ОБ АЛГОРИТМЕ КОЗИНЦА. I*

В. Н. Малоземов v.malozemov@spbu.ru

Н. А. Соловьева 4vinyo@gmail.com

Г. Ш. Тамасян grigoriytamasjan@mail.ru

14 марта 2024 г.

В 1964 году на заре машинного обучения Б. Н. Козинец в работе [2] предложил простой численный метод (алгоритм) для решения следующей экстремальной задачи.

«В n-мерном евклидовом пространстве заданы два конечных множества P_1 и P_2. Предполагается, что соответствующие выпуклые оболочки C_1 и C_2 этих множеств не имеют общих точек. Требуется построить гиперплоскость, разделяющую множества P_1 и P_2, т. е. такую гиперплоскость, которая не имеет общих точек с множествами C_1 и C_2 и, кроме того, множества C_1 и C_2 лежат по разные стороны этой гиперплоскости.

На самом деле, желательно среди всех гиперплоскостей, разделяющих множества P_1 и P_2, найти такую гиперплоскость, расстояние от которой до множества $P_1 \cup P_2$ имеет максимальную величину. Очевидно, что такой гиперплоскостью будет гиперплоскость, проходящая через середину отрезка, соединяющего какие-нибудь две ближайшие точки множеств C_1 и C_2, перпендикулярно к нему».

В алгоритме Козинца используется естественный геометрический вариант критерия оптимальности для данной задачи.

Через 60 лет после публикации работы [2] мы подготовили доклад, в котором дается детальный анализ алгоритма Козинца с современных позиций.

1°. Формализация задачи. Критерий оптимальности. Пусть в пространстве $\mathbb{R}^n$ с евклидовой нормой заданы два конечных множества

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

состоящие из попарно различных точек. Здесь $s \in 1 : m-1$. Обозначим через C_1 и C_2 выпуклые оболочки этих множеств. Задачу, поставленную Б. Н. Козинцом, с учетом указанной им идеи её решения можно формализовать следующим образом:

$$\|x - y\| \to \min_{x \in C_1,\ y \in C_2}. \tag{1}$$

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

Задача (1) имеет решение, вообще говоря, не единственное. Вместе с тем, вектор $w^* = x - y$, где (x, y) — некоторое решение задачи (1), является единственным.

ТЕОРЕМА 1 (Б. Н. Козинец)**.** *Для того чтобы план* (x, y) *задачи* (1) *был оптимальным, необходимо и достаточно, чтобы выполнялись неравенства*

$$\langle p_i - x, y - x\rangle \leqslant 0 \quad \textit{для всех } i \in 1 : s, \tag{2}$$

$$\langle p_j - y, x - y\rangle \leqslant 0 \quad \textit{для всех } j \in (s+1) : m. \tag{3}$$

Рис. 1 поясняет содержание теоремы.

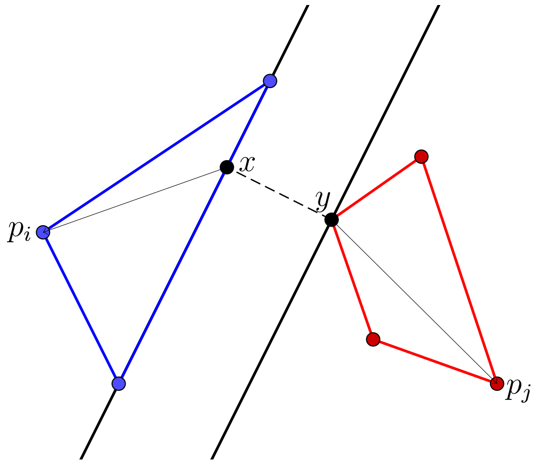

Рис. 1. Пояснение к теореме 1

Доказательство. Достаточность. Пусть для плана (x, y) выполнены условия (2) и (3). Зафиксируем индексы $i \in 1 : s$ и $j \in (s+1) : m$. Сложив (2) и (3), получим

$$\langle x - y - (p_i - p_j), x - y\rangle \leqslant 0.$$

Отсюда следует, что

$$\|x - y\|^2 \leqslant \langle p_i, x - y\rangle - \langle p_j, x - y\rangle. \tag{4}$$

Покажем, что для произвольного плана $(\widetilde{x}, \widetilde{y})$ выполняется неравенство

$$\|x - y\| \leqslant \|\widetilde{x} - \widetilde{y}\|. \tag{5}$$

По определению выпуклой оболочки имеем

$$\widetilde{x} = \sum_{i=1}^{s} u[i]\, p_i, \qquad \widetilde{y} = \sum_{j=s+1}^{m} u[j]\, p_j,$$

