ПРОСТЕЙШАЯ МИНИМАКСНАЯ ЗАДАЧА, СВЯЗАННАЯ С МАШИННЫМ ОБУЧЕНИЕМ*

В. Н. Малоземов А. В. Плоткин
v.malozemov@spbu.ru avplotkin@gmail.com

17 апреля 2025 г.

1°. Пусть в пространстве $\mathbb{R}^n$ заданы N векторов

$$x_1, x_2, \ldots, x_N. \tag{1}$$

Рассмотрим минимаксную задачу

$$\varphi(w) := \max_{x \in 1:N} \langle x_i, w \rangle \to \min_{w \in S}, \tag{2}$$

где S — единичная сфера в $\mathbb{R}^n$, $S = \{w \in \mathbb{R}^n \mid \|w\| = 1\}$. К этой частной задаче сводится общая задача линейного отделения (как жесткого, так и мягкого) двух конечных множеств в евклидовом пространстве [1, пункт 4°].

2°. Отметим некоторые свойства функции $\varphi(w)$. Она выпукла на $\mathbb{R}^n$. Обозначим через C выпуклую оболочку векторов (1). Очевидно, что

$$\varphi(w) = \max_{z \in C} \langle z, w \rangle.$$

ЛЕММА 1. *Для того, чтобы при всех $w \in S$ выполнялось неравенство $\varphi(w) \geqslant 0$, необходимо и достаточно, чтобы $\mathbf{0} \in C$.*

Доказательство. **Необходимость.** Допустим противное, что $\mathbf{0} \notin C$. Обозначим точку из C, ближайшую к началу координат, через z_0. Имеем

$$\langle z, z_0 \rangle \geqslant \langle z_0, z_0 \rangle \text{ при всех } z \in C.$$

Поделив это неравенство на $-\|z_0\|$, получим

$$\langle z, -z_0 \,/\, \|z_0\| \rangle \leqslant -\|z_0\| \text{ при всех } z \in C. \tag{3}$$

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

Вектор $w_0 = -z_0 \,/\, \|z_0\|$ принадлежит S. При $z = z_0$ неравенство (3) обращается в равенство. Значит, $\varphi(w_0) = -\|z_0\|$, что противоречит условию леммы.

Достаточность. Допустим, вопреки утверждению, что $\varphi(w_1) < 0$ при некотором $w_1 \in S$, то есть, что

$$\max_{z\in C}\langle z, w_1\rangle < 0.$$

Но это противоречит условию $\mathbf{0} \in C$.

Лемма доказана. $\square$

СЛЕДСТВИЕ. *Неравенство $\varphi(w_2) < 0$ при некотором $w_2 \in S$ выполняется тогда и только тогда, когда $\mathbf{0} \notin C$.*

Очевидно, что задача (2) имеет решение. Обозначим его через w_*, и пусть $\mu = \varphi(w_*)$.

ЛЕММА 2. *При $\mu < 0$ решение w_* задачи (2) единственно.*

Доказательство. Допустим противное, что существует два различных решения w_1 и w_2 задачи (2). Так как $\|w_1\| = \|w_2\| = 1$, то

$$\langle w_1, w_2\rangle < 1. \tag{4}$$

Возьмем любое $\alpha_0 \in (0, 1)$, при котором вектор

$$w_0 = \alpha_0 w_1 + (1-\alpha_0) w_2$$

отличен от нулевого. Покажем, что $\|w_0\| < 1$. Действительно, в силу (4) имеем

$$\|w_0\|^2 = \|\alpha_0 w_1 + (1-\alpha_0)w_2\|^2 = \alpha_0^2 + 2\alpha_0(1-\alpha_0)\langle w_1, w_2\rangle + (1-\alpha_0)^2 <$$
$$< \alpha_0^2 + 2\alpha_0(1-\alpha_0) + (1-\alpha_0)^2 = 1.$$

Обозначим $\beta = \|w_0\|$. Вектор $\frac{1}{\beta}w_0$ принадлежит S. Воспользуемся выпуклостью функции $\varphi(w)$ и неравенством $\frac{1}{\beta} > 1$. Получим

$$\varphi(\tfrac{1}{\beta}w_0) = \tfrac{1}{\beta}\varphi(w_0) \leqslant \tfrac{1}{\beta}\big(\alpha_0\varphi(w_1) + (1-\alpha_0)\varphi(w_2)\big) = \tfrac{1}{\beta}\mu < \mu.$$

Это противоречит определению и отрицательности μ.

