СТРОГОЕ ЛИНЕЙНОЕ ОТДЕЛЕНИЕ ДВУХ КОНЕЧНЫХ МНОЖЕСТВ И ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ.*

В. Н. Малоземов — v.malozemov@spbu.ru

А. В. Плоткин — avplotkin@gmail.com

24 февраля 2022 г.

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

$$P = \{p_j\}_{j=1}^{m_1} \quad \text{и} \quad Q = \{q_j\}_{j=1}^{m_2}.$$

Они называются строго линейно отделимыми, если найдутся ненулевой вектор $w \in \mathbb{R}^n$ и число $\beta \in \mathbb{R}$, такие, что

$$\begin{aligned} \langle w, p_j \rangle + \beta &\geqslant 1 \quad \text{для всех} \quad j \in 1 : m_1, \\ \langle w, q_j \rangle + \beta &\leqslant -1 \quad \text{для всех} \quad j \in 1 : m_2. \end{aligned} \tag{1}$$

При выполнении условий (1) говорят также, что гиперплоскость, определяемая уравнением $\langle w, x \rangle + \beta = 0$, строго отделяет множество P от Q.

Задача строгого линейного отделения множеств P и Q заключается в решении системы линейных неравенств (1). В докладе будет показано, что решение этой системы можно свести к решению трех вариантов задачи линейного программирования.

2°. Следуя [1, 2], введем функцию

$$f(g) = \frac{1}{m_1} \sum_{j=1}^{m_1} \big[-\langle w, p_j \rangle - \beta + 1 \big]_+ + \frac{1}{m_2} \sum_{j=1}^{m_2} \big[\langle w, q_j \rangle + \beta + 1 \big]_+, \tag{2}$$

где $g = (w, \beta)$ и $[u]_+ = \max\{0, u\}$. Ясно, что $f(g) \geqslant 0$ при всех $g \in \mathbb{R}^{n+1}$.

ПРЕДЛОЖЕНИЕ 1. *Множества P и Q строго линейно отделимы тогда и только тогда, когда существует вектор g_*, на котором $f(g_*) = 0$.*

---

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» `http://www.apmath.spbu.ru/oml/`

Доказательство основано на том, что условие $f(g) = 0$ выполняется только в том случае, когда все слагаемые в правой части формулы (2) равны нулю. Последнее же условие равносильно совместности системы (1), так как равенство $[u]_+ = 0$ эквивалентно неравенству $u \leqslant 0$.

Напомним, что функция $f(g)$ неотрицательна. Предложение 1 показывает, что строгое линейное отделение множеств P и Q сводится к минимизации функции $f(g)$.

Рассмотрим экстремальную задачу

$$f(g) \to \min_{g \in \mathbb{R}^{n+1}}, \tag{3}$$

где $f(g)$ — функция вида (2). Эта задача эквивалентна следующей задаче линейного программирования:

$$\begin{gathered} F(w, \beta, y, z) := \frac{1}{m_1} \sum_{j=1}^{m_1} y_j + \frac{1}{m_2} \sum_{j=1}^{m_2} z_j \to \min, \\ \langle w, p_j \rangle + \beta + y_j \geqslant 1, \quad j \in 1 : m_1, \\ -\langle w, q_j \rangle - \beta + z_j \geqslant 1, \quad j \in 1 : m_2, \\ y_j \geqslant 0,\ j \in 1 : m_1, \quad z_j \geqslant\ 0,\ j \in 1 : m_2. \end{gathered} \tag{4}$$

Множество планов задачи (4) непусто (содержит план с компонентами $w = \mathbb{O}$, $\beta = 0$, $y_j = 1$ при всех $j \in 1 : m_1$ и $z_j = 1$ при всех $j \in 1 : m_2$) и целевая функция ограничена снизу нулем. Значит, задача (4) имеет решение. По эквивалентности существует решение и у задачи (3), причем минимальные значения целевых функция у этих задач равны между собой. Их общее значение обозначим через μ. Отметим, что если (w_*, β_*, y_*, z_*) — решение задачи (4), то $g = (w_*, \beta_*)$ — решение задачи (3) с $f(g_*) = \mu$.

