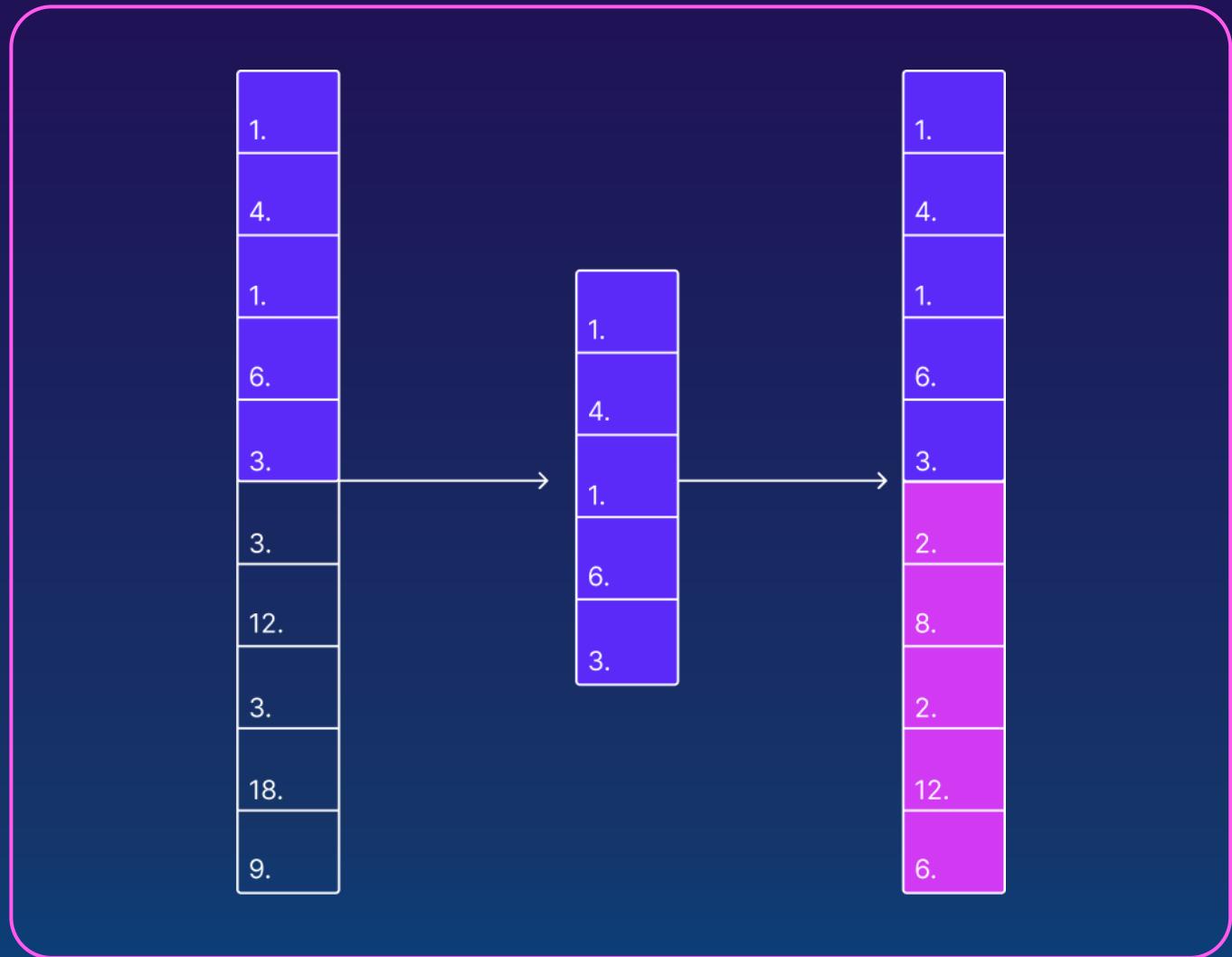


Neural Network





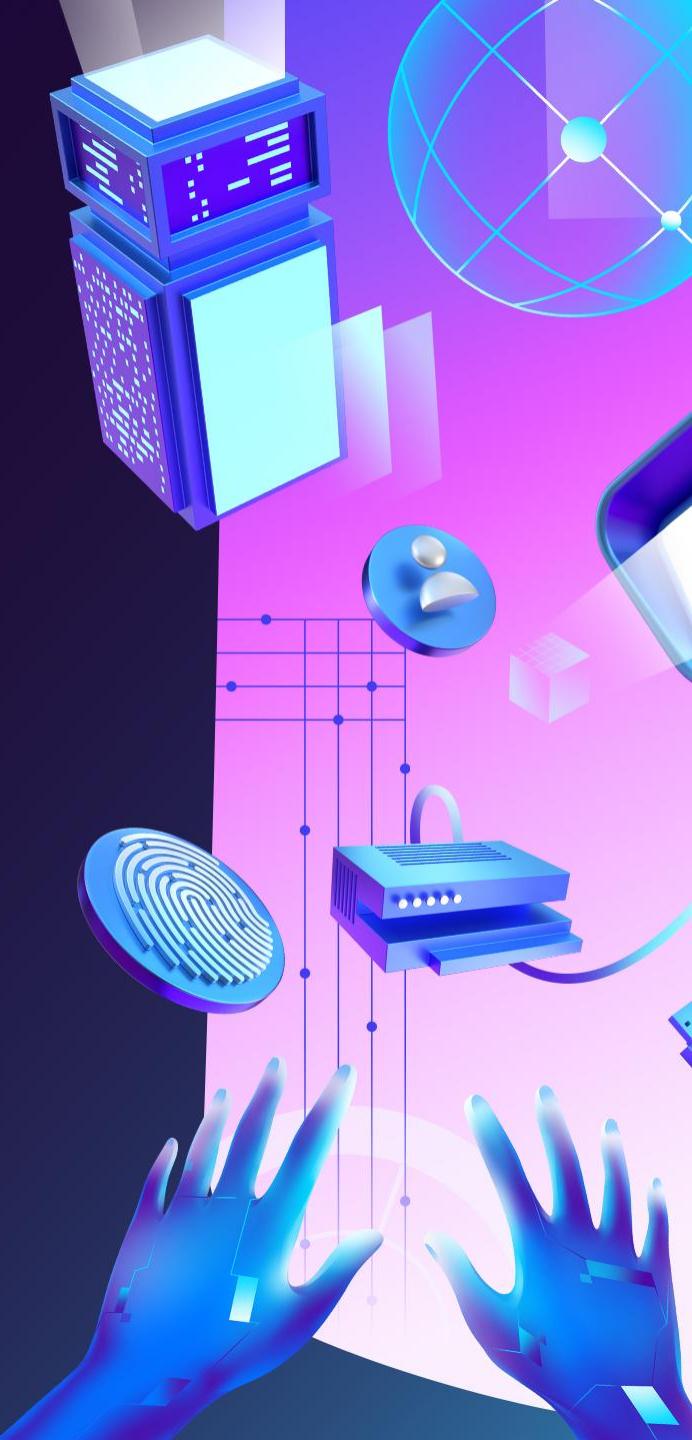




集思广益

Jí sī guǎng yì

kit. «Собирать мысли и мнения для общей пользы»

