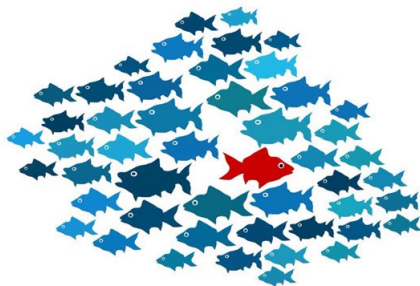


# Обнаружение аномалий

Виктор Китов

[v.v.kitov@yandex.ru](mailto:v.v.kitov@yandex.ru)



## Аномалии (выбросы)

- Аномалия (выброс, outlier) - объект, нетипичный для общего распределения объектов.
- Применения обнаружения аномалий (anomaly detection)
  - очистка данных (убрать ошибочные наблюдения)
  - обнаружение нетипичных объектов:
    - мошеннические транзакции в финансах
    - взлом компьютерной сети
    - мониторинг исправности устройств (станок, вертолет, ядерный реактор)
  - детектирование сдвига модели (concept drift)

## Если есть разметка

- Если выбросы размечены в train, то это imbalanced class classification<sup>1</sup>.
- Пусть  $y = +1$  - редкий класс:  $|n : y_n = +1| \ll |n : y_n = -1|$
- Как решать?

---

<sup>1</sup>Библиотека Python для несбалансированных классов.

## Если есть разметка

- Если выбросы размечены в train, то это imbalanced class classification<sup>1</sup>.
- Пусть  $y = +1$  - редкий класс:  $|n : y_n = +1| \ll |n : y_n = -1|$
- Как решать?
  - дублировать выбросы в выборке
  - обобщение: взвешенная ф-ция потерь ( $w > 1$ )

$$w \sum_{n: y_n = +1} \mathcal{L}(f_{\theta}(x_n), y_n) + \sum_{n: y_n = -1} \mathcal{L}(f_{\theta}(x_n), y_n) \rightarrow \min_{\theta}$$

- исключить часть объектов класса -1
  - Алгоритм NearMiss: оставляем объекты -1 класса, ближайшие к объектам +1 класса
- генерация синтетических объектов для выбросов
- аугментация

---

<sup>1</sup> Библиотека Python для несбалансированных классов.

## Генерация синтетических объектов<sup>2</sup>

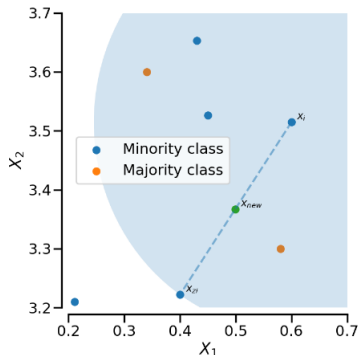
Метод SMOTE генерирует синтетич. объекты класса +1.

- для каждого объекта  $x_n$  с  $y_n = +1$ 
  - ① найдем  $K$  ближ. соседей  $KNN(x_n)$
  - ②  $P$  раз выберем случайные объекты из  $KNN(x_n)$

$$A(x_n) = \{x_{i_1}, \dots, x_{i_K}\}$$

- ③ для каждого  $x' \in A(x_n)$  сгенерируем новый объект класса  $y = +1$

$$x = (1 - \alpha)x_n + \alpha x', \quad \alpha \sim U[0, 1]$$



<sup>2</sup>Как обобщить на категориальные признаки?

## Расширение обучающей выборки

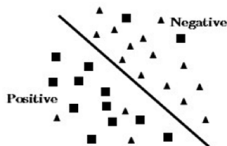
- Расширение обучающей выборки (data augmentation): модификации  $x$ , генерирующие реальные объекты того же класса.
- Как можно расширять выборку для
  - изображений
  - звуков
  - текстов

## Расширение обучающей выборки

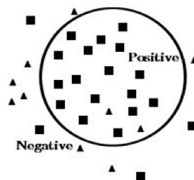
- Расширение обучающей выборки (data augmentation): модификации  $x$ , генерирующие реальные объекты того же класса.
- Как можно расширять выборку для
  - изображений
    - $\uparrow\downarrow$  яркости, контраста
    - сдвиг / поворот с обрезкой
  - звуков
    - $\uparrow\downarrow$  скорости
    - + помехи
    - $\uparrow\downarrow$  тембра (частоты)
  - текстов
    - замена слов синонимами
    - перевод на др. язык и обратно
    - суммаризация (прореживание предложений)