При $\mu = 0$ гиперплоскость, определяемая уравнением $\langle w_*, x \rangle + \beta_* = 0$, строго отделяет множество P от множества Q.

При $\mu > 0$ строгое линейное отделение множества P от множества Q невозможно.

Обозначим через A_1 матрицу со строками $p_1, p_2, \ldots, p_{m_1}$ и через A_2 — матрицу со строками $q_1, q_2, \ldots, q_{m_2}$. Тогда матрицу ограничений задачи (4) можно представить в виде

$$\left(\begin{array}{c:c:c} & 1 & \\ A_1 & \vdots & \\ & 1 & \\ \hdashline & -1 & E_m \\ -A_2 & \vdots & \\ & -1 & \end{array} \right),$$

где E_m — единичная матрица порядка $m = m_1 + m_2$.

3°. Следуя [3], введем функцию

$$\varphi(g) = \max\left\{\max_{j\in 1:m_1}\left[-\langle w, p_j\rangle - \beta + 1\right]_+, \max_{j\in 1:m_2}\left[\langle w, q_j\rangle + \beta + 1\right]_+\right\}. \tag{5}$$

Ясно, что $\varphi(g) \geqslant 0$ при всех $g \in \mathbb{R}^{n+1}$.

ПРЕДЛОЖЕНИЕ 2. *Множества P и Q строго линейно отделимы тогда и только тогда, когда существует вектор g_0, на котором $\varphi(g_0) = 0$.*

Предложение 2 показывает, что строгое линейное отделение множеств P и Q сводится к минимизации функции $\varphi(g)$.

Рассмотрим экстремальную задачу

$$\varphi(g) \to \min_{g\in\mathbb{R}^{n+1}}, \tag{6}$$

где $\varphi(g)$ — функция вида (5). Эта задача эквивалентна следующей задаче линейного программирования

$$\begin{gathered} t \to \min, \\ \langle w, p_j\rangle + \beta + t \geqslant 1, \quad j \in 1:m_1, \\ -\langle w, q_j\rangle - \beta + t \geqslant 1, \quad j \in 1:m_2, \\ t \geqslant 0. \end{gathered} \tag{7}$$

Очевидно, что множество планов задачи (7) непусто и целевая функция ограничена снизу нулем. Значит, задача (7) имеет решение.

По эквивалентности имеет решение и задача (6), причем минимальные значения целевых функций у этих задач равны между собой. Их общее значение обозначим через t_0. Отметим, что если (w_0, β_0, t_0) — решение задачи (7), то $g_0 = (w_0, \beta_0)$ — решение задачи (6) с $\varphi(g_0) = t_0$.

При $t_0 = 0$ гиперплоскость, определяемся уравнением $\langle w_0, x\rangle + \beta_0 = 0$, строго отделяет множество P от множества Q.

В работе [3] показано, что t_0 может принимать только два значение — ноль или единицу. Таким образом, при $t_0 = 1$ строгое линейное отделение множества P от множества Q невозможно.

Матрица ограничений задачи (7) имеет представление

$$\left(\begin{array}{c:c:c} & 1 & 1 \\ A_1 & \vdots & \vdots \\ & 1 & 1 \\ \hdashline & -1 & 1 \\ -A_2 & \vdots & \vdots \\ & -1 & 1 \end{array}\right).$$

4°. Построим задачу линейного программирования, исходя из других соображений. Потребуем, чтобы полоса, разделяющая множества P и Q, имела наибольшую ℓ_1-ширину.

Обозначим через L_1 и L_2 гиперплоскости, определяемые уравнениями

$$\langle w, x\rangle + \beta = 1 \quad \text{и} \quad \langle w, x\rangle + \beta = -1.$$

ПРЕДЛОЖЕНИЕ 3. *ℓ_1-ширина полосы, ограниченной параллельными гиперплоскостями L_1 и L_2, равна* $\dfrac{2}{\|w\|_\infty}$.

Доказательство. Введем величину

$$d = \min_{x_1\in L_1,\, x_2\in L_2} \|x_1 - x_2\|_1 .$$

Нужно показать, что $d = \dfrac{2}{\|w\|_\infty}$.

