НОВЫЕ ФОРМАЛИЗАЦИИ ЛИНЕЙНОЙ ЗАДАЧИ ОТДЕЛЕНИЯ ДВУХ КОНЕЧНЫХ МНОЖЕСТВ*

В. Н. Малоземов v.malozemov@spbu.ru

Н. А. Соловьева 4vinyo@gmail.com

5 декабря 2024 г.

В докладе [1] рассмотрены три варианта сведения задачи строгого (жесткого) линейного отделения двух конечных множеств P_1 и P_2 евклидова пространства к задаче линейного программирования (ЛП). Полученные задачи ЛП различаются только целевыми функциями. Делается предположение о том, что выпуклые оболочки множеств P_1 и P_2 не имеют общих точек. Иначе множество планов задачи ЛП будет пустым.

В данном докладе предлагается еще один вариант сведения линейной задачи отделения множеств P_1 и P_2 к задаче ЛП. Новая задача всегда имеет решение. Экстремальное значение целевой функции неотрицательно и обращается в ноль тогда и только тогда, когда выпуклые оболочки множеств P_1 и P_2 имеют общую точку.

Вторая часть доклада посвящена универсальному подходу к задаче линейного отделения двух конечных множеств P_1 и P_2. Задача отделения сводится к экстремальной задаче, отличающейся от задачи ЛП появлением квадратичного ограничения-равенства. Новая задача всегда имеет решение. Если экстремальное значение целевой функции положительно, то получено мягкое отделение множеств P_1 и P_2 с наименьшей шириной смешанной полосы. Если экстремальное значение целевой функции отрицательно, то получено жесткое отделение множеств P_1 и P_2 с наибольшей шириной нейтральной полосы. Обычно к таким заключениям удается прийти, решая две различные задачи квадратичного программирования (см., например, [2, 3]).

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

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

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

Здесь $s \in 1 : (m-1)$. Обозначим

$$B = \left\{ w \in \mathbb{R}^n \mid -1 \leqslant w[i] \leqslant 1,\ i \in 1:n \right\}$$

и рассмотрим экстремальную задачу

$$\varphi(w) := \min_{j \in 1:s} \langle p_j, w \rangle - \max_{j \in (s+1):m} \langle p_j, w \rangle \to \max_{w \in B}. \tag{1}$$

Очевидно, что задача (1) имеет решение w^*. Принимая во внимание, что $\mathbf{0} \in B$ и $\varphi(\mathbf{0}) = 0$, заключаем, что $\varphi(w^*) \geqslant 0$.

Задача (1) имеет прямое отношение к строгому (жесткому) отделению множеств P_1 и P_2.

ТЕОРЕМА 1. *Если $\varphi(w^*) > 0$, то множества P_1 и P_2 допускают строгое линейное отделение. При этом ширина h^* разделяющей полосы равна следующей величине:*

$$h^* = \varphi(w^*) / \|w^*\|.$$

Отделяющая гиперплоскость определяется уравнением

$$\langle w^*, x - c \rangle = 0,$$

где $c = p_{j''} + \frac{1}{2} h^* \frac{w^*}{\|w^*\|}$ и j'' — индекс из $(s+1) : m$, на котором достигается максимум в представлении $\varphi(w^*)$.

Рис. 1 поясняет содержание теоремы 1.

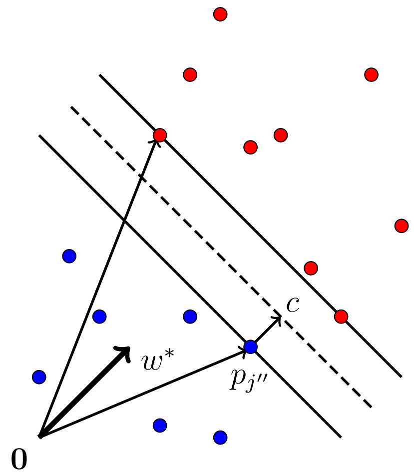

Рис 1. Пояснение к теореме 1

Случай $\varphi(w^*) = 0$ будет рассмотрен после некоторой подготовки.