# Методы обнаружения аномалий

- Обнаружение аномалий (anomaly detection) - обучение без учителя
  - нет разметки выбросов в train
  - но может быть в validation (для оценки)
- Несбалансированная классификация: есть примеры аномалий (есть паттерн)
  - выделяем область каждого класса
- Обнаружение аномалий: сложность в новизне (нет паттерна аномалии)
  - напр. детекция мошеннических действий (всё время новые)
  - выделяем область нормальности, остальное - выбросы



b. Classification



a. Anomaly detection



# Методы обнаружения аномалий

- Методы оценивают степень нетипичности:

$$x - \text{выброс} \iff f(x) > \textit{threshold}$$

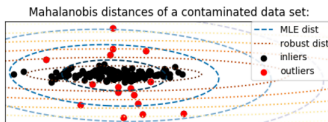
- детекция выбросов (outlier detection): обучающая выборка содержит аномалии.
- детекция новизны (novelty detection): обучающая выборка не содержит аномалий.
  - выше пороги, чем в outlier detection
- Подходы:
  - статистический:  $p(x) < t$
  - метрический: выброс далеко от др. точек
  - модельный: моделируем область нормальности

# Содержание

- 1 Статистические методы
- 2 Метрические методы
- 3 Модельные методы
- 4 Сдвиг модели

## Статистическое обнаружение аномалий

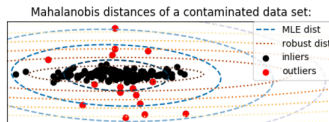
- Выбросы - точки с  $p(x) < threshold$ .
- Можем оценить  $p(x)$  параметрически, например  $\mathcal{N}(x|\mu, \Sigma) \propto e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$
- ❶ Оценим  $\hat{\mu}, \hat{\Sigma}$
- ❷  $outlierness(x) = 1/p_{\hat{\mu}, \hat{\Sigma}}(x)$



- В чем потенциальная проблема?

# Статистическое обнаружение аномалий

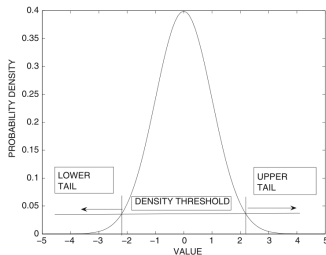
- Выбросы - точки с  $p(x) < threshold$ .
- Можем оценить  $p(x)$  параметрически, например  $\mathcal{N}(x|\mu, \Sigma) \propto e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$
- ❶ Оценим  $\hat{\mu}, \hat{\Sigma}$
- ❷  $outlierness(x) = 1/p_{\hat{\mu}, \hat{\Sigma}}(x)$



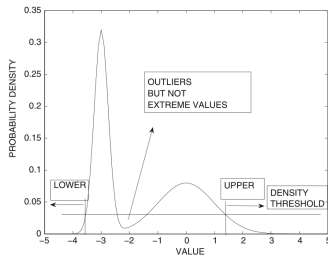
- В чем потенциальная проблема?  $\hat{\mu}, \hat{\Sigma}$  важно оценить устойчивым к выбросам способом.

# Статистическое обнаружение аномалий

- Выбросы не обязательно на границе распределений:



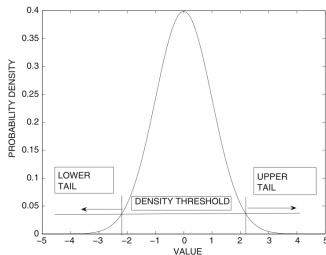
(a) Symmetric distribution



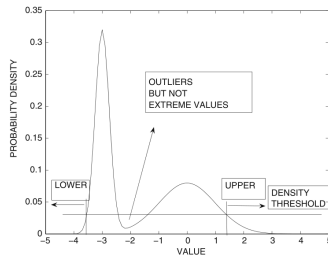
(b) Asymmetric distribution

# Статистическое обнаружение аномалий

- Выбросы не обязательно на границе распределений:



(a) Symmetric distribution



(b) Asymmetric distribution

- $p(x)$  можно оценить смесью распределений или KDE:

$$\hat{p}(x) = \frac{1}{Nh} \sum_{n=1}^N K\left(\frac{x - x_n}{h}\right)$$

# Содержание

1 Статистические методы

2 Метрические методы

3 Модельные методы

4 Сдвиг модели

## K-центров

- Если все точки train нормальные, то можем решить задачу K-покрытия:
  - найти  $z_1, \dots, z_K$  такие, что

$$\min_k \rho(x_n, z_k) \leq R \quad \forall n = 1, 2, \dots, N.$$

$$\text{outlierness}(x) = \min_k \rho(x, z_k) / R$$

- Это метод K-центров (K-centers<sup>3</sup>).
- Как связаны параметры  $K, R$ ?

---

<sup>3</sup>Support Objects for Domain Approximation.



## K-центров

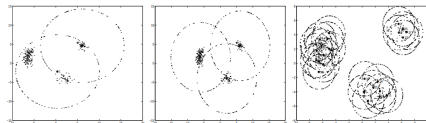
- Если все точки train нормальные, то можем решить задачу K-покрытия:
  - найти  $z_1, \dots, z_K$  такие, что

$$\min_k \rho(x_n, z_k) \leq R \quad \forall n = 1, 2, \dots, N.$$

$$\text{outlierness}(x) = \min_k \rho(x, z_k) / R$$

- Это метод K-центров (K-centers<sup>3</sup>).
- Как связаны параметры  $K, R$ ?

Связь  $K$  и  $R$ :



- Др. применение: опт. расположение  $K$  складов в  $N$  городах.

<sup>3</sup>Support Objects for Domain Approximation.

# Алгоритм К-центров

Жадный алгоритм К-центров:

выбираем  $z_1$  случайным объектом

$k:=1$

ПОКА  $k \leq K$

    выбираем  $z_{k+1}$  самым удалённым объектом от  $\{z_1, \dots, z_k\}$

$k:=k+1$

Перед наращиванием  $k$  можно пробовать улучшить расположение центров

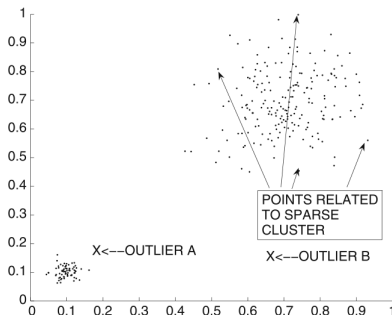
- $z_k$  - наилучший объект в  $R$ -окрестности с точки зрения глобального  $R$  (сложность  $O(N^2)$ )

Но как быть, если обучающая выборка может содержать выбросы?

## Обнаружение аномалий по расстоянию

Простые способы. Объект выброс, если расстояние выше порога

- $\rho(x, NN(x)) > t$  (до ближайшего соседа)
- $\min_k \rho(x, \mu_k) > t$  (до центра ближайшего кластера)



Но тогда выброс А либо пропущен, либо все точки разреженного кластера - выбросы.

## Метод local outlier factor

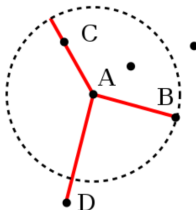
- Пусть  $N_K(x)$  - множество  $K$  ближайших соседей  $x$
- Определим  $NN_K(x)$  -  $K$ -й ближайший сосед  $x$

$$\rho_K(x) = \rho(x, NN_K(x)) \text{ (K-расстояние)}$$

$$rd_K(x, z) = \max\{\rho(x, z), \rho_K(z)\}$$

- k-distance: устойчивее к случайным отдельным точкам.
- max: устойчивость к слишком близким точкам.

$$rd_K(A, B) = rd_K(A, C) < rd_K(A, D) \text{ для } K=3$$



## Метод local outlier factor

- $lrd_K(x)$  (local reachability density)-плотность точек вокруг  $x$ :

$$lrd_K(x) = \frac{1}{\frac{1}{|N_K(x)|} \sum_{z \in N_K(x)} rd_K(x, z)}$$

- Метод local outlier factor - отношение плотности соседей  $x$  к плотности  $x$ :

$$LOF_K(x) = \frac{\frac{1}{|N_K(x)|} \sum_{z \in N_K(x)} lrd_K(z)}{lrd_K(x)}$$

- Это loss или score?