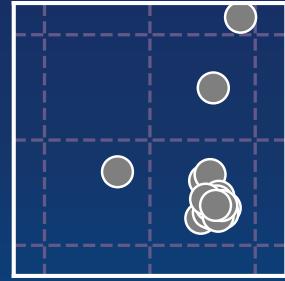
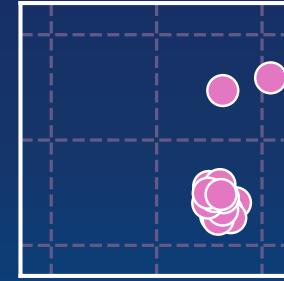
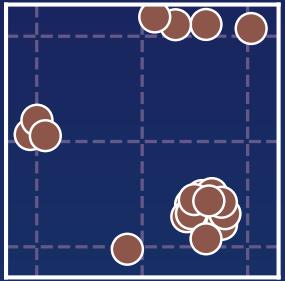
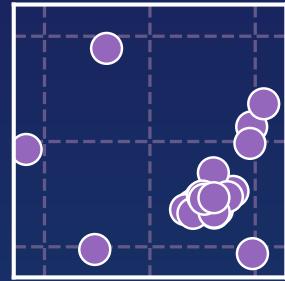
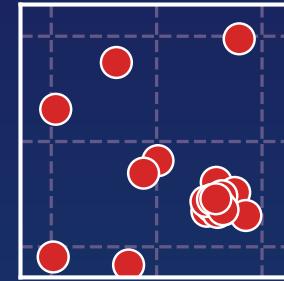
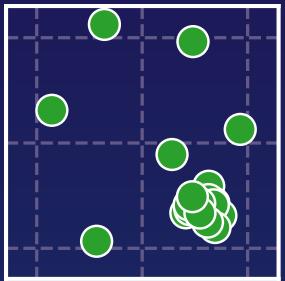
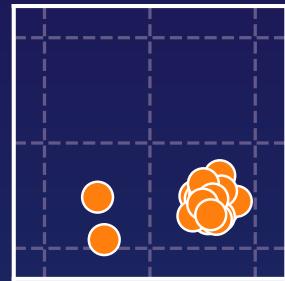
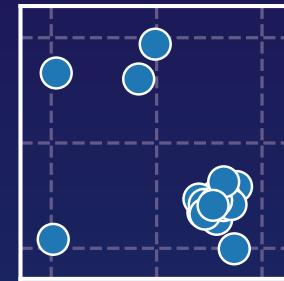
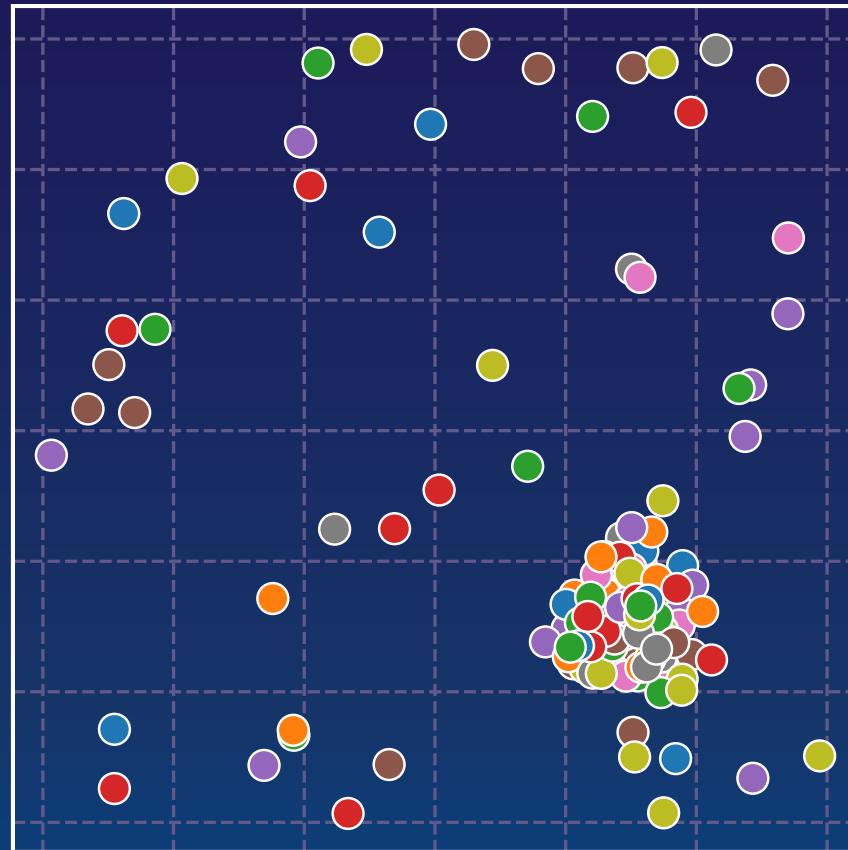


Ensembles



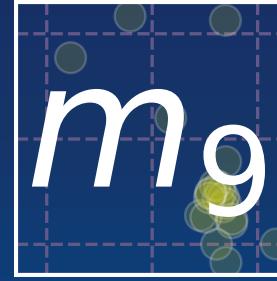
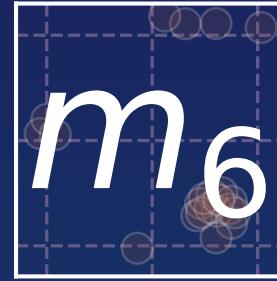
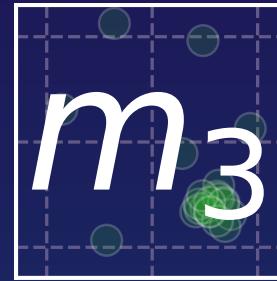
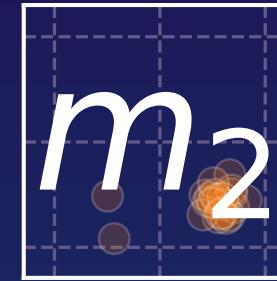
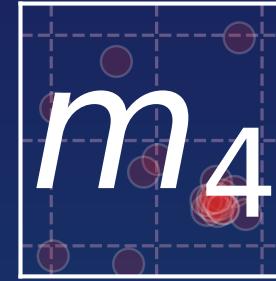
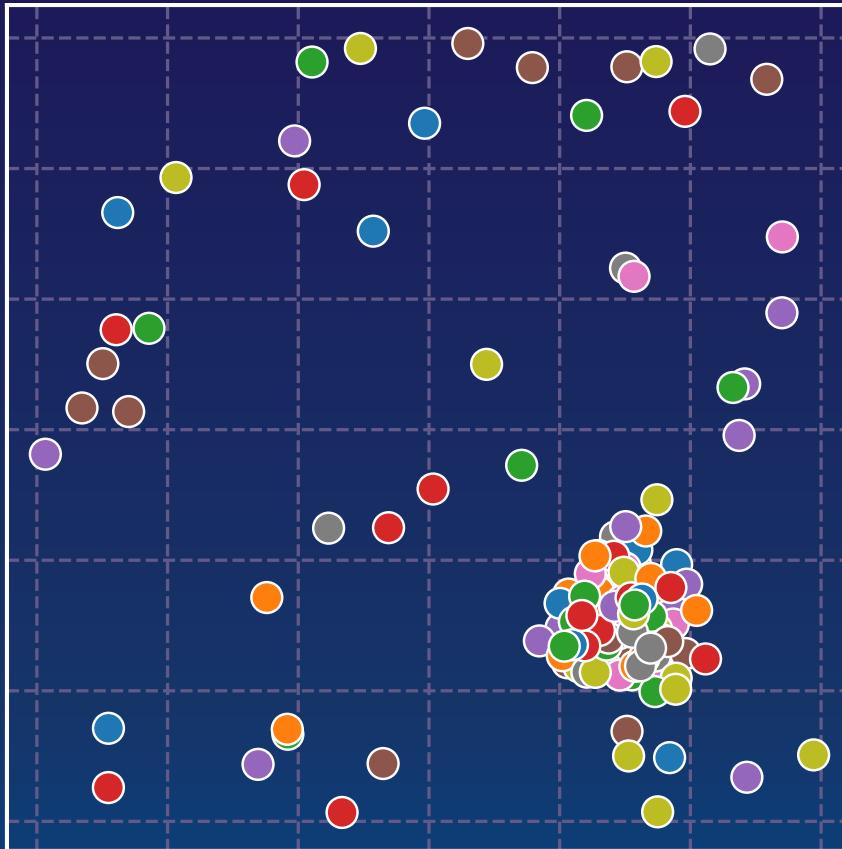
Делим

