
Занятие № 10

Поиск выбросов и
генерация новых
признаков



Содержание

- 1 Что такое выброс
- 2 Как их отлавливать?
- 3 Практика.



Что такое выброс?

Выбросы - точки данных, которые не принадлежат определенной популяции

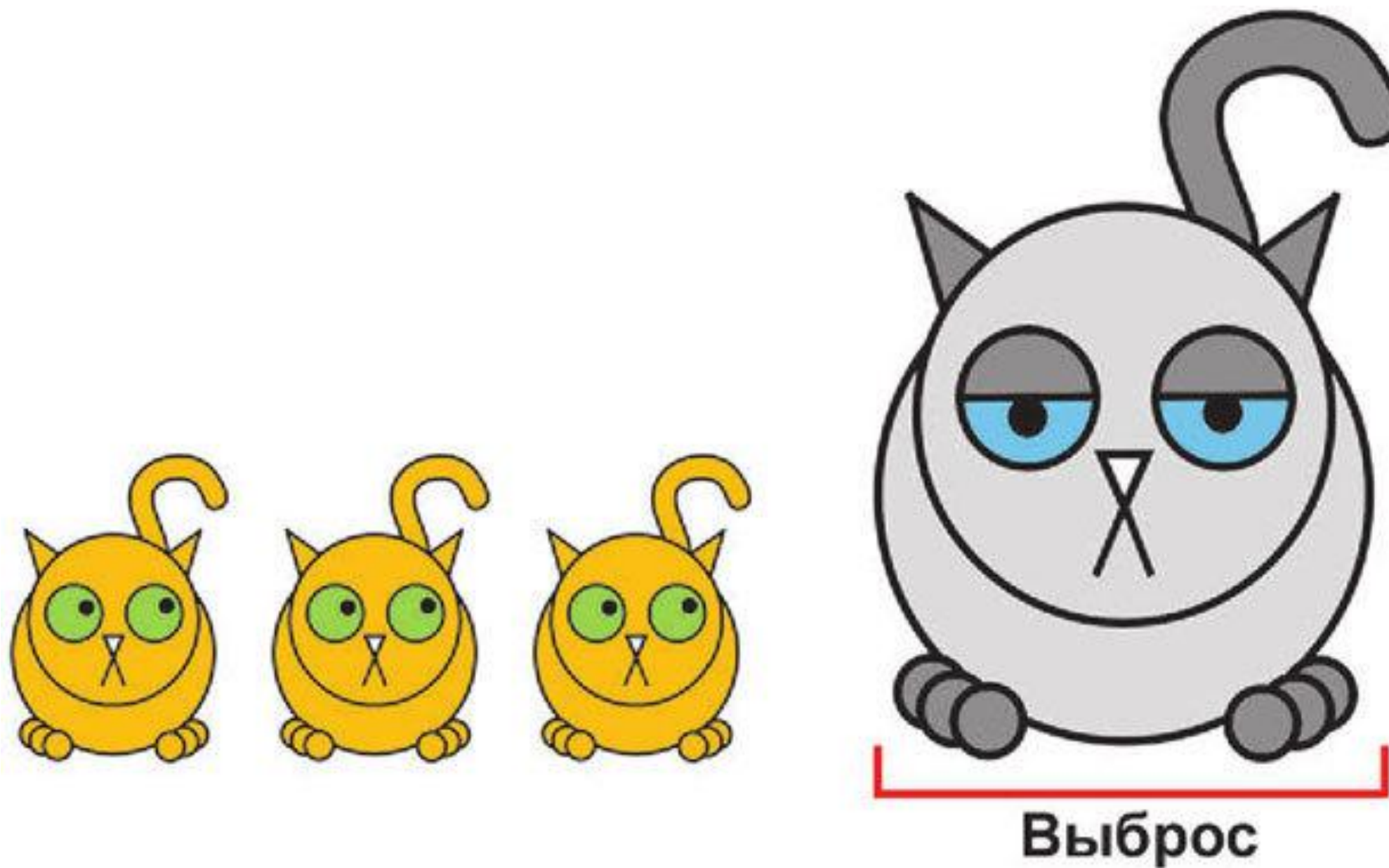
Причины:

- ошибок в данных
- наличия шумовых объектов
- присутствия объектов «других» выборок

