

Подбор параметров в алгоритмах детектирования аномалий

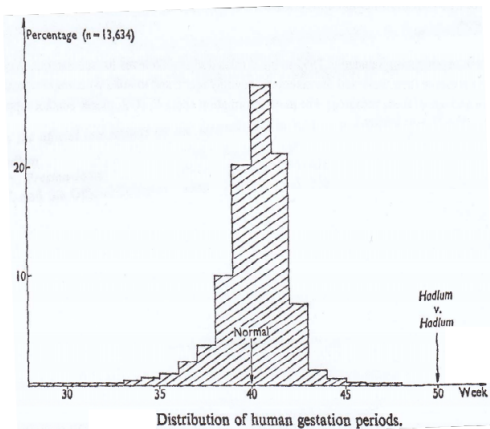
Смоляков Дмитрий

МФТИ, ИППИ РАН

Конференция МФТИ 2014

Пример аномалии

Аномалия – измерение, которое не соответствует некоторому ожидаемому поведению.



Области применения

- Поиск мошенничества на основе необычной активности
- В медицине необычные результаты обследований могут свидетельствовать о потенциальных проблемах со здоровьем
- Аномальные значения могут быть результатом ошибок в измерении
- Раннее обнаружение аномальной работы двигателя может предотвратить неожиданную поломку

Hawkins-Outlier

- Пусть Q некоторый механизм порождения данных
- $X_1 \dots X_n$ порождены механизмом Q
- $\tilde{X}_1 \dots \tilde{X}_m$ отличаются от измерений X_1, \dots, X_n , настолько, что можно предположить, что они порождены другим механизмом.

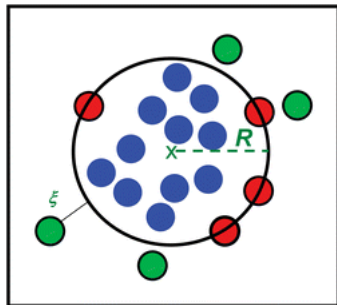
Постановка задачи

- X_1, \dots, X_N $X_i \in S$ – не размеченная выборка
- Хотим построить $f: S \rightarrow \{-1, 1\}$
- $f(X_i) = -1$, если X_i – аномалия
- $f(X_i) = 1$, если X_i – нормальное измерение

Support Vector Data Description

$$\begin{cases} R + \frac{1}{N\nu} \sum_{i=1}^N \xi_i \rightarrow \min_{R, x, \xi} \\ \|\phi(x_i) - x\|^2 \leq R + \xi_i & i = 1, \dots, N \\ \xi_i \geq 0 & i = 1, \dots, N \\ R \geq 0 \end{cases}$$

- ν – верхняя граница количества аномалий и нижняя граница количества опорных векторов
- $\phi(x_i)$ – отображение в пространство более высокой размерности



Двойственная задача

Обычно решают двойственную задачу:

$$\begin{cases} \sum_{i=1}^N \alpha_i \phi(x_i) \cdot \phi(x_i) - \sum_{i,j=1}^N \alpha_i \alpha_j \phi(x_i) \cdot \phi(x_j) \rightarrow \max_{\alpha} \\ \sum_{i=1}^N \alpha_i = 1 \\ 0 \leq \alpha_i \leq \frac{1}{N\nu} \quad i = 1, \dots, N \end{cases}$$

Достаточно знать скалярное произведение.

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Kernel Trick

Необязательно знать явный вид спрямляющего пространства, чтобы использовать $K(x, y)$. Достаточно потребовать:

- $K(x, y) = K(y, x)$ – симметричность
- $\int_X \int_X K(x, x')g(x)g(x')dx dx' \forall g(x): X \rightarrow \mathbb{R}$ – положительная определённость