2°. Воспользуемся формулой

$$\max_{j\in(s+1):m}\langle p_j, w\rangle = -\min_{j\in(s+1):m}\langle -p_j, w\rangle.$$

Она позволяет придать задаче (1) симметричный вид:

$$\varphi(w) := \min_{j\in 1:s}\langle p_j, w\rangle + \min_{j\in(s+1):m}\langle -p_j, w\rangle \to \max_{w\in B}. \tag{2}$$

Запишем задачу линейного программирования (ЛП), эквивалентную задаче (2):

$$\begin{aligned} \psi(w,u,v) &:= u + v \to \max, \\ \langle -p_j, w\rangle + u &\leqslant 0, \quad j \in 1:s, \\ \langle p_j, w\rangle + v &\leqslant 0, \quad j \in (s+1):m, \\ w &\leqslant e, \\ -w &\leqslant e. \end{aligned} \tag{3}$$

Здесь e — вектор из $\mathbb{R}^n$, все компоненты которого равны единице. Эквивалентность проверяется конструктивно. Действительно, обозначим через Ω множество планов задачи (3). Возьмем план (w,u,v). Вектор w является планом задачи (2). При этом

$$\varphi(w) \geqslant u + v = \psi(w,u,v).$$

Наоборот, пусть w — план задачи (2). Положим

$$u = \min_{j\in 1:s}\langle p_j, w\rangle, \quad v = \min_{j\in(s+1):m}\langle -p_j, w\rangle.$$

Тройка (w,u,v) будет планом задачи (3). При этом $\psi(w,u,v) = u + v = \varphi(w)$. Эквивалентность задач (2) и (3) установлена.

По лемме об эквивалентности [4, с. 11] справедливы следующие утверждения:

- если w^* — решение задачи (2), то (w^*,u^*,v^*), где

  $$u^* = \min_{j\in 1:s}\langle p_j, w^*\rangle, \quad v^* = \min_{j\in(s+1):m}\langle -p_j, w^*\rangle,$$

  — решение задачи (3);

- если (w^*,u^*,v^*) — решение задачи (3), то w^* — решение задачи (2);

- справедливо равенство

  $$\max_{w\in B}\varphi(w) = \max_{(w,u,v)\in\Omega}\psi(w,u,v). \tag{4}$$

Задача (3) всегда имеет решение (w^*,u^*,v^*). Если $\psi(w^*,u^*,v^*) > 0$, то вектор w^* является решением задачи (1), (2) и $\varphi(w^*) > 0$. По теореме 1 возможно строгое линейное отделение множеств P_1 и P_2.

3°. Обратимся к случаю $\varphi(w^*) = 0$.

ТЕОРЕМА 2. *Равенство* $\varphi(w^*) = 0$ *выполняется тогда и только тогда, когда выпуклые оболочки множеств* P_1 *и* P_2 *имеют общую точку.*

Доказательство. Пусть $\varphi(w^*) = 0$. Напомним, что w^* — решение задачи (1), (2). На основании (4) запишем

$$\max_{(w,u,v)\in\Omega} \psi(w,u,v) = \varphi(w^*) = 0, \tag{5}$$

то есть наибольшее значение целевой функции задачи (3) равно нулю. В таблице представлены данные задачи ЛП (3).

Таблица. Данные задачи (3)

	$0\dots0$	1	1	
α	$-p_1^T$	1	0	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	$-p_s^T$	1	0	0
	p_{s+1}^T	0	1	0
	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	p_m^T	0	1	0
β	E_n	$\mathbf{0}$	$\mathbf{0}$	e
γ	$-E_n$	$\mathbf{0}$	$\mathbf{0}$	e

Построим двойственную задачу:

$$\begin{gathered} \eta(\alpha,\beta,\gamma) := \sum_{i=1}^{n} \beta[i] + \sum_{i=1}^{n} \gamma[i] \to \min, \\ -\sum_{j=1}^{s} \alpha[j] p_j + \sum_{j=s+1}^{m} \alpha[j] p_j + \beta - \gamma = \mathbf{0}, \\ \sum_{j=1}^{s} \alpha[j] = 1, \quad \sum_{j=s+1}^{m} \alpha[j] = 1, \\ \alpha \geqslant \mathbf{0}, \quad \beta \geqslant \mathbf{0}, \quad \gamma \geqslant \mathbf{0}. \end{gathered} \tag{6}$$