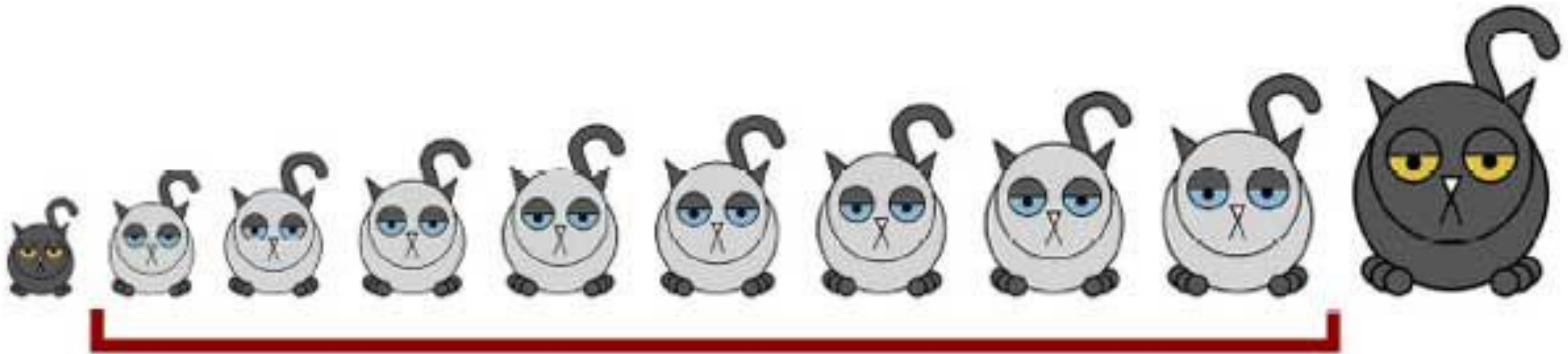


Что такое выброс?

Результат измерения, не подпадающий под общее распределение



Что такое выброс?



Котики для усеченного среднего



Что делать?



Как их отлавливать?

- **Статистические тесты**

 - Стандартное отклонение

 - Боксплот

- **Модельные тесты**

 - построение модель, которая описывает данные (выбросы плохо описываются моделью)

