

Одноклассовый SVM и его применения

Практикум на ЭВМ 2017/2018

Медведев Алексей Владимирович

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

4 декабря 2017 г.

Мы уже знакомы с SVM(Support Vector Machine). И даже использовали метод опорных векторов для решения задачи бинарной классификации. Но оказывается, что и одноклассовый SVM может быть полезен.

Чем может быть интересен One-Class SVM:

- Этот метод решает задачу поиска «новизны» (Novelty Detection). «Новизной» называют нетипичные по каким-то соображениям для выборки объекты. Их нет в обучающей выборке, но они появятся со временем, и их необходимо отследить.

Чем может быть интересен One-Class SVM:

- Этот метод решает задачу поиска «новизны» (Novelty Detection). «Новизной» называют нетипичные по каким-то соображениям для выборки объекты. Их нет в обучающей выборке, но они появятся со временем, и их необходимо отследить.
- Относится к виду обучения без учителя (unsupervised learning).

Постановка задачи

Будем рассматривать задачу с таких позиций:

- Обучающая выборка — нормальные объекты.
- Начало координат — аномалия.
- Необходимо отделить начало координат от обучающей выборки, максимизируя отступ.

Другими словами будем считать, что x «нормальный», если:

$$\langle \omega, x \rangle - \rho > 0$$

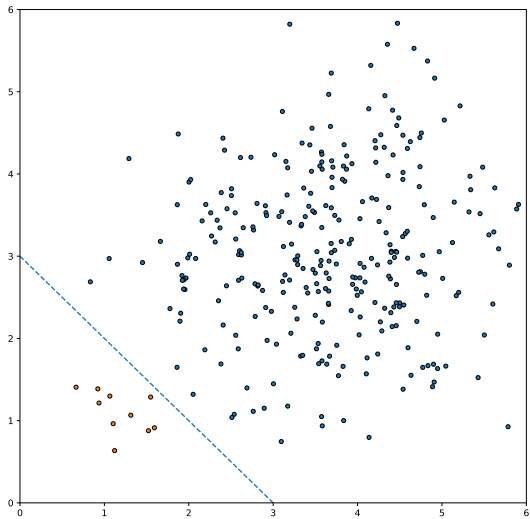
.

Задача оптимизации

Нам необходимо найти нормальный вектор гиперплоскости (перпендикуляр к ней единичной длины) ω и ρ . Гиперплоскость задается уравнением: $\langle \omega, x \rangle - \rho = 0$. Задача оптимизации примет следующий вид:

$$\begin{cases} \frac{1}{2} \|\omega\|_2^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i - \rho \rightarrow \min_{\omega, \rho, \xi} \\ \langle \omega, x_i \rangle \geq \rho - \xi_i, & \xi_i \geq 0 \end{cases}$$

Тогда ξ_i — штраф за выход за разделяющую полосу, а ν — верхняя оценка на долю аномалий на выборке.



Двойственная задача

Такой результат вряд ли нас устроит, не все выбросы лежат только у начала координат. Решение этой проблемы лежит в ядовом переходе. Т.к оптимизационная задача выпукла, то ее решение совпадает с решением двойственной задачи, зависящей от скалярных произведений, а не признаков описаний.

$$\left\{ \begin{array}{l} - \sum_{i=1}^l \sum_{j=1}^l \lambda_i \lambda_j K(x_i, x_j) \rightarrow \max_{\lambda} \\ \sum_{i=1}^l \lambda_i = 1, \quad 0 \leq \lambda_i \leq \frac{1}{\nu l} \end{array} \right.$$

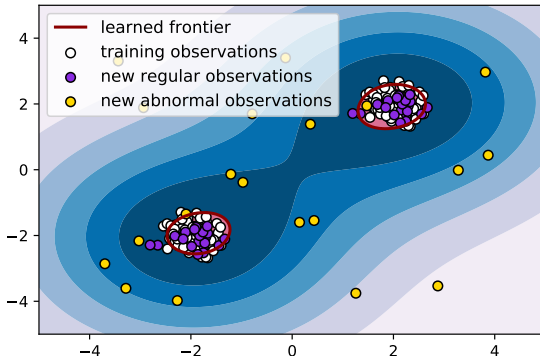
Где $K(x, z) = \exp\left(-\frac{\|x-z\|_2^2}{\sigma^2}\right)$ - ядровая функция, чаще всего «RBF».

Тогда решающая функция записывается так:

$$f(x) = \sum_{i=1}^l \lambda_i K(x_i, x) - \rho.$$

Если $f(x) > 0$ то объект относится к нормальному классу, а величина функции показывает степень уверенности классификатора.

Novelty Detection



error train: 19/200 ; errors novel regular: 5/40 ; errors novel abnormal: 1/40

Зачем вам это?

«Новизна» - это проявление принципиально нового поведения объекта. Гораздо легче составить выборку из стандартных, штатных объектов, а затем по ней отслеживать отклонения от нормы. Существуют следующие приложения:

- Обнаружение неполадок механизмов по показаниям датчиков.
- Обнаружение вирусов и вторжений.
- Подозрительные операции с банковскими счетами или на бирже.

В scikit-learn есть метод `sklearn.svm.OneClassSVM`.