НАИЛУЧШЕЕ ЛИНЕЙНОЕ ОТДЕЛЕНИЕ ДВУХ МНОЖЕСТВ*

В. Н. Малозёмов
malv@math.spbu.ru

Е. К. Чернэуцану
katerinache@yandex.ru

6 октября 2012 г.

На линейном уровне исследуется задача наилучшего приближённого отделения двух конечных множеств. Эта задача сводится к задаче негладкой оптимизации, при анализе которой используется вся мощь теории линейного программирования.

В идейном плане мы следуем работе [1].

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

$$A = \{a_i\}_{i=1}^m \quad \text{и} \quad B = \{b_j\}_{j=1}^k.$$

Множества A и B называются *строго отделимыми*, если существуют ненулевой вектор $w \in \mathbb{R}^n$ и вещественное число γ, такие, что

$$\langle w, a_i \rangle < \gamma \quad \text{при всех } i \in 1:m, \tag{1}$$

$$\langle w, b_j \rangle > \gamma \quad \text{при всех } j \in 1:k. \tag{2}$$

При выполнении условий (1) и (2) говорят также, что гиперплоскость H, определяемая уравнением $\langle w, x \rangle = \gamma$, *строго отделяет* множество A от множества B.

2°. Введём функцию

$$f(g) = \frac{1}{m} \sum_{i=1}^m \big[\langle w, a_i \rangle - \gamma + c\big]_+ + \frac{1}{k} \sum_{j=1}^k \big[-\langle w, b_j \rangle + \gamma + c\big]_+, \tag{3}$$

где $g = (w, \gamma)$, $c > 0$ — параметр и $[u]_+ = \max\{0, u\}$. Ясно, что $f(g) \geqslant 0$ при всех g.

*Семинар по дискретному гармоническому анализу и геометрическому моделированию «DHA & CAGD»: `http://www.dha.spb.ru/`

ТЕОРЕМА 1. *Множества A и B строго отделимы тогда и только тогда, когда существует вектор g_*, на котором $f(g_*) = 0$.*

Доказательство. Пусть $f(g_*) = 0$ на некотором векторе $g_* = (w_*, \gamma_*)$. Покажем прежде всего, что $w_* \neq \mathbb{O}$. В противном случае

$$f(g_*) = (-\gamma_* + c)_+ + (\gamma_* + c)_+ = \begin{cases} -\gamma_* + c & \text{при } \gamma_* \leqslant -c, \\ \ \ 2c & \text{при } \gamma_* \in [-c, c], \\ \ \ \gamma_* + c & \text{при } \gamma_* \geqslant c. \end{cases}$$

Отсюда следует, что $f(g_*) \geqslant 2c$. Это противоречит условию $f(g_*) = 0$.

Далее, условие $f(g_*) = 0$ гарантирует, что все слагаемые

$$\big[\langle w_*, a_i\rangle - \gamma_* + c\big]_+ \quad \text{и} \quad \big[-\langle w_*, b_j\rangle + \gamma_* + c\big]_+$$

равны нулю. Это возможно лишь тогда, когда

$$\langle w_*, a_i\rangle - \gamma_* + c \leqslant 0 \quad \text{при всех } i \in 1:m, \tag{4}$$

$$-\langle w_*, b_j\rangle + \gamma_* + c \leqslant 0 \quad \text{при всех } j \in 1:k. \tag{5}$$

Остаётся отметить, что неравенства (4) и (5) обеспечивают выполнение условий строгой отделимости (1) и (2) с $w = w_*$, $\gamma = \gamma_*$.

Докажем обратное утверждение. Пусть выполнены условия (1) и (2). Обозначим

$$d := \min_{j\in 1:k}\langle w, b_j\rangle - \max_{i\in 1:m}\langle w, a_i\rangle > 0, \tag{6}$$

$$w_* = \left(\tfrac{2c}{d}\right)w, \quad \gamma_* = \tfrac{1}{2}\Big[\min_{j\in 1:k}\langle w_*, b_j\rangle + \max_{i\in 1:m}\langle w_*, a_i\rangle\Big].$$

Согласно (6) и определению w_*

$$\min_{j\in 1:k}\langle w_*, b_j\rangle - \max_{i\in 1:m}\langle w_*, a_i\rangle = 2c.$$

Имеем

$$\max_{i\in 1:m}\langle w_*, a_i\rangle = 2\gamma_* - \min_{j\in 1:k}\langle w_*, b_j\rangle = 2\gamma_* - 2c - \max_{i\in 1:m}\langle w_*, a_i\rangle,$$

