SMO–АЛГОРИТМ КАК ОБОБЩЕНИЕ МДМ–АЛГОРИТМА*

В. Н. Малоземов Н. А. Соловьева
v.malozemov@spbu.ru 4vinyo@gmail.com

16 февраля 2023 г.

МДМ-алгоритм ориентирован на нахождение ближайшей к началу координат точки из выпуклой оболочки конечного числа заданных точек [1, 2, 3]. В докладе [4] предложено естественное обобщение МДМ-алгоритма, которое позволяет решать задачу строгого линейного отделения двух конечных множеств в евклидовом пространстве с наибольшей шириной разделяющей полосы. Такую задачу будем называть SVM-задачей. На самом деле, имеется еще одно, нестандартное обобщение МДМ-алгоритма для решения SVM-задачи. Оно называется SMO-алгоритмом [5,6,7]. В SMO-алгоритме также используется возмущение плана с помощью двух слагаемых и аналогично определяется оценка плана. Данный доклад посвящен детальному анализу построения SMO-алгоритма.

Следует отметить, что SVM-задача сводится к задаче квадратичного программирования. Для ее решения можно использовать конечные методы, например, метод дополнительного базиса [8]. Однако каждая итеарция такого метода довольно сложная, при переходе от итерации к итерации накапливается погрешность. При большом объеме исходных данных это может не привести к требуемому результату.

Альтернативой конечным методам являются бесконечные методы, у которых каждая итерация предельно проста и при переходе от итерации к итерации погрешность не накапливается. МДМ-алгоритм и его обобщения формируют именно такие методы. Важно отметить, что, при условии сходимости, бесконечные методы за конечное число итераций позволяют получить приближенное решение SVM-задачи с требуемой точностью.

---

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» `http://oml.cmlaboratory.com/`

1°. В пространстве $\mathbb{R}^n$ заданы два конечных множества

$$P_1 = \{p_j\}_{j=1}^{s} \quad \text{и} \quad P_2 = \{p_j\}_{j=s+1}^{m},$$

где $s \in 1 : m-1$. Будем считать, что выпуклые оболочки G_1 и G_2 этих множеств не имеют общих точек. Положим $\xi_j = 1$ при $j \in 1 : s$, $\xi_j = -1$ при $j \in (s+1) : m$. Рассмотрим SVM-задачу [9]:

$$\begin{gathered} \tfrac{1}{2}\,\|w\|^2 \to \min, \\ \xi_j\big(\langle w, p_j\rangle + \beta\big) \geqslant 1, \quad j \in 1 : m. \end{gathered} \tag{1}$$

В силу условия $G_1 \cap G_2 = \varnothing$ эта задача имеет решение (w_*, β_*) и оно единственно.

Запишем двойственную задачу, предварительно поменяв знак у ее целевой функции:

$$\begin{gathered} D(v,u) := \tfrac{1}{2}\,\|v\|^2 - \sum_{j=1}^{m} u[j] \to \min, \\ v = \sum_{j=1}^{m} u[j]\xi_j p_j, \\ \sum_{j=1}^{m} u[j]\xi_j = 0, \\ u[j] \geqslant 0, \quad j \in 1 : m. \end{gathered} \tag{2}$$

По первой теореме двойственности в квадратичном программировании задача (2) также имеет решение (v_*, u_*) [10, глава III, параграф 10]. По второй теореме двойственности справедливы соотношения

$$\begin{gathered} w_* = v_*, \\ u_*[j]\Big(\xi_j\big(\langle w_*, p_j\rangle + \beta_*\big) - 1\Big) = 0, \quad j \in 1 : m. \end{gathered} \tag{3}$$

Обозначим через A матрицу со столбцами $\xi_1 p_1, \xi_2 p_2, \ldots, \xi_m p_m$ и через U — множество векторов $u \in \mathbb{R}^m$, удовлетворяющих условиям

$$\begin{gathered} \sum_{j=1}^{m} u[j]\xi_j = 0, \\ u[j] \geqslant 0, \quad j \in 1 : m. \end{gathered}$$

Отметим, что множество планов задачи (2) составляют пары (v, u), у которых $v = Au$, $u \in U$.