Для любых $x_1 \in L_1$, $x_2 \in L_2$ имеем

$$2 = \langle w, x_1 - x_2\rangle \leqslant \|w\|_\infty \cdot \|x_1 - x_2\|_1 .$$

Отсюда следует, что

$$\|x_1 - x_2\|_1 \geqslant \frac{2}{\|w\|_\infty}. \tag{8}$$

Остается проверить, что при некоторых $x_1 \in L_1$, $x_2 \in L_2$ неравенство (8) выполняется как равенство.

Обозначим через $i_0 \in 1 : n$ индекс, на котором

$$\bigl|w[i_0]\bigr| = \max_{i\in 1:n}\bigl|w[i]\bigr|.$$

Положим $\eta = \dfrac{2}{w[i_0]}e_{i_0}$, где e_{i_0} — орт в пространстве $\mathbb{R}^n$. Возьмем произвольный вектор $x_2 \in L_2$ и вектор $x_1 = x_2 + \eta$. Запишем

$$\langle w, x_1\rangle + \beta = \langle w, x_2\rangle + \beta + \langle w, \eta\rangle = -1 + \langle w, \eta\rangle = 1.$$

Получили, что $x_1 \in L_1$. При этом

$$\|x_1 - x_2\|_1 = \|\eta\|_1 = \frac{2}{\bigl|w[i_0]\bigr|} = \frac{2}{\|w\|_\infty}.$$

Предложение доказано. $\square$

5°. Формально задача о строгом линейном отделении множеств P и Q с наибольшей ℓ_1-шириной разделяющей полосы ставится следующим образом:

$$\begin{gathered} \|w\|_\infty \to \min, \\ \langle w, p_j\rangle + \beta \geqslant 1, \quad j \in 1:m_1, \\ -\langle w, q_j\rangle - \beta \geqslant 1, \quad j \in 1:m_2. \end{gathered} \tag{9}$$

Эта задача эквивалентна задаче линейного программирования относительно вектора $\xi = (w, \beta, t)$:

$$\begin{gathered} t \to \min, \\ \langle w, p_j\rangle + \beta \geqslant 1, \quad j \in 1:m_1, \\ -\langle w, q_j\rangle - \beta \geqslant 1, \quad j \in 1:m_2, \\ -w + te \geqslant \mathbb{O}, \\ w + te \geqslant \mathbb{O}. \end{gathered} \tag{10}$$

Здесь e — n-мерный вектор, все координаты которого равны единице.

Матрицу ограничений задачи (10) можно представить в виде

$$\left(\begin{array}{c:c:c} & 1 & 0 \\ A_1 & \vdots & \vdots \\ & 1 & 0 \\ \hdashline & -1 & 0 \\ -A_2 & \vdots & \vdots \\ & -1 & 0 \\ \hdashline & 0 & 1 \\ -E_n & \vdots & \vdots \\ & 0 & 1 \\ \hdashline & 0 & 1 \\ E_n & \vdots & \vdots \\ & 0 & 1 \end{array}\right).$$

Запишем двойственную задачу

$$\sum_{j=1}^{m_1} u_j + \sum_{j=1}^{m_2} v_j \to \max,$$

$$\left(\begin{array}{c:c:c:c} A_1^T & -A_2^T & -E_n & E_n \\ \hdashline 1 \ \dots \ 1 & -1 \ \dots \ -1 & 0 \ \dots \ 0 & 0 \ \dots \ 0 \\ 0 \ \dots \ 0 & 0 \ \dots \ 0 & 1 \ \dots \ 1 & 1 \ \dots \ 1 \end{array}\right) \begin{pmatrix} u \\ v \\ \lambda \\ \nu \end{pmatrix} = \left(\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hdashline 0 \\ 1 \end{array}\right), \tag{11}$$

$$u \geqslant \mathbb{O}, \quad v \geqslant \mathbb{O}, \quad \lambda \geqslant \mathbb{O}, \quad \nu \geqslant \mathbb{O}.$$

Задача (11) имеет каноническую форму. Именно ее и нужно решать модифицированным симплекс методом.