- **Итерационные методы**

 - построение оболочек в n-пространстве

- **Метрические методы**

 - KNN

 - DBScan

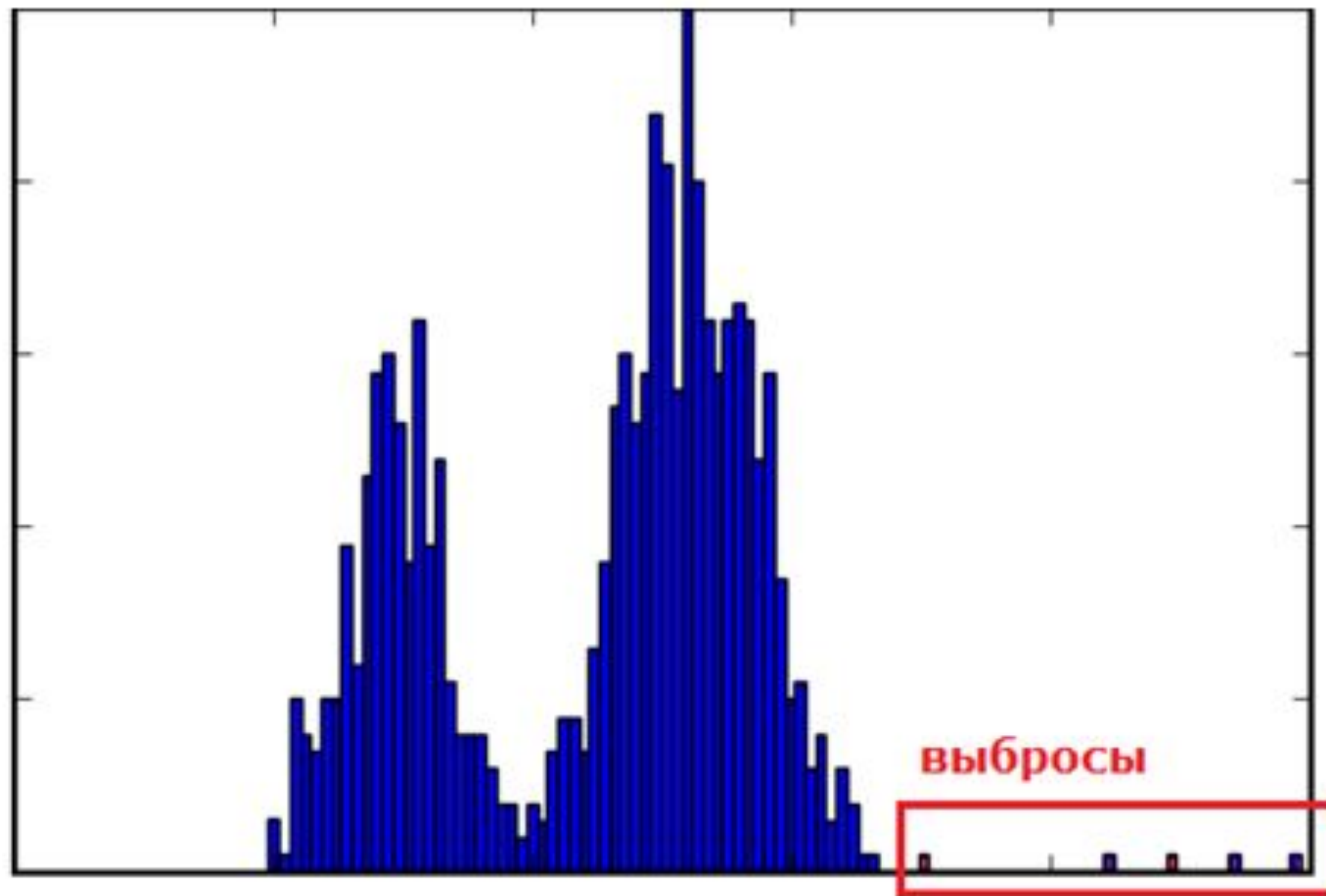
- **Методы машинного обучения**