phd X pt



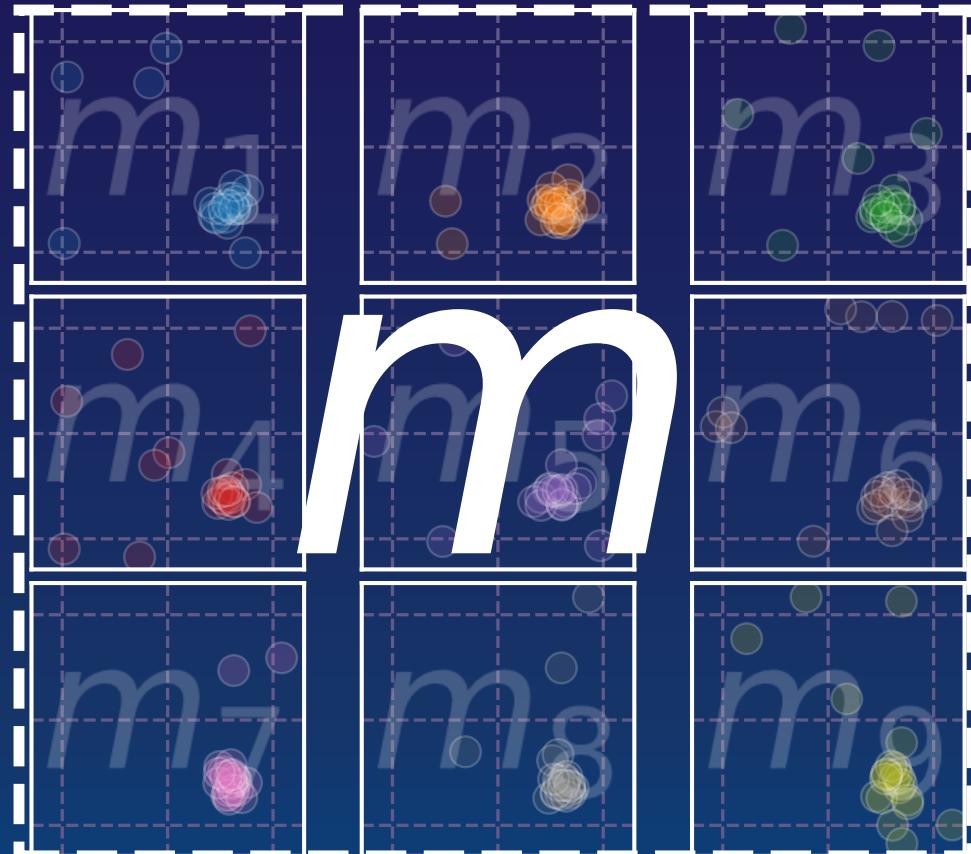
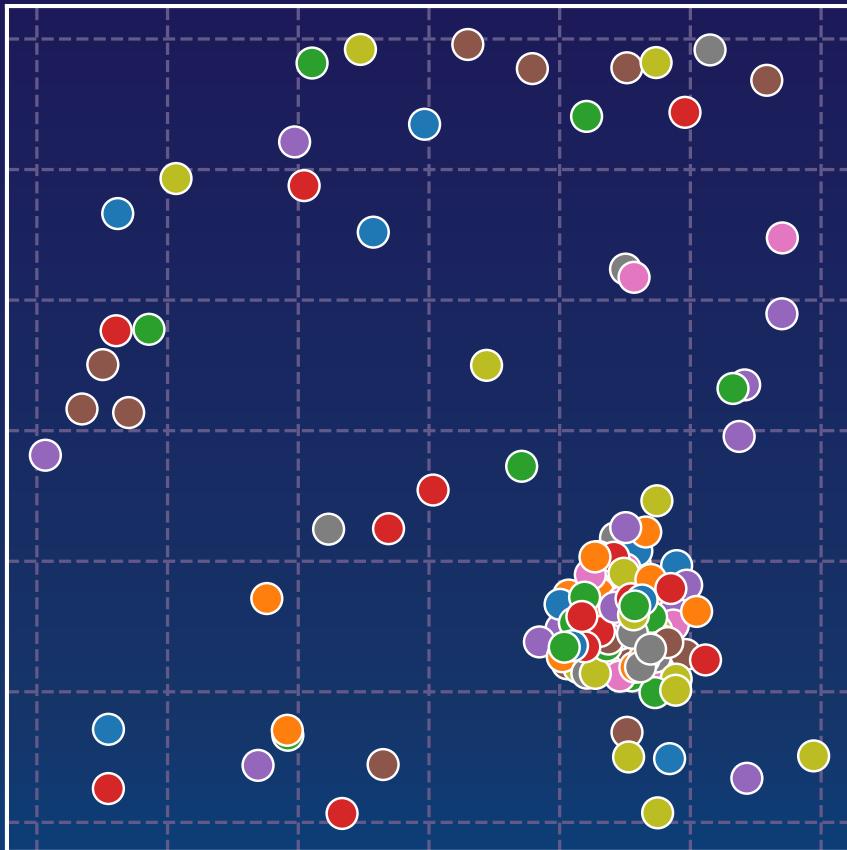
Обучаем