Лемма доказана. $\square$

3°. Если $\mu < 0$, то по следствию из леммы 1 будет $\mathbf{0} \notin C$. Обозначим через z_0 точку из C, ближайшую к началу координат (к нулевому вектору).

ТЕОРЕМА 1. *При $\mu < 0$ справедливы формулы*

$$w_* = -z_0 \,/\, \|z_0\|\,, \quad \mu = -\,\|z_0\|\,.$$

Доказательство. В силу выпуклости и замкнутости множества C при всех $z \in C$ имеем

$$\langle z, z_0\rangle \geqslant \langle z_0, z_0\rangle.$$

Поделив это неравенство на $-\,\|z_0\|$, получим

$$\langle z, -z_0 \,/\, \|z_0\|\rangle \leqslant -\,\|z_0\| \text{ при всех } z \in C. \tag{5}$$

Вектор $w_0 = -z_0 \,/\, \|z_0\|$ принадлежит S, при этом, как следует из (5), справедливо равенство

$$\varphi(w_0) = -\,\|z_0\|\,.$$

Остается отметить, что при всех $w \in S$ выполняется неравенство

$$\varphi(w_0) \leqslant \varphi(w).$$

Действительно, так как $\langle z_0, -w\rangle \leqslant \|z_0\|$, то

$$\varphi(w_0) = -\,\|z_0\| \leqslant \langle z_0, w\rangle \leqslant \max_{z\in C}\langle z, w\rangle = \varphi(w). \tag{6}$$

На основании (6) заключаем, что w_0 — решение задачи (2). В силу единственности решения, $w_* = w_0$.

Теорема доказана. □

4°. Предположим, что $\mu \geqslant 0$. В этом случае $\varphi(w) \geqslant 0$ при всех $w \in S$. Согласно лемме 1, $\mathbf{0} \in C$. Обозначим через r радиус максимального шара с центром в начале координат, вписанного в выпуклую оболочку C, и через z_0 точку внутреннего касания этого шара поверхности многогранника C. Очевидно, что $r = \|z_0\|$.

ТЕОРЕМА 2. *При $\mu \geqslant 0$ справедливы равенства $\mu = r$ и $w_* = z_0 \,/\, \|z_0\|$.*

Доказательство. По условию шар $B_r = \{w \in \mathbb{R}^n \mid \|w\| \leqslant r\}$ содержится в C. При всех $w \in S$ имеем

$$r = \max_{z\in B_r}\langle z, w\rangle \leqslant \max_{z\in C}\langle z, w\rangle = \varphi(w).$$

Отсюда следует, что $r \leqslant \mu$.

Покажем, что $B_\mu \subset C$. В противном случае существует вектор $z_1 \in B_\mu$, который не принадлежит C. Обозначим через z_2 точку из C, ближайшую к z_1. Для всех $z \in C$ получим

$$\langle z - z_1, z_2 - z_1 \rangle \geqslant \langle z_2 - z_1, z_2 - z_1 \rangle.$$

Поделим это неравенство на $-\|z_2 - z_1\|$. Придем к эквивалентному неравенству

$$\langle z - z_1, w_1 \rangle \leqslant -\|z_2 - z_1\|,$$

где $w_1 = -(z_2 - z_1) \,/\, \|z_2 - z_1\|$. Последнее неравенство справедливо при всех $z \in C$, поэтому

$$\varphi(w_1) \leqslant \langle z_1, w_1 \rangle - \|z_2 - z_1\| \leqslant \|z_1\| - \|z_2 - z_1\|.$$

Напомним, что $z_1 \in B_\mu$, то есть $\|z_1\| \leqslant \mu$. Значит,

$$\varphi(w_1) \leqslant \mu - \|z_2 - z_1\| < \mu.$$

Это противоречит определению μ.