## Метод local outlier factor

- $lrd_K(x)$  (local reachability density)-плотность точек вокруг  $x$ :

$$lrd_K(x) = \frac{1}{\frac{1}{|N_K(x)|} \sum_{z \in N_K(x)} rd_K(x, z)}$$

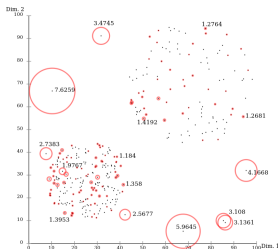
- Метод local outlier factor - отношение плотности соседей  $x$  к плотности  $x$ :

$$LOF_K(x) = \frac{\frac{1}{|N_K(x)|} \sum_{z \in N_K(x)} lrd_K(z)}{lrd_K(x)}$$

- Это loss или score?
- $LOF_K(x) \leq 1$ : типичная точка,  $LOF_K(x) > 1$  - более удалённая.

# Анализ

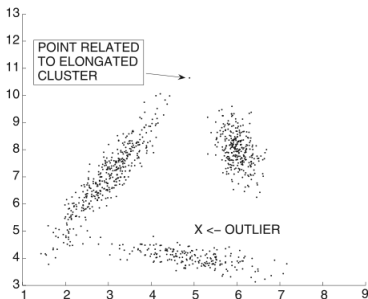
- LOF корректно выделяет выбросы как в контексте густых, так и в контексте разреженных соседей:



- Выброс, если  $LOF_K(x) > t$ : нужно подбирать  $t$  и  $K$ .
- Обобщается на др.  $\rho(x, z)$
- Лучше работает с использованием метода случайных подпространств<sup>4</sup>.

<sup>4</sup>Feature bagging for outlier detection.

## Учет локального распределения точек



- Др. подход: учитывать локальное распределение точек.
- Подходы, учитывающие локальное распределение:
  - смесь Гауссиан
  - метод локального кластера (local cluster)
  - метод локальной окрестности (local neighborhood)



## Метод локального кластера

- 1 Кластеризуем точки на  $K$  кластеров, используя расстояние Махаланобиса:
- 2 Для каждого кластера находим  $\mu_k$  и  $\Sigma_k$ .
- 3 Для объекта  $x$ :
  - 1 находим ближайший кластер:

$$\hat{c} = \arg \min_c \sqrt{(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)}$$

- 2 степень нетипичности:

$$\text{outlierness}(x) = \sqrt{(x - \mu_{\hat{c}})^T \Sigma_{\hat{c}}^{-1} (x - \mu_{\hat{c}})}$$

## Метод локальной окрестности

- ❶ Инициализируем  $L_K(x) = \{x\}$
- ❷ Для  $k = 1, 2, \dots, K$ :
  - ❶  $x_k = \arg \min_z \rho(z, L_K(x))$
  - ❷  $L_K(x) := L_K(x) \cup \{x_k\}$
- ❸ Исключим  $x$ :  $L_K(x) := L_K(x) \setminus \{x\}$
- ❹ Используя  $L_K(x)$  рассчитаем  $\mu(x)$  и  $\Sigma(x)$
- ❺ Степень нетипичности:

$$\text{outlierness}(x) = \sqrt{(x - \mu(x))^T \Sigma(x)^{-1} (x - \mu(x))}$$

Вычислительно сложнее, зато лучше учитывает распределение вокруг  $x$ .

# Содержание

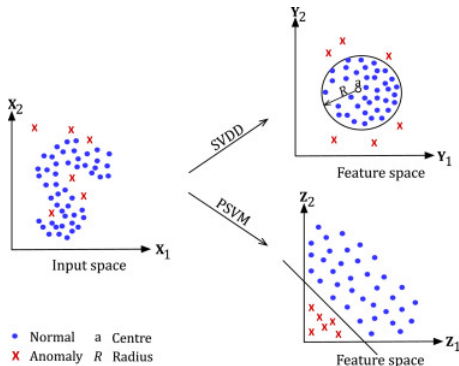
- 1 Статистические методы
- 2 Метрические методы
- 3 **Модельные методы**
  - Одноклассовый метод опорных векторов
  - Изолирующий лес
- 4 Сдвиг модели