Множество планов этой задачи обозначим Λ. По первой теореме двойственности задача (6) имеет решение $(\alpha_*, \beta_*, \gamma_*)$, причем

$$\eta(\alpha_*, \beta_*, \gamma_*) = \psi(w^*, u^*, v^*).$$

Согласно (5), получаем $\eta(\alpha_*, \beta_*, \gamma_*) = 0$. Но тогда

$$\sum_{i=1}^{n} \beta_*[i] + \sum_{i=1}^{n} \gamma_*[i] = 0,$$
$$\beta_* \geqslant \mathbf{0}, \quad \gamma_* \geqslant \mathbf{0}.$$

Это возможно только при $\beta_* = \mathbf{0}$, $\gamma_* = \mathbf{0}$. Из определения Λ следует, что

$$\sum_{j=1}^{s} \alpha_*[j] p_j = \sum_{j=s+1}^{m} \alpha_*[j] p_j,$$
$$\sum_{j=1}^{s} \alpha_*[j] = 1, \quad \sum_{j=s+1}^{m} \alpha_*[j] = 1,$$
$$\alpha_*[j] \geqslant 0, \quad j \in 1:m.$$

Данные соотношения показывают, что выпуклые оболочки множеств P_1 и P_2 имеют общую точку.

Наоборот, пусть выпуклые оболочки множеств P_1 и P_2 имеют общую точку. Тогда множество Λ планов двойственной задачи (6) содержит план (α, β, γ) с $\beta = \mathbf{0}$ и $\gamma = \mathbf{0}$. На нем значение целевой функции равно нулю. Вместе с тем, у прямой задачи (3) нулевой вектор является планом и на нем целевая функция также принимает нулевое значение. По второй теореме двойственности нулевой план будет оптимальным для задачи (3). Таким образом,

$$\max_{(w,u,v) \in \Omega} \psi(w, u, v) = 0.$$

Согласно (4),

$$\varphi(w^*) = \max_{w \in B} \varphi(w) = \max_{(w,u,v) \in \Omega} \psi(w, u, v) = 0,$$

что и требовалось доказать. □

Из теоремы 2 следует, что при $\varphi(w^*) = 0$ множества P_1 и P_2 не могут быть отделены с помощью гиперплоскости.

ПРИМЕР 1. На плоскости $\mathbb{R}^2$ рассмотрим два множества — P_1, состоящее из точек $(1, 0)$ и $(-1, 0)$, и P_2, состоящее из точек $(0, 1)$ и $(0, -1)$. Их нельзя строго отделить с помощью прямой, так как выпуклые оболочки этих множеств имеют общую точку $(0, 0)$.

ПРИМЕР 2. Пусть $s = 3$, $m = 6$. Сформируем в $\mathbb{R}^2$ множества $P_1 = \{p_1, p_2, p_3\}$ и $P_2 = \{p_4, p_5, p_6\}$, где

$$p_1 = \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad p_3 = \begin{pmatrix} 6 \\ -5 \end{pmatrix},$$
$$p_4 = \begin{pmatrix} 6 \\ -2 \end{pmatrix}, \quad p_5 = \begin{pmatrix} 11 \\ -2 \end{pmatrix}, \quad p_6 = \begin{pmatrix} 11 \\ 2 \end{pmatrix}.$$

Для них задача ЛП (3) принимает вид

$$\begin{gathered}\psi(w,u,v)=u+v\to\max,\\ -w[1]+4w[2]+u\leqslant 0,\\ -2w[1]-2w[2]+u\leqslant 0,\\ -6w[1]+5w[2]+u\leqslant 0,\\ 6w[1]-2w[2]+v\leqslant 0,\\ 11w[1]-2w[2]+v\leqslant 0,\\ 11w[1]+2w[2]+v\leqslant 0,\\ w[1]\leqslant 1,\quad -w[1]\leqslant 1,\\ w[2]\leqslant 1,\quad -w[2]\leqslant 1.\end{gathered}$$

Найдем решение этой задачи:

$$w^*=\left(-1,-\frac{4}{7}\right)^T,\quad u^*=-\frac{22}{7},\quad v^*=\frac{34}{7}.$$