 - Метод опорных векторов для одного класса

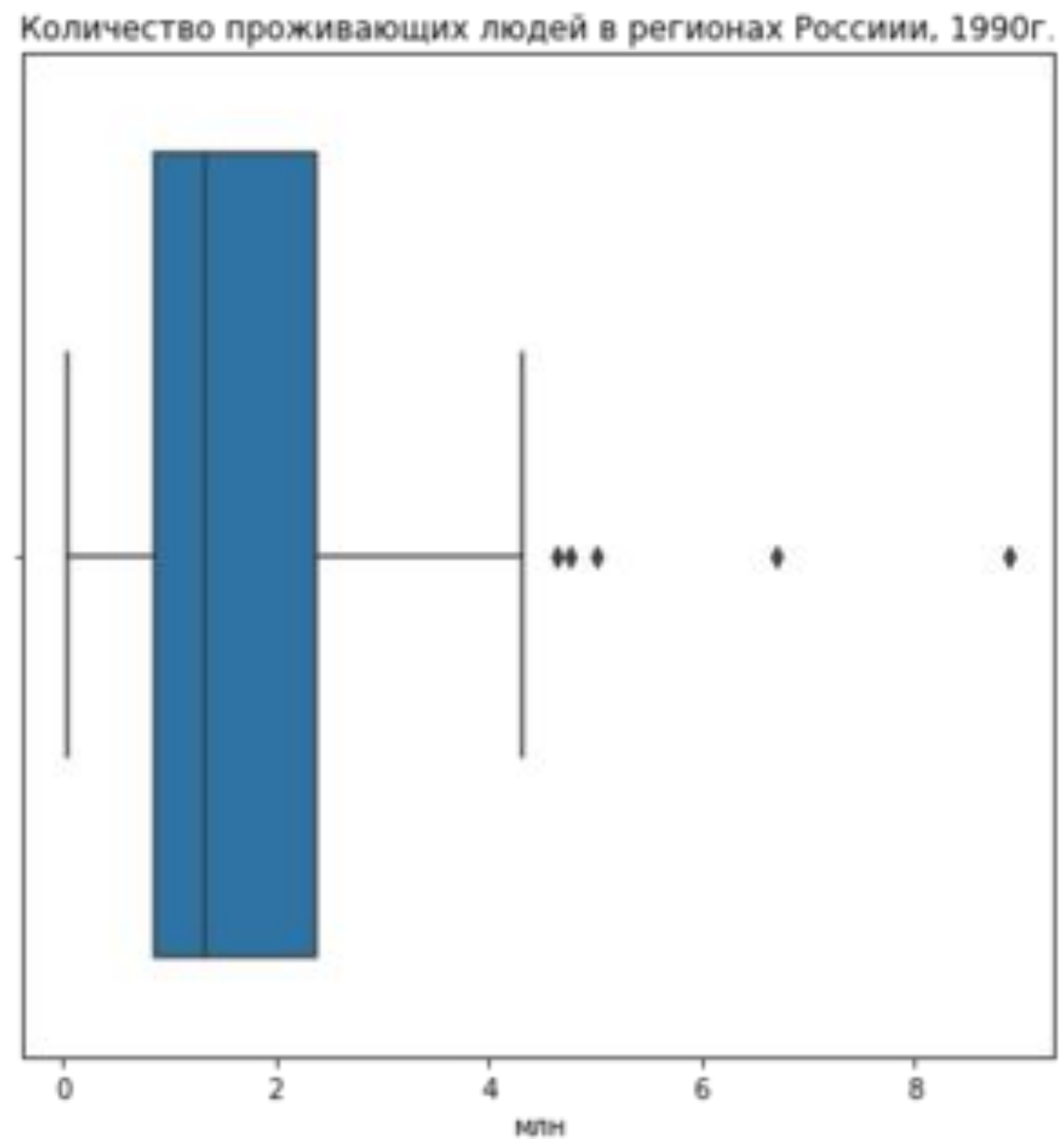
 - Изолирующий лес



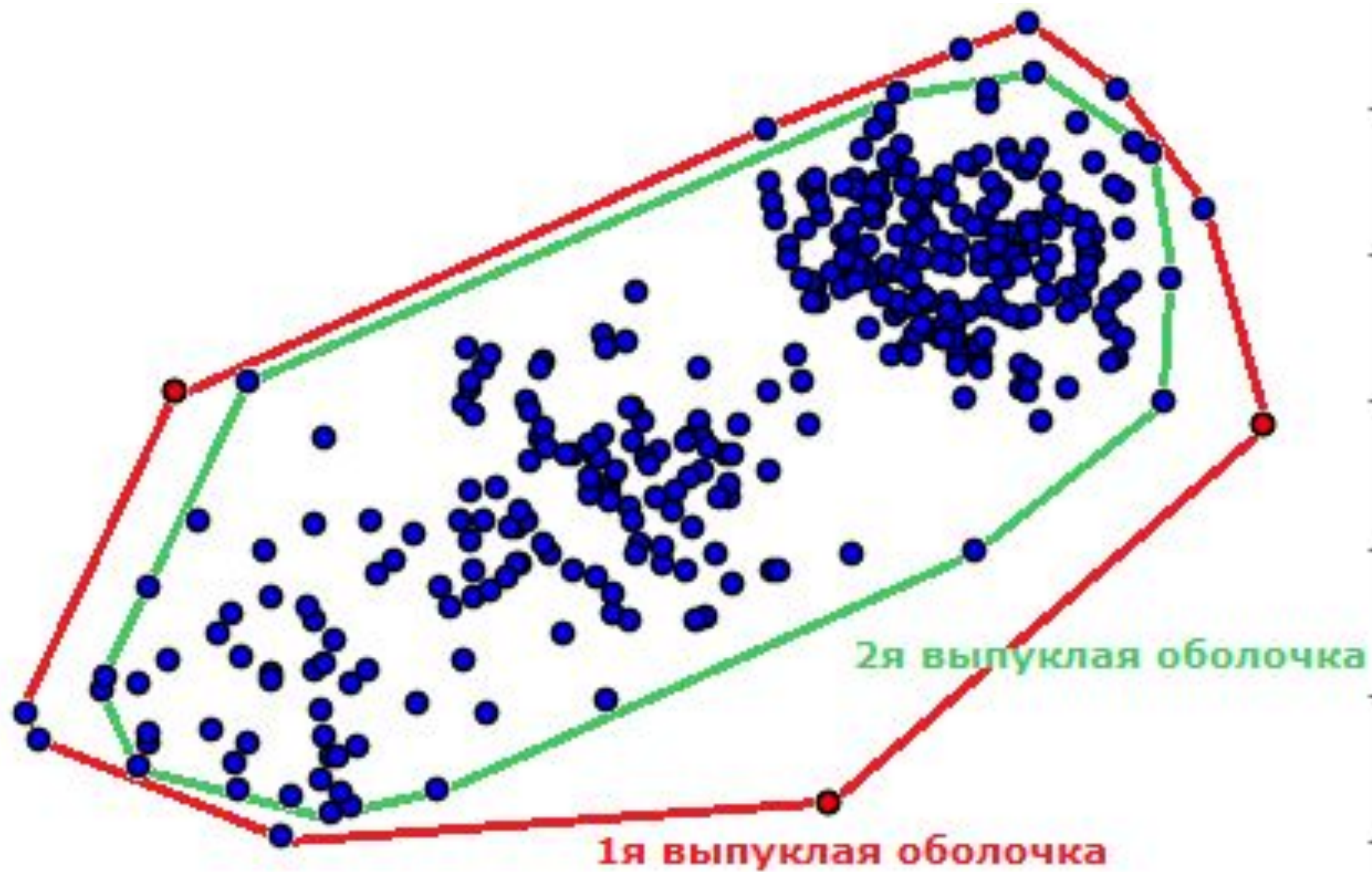
Как их отлавливать?