### 3 Модельные методы

- Одноклассовый метод опорных векторов
- Изолирующий лес

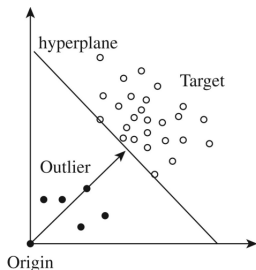
# Одноклассовый метод опорных векторов<sup>5</sup>

- Преобразование пр-ков:  $x \rightarrow \Phi(x) \in F$ .
- Отделим в пространстве  $\Phi(x)$  нормальные точки от остальных
  - гиперплоскостью, максимально отдалённой от нуля
  - шаром минимального радиуса



<sup>5</sup> Estimating the Support of a High-Dimensional Distribution.

# Отделение гиперплоскостью



$$\begin{cases} \frac{1}{2} \|w\|^2 + \frac{1}{\nu N} \sum_{n=1}^N \xi_n - \rho \rightarrow \min_{w, \xi \in \mathbb{R}^N, \rho \in \mathbb{R}} \\ \langle w, \Phi(x_n) \rangle \geq \rho - \xi_n; \quad \xi_n \geq 0, \quad n = 1, 2, \dots, N. \end{cases}$$

$$f(x) = \text{sign}(\langle w, \Phi(x) \rangle - \rho) \quad -1 \text{ для выброса}$$

Максимизируем расстояние от нуля до гиперплоскости  $\frac{\rho}{\|w\|}$ .

Гиперпараметр  $\nu \in (0, 1)$  - макс. доля выбросов в выборке.

## Отделение гиперплоскостью - решение

В терминах ядер  $K(x, z) = \langle \Phi(x), \Phi(z) \rangle$ :

$$f(x) = \text{sign} \left( \sum_n \alpha_n K(x_n, x) - \rho \right)$$

где  $\{\alpha_n\}$  находятся из решения двойственной задачи:

$$\begin{cases} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \rightarrow \min_{\alpha} \\ 0 \leq \alpha_i \leq \frac{1}{\nu N}; \quad \sum_i \alpha_i = 1 \end{cases}$$

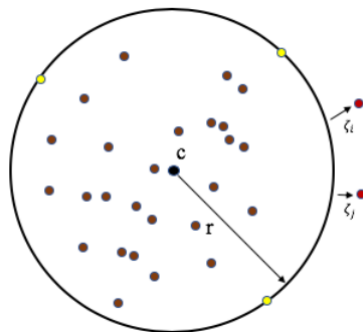
- $\alpha_i = 0$ : обычные объекты
- $\alpha_i \in (0, \frac{1}{\nu N})$ : на гиперплоскости, по 1 такому  $x_i$  находим:

$$\rho = \langle w, \Phi(x_i) \rangle = \sum_n \alpha_n K(x_n, x_i)$$

- $\alpha_i = \frac{1}{\nu N}$ : выбросы<sup>6</sup>

<sup>6</sup>Как отсюда следует, что  $\nu$ -макс. доля выбросов в выборке?

# Отделение шаром



$$\begin{cases} R^2 + \frac{1}{\nu N} \sum_n \xi_n \rightarrow \min_{R \in \mathbb{R}, \xi \in \mathbb{R}^N, c \in F} \\ \|\Phi(x_n) - c\|^2 \leq R^2 + \xi_n; \quad \xi_n \geq 0; \quad n = 1, 2, \dots, N. \end{cases}$$

$$f(x) = \text{sign} \left( R^2 - \|\Phi(x_n) - c\|^2 \right) \quad -1 \text{ для выброса}$$



## Отделение шаром - решение

В терминах ядер  $K(x, z) = \langle \Phi(x), \Phi(z) \rangle$ :

$$f(x) = \text{sign} \left( R^2 - \sum_{i,j} \alpha_i \alpha_j L(x_i, x_j) + 2 \sum_i \alpha_i K(x_i, x) - K(x, x) \right)$$

где  $\{\alpha_n\}$  находятся из решения двойственной задачи:

$$\begin{cases} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) - \sum_i \alpha_i K(x_i, x_i) \rightarrow \min_{\alpha} \\ 0 \leq \alpha_i \leq \frac{1}{\nu N}; \quad \sum_i \alpha_i = 1 \end{cases}$$