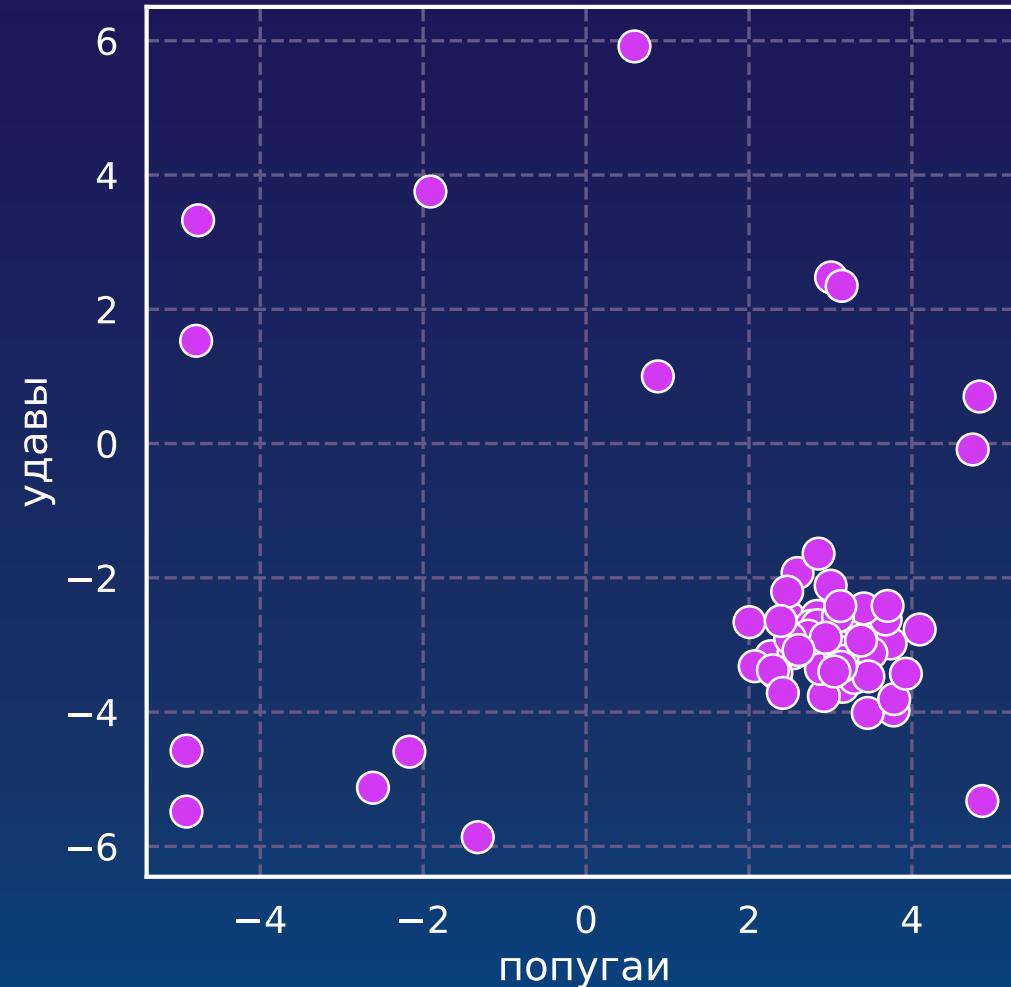
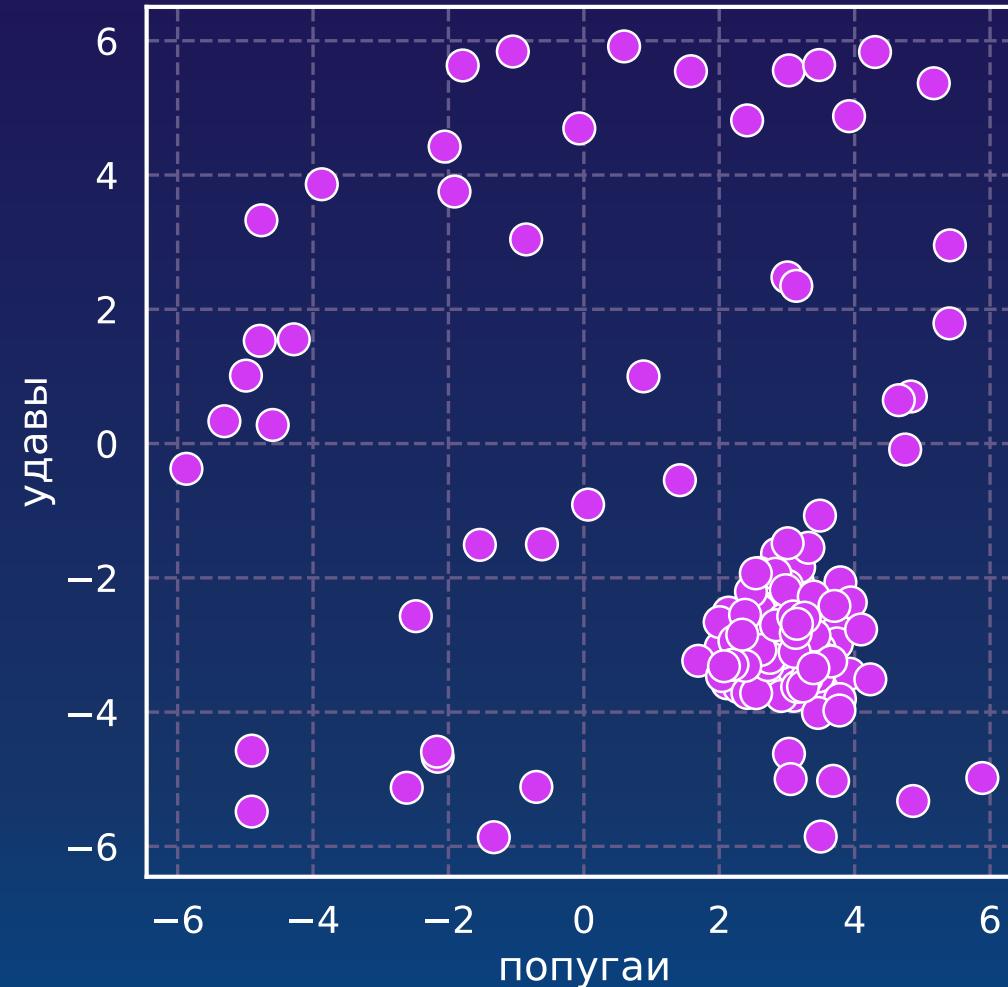
phd X pt