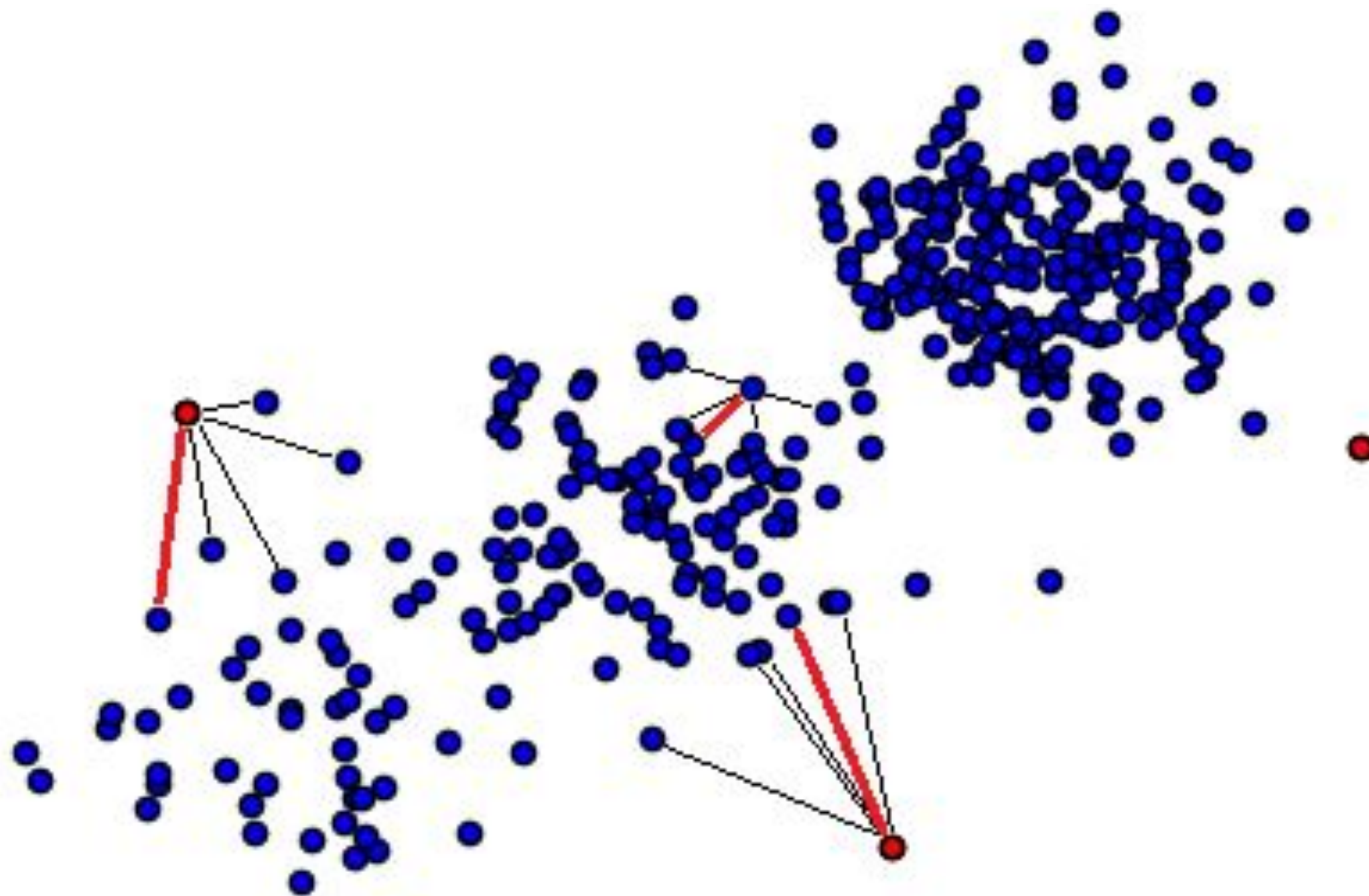
Как их отлавливать?