- $\alpha_i = 0$ : обычные объекты
- $\alpha_i \in (0, \frac{1}{\nu N})$ : на гиперплоскости, по 1 такому  $x_i$  находим:

$$R^2 = \|\Phi(x_i) - c\|^2 = K(x_i, x_i) - 2 \sum_n \alpha_n K(x_n, x_i) + \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j)$$

- $\alpha_i = \frac{1}{\nu N}$ : выбросы<sup>7</sup>

<sup>7</sup>Как отсюда следует, что  $\nu$ -макс. доля выбросов в выборке?

## Ядерное обобщение с RBF ядром

популярные ядра:  $K(x, z) = e^{-\frac{\|x-z\|^2}{2\sigma^2}}$ ,  $K(x, z) = (\langle x, z \rangle + 1)^K$

- Стационарные ядра<sup>8</sup>:  $K(x, z) = G(x - z)$ 
  - инвариантны к сдвигу  $K(x, z) = K(x + \Delta, z + \Delta)$
  - переводят  $\{x_n\}$  на сферу

$$\|\Phi(x_n)\|^2 = \langle \Phi(x_n), \Phi(x_n) \rangle = K(x_n, x_n) = G(0) = \text{const}$$

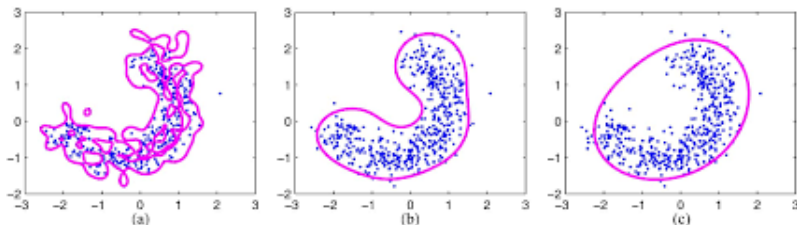
- для данных на сфере минимизация сегмента (отделение гиперплоскостью) и минимизация шаром эквивалентны  
=> 2 последних метода дают одинаковый результат.

---

<sup>8</sup>Будут ли Гауссово и полиномиальное ядра стационарными?

# Параметры

Одноклассовый SVM с RBF ядром ( $\sigma \uparrow$ )



$\gamma$  контролирует размер фигуры (и долю выбросов).

### 3 Модельные методы

- Одноклассовый метод опорных векторов
- Изолирующий лес

# Изолирующее дерево

## Построение изолирующего дерева (isolation tree).

инициализировать корень всеми наблюдениями

ПОКА (существуют узлы с несовпадающими наблюдениями  
глубины  $< S$ ): # рекомендуется  $S=8$

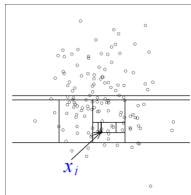
выбрать такой узел

выбрать случайный неконстантный признак  $f \in [f_{min}, f_{max}]$

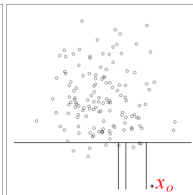
выбрать случайный порог  $t \in (f_{min}, f_{max})$

разбить узел на 2 подузла по правилу  $f \leq t$

- Дерево строится без учителя.
- Как по нему оценить типичность  $x$ ?



(a) Isolating  $x_i$



(b) Isolating  $x_o$

# Изолирующий лес

- Типичность объекта в дереве

$$h(x) = p(x) + c(m)$$

- $p(x)$  - глубина пути в дереве
  - $m = \#$ др. объектов в листе
  - $c(m) = 2(\ln(m-1) + 0.57) - 2(m-1)/m$  - оценка доп. пути до  $x$ , если бы дерево строилось до конца.
- По одному дереву считать нельзя (много случайности).