SMO-алгоритм решает двойственную задачу (2). По формулам (3) восстанавливается решение прямой задачи (1).

2°. Возьмем план (v, u) двойственной задачи (2). Рассмотрим возмущение специального вида $\widetilde{v}$ вектора v:

$$\widetilde{v} = v + \delta' \xi_{i'} p_{i'} + \delta'' \xi_{i''} p_{i''}.$$

Оно порождается вектором $\widetilde{u}$ с компонентами

$$\widetilde{u}[i] = \begin{cases} u[i'] + \delta' & \text{при } i = i', \\ u[i''] + \delta'' & \text{при } i = i'', \\ u[i] & \text{при остальных } i. \end{cases}$$

Легко проверить, что $\widetilde{v} = A\widetilde{u}$.

Мы заинтересованы в том, чтобы пара $(\widetilde{v}, \widetilde{u})$ была планом двойственной задачи и чтобы выполнялось неравенство $D(\widetilde{v}, \widetilde{u}) < D(v, u)$.

Для справедливости включения $\widetilde{u} \in U$ необходимо, в частности, чтобы сумма $\sum_{i=1}^{m} \xi_i \widetilde{u}[i]$ равнялась нулю. Это приводит к условию

$$\xi_{i'} \delta' + \xi_{i''} \delta'' = 0,$$

согласно которому

$$\delta'' = -\xi_{i'} \xi_{i''} \delta'.$$

Если обозначить $\delta = \delta'$, то получим

$$\widetilde{u}[i] = \begin{cases} u[i'] + \delta & \text{при } i = i', \\ u[i''] - \xi_{i'} \xi_{i''} \delta & \text{при } i = i'', \\ u[i] & \text{при остальных } i. \end{cases}$$

При этом вектор $\widetilde{v}$ принимает вид

$$\widetilde{v} = v + \delta \xi_{i'} \big(p_{i'} - p_{i''}\big).$$

Запишем разложение целевой функции двойственной задачи (2) в точке $(\widetilde{v}, \widetilde{u})$ по степеням параметра δ:

$$D(\widetilde{v}, \widetilde{u}) = \tfrac{1}{2} \left\| \widetilde{v} \right\|^2 - \sum_{i=1}^{m} \widetilde{u}[i] =$$

$$= \tfrac{1}{2} \left\| v + \delta \xi_{i'} \big(p_{i'} - p_{i''}\big) \right\|^2 - \sum_{i=1}^{m} u[i] - \delta + \xi_{i'} \xi_{i''} \delta =$$

$$= \tfrac{1}{2} \left\| v \right\|^2 + \delta \xi_{i'} \langle v, p_{i'} - p_{i''} \rangle + \tfrac{1}{2} \delta^2 \left\| p_{i'} - p_{i''} \right\|^2 - \sum_{i=1}^{m} u[i] - \delta + \xi_{i'} \xi_{i''} \delta =$$

$$= D(v,u) + \delta\xi_{i'}\langle v, p_{i'} - p_{i''}\rangle + \tfrac{1}{2}\delta^2 \left\|p_{i'} - p_{i''}\right\|^2 - \delta\xi_{i'}\big(\xi_{i'} - \xi_{i''}\big) =$$

$$= D(v,u) + \delta\xi_{i'}\Big(\langle v, p_{i'} - p_{i''}\rangle - \big(\xi_{i'} - \xi_{i''}\big)\Big) + \tfrac{1}{2}\delta^2 \left\|p_{i'} - p_{i''}\right\|^2.$$

Введем обозначение

$$\Delta(i', i'') = \langle v, p_{i''} - p_{i'}\rangle - \big(\xi_{i''} - \xi_{i'}\big) = \big(\langle v, p_{i''}\rangle - \xi_{i''}\big) - \big(\langle v, p_{i'}\rangle - \xi_{i'}\big). \tag{4}$$