Итак, $B_\mu \subset C$. По определению r имеем $r \geqslant \mu$. Вместе с ранее доказанным неравенством $r \leqslant \mu$ приходим к равенству $\mu = r$.

Для вектора $w_* = z_0 \,/\, \|z_0\|$ по построению имеем

$$\max_{z \in C} \langle z, w_* \rangle = \langle z_0, w_* \rangle = \langle r w_*, w_* \rangle = r = \mu.$$

Это значит, что вектор w_* является решением задачи (2).

Теорема доказана. $\square$

Отметим, что при $\mu \geqslant 0$ решение задачи (2), вообще говоря, не единственно.

Геометрические идеи, которые использовались при доказательстве теорем 1 и 2 являются достоянием общей теории минимаксных задач [2, глава III, § 3].

5°. Теперь рассмотрим в пространстве $\mathbb{R}^n$ два конечных множества

$$P_1 = \{a_i\}_{i \in I_1} \quad \text{и} \quad P_2 = \{a_j\}_{j \in I_2}.$$

В докладе [1, пункт 4°] предложена следующая формализация задачи о наилучшем линейном отделении таких множеств:

$$\varphi(w) := \max_{i \in I_1} \langle w, a_i \rangle - \min_{j \in I_2} \langle w, a_j \rangle \to \min_{w \in S}. \tag{7}$$

Нас интересует мягкое отделение. Покажем, как задача (7) связана с мягким линейным отделением множеств P_1 и P_2, и как она сводится к простейшей минимаксной задаче вида (2).

Очевидно, что задача (7) имеет решение w_*. При этом

$$\varphi(w_*) = \langle w_*, a_{i'} - a_{j'} \rangle,$$

где $i' \in I_1$ и $j' \in I_2$ — индексы, на которых достигаются максимум и минимум в определении $\varphi(w_*)$. Обозначим через L_1 и L_2 гиперплоскости, определяемые уравнениями

$$\langle w_*, x \rangle = \langle w_*, a_{i'} \rangle \quad \text{и} \quad \langle w_*, x \rangle = \langle w_*, a_{j'} \rangle$$

соответственно. Справедливо следующее утверждение о мягком отделении множеств P_1 и P_2.

ТЕОРЕМА 3. *Пусть $\varphi(w_*) > 0$. Тогда полосе, ограниченной гиперплоскостями L_1 и L_2, принадлежат некоторые точки как множества P_1, так и множества P_2. Ширина смешанной полосы равна $\varphi(w_*)$ и является наименьшей из возможных. Полупространство $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle > \langle w_*, a_{i'} \rangle\}$ содержит точки множества P_2 и не содержит точки множества P_1, а полупространство $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle < \langle w_*, a_{j'} \rangle\}$ содержит точки множества P_1 и не содержит точки множества P_2.*

Рисунок иллюстрирует содержание теоремы.

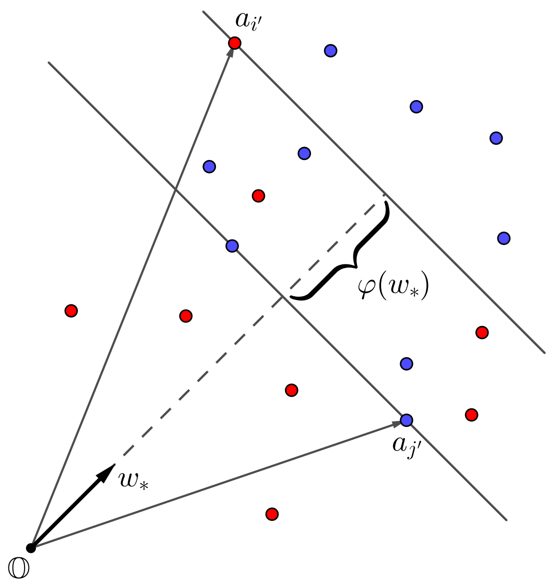

Рис. 1. Иллюстрация к теореме

6°. Покажем, как задача (7) сводится к задаче вида (2).