# Изолирующий лес

- Типичность объекта в дереве

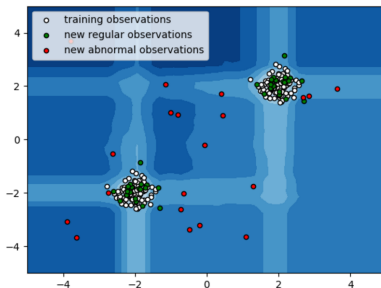
$$h(x) = p(x) + c(m)$$

- $p(x)$  - глубина пути в дереве
  - $m = \#$ др. объектов в листе
  - $c(m) = 2(\ln(m-1) + 0.57) - 2(m-1)/m$  - оценка доп. пути до  $x$ , если бы дерево строилось до конца.
- По одному дереву считать нельзя (много случайности).
- Изолирующий лес (isolation forest) - ансамбль  $K$  независимых изолирующих деревьев (рекоменд.  $K = 100$ ).

$$\text{outlierness}(x) = 2^{-\frac{\mathbb{E}\{h(x)\}}{c(N)}}, \quad N = \# \text{объектов обуч. выборки}$$

- $\text{outlierness}(x) \approx 1$ : выброс
- $\text{outlierness}(x) \approx 0.5$ : обычный объект

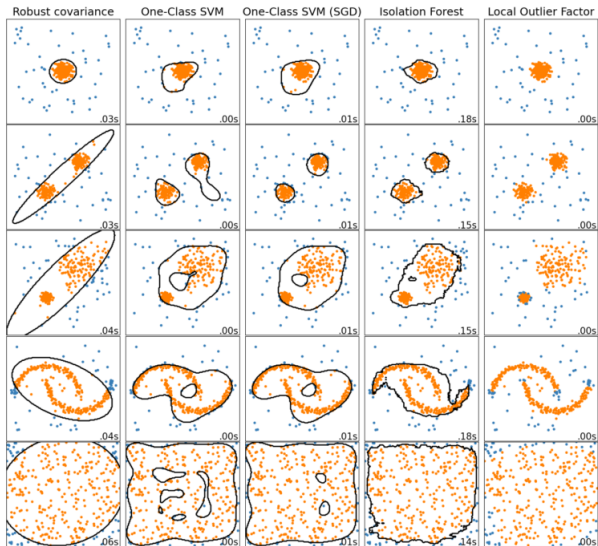
# Преимущества



- ⊕ : Работает с вещественными, порядковыми, бинарными признаками.
- ⊕ : Быстрый, интерпретируемый алгоритм.
- ⊕ : Интерпретируемая  $\text{outlierness}(x) \in (0, 1)$
- ⊕ : Обучается, даже если в  $X$  нет выбросов.
  - в отличие от одноклассового SVM
- ⊕ : Хорошо учится на малой подвыборке типичных объектов.



# Сравнение методов



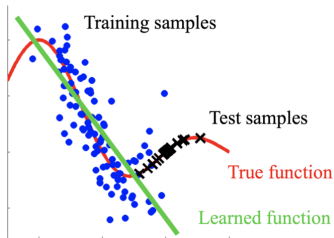
# Содержание

- 1 Статистические методы
- 2 Метрические методы
- 3 Модельные методы
- 4 Сдвиг модели**

## Сдвиг модели

Связанно с обнаружением аномалий - сдвиг модели (concept drift): изменяется целевая зависимость  $y = f(x)$

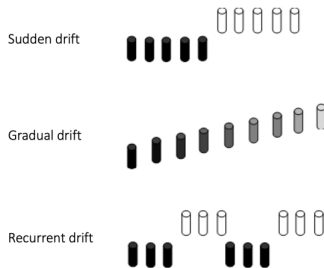
- изменения носят устойчивый характер во времени
  - у клиента изменились вкусы, женился, переехал
  - изменилась внешняя ситуация (пандемия)



- Будем наблюдать устойчивые изменения в ошибках модели.

## Сдвиг модели

- Каждый тип сдвига детектируется и обрабатывается по-своему:



- В отличие от аномалий, сдвиги носят устойчивых характер.
- Детекция - на основе статистик сравнения распределений ошибок во времени.

# Заключение

- Детекция выбросов - задача обучения без учителя
  - если с учителем - то это классификация несбалансированных классов
    - генерация объектов для редкого класса
    - удаление объектов для частого класса
- Оценка - по размеченной валидации, используя ROC, AUC.
- Методы обнаружения аномалий:
  - статистические:  $p(x) < t$
  - метрические: выброс далеко от др. точек
  - модельные: моделируем область нормальности
    - one-class SVM
    - isolation forest