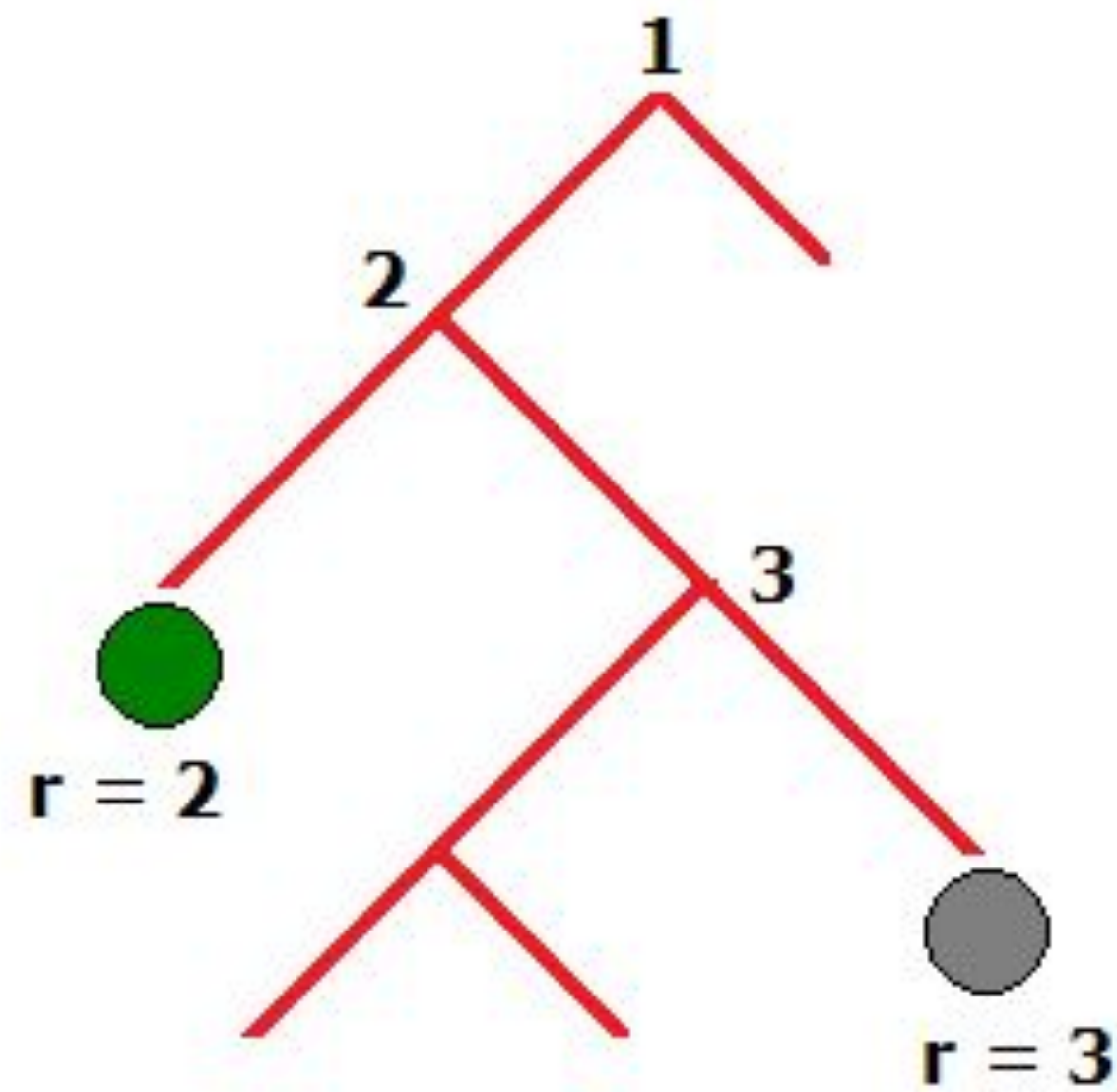
Как их отлавливать?



Как их отлавливать?



Как их отлавливать?



ПРАКТИКА



Спасибо за
внимание!