Равенство
$$\min_{j\in I_2}\langle w,a_j\rangle = -\max_{j\in I_2}\langle w,-a_j\rangle$$
позволяет представить задачу (7) в симметричной форме
$$\varphi(w) := \max_{i\in I_1}\langle w,a_i\rangle + \max_{j\in I_2}\langle w,-a_j\rangle \to \min_{w\in S}. \tag{8}$$

Преобразуем функцию $\varphi(w)$. Воспользуемся тем, что из максимума можно вынести слагаемое, не зависящее от индекса, по которому берется максимум. Запишем
$$\varphi(w) = \max_{i\in I_1}\big\{\langle w,a_i\rangle + \max_{j\in I_2}\langle w,-a_j\rangle\big\} =$$
$$= \max_{i\in I_1}\max_{j\in I_2}\langle w,a_i-a_j\rangle.$$
Разности $a_i - a_j$, $i\in I_1$, $j\in I_2$, линейно упорядочим. Получим последовательность вида (1). После этого задача (2) будет эквивалентной задаче (8), а значит и задаче (7).

7°. Приведем алгоритм поиска приближенного решения задачи (7). Запишем задачу в упрощенном виде:
$$\varphi(w) = \max_{i\in I_1}\max_{j\in I_2}\langle w,a_i-a_j\rangle \to \min_{w\in S}.$$
Возьмем начальное приближение $w_0\in S$. Пусть уже имеется k-е приближение w_k. Опишем k-ю итерацию алгоритма, на которой происходит построение w_{k+1}.

Обозначим
$$I_1^k = \{i\in I_1 \mid \langle w_k,a_i\rangle = \max_{i\in I_1}\langle w_k,a_i\rangle\} \quad \text{и} \quad I_2^k = \{j\in I_2 \mid \langle w_k,a_j\rangle = \min_{j\in I_2}\langle w_k,a_j\rangle\}$$
и рассмотрим множество векторов
$$X_k = \{a_i - a_j \mid i\in I_1^k, j\in I_2^k\}.$$
Через C_k обозначим выпуклую оболочку векторов X_k.

Из общей теории минимаксных задач [2, глава III, §3] известно, что, если $\mathbf{0}\in C_k$, то точка w_k является минимумом задачи безусловной минимизации
$$\varphi(w) \to \min_{w\in\mathbb{R}^n}. \tag{9}$$
А так как $w_k\in S$, то найдено точное решение задачи (2). Алгоритм завершается.

В противном случае, найдем направление *наискорейшего спуска* из точки w_k в задаче (9). Оно определяется как решение задачи

$$\max_{x \in X_k} \langle x, w \rangle \to \min_{w \in S}. \tag{10}$$

Так как $\mathbf{0} \notin C_k$, то воспользуемся теоремой 1 для нахождения решения этой задачи. Точку из C_k, ближайшую к началу координат, будем искать МДМ-алгоритмом [3].

Обозначим через g_k решение задачи (10). Спроектируем этот вектор на касательное пространство множества S в точке w_k. Получим направление

$$d_k = g_k - \langle g_k, w_k \rangle w_k.$$

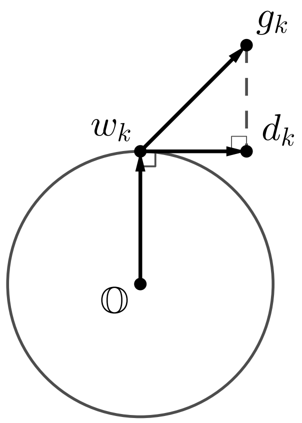

Рис. 2. Проектирование направления наискорейшего спуска

Если $d_k = \mathbf{0}$, то вектор w_k является точкой локального минимума в задаче (2). Вычисления прекращаются.