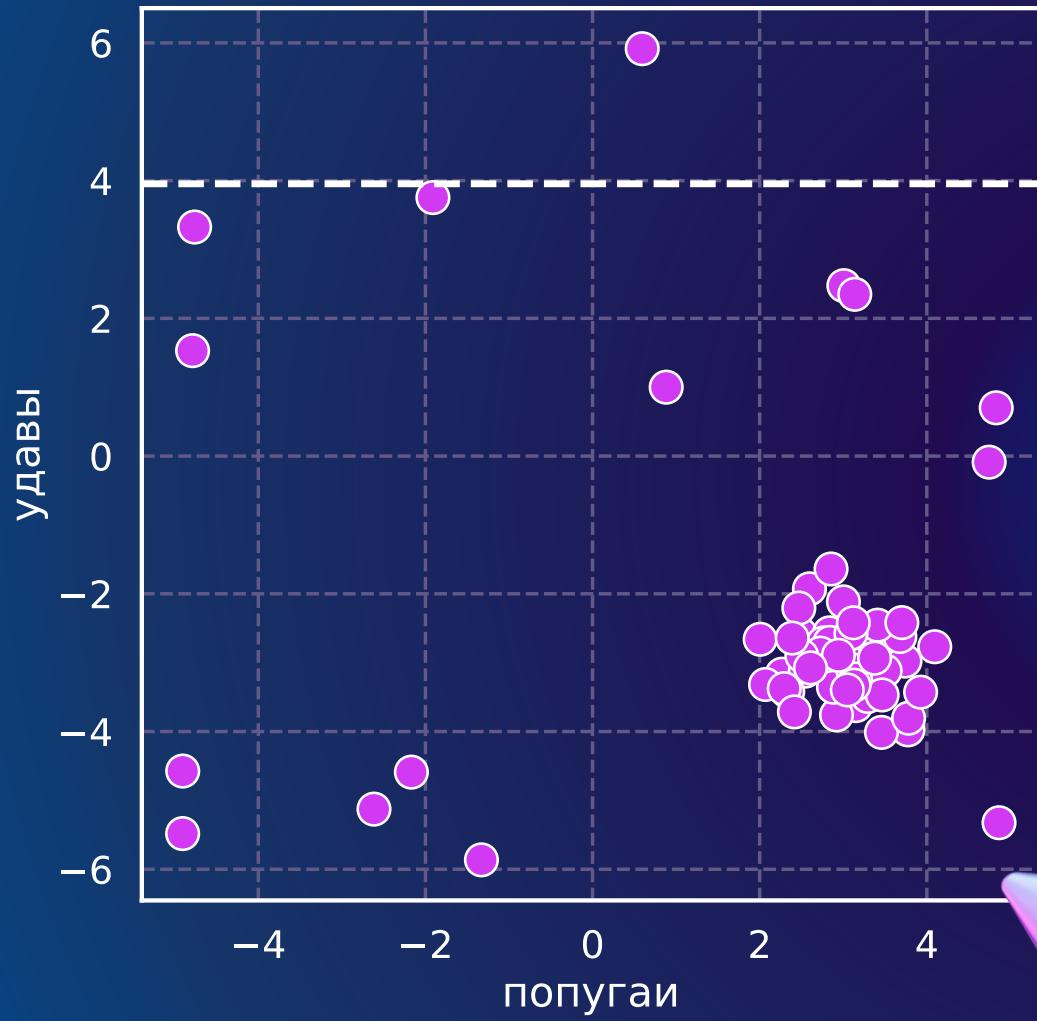


Усредняем

$\bar{p}hd$ X pt







In [1]: import random

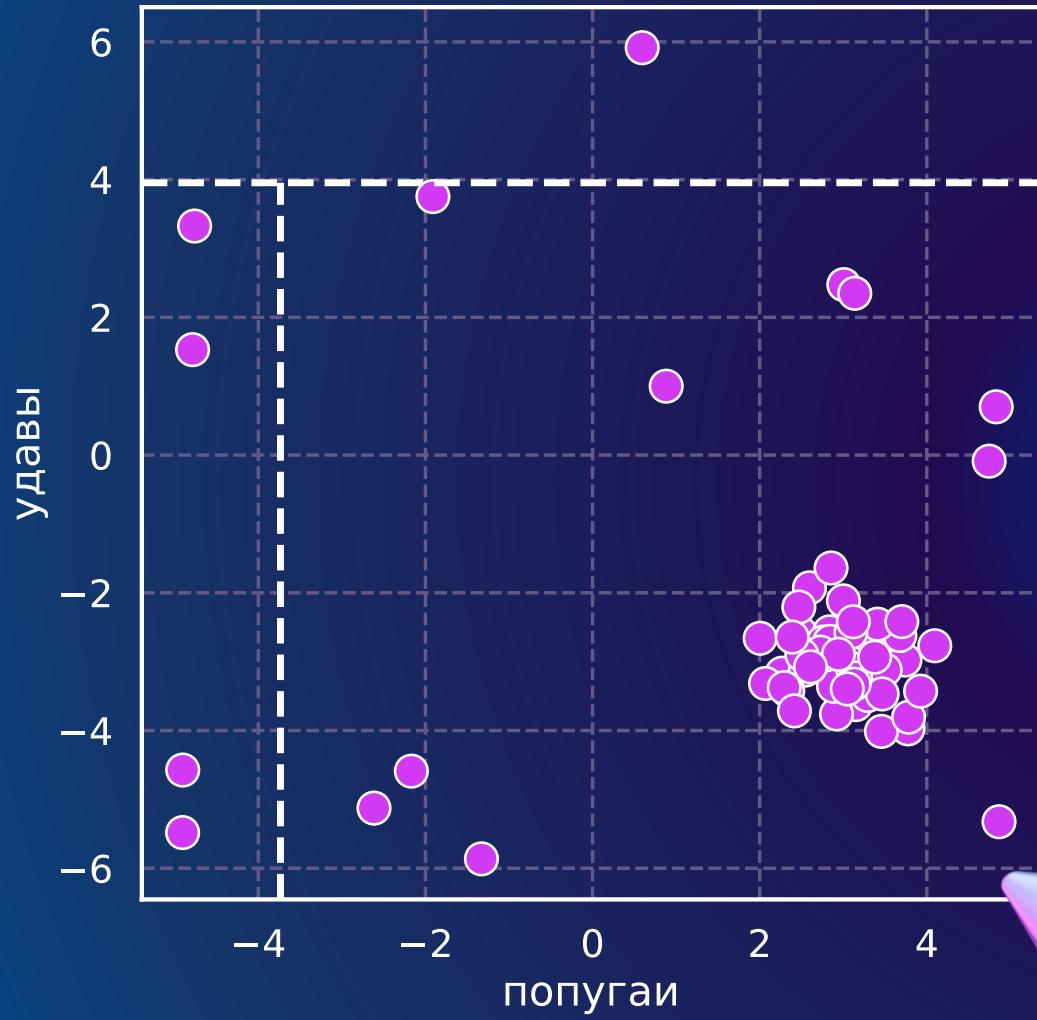
In [2]: axes = ['попугай', 'удавы']

In [3]: random.choice(axes)

Out[3]: 'удавы'

In [4]: random.uniform(-6.0, 6.0)