Очевидно, что

$$\psi(w^*,u^*,v^*)=u^*+v^*=\frac{12}{7}.$$

Как отмечалось в разд. 2°, вектор $\left(-1,-\frac{4}{7}\right)^T$ является оптимальным планом задачи жесткого линейного отделения множеств P_1 и P_2. При этом

$$\varphi(w^*)=\psi(w^*,u^*,v^*)=\frac{12}{7}>0.$$

По теореме 1 ширина разделяющей полосы имеет следующее значение:

$$h^*=\frac{\varphi(w^*)}{\|w^*\|}=\frac{12}{\sqrt{65}}.$$

Для того, чтобы записать уравнение разделяющей прямой, нужно найти индекс i'', на котором достигается максимум скалярных произведений $\langle w^*,p_j\rangle$ при $j\in 4:6$, и вектор c. Получаем $i''=4$ и

$$c=p_4+\tfrac{1}{2}\,h^*\frac{w^*}{\|w^*\|}=\begin{pmatrix}348/65\\-154/65\end{pmatrix}.$$

После простых преобразований приходим к уравнению разделяющей прямой

$$x[2]=-\frac{7}{4}\,x[1]+7.$$

На рис. 2 представлен результат строгого отделения множеств P_1 и P_2.

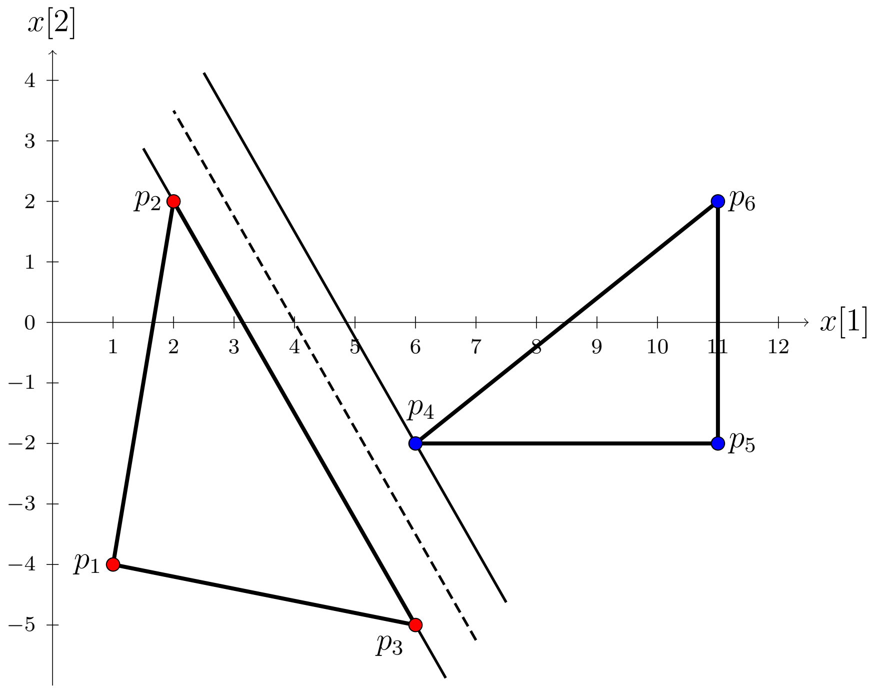

Рис. 2. Строгое линейное отделение

ПРИМЕР 3. Пусть $s = 3$, $m = 6$. Сформируем в $\mathbb{R}^2$ множества $P_1 = \{p_1, p_2, p_3\}$ и $P_2 = \{p_4, p_5, p_6\}$, где

$$p_1 = \begin{pmatrix} 1 \\ -4 \end{pmatrix}, \quad p_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad p_3 = \begin{pmatrix} 6 \\ -5 \end{pmatrix},$$
$$p_4 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \quad p_5 = \begin{pmatrix} 8 \\ -2 \end{pmatrix}, \quad p_6 = \begin{pmatrix} 8 \\ 2 \end{pmatrix}.$$

Для них задача ЛП (3) принимает вид

$$\begin{gathered} \psi(w,u,v) = u + v \to \max, \\ -w[1] + 4w[2] + u \leqslant 0, \\ -2w[1] - 2w[2] + u \leqslant 0, \\ -6w[1] + 5w[2] + u \leqslant 0, \\ 3w[1] - 2w[2] + v \leqslant 0, \\ 8w[1] - 2w[2] + v \leqslant 0, \\ 8w[1] + 2w[2] + v \leqslant 0, \\ w[1] \leqslant 1, \quad -w[1] \leqslant 1, \\ w[2] \leqslant 1, \quad -w[2] \leqslant 1. \end{gathered}$$

Найдем решение этой задачи:

$$w^* = \big(0, 0\big)^T, \quad u^* = 0, \quad v^* = 0.$$

Очевидно, что

$$\varphi(w^*) = \psi(w^*, u^*, v^*) = u^* + v^* = 0.$$

Как отмечалось в теореме 2, в этом случае множества P_1 и P_2 не могут быть разделены с помощью прямой. Рис. 3 подтверждает такой вывод.