В противном случае, найдем следующее приближение w_{k+1} по следующей итеративной процедуре. Положим $t_k = 1$. Рассмотрим вектор $v_k = w_k + t_k \frac{d_k}{\|d_k\|}$. Спроектируем его на S и получим вектор $u_k = \frac{v_k}{\|v_k\|}$. Если $\varphi(u_k) < \varphi(w_k)$, то положим $w_{k+1} = u_k$ и перейдем к следующей итерации нашего алгоритма. Иначе, уменьшим t_k вдвое и повторно выполним процедуру.

Описание алгоритма завершено.

8°. Перейдем к численным экспериментам.

Покажем, как можно выбрать начальное приближение $w_0 \in S$ в общем случае. Для этого рассмотрим широкоизвестную формализацию задачи о мягком отделении множеств P_1 и P_2 [4]:

$$\begin{gathered}
\tfrac{1}{2}\,\|w\|^2 + C\Bigl(\sum_{i\in I_1}\eta_i + \sum_{j\in I_2}\eta_j\Bigr) \to \min, \\
\langle w, a_i\rangle + \beta \geqslant 1 - \eta_i, \quad i \in I_1, \\
\langle w, a_j\rangle + \beta \leqslant \eta_j - 1, \quad j \in J_1 \\
\eta_i \geqslant 0,\ i \in I_1, \quad \eta_j \geqslant 0,\ j \in I_2,
\end{gathered} \tag{11}$$

где C — штрафной множитель. Пусть w'_* — решение этой задачи. Тогда в качестве начального приближения для нашего алгоритма возьмем $w_0 = w'_* \ / \ \|w'_*\|$.

Отметим, что нахождение решения задачи (11) не составляет большой сложности, так как тема глубоко изучена, и разработаны эффективные программные реализации [5].

На рисунках 3–5 представлены примеры разделения точек на плоскости, слева — с использованием w_0, а справа — на основе полученного алгоритмом решения w_*.

ЛИТЕРАТУРА

1. Малоземов В. Н., Соловьева Н. А. *Новые формализации линейной задачи отделения двух конечных множеств* // Семинар «O&ML». Избранные доклады. 5 декабря 2024 г.
(`http://oml.cmlaboratory.com/reps24.shtml#1205`)

2. Демьянов В. Ф. Малоземов В. Н. *Введение в минимакс.* М.: Наука, 1972. 368 с.

3. Малоземов В. Н. *МДМ-методу — 50 лет* // Семинар «O&ML». Литература. Публикации участников семинара.
(`http://oml.cmlaboratory.com/literatura.shtml`)

4. Малоземов В. Н., Плоткин А. В. *SVM-метод: мягкое отделение* // Семинар «O&ML». Избранные доклады. 30 марта 2022 г.
(`http://oml.cmlaboratory.com/reps22.shtml#0406`)

5. Chang C.-C., Lin C.-J. *LIBSVM: A library for support vector machines.* ACM Transactions on Intelligent Systems and Technology. 2 (3): 27 (2011), pp. 1–27.