Out[4]: 3.954197818641566

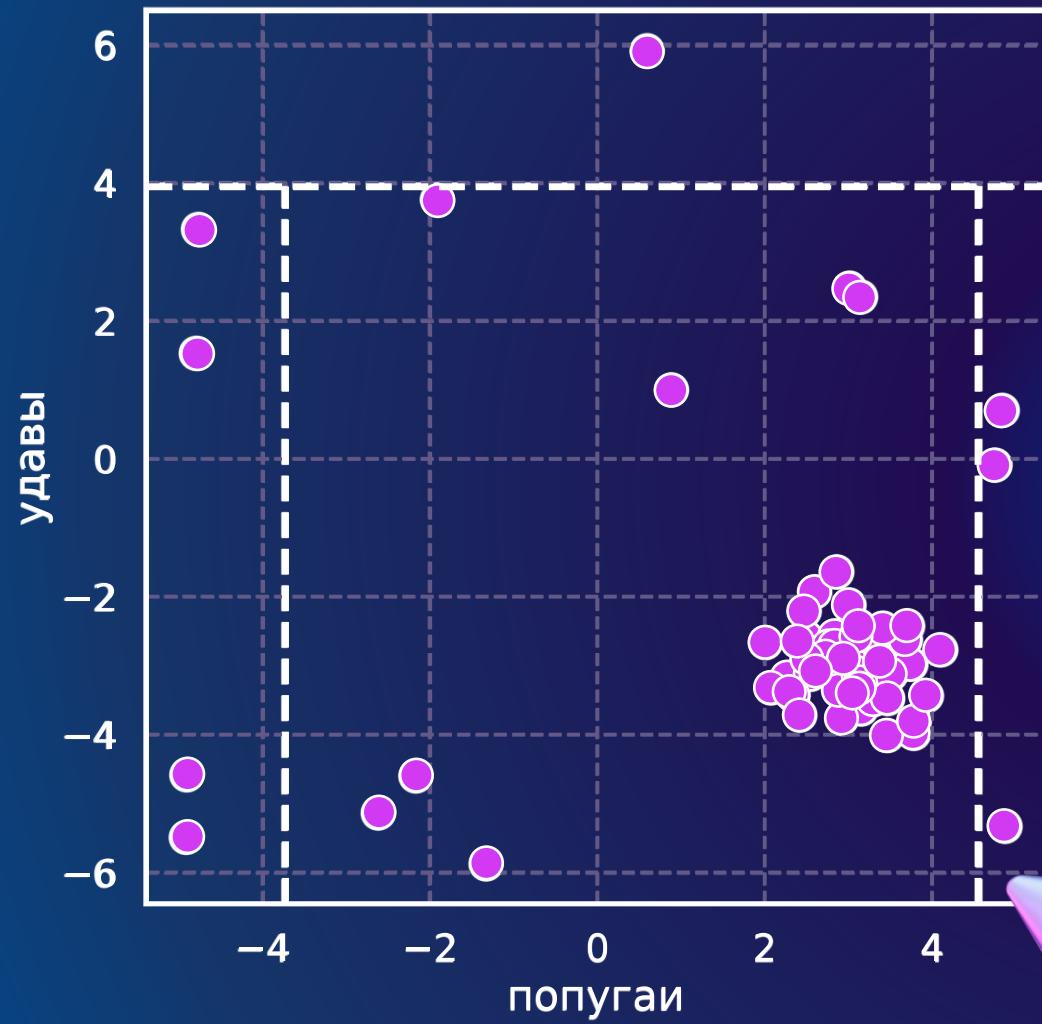


In [5]: random.choice(axes)

Out[5]: 'попугай'

In [6]: random.uniform(-6.0, 3.95)

Out[6]: -3.7345546743319455

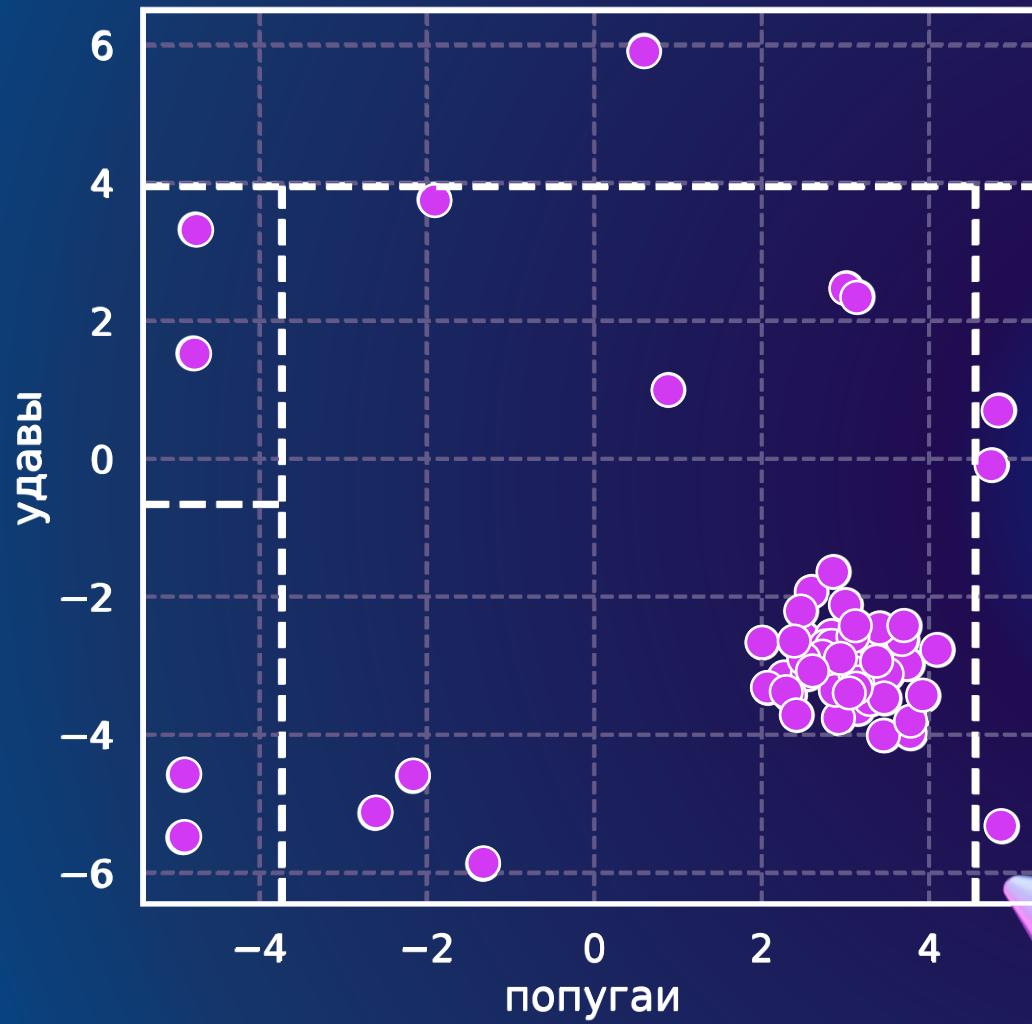


In [7]: random.choice(axes)

Out[7]: 'попугай'

In [8]: random.uniform(-3.73, 6.0)