где все $u[i]$, $i \in 1 : m$, неотрицательны и

$$\sum_{i=1}^{s} u[i] = 1, \qquad \sum_{j=s+1}^{m} u[j] = 1.$$

Умножим (4) на $u[i]$ и просуммируем по $i \in 1:s$. Получим

$$\|x-y\|^2 \leqslant \langle \widetilde{x}, x-y\rangle - \langle p_j, x-y\rangle, \quad j \in (s+1):m. \tag{6}$$

Умножим (6) на $u[j]$ и просуммируем по $j \in (s+1):m$. Получим

$$\|x-y\|^2 \leqslant \langle \widetilde{x}, x-y\rangle - \langle \widetilde{y}, x-y\rangle = \langle \widetilde{x}-\widetilde{y}, x-y\rangle \leqslant \|\widetilde{x}-\widetilde{y}\| \cdot \|x-y\|. \tag{7}$$

Если $x=y$, то неравенство (5) тривиально. В случае $x \neq y$ к неравенству (5) придем, поделив (7) на $\|x-y\|$.
Необходимость. Пусть (x,y) — оптимальный план задачи (1). Покажем, что выполняются неравенства (2) и (3).

Разберемся сначала с неравенством (2). Допустим противное, что найдется индекс $i \in 1:s$, на котором

$$\langle p_i - x, y-x\rangle > 0. \tag{8}$$

Отсюда, в частности, следует, что $x \neq p_i$. Обозначим через $\widetilde{x}_i$ ближайшую к y точку отрезка $[x, p_i]$ и покажем, что

$$\|\widetilde{x}_i - y\| < \|x-y\|. \tag{9}$$

Это будет противоречить оптимальности плана (x,y).

Точки отрезка $[x,p_i]$ допускают представление $x+\lambda(p_i-x)$, где $\lambda \in [0,1]$. Нахождение точки $\widetilde{x}_i$ сводится к решению задачи

$$\varphi_i(\lambda) := \|(x+\lambda(p_i-x)) - y\|^2 \to \min_{\lambda\in[0,1]}.$$

Имеем

$$\varphi_i(\lambda) = \|x-y\|^2 - 2\lambda\langle p_i-x, y-x\rangle + \lambda^2\|p_i-x\|^2,$$

так что

$$\varphi_i'(\lambda) = -2\langle p_i-x, y-x\rangle + 2\lambda\|p_i-x\|^2.$$

Так как, в силу (8), $\varphi_i'(0)<0$ и $\varphi_i'(\widetilde{\lambda}_i)=0$ при

$$\widetilde{\lambda}_i = \frac{\langle p_i-x, y-x\rangle}{\|p_i-x\|^2} > 0,$$

то $\widetilde{x}_i = x+\lambda_i(p_i-x)$, где

$$\lambda_i = \begin{cases} \widetilde{\lambda}_i, & \text{если } \widetilde{\lambda}_i \leqslant 1,\\ 1, & \text{если } \widetilde{\lambda}_i > 1.\end{cases}$$

Очевидно, что $\widetilde{x}_i \in C_1$. По определению $\varphi_i(\lambda)$ получаем

$$\|\widetilde{x}_i - y\|^2 = \|x-y\|^2 - 2\lambda_i\langle p_i-x, y-x\rangle + \lambda_i^2\|p_i-x\|^2. \tag{10}$$

В случае $\widetilde{\lambda}_i \leqslant 1$, то есть при $\lambda_i = \widetilde{\lambda}_i$, из (10) следует, что

$$\|\widetilde{x}_i - y\|^2 = \|x - y\|^2 - \left(\frac{\langle p_i - x, y - x\rangle}{\|p_i - x\|}\right)^2 < \|x - y\|^2 .$$

Условие $\widetilde{\lambda}_i > 1$ можно переписать в эквивалентном виде

$$\langle p_i - x, y - x\rangle > \|p_i - x\|^2 . \tag{11}$$

Поэтому при $\widetilde{\lambda}_i > 1$, то есть при $\lambda_i = 1$, из (10) и (11) следует, что

$$\begin{aligned}\|\widetilde{x}_i - y\|^2 = \|x - y\|^2 - 2\langle p_i - x, y - x\rangle + \|p_i - x\|^2 <\\ < \|x - y\|^2 - \|p_i - x\|^2 < \|x - y\|^2 .\end{aligned}$$

В обоих случаях установлено, что выполняется неравенство (9).

