

Современные информационные технологии

Александр Широков ПМ-1701

Преподаватель:

ЛЕОРА

Санкт-Петербург
2020 г., 7 семестр

Список литературы

[1]

Содержание

1	03.09.2020	2
1.1	Типы аномальных данных	2

1 03.09.2020

1.1 Типы аномальных данных

Существует два основных направления занимающихся поиском аномалий в данных:

- Выявление выбросов (Outlier Detection)
- Выявление нетипичных объектов (Novelty Detection)

Постановка задачи выявления аномалий.

В дальнейшем будем полагать, что данные имеют признаковое представление, то есть каждый объект x задан в виде некоторого вектора из R^n . В заданном множестве X решить задачу классификации на 0 или 1 - аномалия ли или нет.

Все алгоритмы детектирования аномалий сводятся к построению по некоторой функции $anomaly_score(x)$, которая по данному объекту выдает некоторый рейтинг аномальности. После этого разделение на класс аномалий и класс нормальных данных производится бинаризацией по некоторому порогу.

1. Статистические тесты

Как правило, статистические тесты применяют для отдельных признаков и отлавливают экстремальные значения. Для этого используют $Z - value$ или $Kurtozis$:

$$Z - value Z_i = \frac{|x_i - \mu|}{\sigma}$$

$$Kurtozis = \frac{1}{n} \sum_{i=1}^n Z_i^4$$

Куртозис отвечает за то, насколько вытянуто наше пространство признаков.