

# Лекция 18

## Одноклассовые методы и обнаружение аномалий

Е. А. Соколов  
ФКН ВШЭ

23 апреля 2017 г.

В задачах кластеризации, о которых шла речь ранее, требуется разделить выборку на группы так, чтобы внутри каждой группы объекты были похожи друг на друга. Теперь мы изучим немного другую постановку — поиск аномалий. В ней даётся выборка «нормальных» объектов, и требуется построить некоторую модель, описывающую данную выборку. Далее для новых объектов требуется определять, принадлежат ли они тому же распределению, что и эта выборка, или же являются выбросами или аномалиями. Такие методы применяются, например, в задачах обнаружения мошеннического поведения или раннего обнаружения неполадок оборудования.

### 1 Несбалансированная классификация

В некоторых задачах примеры аномалий могут быть даны, но в небольших объёмах — например, при анализе данных систем самолёта может быть известно несколько аномальных ситуаций из прошлого. Такую задачу можно рассматривать как классификацию с несбалансированными классами. При решении обычными методами классификатору оказаться выгоднее относить все объекты к одному классу, поэтому имеет смысл модифицировать процедуру обучения.

Самые простые методы борьбы с несбалансированностью — *undersampling* и *oversampling*. Первый из них удаляет случайные объекты доминирующего класса до тех пор, пока соотношение классов не станет приемлемым; второй дублирует случайные объекты минорного класса. Оптимальное число объектов для удаления или дублирования следует подбирать с помощью кросс-валидации. Отметим, что данные методы применяются лишь к обучающей выборке, а контрольная выборка остается без изменений.

Более сложный метод SMOTE [1] заключается в дополнении минорного класса синтетическими объектами. Генерация нового объекта производится следующим образом. Выбирается случайный объект  $x_1$  минорного класса, для него выделяются  $k$  ближайших соседей из этого же класса ( $k$  — настраиваемый параметр), из этих соседей выбирается один случайный  $x_2$ . Новый объект вычисляется как точка на отрезке между  $x_1$  и  $x_2$ :  $\alpha x_1 + (1 - \alpha)x_2$ , для случайного  $\alpha \in (0, 1)$ .

## 2 Одноклассовая классификация

Ниже мы будем обсуждать обнаружение точечных аномалий — объектов, которые существенно отличаются от заданной выборки. При этом выделяют и другие типы. Например, контекстными аномалиями называют наблюдения, отличающиеся от наблюдений, близких по некоторому параметру. Например, температура  $-10^\circ$  является нормальной в январе, но аномальной в июне.

### §2.1 Статистические методы

В статистических методах предлагается восстановить плотность выборки  $p(x)$ , и затем определять аномальность объекта на основе того, насколько вероятно его получить из данной плотности. Например, это можно делать через отклонение от среднего  $[\rho(x, \mu) > d]$  (порог может подбираться, если известно некоторое количество примеров аномалий) или с помощью статистических тестов. Существует два подхода к восстановлению плотности: параметрический и непараметрический.

#### 2.1.1 Непараметрический подход

Начнём с одномерных величин. Согласно одному из определений неотрицательная функция  $p(x)$  является плотностью распределения случайной величины  $\xi$ , если её значение в каждой точке равно пределу

$$p(x) = \lim_{h \rightarrow 0} \mathbb{P}(\xi \in [x - h, x + h]).$$

Воспользуемся этим определением и построим эмпирическую оценку плотности:

$$\hat{p}(x) = \frac{1}{2h} \frac{1}{\ell} \sum_{i=1}^{\ell} [|x - x_i| < h] = \frac{1}{\ell h} \sum_{i=1}^{\ell} \frac{1}{2} \left[ \frac{|x - x_i|}{h} < 1 \right],$$

где  $h$  — ширина окна, регулирующая гладкость эмпирической плотности. Чем больше объектов обучающей выборки в окрестности точки, тем выше будет плотность.

В указанной оценке используется индикатор, что приводит к отсутствию гладкости. Чтобы устранить это, заменим индикатор того, что расстояние меньше ширины окна, на некоторую гладкую функцию  $K(z)$ :

$$\hat{p}(x) = \frac{1}{\ell h} \sum_{i=1}^{\ell} K\left(\frac{x - x_i}{h}\right).$$

Здесь  $K(z)$  — ядро (не путайте с ядрами Мерсера!), которое должно удовлетворять четырём требованиям:

- чётность:  $K(-z) = K(z)$ ;
- нормированность:  $\int K(z) dz = 1$ ;
- неотрицательность:  $K(z) \geq 0$ ;
- невозрастание при  $z > 0$ .