Сделаем три замечания.

1) При построении начального базисного плана можно ограничиться двумя неотрицательными искусственными переменными. Достаточно в качестве начального базиса индексов $N_+^{(0)}$ взять множество

$$(m+n+1) : (m+2n+2),$$

к правому блоку матрицы ограничений добавить указанные ниже два столбца и учесть, что

$$\left(\begin{array}{c:c:c} & 0 & 0 \\ E_n & \vdots & \vdots \\ & 0 & 0 \\ \hdashline 0 \ \dots \ 0 & 1 & 0 \\ \hdashline 1 \ \dots \ 1 & 1 & 1 \end{array}\right)^{-1} = \left(\begin{array}{c:c:c} & 0 & 0 \\ E_n & \vdots & \vdots \\ & 0 & 0 \\ \hdashline 0 \ \dots \ 0 & 1 & 0 \\ \hdashline -1 \ \dots \ -1 & -1 & 1 \end{array}\right).$$

2) Если при некотором k у вектора $z_k[N_+^{(k)}]$ нет положительных компонент, то целевая функция задачи (11) не ограничена сверху. Целевая функция t задачи (10) ограничена снизу нулем. Из неограниченности целевой функции двойственной задачи (11) следует, что множество планов задачи (10) пусто. Пустым будет и множество планов задачи (9). Иначе при большом t получили бы план задачи (10).

3) По решению задачи (11) решение задачи (10) восстанавливается с помощью условий дополнительности. Используется обратная базисная матрица, соответствующая оптимальному базисному плану задачи (11).

6°. На базе `MATLAB` была разработана демонстрационная система строгого линейного отделения двух конечных множеств в трехмерном пространстве. Данная система включает три части: создание тестовых данных, решение задач линейного программирования, визуализацию полученных результатов.

Наибольший интерес представляет первая часть. Все тестовые точки берутся из куба $[-1,1]^3$. На вход система получает три параметра: натуральные m_1, m_2 и неотрицательное h. Параметры m_1 и m_2 задают количество точек в множествах P и Q. Сами множества P и Q формируются следующим образом.

В рассматриваемом кубе случайно выбираются три точки, не лежащие на одной прямой. Через них проводится плоскость L. Эта плоскость в дальнейшем будет разделять множества P и Q. Параметр h определяет евклидово расстояние до плоскости L, в пределах которого должны отсутствовать точки обоих множеств. Таким образом, мы гарантируем, что разделяющая полоса будет иметь евклидову ширину не меньше $2h$.

Далее поступаем так. Выбираем случайную точку x из данного куба. Если евклидово расстояние от точки p до плоскости L меньше h, то точка пропускается. Иначе помещаем точку x в одно из множеств P или Q в зависимости от ее расположения относительно плоскости L при условии, что соответствующее множество еще не заполнено. В противном случае точка пропускается. Процесс повторяется, пока множества P и Q не будут содержать m_1 и m_2 точек соответственно.

Во второй части системы формируются матрицы ограничений задач линейного программирования (4), (7), (10) и осуществляется решение этих задач с помощью встроенной процедуры `linprog`.

Третья часть обеспечивает визуализацию построенных множеств и разделяющих плоскостей. Используются процедуры `scatter3` и `surf`.

7°. Приведем четыре примера, полученных с помощью демонстрационной системы. На всех иллюстрациях разделяющая плоскость красного цвета является решением задачи (4), зеленого цвета — задачи (7), синего цвета — задачи (10).

Напомним, что в системе используются только три параметра m_1, m_2 и h. Остальное делается автоматически.