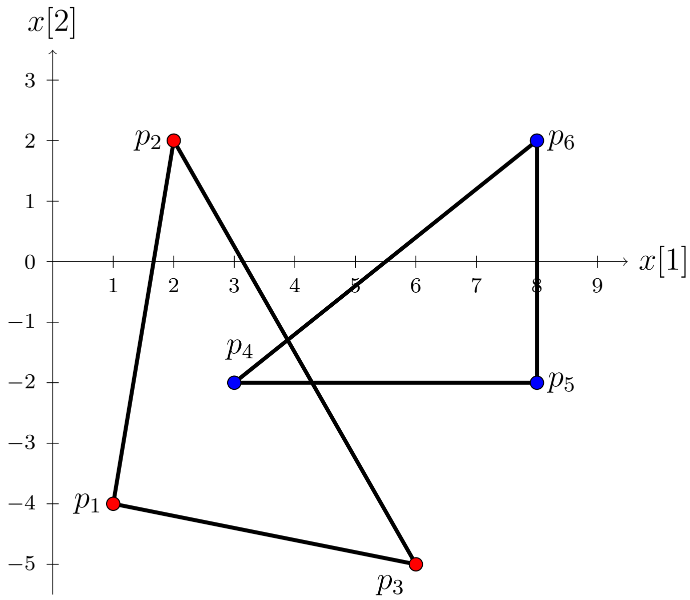

Рис. 3. Иллюстрация к примеру 3

4°. Обозначим $S = \{w \in \mathbb{R}^n \mid \langle w, w\rangle = 1\}$ и рассмотрим еще один вариант формализации задачи о линейном отделении двух конечных множеств P_1 и P_2. А именно,

$$\varphi(w) := \max_{j \in (s+1):m} \langle p_j, w\rangle - \min_{j \in 1:s} \langle p_j, w\rangle \to \min_{w \in S}. \tag{7}$$

Воспользуемся формулой

$$\min_{j \in 1:s} \langle p_j, w\rangle = -\max_{j \in 1:s} \langle -p_j, w\rangle.$$

С ее помощью задача (7) приводится к симметричному виду:

$$\varphi(w) := \max_{j \in (s+1):m} \langle p_j, w\rangle + \max_{j \in 1:s} \langle -p_j, w\rangle \to \min_{w \in S}. \tag{8}$$

Очевидно, что задача (7), (8) имеет решение w_*.

Запишем экстремальную задачу, эквивалентную задаче (8):

$$
\begin{aligned}
\psi(w,u,v) &:= u+v \to \min, \\
\langle -p_j, w\rangle + u &\geqslant 0, \quad j \in (s+1):m, \\
\langle p_j, w\rangle + v &\geqslant 0, \quad j \in 1:s, \\
\langle w, w\rangle &= 1.
\end{aligned}
\tag{9}
$$

Эквивалентность проверяется конструктивно. Действительно, возьмем план (w,u,v) задачи (9). Вектор w является планом задачи (8). При этом

$$\varphi(w) \leqslant u+v = \psi(w,u,v).$$

Наоборот, пусть $w \in S$. Положим

$$u = \max_{j\in(s+1):m} \langle p_j, w\rangle, \qquad v = \max_{j\in 1:s} \langle -p_j, w\rangle.$$

Тройка (w,u,v) является планом задачи (9). При этом

$$\psi(w,u,v) = u+v = \varphi(w).$$

Эквивалентность задач (8) и (9) установлена.