Аналогично показывается, что для оптимального плана (x, y) справедливы неравенства (3). Действительно, допустим вопреки утверждению, что найдется индекс $j \in (s+1) : m$, такой, что

$$\langle p_j - y, x - y\rangle > 0. \tag{12}$$

Отсюда, в частности, следует, что $y \neq p_j$. Обозначим через $\widetilde{y}_j$ ближайшую к x точку отрезка $[y, p_j]$ и покажем, что

$$\|x - \widetilde{y}_j\| < \|x - y\| . \tag{13}$$

Нахождение $\widetilde{y}_j$ сводится к решению задачи

$$\varphi_j(\lambda) := \|(y + \lambda(p_j - y)) - x\|^2 \to \min_{\lambda\in[0,1]} .$$

Имеем

$$\varphi_j(\lambda) = \|x - y\|^2 - 2\lambda\langle p_j - y, x - y\rangle + \lambda^2 \|p_j - y\|^2 ,$$

так что

$$\varphi_j'(\lambda) = -2\langle p_j - y, x - y\rangle + 2\lambda \|p_j - y\|^2 .$$

Так как, в силу (12), $\varphi_j'(0) < 0$ и $\varphi_j'(\widetilde{\lambda}_j) = 0$ при

$$\widetilde{\lambda}_j = \frac{\langle p_j - y, x - y\rangle}{\|p_j - y\|^2} > 0, \tag{14}$$

то $\widetilde{y}_j = y + \lambda_j(p_j - y)$, где

$$\lambda_j = \begin{cases}\widetilde{\lambda}_j, & \text{если } \widetilde{\lambda}_j \leqslant 1,\\ 1, & \text{если } \widetilde{\lambda}_j > 1.\end{cases}$$

Очевидно, что $\widetilde{y}_j \in C_2$. По определению $\varphi_j(\lambda)$ получаем

$$\|x - \widetilde{y}_j\|^2 = \|x - y\|^2 - 2\lambda_j\langle p_j - y, x - y\rangle + \lambda_j^2 \|p_j - y\|^2 . \tag{15}$$

В случае $\widetilde{\lambda}_j \leqslant 1$, то есть при $\lambda_j = \widetilde{\lambda}_j$, из (15) следует, что

$$\|x - \widetilde{y}_j\|^2 = \|x - y\|^2 - \left(\frac{\langle p_j - y, x - y \rangle}{\|p_j - y\|} \right)^2 < \|x - y\|^2 .$$

Условие $\widetilde{\lambda}_j > 1$ равносильно неравенству

$$\langle p_j - y, x - y \rangle > \|p_j - y\|^2 . \tag{16}$$

Поэтому при $\widetilde{\lambda}_j > 1$, то есть при $\lambda_j = 1$, из (15) и (16) следует, что

$$\|x - \widetilde{y}_j\|^2 < \|x - y\|^2 - \|p_j - y\|^2 < \|x - y\|^2 .$$

В обоих случаях выполняется неравенство (13), что противоречит оптимальности плана (x, y).

Теорема доказана. $\square$

2°. Оценка плана задачи (1). Пусть (x, y) — план задачи (1). Введем обозначения

$$\Delta^{(1)}(x, y) = \max_{i \in 1:s} \langle p_i - x, y - x \rangle,$$
$$\Delta^{(2)}(x, y) = \max_{j \in (s+1):m} \langle p_j - y, x - y \rangle.$$

ЛЕММА 1. *Справедливы неравенства*

$$\Delta^{(1)}(x, y) \geqslant 0, \quad \Delta^{(2)}(x, y) \geqslant 0.$$

Доказательство. Проверим первое неравенство. Допустим противное, что $\Delta^{(1)}(x, y) < 0$. Тогда

$$\langle p_i - x, y - x \rangle < 0 \quad \text{при всех } i \in 1 : s. \tag{17}$$

Точка x принадлежит выпуклой оболочке C_1. Она допускает представление

$$x = \sum_{i=1}^{s} u[i]\, p_i, \tag{18}$$

где все $u[i]$, $i \in 1 : s$, неотрицательны и в сумме равны единице. Умножим неравенство (17) на $u[i]$ и сложим по $i \in 1 : s$. Так как среди коэффициентов $u[i]$ есть положительные, то придем к неравенству

$$\left\langle \sum_{i=1}^{s} u[i]\, p_i - x, y - x \right\rangle < 0$$

или, в силу (18), $0 < 0$. Получили противоречие.

Аналогично проверяется второе неравенство $\Delta^{(2)}(x, y) \geqslant 0$.