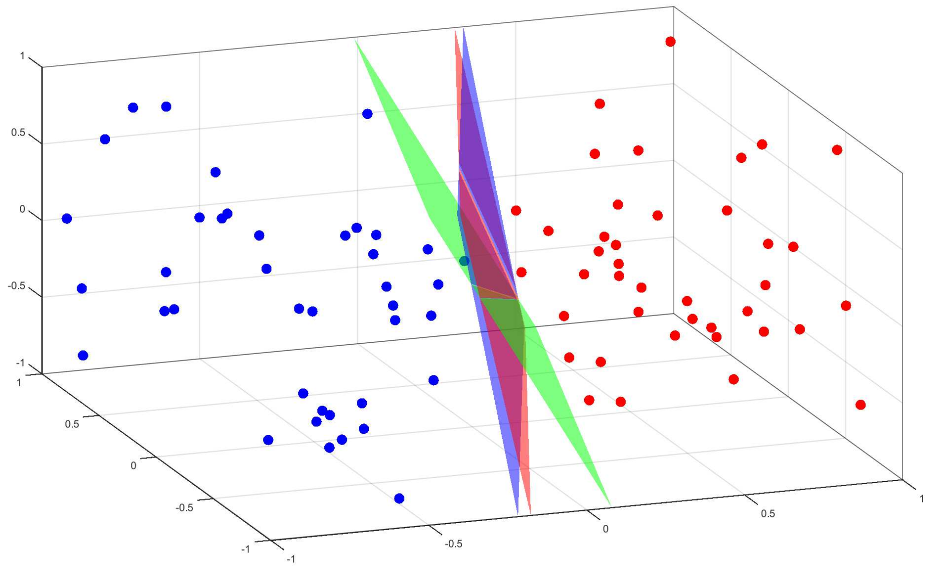

Рис. 1. $m_1 = m_2 = 40$, $h = 0.1$.

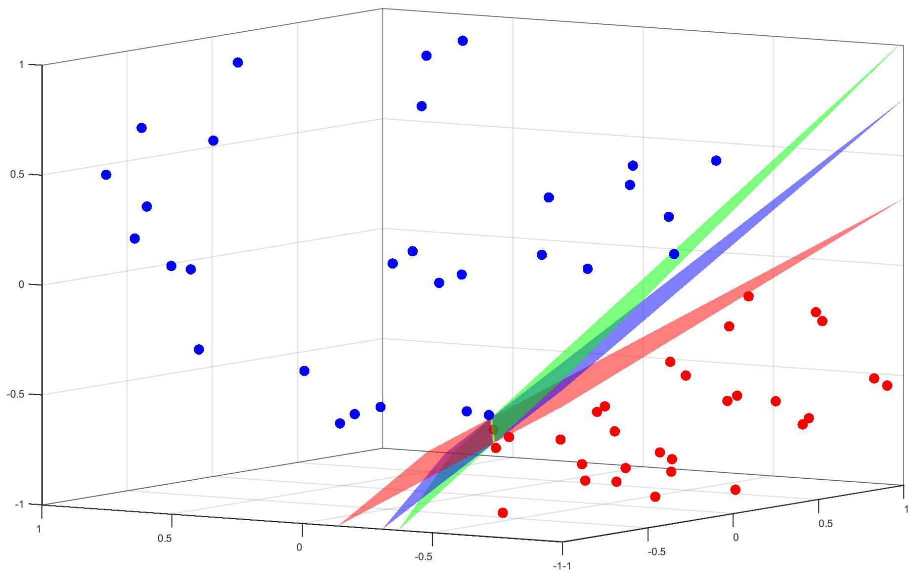

Рис. 2. $m_1 = m_2 = 30$, $h = 0$.