С его помощью перепишем последнюю формулу более компактно

$$\Phi(\delta) := D(\widetilde{v}, \widetilde{u}) = D(v,u) - \delta\xi_{i'}\Delta(i', i'') + \tfrac{1}{2}\delta^2 \left\|p_{i'} - p_{i''}\right\|^2. \tag{5}$$

Функция $\Phi(\delta)$ является квадратным трехчленом относительно δ. Желая уменьшить значение $D(\widetilde{v}, \widetilde{u})$, найдем точку минимума функции $\Phi(\delta)$. Из уравнения $\Phi'(\delta) = 0$ находим

$$\delta = \xi_{i'} \frac{\Delta(i', i'')}{\left\|p_{i'} - p_{i''}\right\|^2}.$$

Подставим это значение δ в формулу для $D(\widetilde{v}, \widetilde{u})$. Получим

$$D(\widetilde{v}, \widetilde{u}) = D(v,u) - \frac{1}{2} \frac{\big[\Delta(i', i'')\big]^2}{\left\|p_{i'} - p_{i''}\right\|^2}.$$

Теперь займемся увеличением значения $\Delta(i', i'')$ за счет выбора индексов i' и i''. Рассмотрим предварительный вариант, когда эти индексы находятся из условий

$$\langle v, p_{i''}\rangle - \xi_{i''} = \max_{i \in 1:m}\big\{\langle v, p_i\rangle - \xi_i\big\},$$
$$\langle v, p_{i'}\rangle - \xi_{i'} = \min_{i \in 1:m}\big\{\langle v, p_i\rangle - \xi_i\big\}.$$

Обозначим

$$\widetilde{\lambda} = \frac{\Delta(i', i'')}{\left\|p_{i'} - p_{i''}\right\|^2}.$$

Тогда $\delta = \xi_{i'}\widetilde{\lambda}$ и, согласно (5),

$$D(\widetilde{v}, \widetilde{u}) = D(v,u) - \frac{1}{2}\widetilde{\lambda}\Delta(i', i'').$$

При этом

$$\widetilde{u}[i] = \begin{cases} u[i'] + \xi_{i'}\widetilde{\lambda} & \text{при } i = i', \\ u[i''] - \xi_{i''}\widetilde{\lambda} & \text{при } i = i'', \\ u[i] & \text{при остальных } i. \end{cases}$$

Остается обеспечить неотрицательность компонент вектора $\widetilde{u}$. Для этого прежде всего уточним выбор индексов i' и i''. При выборе i' исключим те $i \in 1:m$, при которых $u[i]=0$ и $\xi_i=-1$, то есть $i \in (s+1):m$ с $u[i]=0$. Аналогично при выборе i'' исключим те $i \in 1:m$, при которых $u[i]=0$ и $\xi_i=1$, то есть $i \in 1:s$ с $u[i]=0$. Введем индексные множества

$$\begin{aligned} I'(u) &= \big\{ i \in 1:s \big\} \cup \big\{ i \in (s+1):m \mid u[i]>0 \big\}, \\ I''(u) &= \big\{ i \in 1:s \mid u[i]>0 \big\} \cup \big\{ i \in (s+1):m \big\}. \end{aligned}$$

Найдем i' и i'' из уточненных условий

$$\begin{aligned} \langle v, p_{i''} \rangle - \xi_{i''} &= \max_{i \in I''} \big\{ \langle v, p_i \rangle - \xi_i \big\}, \\ \langle v, p_{i'} \rangle - \xi_{i'} &= \min_{i \in I'} \big\{ \langle v, p_i \rangle - \xi_i \big\}. \end{aligned} \tag{6}$$

Более того, для неотрицательности компонент вектора $\widetilde{u}$ придется, вообще говоря, уменьшить $\widetilde{\lambda}$, заменив его следующей величиной:

$$\lambda = \begin{cases} \min\{\widetilde{\lambda}, u[i']\}, & \text{если } i' \in (s+1):m \text{ и } i'' \in (s+1):m, \\ \min\{\widetilde{\lambda}, u[i'']\}, & \text{если } i' \in 1:s \text{ и } i'' \in 1:s, \\ \min\{\widetilde{\lambda}, u[i'], u[i'']\}, & \text{если } i' \in (s+1):m \text{ и } i'' \in 1:s, \\ \widetilde{\lambda}, & \text{если } i' \in 1:s,\ i'' \in (s+1):m. \end{cases}$$