Лемма доказана. $\square$

Величину

$$\Delta(x,y) = \max\bigl\{\Delta^{(1)}(x,y),\ \Delta^{(2)}(x,y)\bigr\}$$

назовем *оценкой плана* (x,y) задачи (1). Согласно лемме 1, $\Delta(x,y) \geqslant 0$ на любом плане (x,y).

ТЕОРЕМА 2. *Для того чтобы план (x,y) задачи (1) был оптимальным, необходимо и достаточно, чтобы выполнялось равенство $\Delta(x,y) = 0$.*

Доказательство. Необходимость. Пусть (x,y) — оптимальный план. По теореме 1, $\Delta(x,y) \leqslant 0$. Вместе с тем, всегда $\Delta(x,y) \geqslant 0$. Значит, $\Delta(x,y) = 0$. Достаточность. Из условия $\Delta(x,y) = 0$ следует, что $\Delta^{(1)}(x,y) \leqslant 0$ и $\Delta^{(2)}(x,y) \leqslant 0$. Последние два неравенства эквивалентны неравенствам (2) и (3), а это по теореме 1 гарантирует оптимальность плана (x,y).

Теорема доказана. □

3°. Алгоритм Козинца (принципиальная схема). Для решения задачи (1) Б. Н. Козинец в 1964 году предложил простой алгоритм, идейно близкий методу условного градиента. Приведем принципиальную схему алгоритма Козинца.

В качестве начального приближения возьмем произвольный план (x_0, y_0) задачи (1).

Пусть уже имеется k-е приближение — план (x_k, y_k). Проверим его на оптимальность. Для этого вычислим величины

$$\Delta_k^{(1)} = \max_{i\in 1:s}\langle p_i - x_k, y_k - x_k\rangle, \qquad \Delta_k^{(2)} = \max_{j\in(s+1):m}\langle p_j - y_k, x_k - y_k\rangle,$$
$$\Delta_k = \max\bigl\{\Delta_k^{(1)}, \Delta_k^{(2)}\bigr\}.$$

Если $\Delta_k = 0$, то (x_k, y_k) — оптимальный план. Процесс закончен.

Пусть $\Delta_k > 0$.

В случае $\Delta_k = \Delta_k^{(1)} = \langle p_{i_k} - x_k, y_k - x_k\rangle$, где $i_k \in 1:s$, вычисляем

$$\widetilde{\lambda}_k = \frac{\Delta_k}{\|p_{i_k} - x_k\|^2}\,.$$

Полагаем $y_{k+1} = y_k$ и $x_{k+1} = x_k + \lambda_k(p_{i_k} - x_k)$, где

$$\lambda_k = \begin{cases} \widetilde{\lambda}_k, & \text{если } \widetilde{\lambda}_k \leqslant 1,\\ 1, & \text{если } \widetilde{\lambda}_k > 1. \end{cases}$$

На рис. 2 и рис. 3 демонстрируется шаг алгоритма при $\widetilde{\lambda}_k \leqslant 1$ и $\widetilde{\lambda}_k > 1$ соответственно.

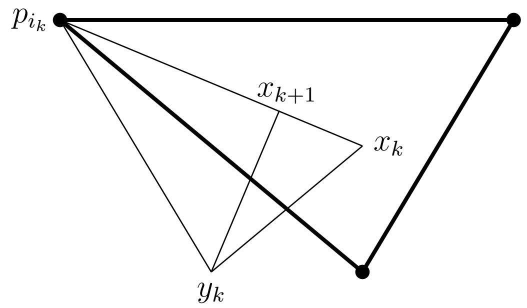

Рис. 2. Шаг алгоритма при $\widetilde{\lambda}_k \leqslant 1$

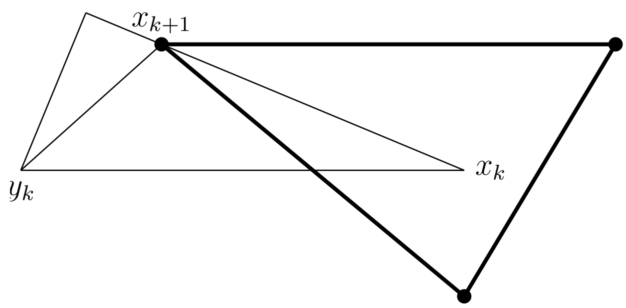

Рис. 3. Шаг алгоритма при $\widetilde{\lambda}_k > 1$

В случае $\Delta_k = \Delta_k^{(2)} = \langle p_{j_k} - y_k, x_k - y_k \rangle$, где $j_k \in (s+1) : m$, вычисляем

$$\widetilde{\lambda}_k = \frac{\Delta_k}{\|p_{j_k} - y_k\|^2}\,.$$

Полагаем $x_{k+1} = x_k$ и $y_{k+1} = y_k + \lambda_k(p_{j_k} - y_k)$, где

$$\lambda_k = \begin{cases} \widetilde{\lambda}_k, & \text{если } \widetilde{\lambda}_k \leqslant 1, \\ 1, & \text{если } \widetilde{\lambda}_k > 1. \end{cases}$$

Описание алгоритма завершено.

