

Задача кластеризации. Поиск аномальных данных

(Часть 1)

Задание 1. Истоки поиска аномальных данных. Примеры из реальной жизни

Исследуйте, откуда берет начало задача поиска аномальных данных (анализ выбросов), и приведите три примера применения методов поиска аномалий в реальной жизни.

- 1) Изучите происхождение задачи поиска аномалий.
- 2) Приведите три примера реальных задач, где поиск аномалий используется на практике (например, в финансах, медицине и т.д.).
- 3) Объясните, почему обнаружение аномалий важно в каждом примере.

Задание 2. Вероятностный подход к поиску аномалий. Применение метода максимального правдоподобия (МП)

Используя вероятностный подход и метод максимального правдоподобия (МП), решите задачу поиска аномалий в наборе данных. Примените модель нормального распределения для обнаружения выбросов.

- 1) Выполните оценку параметров нормального распределения для данного набора данных.
- 2) Рассчитайте вероятность для каждой точки данных и определите выбросы как те, которые имеют низкую вероятность.
- 3) Визуализируйте распределение данных и выбросы.

Задание 3. Модель смеси распределений (Gaussian Mixture Model) для поиска аномалий

Используя модель смеси нормальных распределений (GMM), выполните кластеризацию и выявите выбросы как те точки, которые имеют низкую вероятность принадлежности к любому из кластеров.

- 1) Примените GMM для кластеризации данных.
- 2) Вычислите вероятности принадлежности каждой точки к кластерам и определите выбросы.
- 3) Визуализируйте кластеры и аномалии.

Критерии оценивания

K1 Выполнено задание 1 2 балл

K2 Выполнено задание 2 2 балл

K3 Выполнено задание 3 1 балл

Максимальное количество баллов 5 баллов

Минимальное количество баллов чтобы преподаватель смог зачесть вашу работу 2 балла