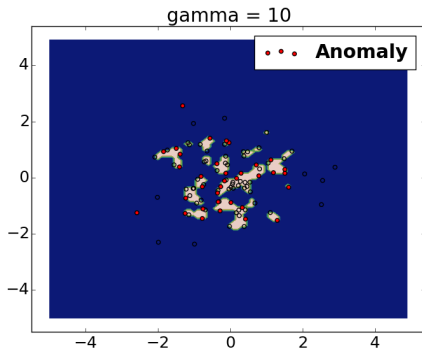
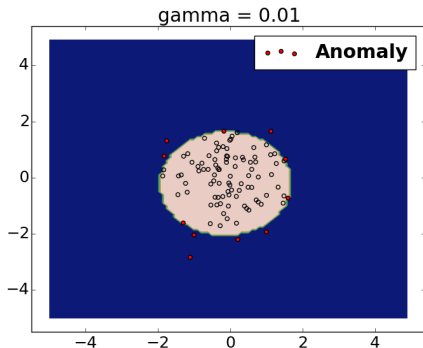
Тогда $K(x, y)$ – скалярное произведение в некотором пространстве.

Примеры ядер

Название	Формула	Параметры
Линейное	$x \cdot y$	—
Полиномиальное	$(\gamma x \cdot y + d)^k$	γ, d, k
Гауссово	$\exp(-\gamma \ x - y\ ^2)$	γ
Сигмоидное	$\text{th}(\gamma x \cdot y + d)$	γ, d

Выбор ядерной функции

Параметры ядра могут существенно влиять на результат



Хочется получить функцию риска для эффективного подбора параметров ядра.

В приложениях часто пытаются достигнуть границы значений количества аномалий и опорных векторов¹

- $f_{anomaly}$ – доля объектов отмеченных, как аномалии
- $f_{support}$ – доля объектов используемых, как опорные векторы

$$r = (\nu - f_{anomaly})^2 + (\nu - f_{support})^2$$

¹Lukashevich H., Nowak S., Dunker P. Using one-class SVM outliers detection for verification of collaboratively tagged image training sets

Оптимизация функции от ядерной матрицы

Мерой риска может послужить отношение квадрата стандартного отклонения к среднему значению²

$$r = \frac{s^2}{\bar{k} + \varepsilon}$$

$$\bar{k} = \frac{\sum_{i=1}^N \sum_{j=1}^N K(x_i, x_j)}{N(N-1)}; \quad s^2 = \frac{\sum_{i=1}^N \sum_{j=1}^N (K(x_i, x_j) - \bar{k})^2}{N(N-1) - 1}$$

²Evangelista P. F., Embrechts M. J., Szymanski B. K. Some properties of the Gaussian kernel for one class learning

Обнаружение уровня плотности

Функция риска может в явном виде учитывать распределения аномальных точек³:

$$r = \frac{1}{(1 + \rho)|S|} \sum_{i \in S} l(1, \text{sign}(f(x_i))) + \frac{\rho}{1 + \rho} \mathbb{E}_{\mu} l(-1, \text{sign}(f(x_i)))$$

$$l(y, t) = \max\{0, 1 - yt\}; \quad \mu - \text{распределение аномалий}$$

Здесь ρ подбирается из априорных соображений.

³Steinwart I., Hush D. R., Scovel C. A classification framework for anomaly detection

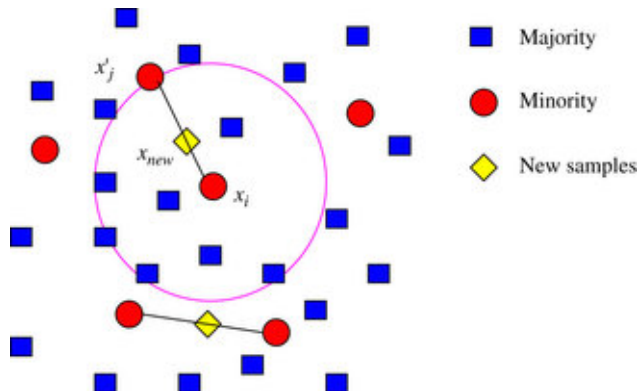
Генерация тестовой выборки

Искусственно сгенерируем примеры аномальных и нормальных точек.