С помощью этого алгоритма строится последовательность планов задачи (1):

$$(x_0, y_0),\ (x_1, y_1), \ldots, (x_k, y_k), \ldots \tag{19}$$

Она конечна, если при некотором k_0 выполнится равенство $\Delta_{k_0} = 0$. В этом случае план (x_{k_0}, y_{k_0}) будет оптимальным. Если $\Delta_k > 0$ при всех $k = 0, 1, \ldots$, то последовательность (19) бесконечна. Ниже будет доказана ее сходимость по норме к решению задачи (1).

4°. **Оценка погрешности алгоритма.** Пусть (x_k, y_k) — план из последовательности (19) и (x^*, y^*) — оптимальный план. Обозначим

$$w_k = x_k - y_k, \quad w^* = x^* - y^*.$$

Отметим, что

$$\langle w_k, w^* \rangle \geqslant \langle w^*, w^* \rangle. \tag{20}$$

Действительно, в силу выпуклости множеств C_1 и C_2 при $t \in (0, 1)$ имеем

$$\|w^* + t(w_k - w^*)\|^2 = \|tw_k + (1-t)w^*\|^2 =$$

$$= \|(tx_k + (1-t)x^*) - (ty_k + (1-t)y^*)\|^2 \geqslant \|x^* - y^*\|^2 = \|w^*\|^2.$$

Отсюда следует, что

$$2t\langle w^*, w_k - w^* \rangle + t^2 \|w_k - w^*\|^2 \geqslant 0.$$

Поделим это неравенство на $t \in (0, 1)$, после чего перейдем к пределу при $t \to +0$. Получим (20).

ТЕОРЕМА 3. *Справедлива следующая оценка погрешности алгоритма Козинца:*

$$\|w_k - w^*\|^2 \leqslant 2\Delta_k,$$

где $\Delta_k = \Delta(x_k, y_k)$ *— оценка плана* (x_k, y_k).

Доказательство. В силу (20) имеем

$$\|w_k - w^*\|^2 = \big(\|w_k\|^2 - \langle w_k, w^* \rangle\big) + \big(\|w^*\|^2 - \langle w_k, w^* \rangle\big) \leqslant \|w_k\|^2 - \langle w_k, w^* \rangle.$$

Воспользуемся тем, что вектор w^* допускает представление

$$w^* = x^* - y^* = \sum_{i=1}^{s} u^*[i]\, p_i - \sum_{j=s+1}^{m} u^*[j]\, p_j,$$

где как $u^*[i]$, $i \in 1 : s$, так и $u^*[j]$, $j \in (s+1) : m$, неотрицательны и в сумме равны единице. Получим

$$\|w_k - w^*\|^2 \leqslant \|w_k\|^2 + \sum_{i=1}^{s} u^*[i]\big(\langle p_i - x_k, -w_k \rangle - \langle x_k, w_k \rangle\big) +$$

$$+ \sum_{j=s+1}^{m} u^*[j]\big(\langle p_j - y_k, w_k \rangle + \langle y_k, w_k \rangle\big) = \|w_k\|^2 - \langle x_k - y_k, w_k \rangle +$$

$$+ \sum_{i=1}^{s} u^*[i]\langle p_i - x_k, y_k - x_k \rangle + \sum_{j=s+1}^{m} u^*[j]\langle p_j - y_k, x_k - y_k \rangle \leqslant$$

$$\leqslant \Delta^{(1)}(x_k, y_k) + \Delta^{(2)}(x_k, y_k) \leqslant 2\Delta_k.$$

Теорема доказана. □

ЛИТЕРАТУРА

1. Козинец Б. Н. *Об одном алгоритме обучения линейного персептрона.* В сб. «Вычислительная техника и вопросы программирования». Изд. ВЦ ЛГУ, 1964, вып. 3. С. 80–83.

2. Козинец Б. Н. *Рекуррентный алгоритм разделения выпуклых оболочек двух множеств* // Алгоритмы обучения распознаванию образов / Под. ред. В. Н. Вапника. М.: Сов. Радио. 1973. С. 43–50.