так что

$$\max_{i\in 1:m}\langle w_*, a_i\rangle = \gamma_* - c. \tag{7}$$

Аналогично

$$\min_{j\in 1:k}\langle w_*, b_j\rangle = 2\gamma_* - \max_{i\in 1:m}\langle w_*, a_i\rangle = 2\gamma_* + 2c - \min_{j\in 1:k}\langle w_*, b_j\rangle,$$

так что

$$\min_{j\in 1:k}\langle w_*, b_j\rangle = \gamma_* + c. \tag{8}$$

Положим $g_* = (w_*, \gamma_*)$. На основании (7) и (8) получим $f(g_*) = 0$.

Теорема доказана. $\square$

3°. При доказательстве теоремы 1 описано преобразование вектора $g = (w, \gamma)$, порождающего строго отделяющую гиперплоскость $H = \{x \mid \langle w, x\rangle = \gamma\}$, в вектор $g_* = (w_*, \gamma_*)$, на котором $f(g_*) = 0$. Дело в том, что на само́м векторе g значение $f(g)$ может быть положительным (это зависит от параметра c).

ПРИМЕР 1. В качестве A и B возьмём одноточечные множества на плоскости $\mathbb{R}^2$: $A = \{a\}$, $B = \{b\}$, где $a = (0, 0)$ и $b = (0, 2)$. Вектор $g = (w, \gamma)$ с компонентами $w = (0, 1)$, $\gamma \in (0, 2)$ порождает прямую $x_2 = \gamma$, строго отделяющую точку a от точки b (см. рис. 1).

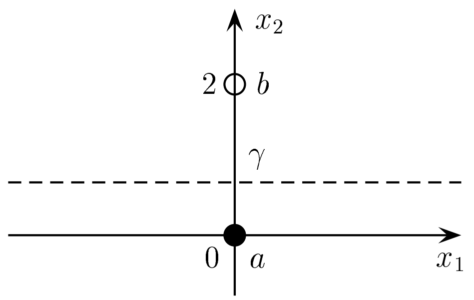

Рис. 1

Вместе с тем,

$$f(g) = [-\gamma + c]_+ + [-2 + \gamma + c]_+.$$

На рис. 2 представлен график $f(g)$ как функции от γ при $c \in (0, 1]$.

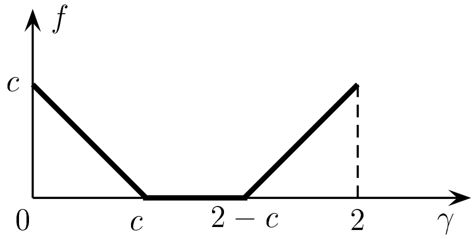

Рис. 2

Видим, что $f(g) = 0$ при $\gamma \in [c, 2 - c]$. При $\gamma \in (0, c) \cup (2 - c, 2)$ прямая $x_2 = \gamma$ по-прежнему строго отделяет точку a от точки b, но $f(g) > 0$.

4°. Рассмотрим экстремальную задачу

$$f(g) \to \min, \tag{9}$$

где $f(g)$ — функция вида (3). Эта задача эквивалентна задаче линейного программирования

$$\begin{gathered} \frac{1}{m}\sum_{i=1}^{m} y_i + \frac{1}{k}\sum_{j=1}^{k} z_j \to \min, \\ -\langle w, a_i\rangle + \gamma + y_i \geqslant c, \quad i \in 1:m; \\ \langle w, b_j\rangle - \gamma + z_j \geqslant c, \quad j \in 1:k; \\ y_i \geqslant 0,\ i \in 1:m; \quad z_j \geqslant 0,\ j \in 1:k. \end{gathered} \tag{10}$$

Множество планов задачи (10) непусто (планом является вектор с компонентами $w = \mathbb{O}$, $\gamma = 0$, $y_i \equiv c$, $z_j \equiv c$) и целевая функция ограничена снизу нулём. Значит, задача (10) имеет решение. По эквивалентности существует решение и у задачи (9), причём минимальные значения целевых функций у этих задач равны между собой. Это общее значение обозначим через μ. Отметим также, что если $\big(w_*, \gamma_*, \{u_i^*\}, \{v_j^*\}\big)$ — решение задачи (10), то $g_* = \{w_*, \gamma_*\}$ — решение задачи (9).