Out[8]: 4.55352143693694

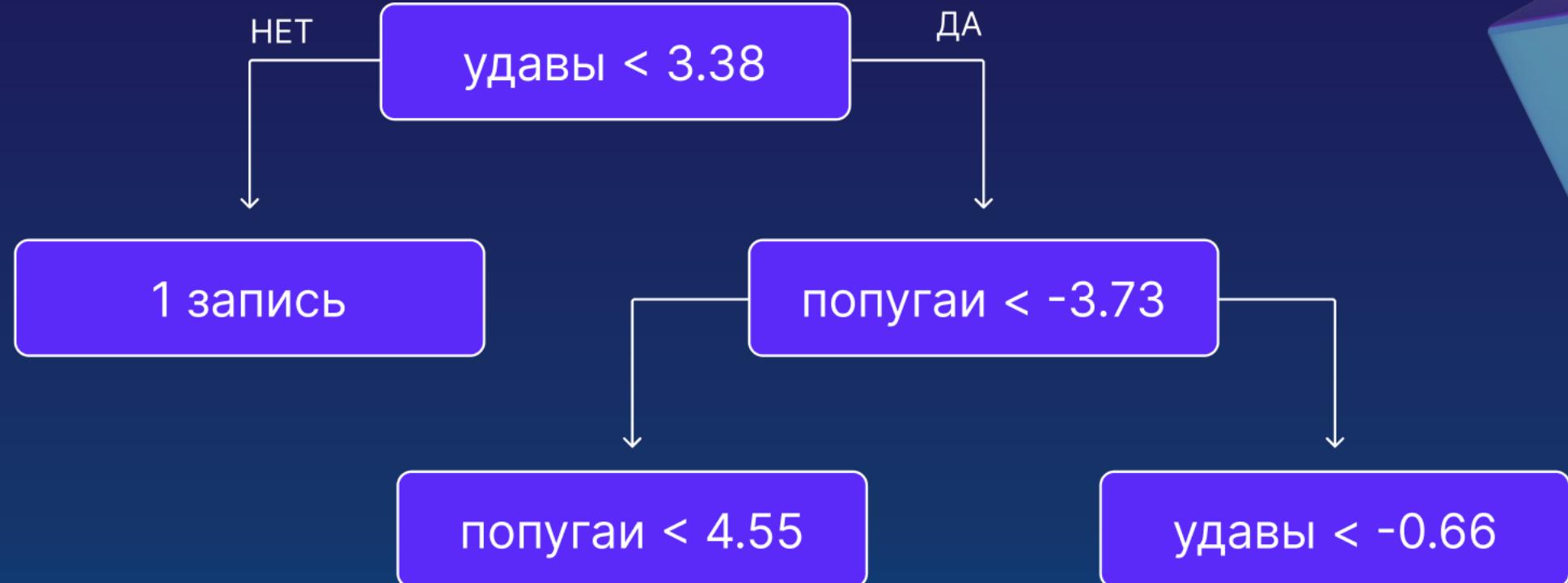


In [9]: random.choice(axes)

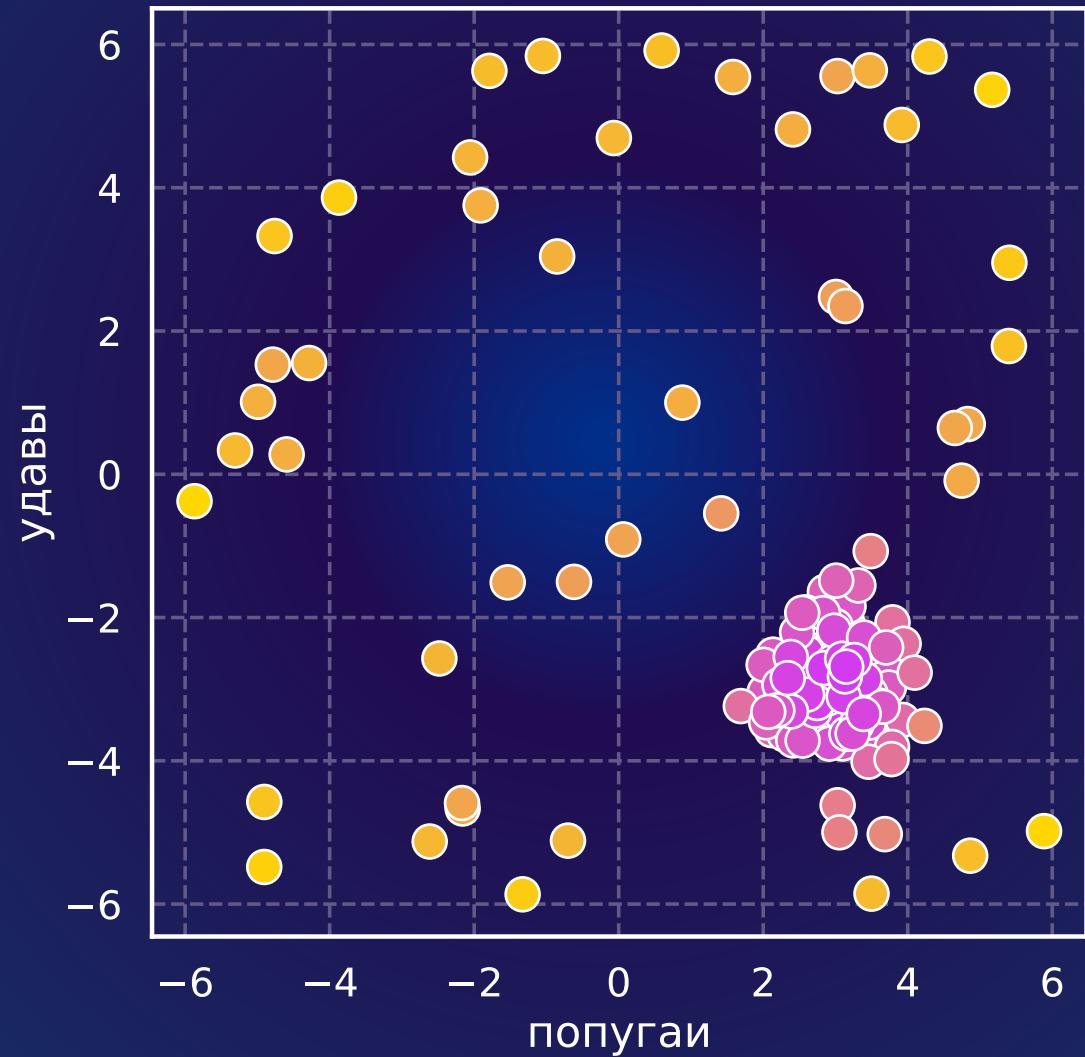
Out[9]: 'удавы'

In [10]: random.uniform(-6.00, -3.73)