- Аномалии генерируются из равномерного распределения
- Нормальные данные генерируются на основе тестовой выборки, например при помощи SMOTE (Synthetic Minority Over-sampling Technique)

$$r = \frac{1}{|X_{Normal}|} \sum_{x_i \in X_{Normal}} [f(x_i) = -1] + \frac{1}{|X_{Anomaly}|} \sum_{x_i \in X_{Anomaly}} [f(x_i) == 1]$$

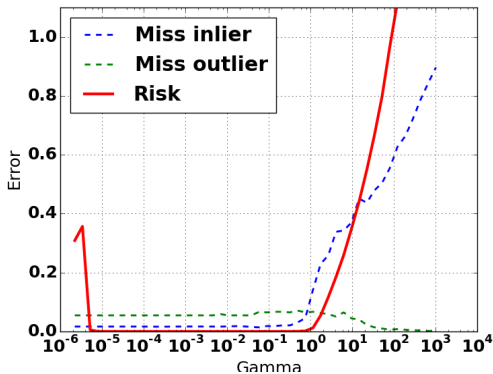
Synthetic Minority Over-sampling Technique



Постановка эксперимента

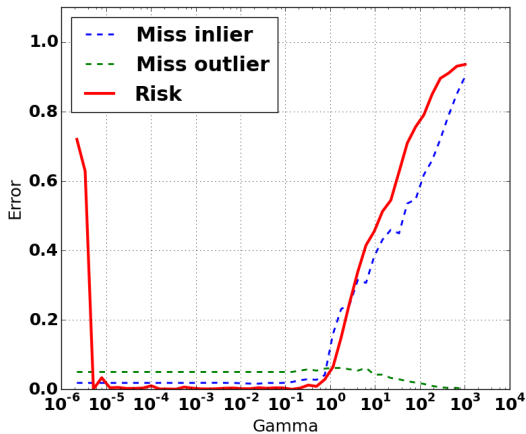
- Нормальные данные генерируются из стандартного многомерного нормального распределения
- Аномалии генерируются из равномерного распределения, причем границы подбираются так, чтобы все нормальные данные попали в заданные границы.
- Получившиеся данные делятся на обучающую и тестовую выборку. На первой происходит подбор параметров и обучение модели, на второй проверка результата.

Результаты



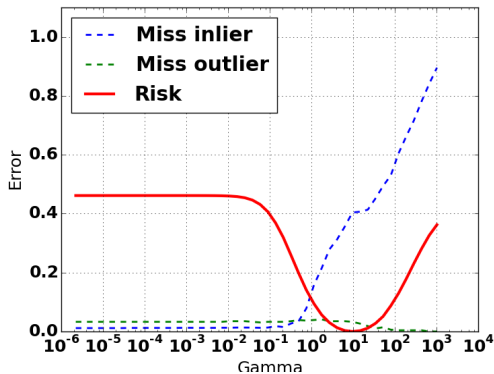
Оптимизация количества аномалий и опорных векторов

Результаты



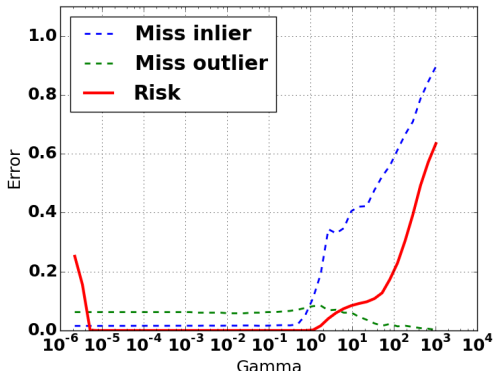
Учет распределения аномалий

Результаты



Оптимизация ядерной матрицы

Результаты



Случайная генерация тестовой выборки

Выводы

- Только в методе оптимизации ядерной матрицы существует единственный минимум
- Критерий оптимальности может быть различным в зависимости от важности ошибки на нормальных и аномальных данных
- Только критерий на основе случайной генерации тестовой выборки способен учитывать различия ошибок

Спасибо за внимание