Примером может служить гауссово ядро  $K(z) = (2\pi)^{-1/2} \exp(-0.5z^2)$ .

Оценку плотности легко обобщить на многомерный случай, заменив разность  $|x - x_i|$  на некоторую метрику  $\rho(x, x_i)$ :

$$\hat{p}(x) = \frac{1}{\ell V(h)} \sum_{i=1}^{\ell} K\left(\frac{\rho(x, x_i)}{h}\right),$$

где  $V(h) = \int K\left(\frac{\rho(x, x_i)}{h}\right) dx$  — нормировочная константа. Следует помнить, что число объектов, необходимое для качественной оценки плотности, растёт экспоненциально по мере роста числа признаков. Из-за этого непараметрические методы подходят только для обнаружения аномалий в маломерных пространствах.

### 2.1.2 Параметрический подход

Параметрический подход состоит в приближении плотности с помощью распределения  $p(x | \theta)$  из некоторого семейства  $\{p(x | \theta) | \theta \in \Theta\}$  с помощью метода максимального правдоподобия:

$$\sum_{i=1}^{\ell} \log p(x_i | \theta) \rightarrow \max_{\theta}$$

В качестве распределений могут выступать, например, нормальные или смеси нормальных. В пространствах большой размерности может иметь смысл наивное байесовское предположение, о котором пойдёт речь на семинарах.

## §2.2 Метрические методы

Метрический подход основан на выделении объектов, которые расположены от других существенно дальше, чем объекты в среднем удалены друг от друга. А именно, объект  $x$  объявляется аномальным, если  $p$  или меньше процентов объектов имеют до него расстояние меньше  $\varepsilon$ :

$$\frac{1}{\ell} \sum_{i=1}^{\ell} [\rho(x, x_i) < \varepsilon] \leq p.$$

Пороги  $p$  и  $\varepsilon$  являются параметрами, которые должны настраиваться по известным примерам аномалий или исходя из априорных предположений.

### §2.3 Одноклассовый метод опорных векторов

Для обнаружения аномалий, по сути, необходимо построить некоторую функцию  $a(x)$ , которая принимает значение 1 на области как можно меньшего объёма, содержащей как можно больше объектов выборки; во всех остальных точках она должна иметь значение 0. Такая функция будет компактно описывать обучающую выборку, и можно рассчитывать, что на аномальных объектах она будет равна нулю.

Будем строить линейную функцию  $a(x) = \text{sign}\langle w, x \rangle$ , и потребуем, чтобы она разделяла выборку от начала координат с максимальным отступом. Соответствующая оптимизационная задача будет иметь вид [2]

$$\begin{cases} \frac{1}{2}\|w\|^2 + \frac{1}{\nu\ell} \sum_{i=1}^{\ell} \xi_i - \rho \rightarrow \min_{w, \xi, \rho} \\ \langle w, x_i \rangle \geq \rho - \xi_i, \quad i = 1, \dots, \ell, \\ \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{cases}$$

Здесь гиперпараметр  $\nu$  отвечает за корректность на обучающей выборке — можно показать, что он является верхней границей на число аномалий (объектов выборки, на которых  $a(x) = 0$ ).

Для данной задачи можно выписать двойственную и сделать ядровой переход в ней:

$$\begin{cases} \frac{1}{2} \sum_{i,j=1}^{\ell} \lambda_i \lambda_j K(x_i, x_j) \rightarrow \min_{\lambda} \\ 0 \leq \lambda_i \leq \frac{1}{\nu\ell}, \quad i = 1, \dots, \ell, \\ \sum_{i=1}^{\ell} \lambda_i = 1. \end{cases}$$

Модель при этом будет иметь вид

$$a(x) = \text{sign} \left( \sum_{i=1}^{\ell} \lambda_i K(x, x_i) - \rho \right).$$

При использовании подходящих ядер можно действительно получить функцию, которая точно описывает обучающую выборку в исходном пространстве. Также можно показать, что объекты из того же распределения, из которого сгенерирована обучающая выборка, будут с не очень большой вероятностью попадать в область с нулевым значением  $a(x)$ .

## Список литературы

- [1] Chawla N., Bowyer K., Hall L., Kegelmeyer W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. // Journal of Artificial Intelligence Research, Vol. 16, Pp. 321–357.
- [2] Schölkopf, Bernhard and Williamson, Robert and Smola, Alex and Shawe-Taylor, John and Platt, John (1999). Support Vector Method for Novelty Detection. // NIPS'99.