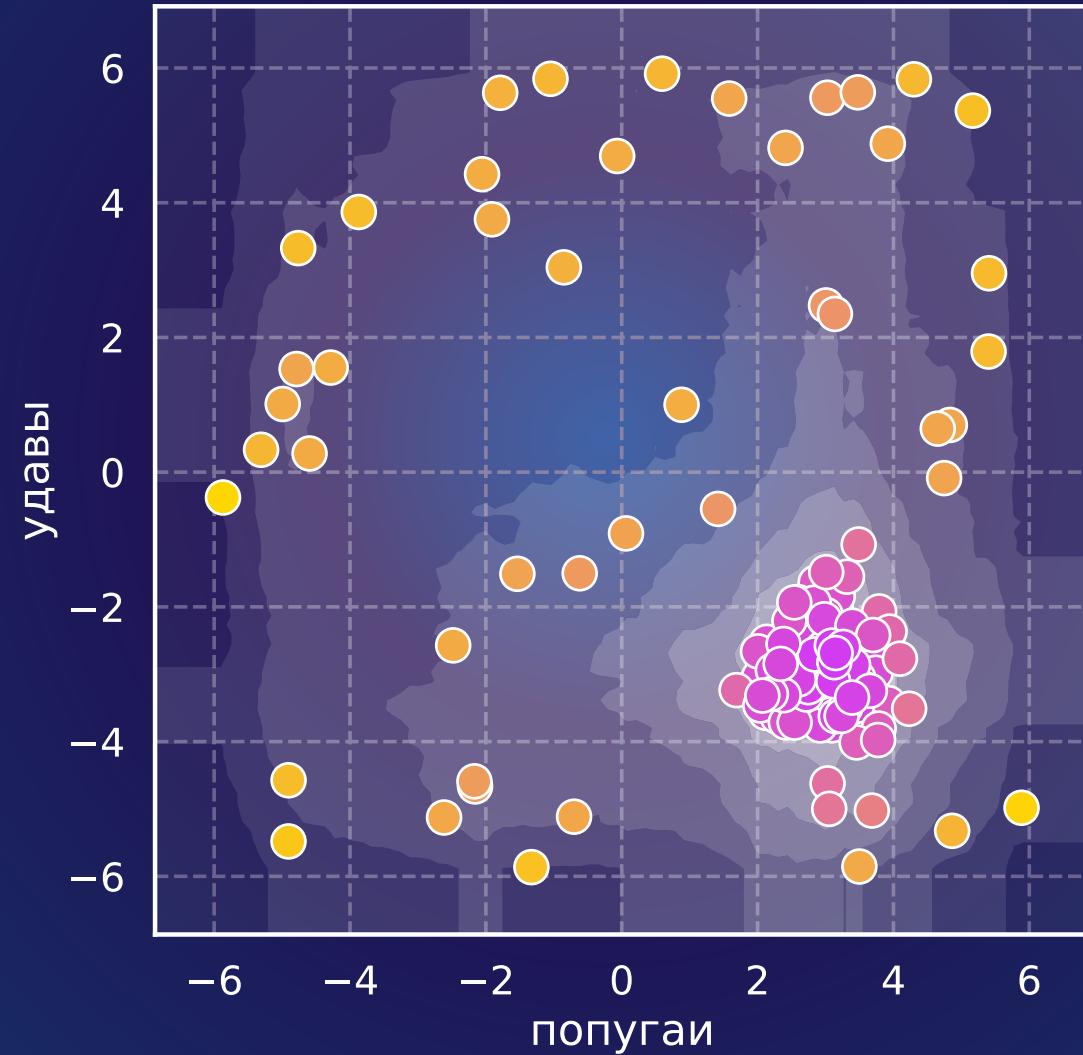
Out[10]: -0.6572094533790986



Isolation Forest



Isolation Forest





Dancu diabloj,
sed ne en mia arbaro

proverbo

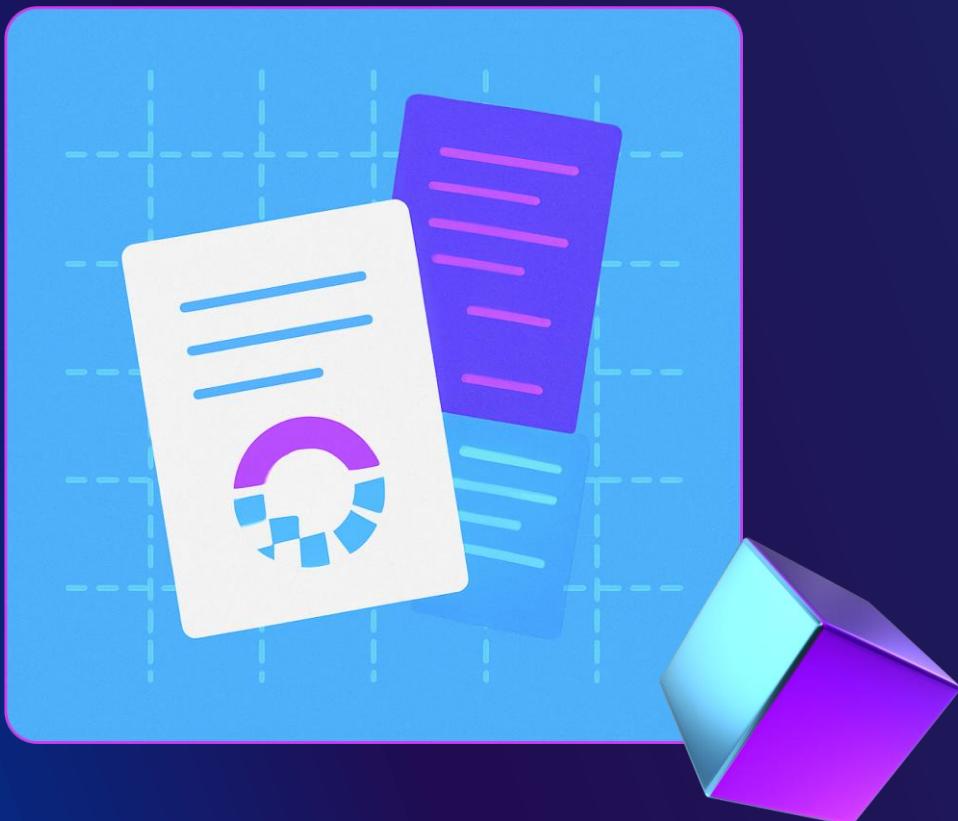
Наша задача



phd X pt

Ограничения

phd X pt



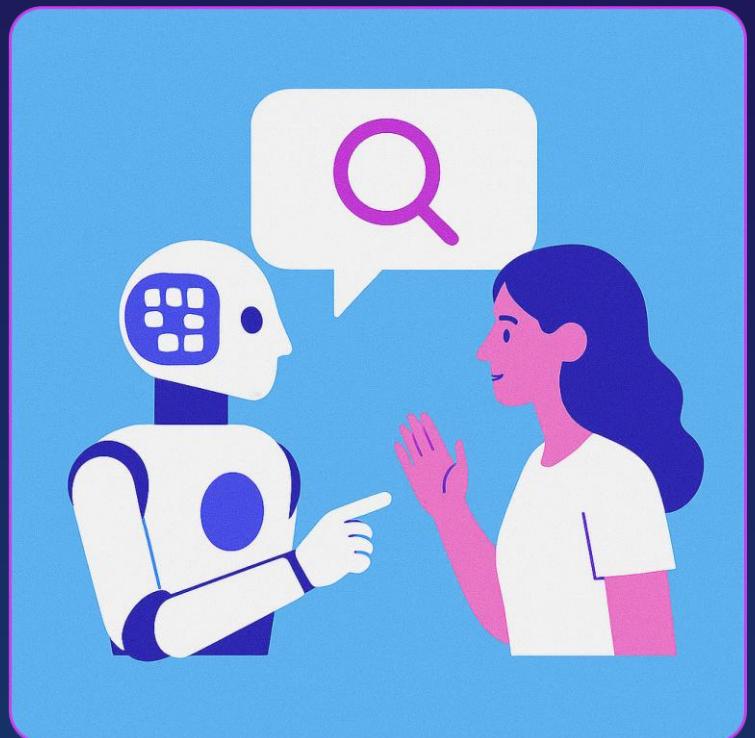
Ограничения

phd X pt



Ограничения

phd X pt





10 000 000 000

событий



3 000 000 000

событий

Применение

phd X pt





Резюме

Литература



1. *Charu C. Aggarwal.* Outlier Analysis (англ.). – Second Edition. – Springer, 2017. – ISBN 978-3-319-47578-3.
2. *Kishan G. Mehrotra, Chilukuri K. Mohan, Huaming Huang.* Anomaly Detection Principles and Algorithms (англ.). – Springer, 2017. – ISBN 978-3-319-67526-8.
3. *Jiawei Han, Micheline Kamber, Jian Pei.* Chapter 12. Outlier Detection // Data Mining Concepts and Techniques (англ.). – 3rd ed. – Morgan Kaufmann, 2012. – P. 543–584. – ISBN 978-0-12-381479-1.
4. *Chris Kuo.* Handbook of Anomaly Detection: Cutting-edge Methods and Hands-On Code Examples (англ.). – 2nd edition. – Innovation Press, 2024. – ISBN 979-8990781016.
5. *Brett Kennedy.* Outlier Detection in Python (англ.). – Manning Publications, 2025. – ISBN 978-1633436473.

Вопросы?