5°. При $\mu = 0$ получим $f(g_*) = 0$. По теореме 1 вектор $g_* = (w_*, \gamma_*)$ порождает гиперплоскость $H = \big\{x \mid \langle w_*, x\rangle = \gamma_*\big\}$, строго отделяющую множество A от множества B.

Вектор g_* можно улучшить, пользуясь неединственностью решения задачи (9). Положим

$$\begin{gathered} w_0 = w_*/\|w_*\|, \\ \gamma_0 = \tfrac{1}{2}\big[\min_{j\in 1:k}\langle w_0, b_j\rangle + \max_{i\in 1:m}\langle w_0, a_i\rangle\big], \\ c_0 = \tfrac{1}{2}\big[\min_{j\in 1:k}\langle w_0, b_j\rangle - \max_{i\in 1:m}\langle w_0, a_i\rangle\big], \\ g_0 = (w_0, \gamma_0). \end{gathered}$$

Тогда при всех $i \in 1:m$

$$\langle w_0, a_i\rangle - \gamma_0 + c_0 = \langle w_0, a_i\rangle - \max_{i\in 1:m}\langle w_0, a_i\rangle \leqslant 0,$$

и при всех $j \in 1:k$

$$-\langle w_0, b_j\rangle + \gamma_0 + c_0 = -\langle w_0, b_j\rangle + \min_{j\in 1:k}\langle w_0, b_j\rangle \leqslant 0.$$

Значит, $f(g_0) = 0$ при $c = c_0$. Гиперплоскость $H_0 = \big\{x \mid \langle w_0, x\rangle = \gamma_0\big\}$ строго отделяет множество A от множества B, причём ширина отделяющей полосы равна $2c_0$.

6°. Как отмечалось в п. 4°, задача (9) всегда имеет решение. При $\mu > 0$ по теореме 1 множества A и B не допускают строгого линейного отделения. В этом случае будем говорить, что гиперплоскость $H_* = \big\{x \mid \langle w_*, x\rangle = \gamma_*\big\}$, порождаемая решением $g_* = (w_*, \gamma_*)$ задачи (9), является *наилучшей гиперплоскостью, приближённо отделяющей множество A от множества B* (при данном значении параметра c).

Здесь, однако, имеется тонкость: нет гарантии, что компонента w_* вектора g_* отлична от нулевой. Разберёмся в этой ситуации.

ТЕОРЕМА 2. *Для того чтобы задача* (9) *имела решение $g_* = (w_*, \gamma_*)$ с $w_* = \mathbb{O}$, необходимо и достаточно, чтобы выполнялось условие*

$$\frac{1}{m}\sum_{i=1}^{m} a_i = \frac{1}{k}\sum_{j=1}^{k} b_j. \tag{11}$$

Доказательство. Необходимость. При $w_* = \mathbb{O}$ легко вычисляется экстремальное значение целевой функции у задачи линейного программирования (10). Действительно,

$$\mu = f(g_*) = \min_{\gamma}\big\{[-\gamma + c]_+ + [\gamma + c]_+\big\} = 2c.$$

Такое же экстремальное значение имеет задача линейного программирования, двойственная к задаче (10). В силу разрешимости двойственной задачи совместна система

$$c\Bigg(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j\Bigg) = 2c, \tag{12}$$

$$-\sum_{i=1}^{m} u_i a_i + \sum_{j=1}^{k} v_j b_j = \mathbb{O}, \tag{13}$$

$$\sum_{i=1}^{m} u_i - \sum_{j=1}^{k} v_j = 0, \tag{14}$$

$$0 \leqslant u_i \leqslant \tfrac{1}{m},\ i \in 1:m; \quad 0 \leqslant v_j \leqslant \tfrac{1}{k},\ j \in 1:k. \tag{15}$$

Из (12) и (14) следует, что

$$\sum_{i=1}^{m} u_i = 1, \quad \sum_{j=1}^{k} v_j = 1.$$

Принимая во внимание (15), заключаем, что все u_i равны $\frac{1}{m}$ и все v_j равны $\frac{1}{k}$. Теперь формула (13) становится эквивалентной равенству (11).

Д о с т а т о ч н о с т ь. Запишем задачу, двойственную к (10):

$$c\left(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j\right) \to \max$$

при ограничениях (13)–(15). В силу (11) набор $u_i \equiv \frac{1}{m}$, $v_j \equiv \frac{1}{k}$ является планом этой задачи. Значение целевой функции на нём равно $2c$.