Приходим к окончательным результатам

$$\begin{aligned} \widetilde{u}[i] &= \begin{cases} u[i'] + \xi_{i'}\lambda & \text{при } i = i', \\ u[i''] - \xi_{i''}\lambda & \text{при } i = i'', \\ u[i] & \text{при остальных } i \in 1:m; \end{cases} \\ \widetilde{v} &= v + \lambda\big(p_{i'} - p_{i''}\big), \end{aligned}$$

где индексы i' и i'' определяются формулами (6). Формула (4) определяет величину $\Delta(i', i'')$. В дальнейшем вместо $\Delta(i', i'')$ будем писать $\Delta(v, u)$.

Приведем без доказательства принципиальный результат — критерий оптимальности для двойственной задачи (2) в терминах $\Delta(v, u)$.

ТЕОРЕМА 1. *Для любого плана (v, u) задачи (2) величина $\Delta(v, u)$ неотрицательна. Равенство $\Delta(v, u) = 0$ выполняется тогда и только тогда, когда (v, u) — оптимальный план.*

3°. Переходим к описанию SMO-алгоритма решения задачи (2).

В качестве начального приближения берем план (v_0, u_0), где $v_0 = Au_0$ и u_0 — произвольный вектор с неотрицательными компонентами, удовлетворяющий простому условию

$$\sum_{j=1}^{s} u_0[j] = \sum_{j=s+1}^{m} u_0[j]$$

(случай $u_0 = \mathbb{O}$ не исключается).

Пусть уже имеется k-е приближение — план (v_k, u_k). Обозначим

$$\begin{aligned} I'(u_k) &= \big\{i \in 1:s\big\} \cup \big\{i \in (s+1):m \mid u_k[i] > 0\big\}, \\ I''(u_k) &= \big\{i \in 1:s \mid u_k[i] > 0\big\} \cup \big\{i \in (s+1):m\big\}. \end{aligned}$$

Найдем индексы $i'_k \in I'(u_k)$ и $i''_k \in I''(u_k)$, такие, что

$$\begin{aligned} \langle v_k, p_{i'_k}\rangle - \xi_{i'_k} &= \min_{i \in I'(u_k)} \big\{\langle v_k, p_i\rangle - \xi_i\big\}, \\ \langle v_k, p_{i''_k}\rangle - \xi_{i''_k} &= \max_{i \in I''(u_k)} \big\{\langle v_k, p_i\rangle - \xi_i\big\}. \end{aligned}$$

Вычислим величину

$$\Delta_k := \Delta(v_k, u_k) = \langle v_k, p_{i''_k} - p_{i'_k}\rangle - \big(\xi_{i''_k} - \xi_{i'_k}\big).$$

Если $\Delta_k = 0$, то по теореме 1 пара (v_k, u_k) является решением задачи (2). Процесс закончен.

Пусть $\Delta_k > 0$. Вычисляем

$$\widetilde{\lambda}_k = \frac{\Delta_k}{\big\|p_{i'_k} - p_{i''_k}\big\|^2}$$

и

$$\lambda_k = \begin{cases} \min\{\widetilde{\lambda}_k, u[i'_k]\}, & \text{если } i'_k \in (s+1):m \text{ и } i''_k \in (s+1):m, \\ \min\{\widetilde{\lambda}_k, u[i''_k]\}, & \text{если } i'_k \in 1:s \text{ и } i''_k \in 1:s, \\ \min\{\widetilde{\lambda}_k, u[i'_k], u[i''_k]\}, & \text{если } i'_k \in (s+1):m \text{ и } i''_k \in 1:s, \\ \widetilde{\lambda}_k, & \text{если } i'_k \in 1:s,\ i''_k \in (s+1):m. \end{cases}$$

Очевидно, что $\lambda_k > 0$. Если $\lambda_k = \widetilde{\lambda}_k$, то k-ю итерацию назовем *неусеченной*, если же $\lambda_k < \widetilde{\lambda}_k$, то *усеченной*.