По эквивалентноости, задача (9), так же, как и задача (8), имеет решение. Кроме того,

- если w_* — решение задачи (7), то (w_*, u_*, v_*), где

  $$u_* = \max_{j\in(s+1):m} \langle p_j, w_*\rangle, \qquad v_* = \max_{j\in 1:s} \langle -p_j, w_*\rangle,$$

  — решение задачи (9),

- если (w_*, u_*, v_*) — решение задачи (9), то w_* — решение задачи (7),(8),

- справедливо равенство

  $$\min_{w\in S} \varphi(w) = \min_{(w,u,v)\in\Omega} \psi(w,u,v), \tag{10}$$

  где Ω — множество планов задачи (9).

5°. Допустим, что удалось найти решение (w_*, u_*, v_*) задачи (9) (методы решения этой задачи в докладе не обсуждаются). Какой можно сделать вывод об отделимости множеств P_1 и P_2? Это зависит от величины $\psi(w_*, u_*, v_*)$, которая в силу (10) равна $\varphi(w_*)$. Отметим, что

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

где $j'' \in (s+1) : m$ и $j' \in 1 : s$ — индексы, на которых достигаются максимум и минимум в представлении $\varphi(w_*)$ (см. (7)). Обозначим через L_1 и L_2 гиперплоскости, определяемые уравнениями

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

соответственно.

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 $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle > \langle w_*, p_{j''} \rangle\}$ содержит точки множества P_1 и не содержит точек множества P_2, а полупространство $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle < \langle w_*, p_{j'} \rangle\}$ содержит точки множества P_2 и не содержит точек множества P_1.*

Доказательство. Принять во внимание, что согласно (7) справедливо равенство

$$\max_{j \in (s+1):m} \langle w_*, p_j \rangle = \min_{j \in 1:s} \langle w_*, p_j \rangle + \varphi(w_*).$$

ТЕОРЕМА 4. *Пусть $\varphi(w_*) < 0$. Тогда в открытой полосе, ограниченной гиперплоскостями L_2 и L_1, отсутствуют точки как множества P_1, так и множества P_2. Ширина разделяющей полосы равна $|\varphi(w_*)|$ и является наибольшей из возможных. Полупространство $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle \geqslant \langle w_*, p_{j'} \rangle\}$ содержит все точки множества P_1, а полупространство $\{x \in \mathbb{R}^n \mid \langle w_*, x \rangle \leqslant \leqslant \langle w_*, p_{j''} \rangle\}$ содержит все точки множества P_2.*

Доказательство. Принять во внимание, что согласно (7) справедливо равенство

$$\min_{j \in 1:s} \langle w_*, p_j \rangle = \max_{j \in (s+1):m} \langle w_*, p_j \rangle + \big|\varphi(w_*)\big|.$$

Осталось выяснить, чему соответствует условие $\varphi(w_*) = 0$, которое в подробной записи выглядит так:

$$\min_{j \in 1:s} \langle w_*, p_j \rangle = \max_{j \in (s+1):m} \langle w_*, p_j \rangle =: \mu.$$

ТЕОРЕМА 5. *Пусть* $\varphi(w_*) = 0$. *Тогда гиперплоскости* $\{x \in \mathbb{R}^n \mid \langle w_*, x\rangle = \mu\}$ *принадлежат некоторые точки как множества* P_1, *так и множества* P_2. *Полупространство* $\{x \in \mathbb{R}^n \mid \langle w_*, x\rangle > \mu\}$ *содержит точки множества* P_1 *и не содержит точек множества* P_2. *Полупространство* $\{x \in \mathbb{R}^n \mid \langle w_*, x\rangle < \mu\}$ *содержит точки множества* P_2 *и не содержит точек множества* P_1.

Замечание. Теорема 3 связана с мягким линейным отделением множеств P_1 и P_2, когда появляется смешанная полоса. Теорема 4 демонстрирует результат нестандартного подхода к жесткому отделению.

ЛИТЕРАТУРА

1. Малоземов В. Н., Плоткин А. В. *Строгое линейное отделение двух конечных множеств и линейное программирование* // Семинар «O&ML». Избранные доклады. 24 февраля 2022 г. (`http://oml.cmlaboratory.com/reps22.shtml#0224`)

2. Малоземов 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`)

3. Малоземов В. Н., Плоткин А. В. *SVM-метод: мягкое отделение* // Семинар «O&ML». Избранные доклады. 6 апреля 2022 г. (`http://oml.cmlaboratory.com/reps22.shtml#0406`)

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