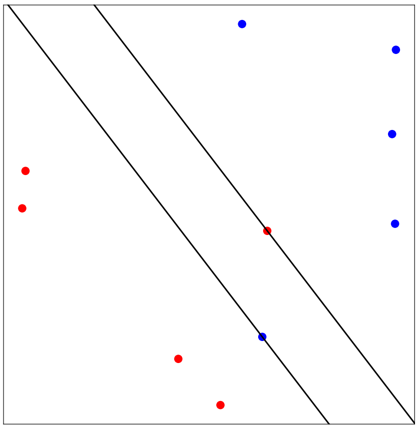

Рис. 3а. $\varphi(w_0) \approx 1.40$

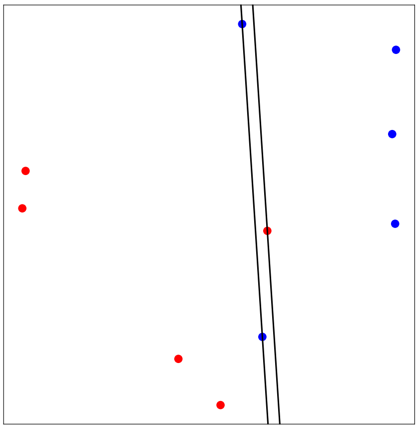

Рис. 3б. $\varphi(w_*) \approx 0.24$

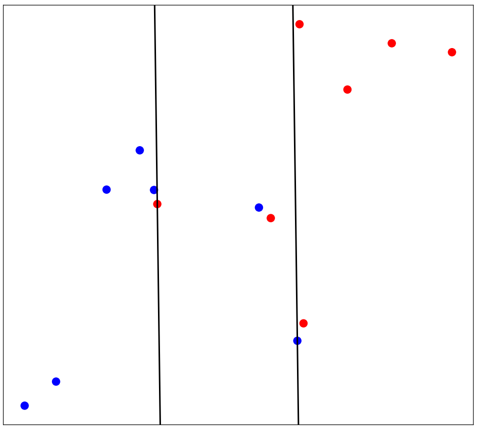

Рис. 4а. $\varphi(w_0) \approx 2.90$

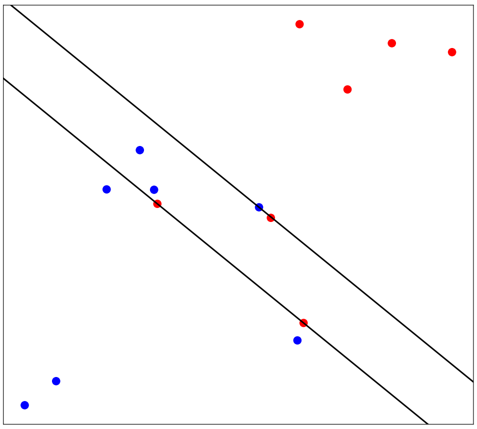

Рис. 4б. $\varphi(w_*) \approx 1.29$

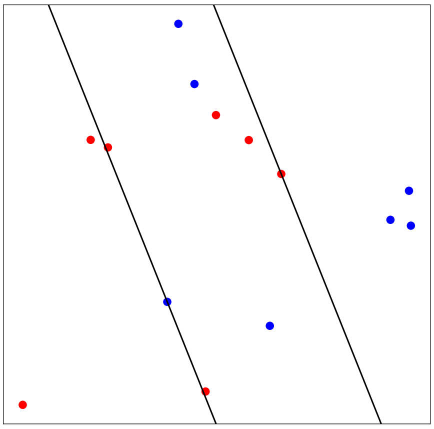

Рис. 5а. $\varphi(w_0) \approx 3.61$

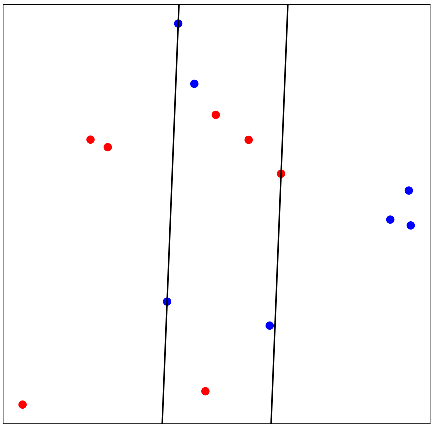

Рис. 5б. $\varphi(w_*) \approx 2.56$