Сформируем очередное приближение (v_{k+1}, u_{k+1}) следующим образом:

$$u_{k+1}[i] = \begin{cases} u_k[i'_k] + \xi_{i'_k}\lambda_k & \text{при } i = i'_k, \\ u_k[i''_k] - \xi_{i''_k}\lambda_k & \text{при } i = i''_k, \\ u_k[i] & \text{при остальных } i \in 1:m; \end{cases}$$

$$v_{k+1} = v_k + \lambda_k\big(p_{i'_k} - p_{i''_k}\big).$$

Описание SMO-алгоритма завершено.

Приведем без доказательства теорему о сходимости SMO-алгоритма.

ТЕОРЕМА 2. *Пусть* $G_1 \cap G_2 = \varnothing$ *и* $\Delta_k > 0$ *при всех* $k = 0, 1, \ldots$ *Тогда последовательность* $\{v_k\}$, *построенная SMO-алгоритмом, сходится к* w_* *— элементу оптимального плана задачи* (1).

Вопрос об определении β_* решается просто. При известном w_* ограничения задачи (1) принимают вид

$$\begin{aligned} \beta &\geqslant \ \ 1 - \langle w_*, p_j \rangle, \quad j \in 1:s, \\ \beta &\leqslant -1 - \langle w_*, p_j \rangle, \quad j \in (s+1):m, \end{aligned}$$

так что

$$\max_{j \in 1:s} \bigl\{ 1 - \langle w_*, p_j \rangle \bigr\} \leqslant \beta_* \leqslant \min_{j \in (s+1):m} \bigl\{ -1 - \langle w_*, p_j \rangle \bigr\}.$$

ЛИТЕРАТУРА

1. Митчелл Б. Ф., Демьянов В. Ф., Малоземов В. Н. *Нахождение ближайшей к началу координат точки многогранника* // Вестник ЛГУ. 1971. № 19. С. 38–45.

2. B. F. Mitchell, V. F. Dem'yanov, and V. N. Malozemov. *Finding the point of a polyhedron closest to the origin.* SIAM J. Control, 1974. Vol. 12, No. 1, pp. 19–26.

3. Малоземов В. Н. *МДМ-методу — 50 лет* // Семинар «CNSA & NDO». Избранные доклады. 10 ноября 2021 г. (`http://cnsa.cmlaboratory.com/reps21.shtml#1110`)

4. Малоземов В. Н., Соловьева Н. А. *МДМ-метод для решения общей квадратичной задачи математической диагностики*// Семинар «O&ML». Избранные доклады. 1 июня 2022 г. (`http://oml.cmlaboratory.com/reps22.shtml#0601`)

5. J. C. Platt. *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines.* Technical Report MSR-TR-98-14. April 21, 1998.

6. J. Lopez, J. R. Dorronsoro. *A Simple Proof of the Convergence of the SMO Algorithm for Linearly Separable Problems.* ICANN 2009, Part I, LNCS 5768, pp. 904–912, 2009.

7. J. L. Lazaro. *Analysis and Convergence of SMO-like Decomposition and Geometrical Algorithms for Support Vector Machines.* A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy. Universidad Autonoma de Madrid, 2011. 252 pp.

8. Даугавет В. А. *Численные методы квадратичного программирования.* СПб.: Изд-во СПбГУ, 2004. 128 с.

9. Малозёмов В. Н., Плоткин А. В. *SVM-метод строгого линейного отделения двух конечных множеств*// Семинар «O&ML». Избранные доклады. 30 марта 2022 г. (`http://oml.cmlaboratory.com/reps22.shtml#0330`)

10. Гавурин М. К., Малозёмов В. Н. *Экстремальные задачи с линейными ограничениями.* Л.: Изд-во ЛГУ, 1984. 176 с.