

МЕТОД ОПОРНЫХ ВЕКТОРОВ С НЕОТРИЦАТЕЛЬНЫМИ КОЭФФИЦИЕНТАМИ И ЕГО ПРИМЕНЕНИЯ

Ахуньянов И. Х., Воронцов К. В.

Московский Государственный Университет, факультет ВМиК

e-mail: IAkhunianov@gmail.com

В данной работе предлагается модификация метода опорных векторов (SVM) [1] — SVM с неотрицательными коэффициентами (non-negative SVM, nnSVM) и рассматриваются его применения.

Пусть X — множество объектов, $Y = \{-1, +1\}$ — множество меток классов. Строится линейный классификатор $a(x) = \text{sgn}\left(\sum_{j=1}^n w_j x^j - w_0\right)$, где $x = (x^1, \dots, x^n)$ — признаковое описание объекта x . Аналогично [1] настройка весов $w = (w_1, \dots, w_n)$ и w_0 по выборке $\{x_i, y_i\}_{i=1}^\ell \subset X \times Y$ сводится к решению задачи квадратичного программирования. Однако, в отличие от стандартного SVM, вводятся дополнительные ограничения на веса: $w_j \geq 0$, $j = 1, \dots, n$:

$$\begin{cases} \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, \xi}; \\ y_i (\langle w, x_i \rangle - w_0) \geq 1 - \xi_i; \quad \xi_i \geq 0, i = 1, \dots, \ell; \quad w_j \geq 0, j = 1, \dots, n; \end{cases}$$

Решение данной задачи имеет несколько важных применений.

Отбор признаков. Метод nnSVM может быть непосредственно применён к задачам с априорным ограничением монотонности, когда заранее известно, что чем больше значение признака, тем ближе объект к классу +1. Благодаря ограничению $w_j \geq 0$ и условиям Куна-Таккера веса некоторых признаков обнуляются, что приводит к отсеву неинформативных признаков. Метод легко обобщается на тот случай, когда ограничение монотонности вводится только для части признаков.

Построение линейных композиций классификаторов. Пусть имеется T базовых алгоритмов $a_t : X \rightarrow \mathbb{R}$, $t = 1, \dots, T$, из которых строится линейная композиция $a(x) = \text{sgn}(w_1 a_1(x) + \dots + w_T a_T(x))$. Обычно требуется, чтобы веса w_t были неотрицательны, поэтому для их настройки можно применять nnSVM. Обнуление веса w_t означает, что алгоритм $a_t(x)$ исключается из композиции. Таким образом, nnSVM позволяет упрощать линейные композиции.

Взвешенное голосование закономерностей (правил) в логических алгоритмах классификации. Это применение является частным случаем

предыдущего, если предположить, что $a_i(x)$ — это закономерности (правила) двух классов, т. е. либо $a_i(x) \in \{-1, 0\}$, либо $a_i(x) \in \{0, +1\}$.

Отбор эталонных объектов в методе ближайших соседей. Пусть на X задана функция расстояния $\rho(x, x')$. Для произвольного объекта $u \in X$ обозначим через $x_{i,u}$ его i -го соседа из обучающей выборки, $y_{i,u}$ — соответствующий ему ответ. Рассмотрим алгоритм k ближайших соседей (kNN) с весами: $a(u) = \text{sgn}\left(\sum_{i=1}^k w_{i,u} y_{i,u}\right)$, где $w_{i,u} \geq 0$ — вес объекта $x_{i,u}$. Нетрудно показать, что для настройки весов w_i можно применить nnSVM, причём в роли признаков выступают объекты, $n = \ell$. В стандартном SVM все обучающие объекты делятся на три типа: опорные, нарушители (переходящие через границу разделяющей полосы) и неинформативные (не влияющие на вектор w). В nnSVM оказывается уже 6 типов объектов, так как для i -го объекта появляются две новые возможности: либо $w_i > 0$, тогда объект является *эталонным*; либо $w_i = 0$, тогда объект не участвует в классификациях, и его можно не хранить. Отметим, что понятия *опорных* и *эталонных* объектов существенно отличаются. Отбор эталонов соответствует отбору признаков в nnSVM.

Эксперименты на модельных данных показали, что при определённых значениях параметров C и k метод nnSVM отбирает небольшое число эталонных объектов. При этом качество классификации методом kNN с весами по отобранным эталонам часто оказывается выше, чем у обычного kNN по полной выборке. Это можно объяснить тем, что nnSVM, как и стандартный SVM, основан на принципе максимизации зазора (maximum margin), что способствует повышению обобщающей способности.

Рассматриваются также обобщения nnSVM на случаи, когда число классов произвольно, и когда априори заданы веса объектов, что позволяет применять его в качестве базового алгоритма в бустинге.

Работа выполнена при поддержке РФФИ, проект 08-07-00422.

Литература

1. Burges C. J. C. *A tutorial on support vector machines for pattern recognition // Data Mining and Knowledge Discovery. — 1998. — Vol. 2, no. 2. — Pp. 121–167.*