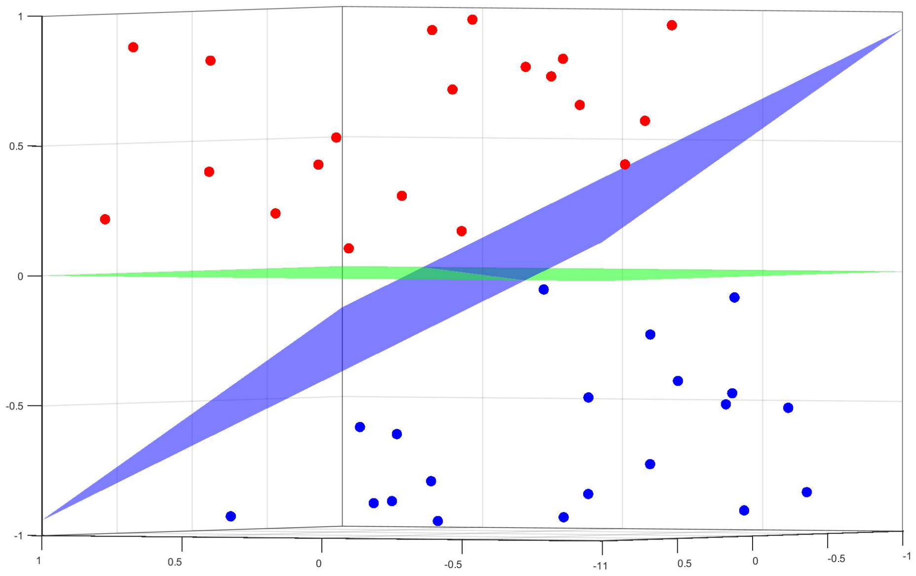

Рис. 3. $m_1 = m_2 = 20$, $h = 0.1$.
Решения задач (4) и (7) совпали ($z = 0$).

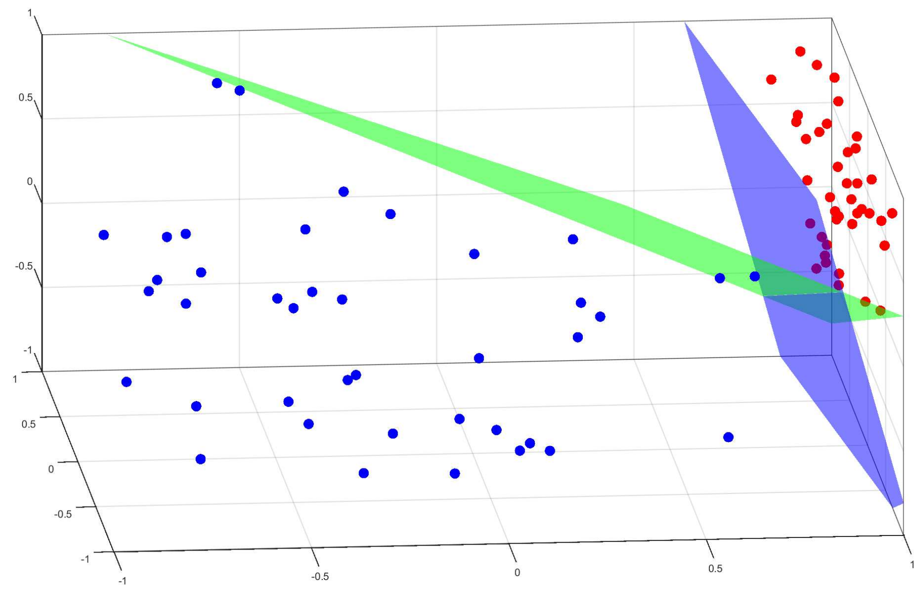

Рис. 4. $m_1 = m_2 = 40$, $h = 0.1$.
Решения задач (4) и (7) совпали.

В таблице 1 представлены данные об евклидовом расстоянии от полученных разделяющих плоскостей до множества точек $P \cup Q$ для каждого из рассмотренных примеров.

Таблица 1. Евклидово расстояние от разделяющих плоскостей до $P \cup Q$.

Пример	Задача (4)	Задача (7)	Задача (10)
1	0.128	0.081	**0.131**
2	**0.046**	0.038	0.041
3	0.059	0.059	**0.163**
4	0.027	0.027	**0.138**

ЛИТЕРАТУРА

1. Bennett K. P., Mangasarian O. L. *Robust linear programming discrimination of two linearly inseparable sets* // Optimization Methods and Software. 1992. Vol. 1. P. 23–34.

2. Малоземов В. Н., Чернэуцану К. К. *О математической диагностике (линейная модель)* // Семинар «DHA & CAGD». Избранные доклады. 17 апреля 2010 г. (`http://dha.spb.ru/reps10.shtml#0417`)

3. Малоземов В. Н., Плоткин А. В. *Строгое полиномиальное отделение двух множеств* // Вестник СПбГУ. Сер. 1. 2019. Т. 6 (64). Вып. 2. С. 232–240.