Вместе с тем, на плане

$$w = \mathbb{O}, \quad \gamma = 0, \quad y_i \equiv c, \quad z_j \equiv c \tag{16}$$

задачи (10) значение целевой функции также равно $2c$. Отсюда следует, что план (16) задачи (10) с $w = \mathbb{O}$ является оптимальным.

Теорема доказана. $\square$

ПРИМЕР 2. Рассмотрим на плоскости $\mathbb{R}^2$ два двухточечных множества

$$A = \{(0,0),(1,1)\}, \quad B = \{(1,0),(0,1)\}$$

(см. рис. 3). В данном случае выполняется условие (11). По теореме 2 задача (9) имеет решение $g_* = (w_*, \gamma_*)$ с $w_* = \mathbb{O}$. При этом $\mu = 2c$.

Покажем, что у задачи (9) существует другое решение $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \neq \mathbb{O}$.

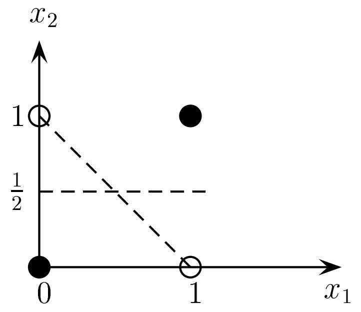

Рис. 3

Согласно (3)

$$f(g) = \tfrac{1}{2}\big\{[-\gamma + c]_+ + [w^1 + w^2 - \gamma + c]_+\big\} + \tfrac{1}{2}\big\{[-w^1 + \gamma + c]_+ + [-w^2 + \gamma + c]_+\big\}.$$

Здесь $w = (w^1, w^2)$. Положим

$$w_0 = (c,c), \quad \gamma_0 = c, \quad g_0 = (w_0, \gamma_0).$$

Тогда $f(g_0) = 2c$. Значит, на векторе g_0 достигается минимум функции $f(g)$. Гиперплоскость $H_0 = \{x \mid x_1 + x_2 = 1\}$ является наилучшей гиперплоскостью, приближённо отделяющей множество A от множества B.

Таким же свойством обладают вектор $g_1 = (w_1, \gamma_1)$ с $w_1 = (0, c)$, $\gamma_1 = \frac{c}{2}$ и гиперплоскость $H_1 = \{x \mid x_2 = \frac{1}{2}\}$ (см. рис. 3).

7°. Особенность,отмеченная в примере 2, имеет общий характер.

ТЕОРЕМА 3. *При $\mu > 0$ у задачи* (9) *существует решение $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$.*

Доказательство. Допустим, что у решения $g_* = (w_*, \gamma_*)$ задачи (9) компонента w_* оказалась нулевой. Построим другое решение $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$.

По теореме 2 выполняется соотношение (11) и $\mu = 2c$. Возьмём произвольный ненулевой вектор $h \in \mathbb{R}^n$ и рассмотрим задачу линейного программирования

$$\begin{gathered}\langle h, w\rangle \to \min, \qquad (17)\\ -\frac{1}{m}\sum_{i=1}^{m} y_i - \frac{1}{k}\sum_{j=1}^{k} z_j = -2c;\\ -\langle w, a_i\rangle + \gamma + y_i \geqslant c, \quad i \in 1:m;\\ \langle w, b_j\rangle - \gamma + z_j \geqslant c, \quad j \in 1:k;\\ y_i \geqslant 0,\ i \in 1:m; \quad z_j \geqslant 0,\ j \in 1:k.\end{gathered}$$

Вектор (16) удовлетворяет ограничениям задачи (17), то есть является её планом. Покажем, что этот план не может быть оптимальным.

В случае оптимальности плана (16) у задачи, двойственной к (17), должен существовать план с таким же (нулевым) значением целевой функции. Таким образом, должна быть совместной система

$$c\left(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j - 2\zeta\right) = 0, \qquad (18)$$

$$-\sum_{i=1}^{m} u_i a_i + \sum_{j=1}^{k} v_j b_j = h, \qquad (19)$$

$$\sum_{i=1}^{m} u_i - \sum_{j=1}^{k} v_j = 0, \qquad (20)$$

$$0 \leqslant u_i \leqslant \tfrac{1}{m}\zeta,\ i \in 1:m; \quad 0 \leqslant v_j \leqslant \tfrac{1}{k}\zeta,\ j \in 1:k. \qquad (21)$$

Покажем, однако, что эта система несовместна.

Из (18) и (20) следует, что

$$\sum_{i=1}^{m} u_i = \zeta, \quad \sum_{j=1}^{k} v_j = \zeta.$$

В силу (21) получаем $u_i \equiv \frac{1}{m}\zeta$, $v_j \equiv \frac{1}{k}\zeta$. Равенство (19) принимает вид

$$\zeta\Big(-\frac{1}{m}\sum_{i=1}^{m} a_i + \frac{1}{k}\sum_{j=1}^{k} b_j\Big) = h.$$

Но это противоречит условию (11) (напомним, что $h \neq \mathbb{O}$).

Установлено, что план (16) задачи (17) с нулевым значением целевой функции не является оптимальным. Значит, существует план

$$\big(w_0,\ \gamma_0,\ \{u_i^0\},\ \{v_j^0\}\big) \tag{22}$$

с отрицательным значением целевой функции. У такого плана должно быть $w_0 \neq \mathbb{O}$.

Теперь отметим, что план (22) задачи (17) удовлетворяет ограничениям задачи (10) и на нём целевая функция задачи (10) принимает наименьшее возможное значение, равное $2c$ (напомним, что $\mu = 2c$). В силу эквивалентности задач (9) и (10) вектор $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$ будет решением задачи (9).

Теорема доказана. $\square$

З а м е ч а н и е. В качестве ненулевого вектора h можно взять, например, любую ненулевую разность $b_{j_0} - a_{i_0}$. В этом случае множество планов задачи, двойственной к (17), которое определяется условиями (19)– (21), будет непустым. Вместе с непустотой множества планов само́й задачи (17) это гарантирует наличие у задачи (17) оптимального плана.

8°. При $\mu > 0$ решение $g_0 = (w_0, \gamma_0)$ задачи (9) с $w_0 \neq \mathbb{O}$ можно привести к каноническому виду. Как и в п. 5° положим

$$\begin{aligned} w_1 &= w_0/\|w_0\|,\\ \gamma_1 &= \tfrac{1}{2}\big[\min_{j\in 1:k}\langle w_1, b_j\rangle + \max_{i\in 1:m}\langle w_1, a_i\rangle\big],\\ c_1 &= \tfrac{1}{2}\big[\min_{j\in 1:k}\langle w_1, b_j\rangle - \max_{i\in 1:m}\langle w_1, a_i\rangle\big],\\ g_1 &= (w_1, \gamma_1). \end{aligned}$$

В данном случае $c_1 \leqslant 0$. При $c_1 = 0$ гиперплоскость $H_1 = \big\{x \mid \langle w_1, x\rangle = \gamma_1\big\}$ нестрого отделяет множество A от множества B. При $c_1 < 0$ та же гиперплоскость H_1 является наилучшей, приближённо отделяющей множество A от множества B.

Согласно определению w_1, γ_1, c_1 имеем

$$\begin{aligned} \langle w_1, a_i\rangle - \gamma_1 + c_1 &\leqslant 0, \quad i \in 1:m\\ -\langle w_1, b_j\rangle + \gamma_1 + c_1 &\leqslant 0, \quad j \in 1:k. \end{aligned}$$

При $c_1 < 0$ эти неравенства определяют “смешанную полосу”

$$c_1 \leqslant \langle w_1, x \rangle - \gamma_1 \leqslant -c_1,$$

которая содержит как точки множества A, так и точки множества B. Ширина смешанной полосы равна $2|c_1|$.

9°. На рис. 4 приведён пример наилучшего приближённого отделения двух множеств.

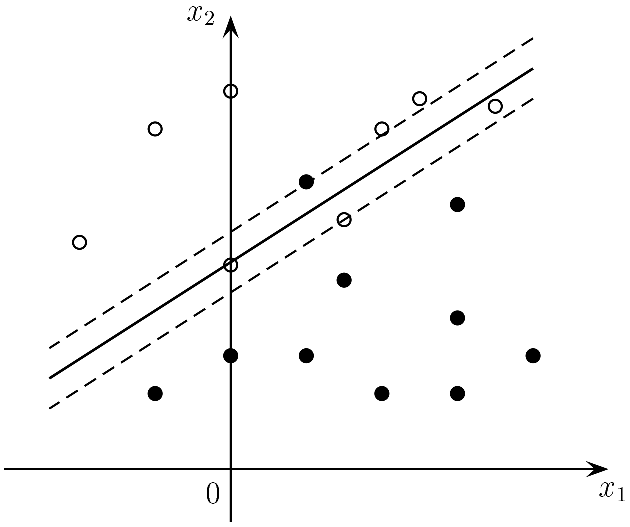

Рис. 4

ЛИТЕРАТУРА

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