Алгоритм приближенного разделения множеств с пересекающимися выпуклыми оболочками

Е. А. Марзель, Санкт-Петербургский государственный университет,
г. Санкт-Петербург

19 февраля 2026

Введение и актуальность

В различных областях, таких как классификация данных, обработка изображений и анализ многомерных данных, часто возникает задача оптимального разделения двух классов. Для ее решения требуются новые методы, это обусловлено тем, что в реальных приложениях часто требуется решать задачу оптимального разделения классов в условиях неполных данных, шума или неопределенности, что усложняет задачу.

Данная работа посвящена развитию модификации алгоритма Б. Н. Козинца[a], которая позволяет оптимально разделять множества с пересекающимися выпуклыми оболочками.

[a] Алгоритмы обучения распознаванию образов / Под ред. В. Н. Вапника М.: Сов. Радио. 1973. С. 43-50.

Постановка задачи

Даны конечные множества $A, B \in R^n$:

$$A = \{x_i, i = 1, ..., N\},$$

$$B = \{x_i, i = N+1, ..., N+M\},$$

где x_i - элементы множества A, B, N - количество точек в A, M - количество точек в B.

Требуется найти гиперплоскость, оптимально разделяющую A и B.

Постановка задачи

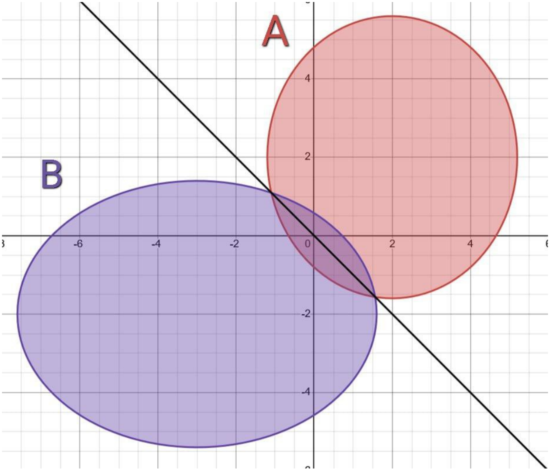

Figure 1: Постановка задачи.

Алгоритм Б. Н. Козинца

Рассмотрим два конечных множества в евклидовом пространстве:

$$A = \{x_i \in \mathbb{R}^n,\ i = 1, \ldots, N\}, \quad B = \{x_i \in \mathbb{R}^n,\ i = N+1, \ldots, N+M\}.$$

Требуется найти гиперплоскость, оптимально разделяющую A и B.

Для решения данной задачи предлагается использовать итерационный алгоритм Б. Н. Козинца, основанный на построении *кратчайшего моста* между точками множеств A и B.

Кратчайший мост

Определение

[b] *Пусть A, B — непустые подмножества $\mathbb{R}^n$. Кратчайшим мостом между ними называется отрезок*

$$[a, b], \quad a \in A,\ b \in B,$$

такой, что

$$\|a - b\| = \inf_{x \in A,\, y \in B} \|x - y\|.$$

[b] Гелиг А. Х., Матвеев А. С. Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. - Санкт-Петербург: Изд-во СПбГУ, 2014.

Кратчайший мост существует тогда и только тогда, когда инфимум достигается, при этом длина этого моста равна расстоянию между множествами.

Критерий кратчайшего моста

Теорема

[b] *$[a_*, b_*]$ - кратчайший мост между A и B $\iff$ $\forall a \in A, b \in B$ выполнены неравенства*

$$\langle a_* - b_*, a - a_* \rangle \geq 0, \quad \langle b_* - a_*, b - b_* \rangle \geq 0.$$

Первое утверждение означает, что вектор, направленный из $a_* \in A$ в произвольную точку множества A, образует либо острый, либо прямой угол с направлением отрезка $[a_*, b_*]$. Второе неравенство имеет сходный смысл.

[b] Гелиг А. Х., Матвеев А. С. Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. - Санкт-Петербург: Изд-во СПбГУ, 2014.

Идея алгоритма

Пусть текущее приближение моста задано отрезком $[a_0, b_0]$. При появлении новой точки $x \in A$, строится её ортогональная проекция на прямую, проходящую через a_0 и b_0. Если x лежит за пределами $[a_0, b_0]$, то в качестве нового приближения a_{new} берется ближайшая граничная точка отрезка, т. е. либо a_0, либо b_0. Иначе новая опорная точка a_{new} вычисляется по формуле:

$$a_{\text{new}} = a_0 + \lambda(x - a_0), \quad \text{где } \lambda = \frac{\langle x - a_0,\, b_0 - a_0\rangle}{\|x - a_0\|^2}, \quad 0 < \lambda < 1.$$

Аналогично, при появлении новой точки $x \in B$, выполняется:

$$b_{\text{new}} = b_0 + \lambda(x - b_0), \quad \lambda = \frac{\langle x - b_0,\, a_0 - b_0\rangle}{\|x - b_0\|^2}, \quad 0 < \lambda < 1.$$

Максиминное отделение от 0

Предлагается свести данную задачу разделения двух конечных множеств к задаче отделения конечного множества от 0. Для этого элементы множества B отражаются следующим образом:

$$x_i = -x_i, i = N+1, ..., N+M.$$

Это равносильно тому, что

$$A = \{y_i x_i, y_i = 1, i = 1, ..., N\},$$

$$B = \{y_i x_i, y_i = -1, i = N+1, ..., N+M\}.$$

Таким образом, разделение двух исходных множеств эквивалентно отделению от 0 выпуклой оболочки конечного множества C, равного объединению A и B.

Максиминное отделение от 0

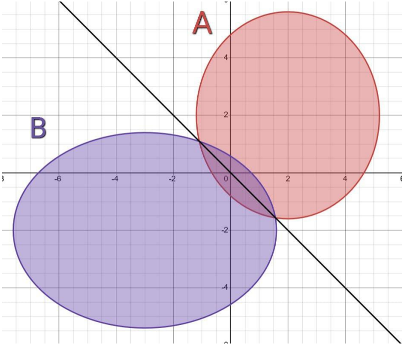

Figure 2: Общая задача.

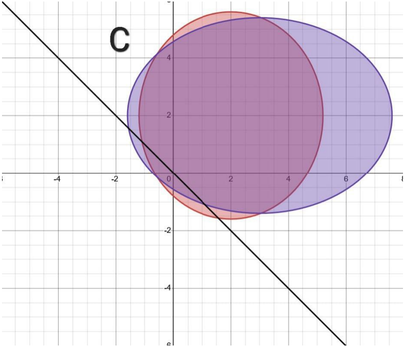

Figure 3: Отделение множества от 0.

Описание алгоритма

Сведем задачу к построению гиперплоскости, проходящей через начало координат, для этого размерность пространства увеличиваем на 1:

$$x_i^{'} = (x_{i1}, ..., x_{ik}, 1).$$

Пусть w_k - текущий вектор весов (коэффициентов гиперплоскости), дана начальная гиперплоскость:

$$w_0^T x = 0.$$

Определим функцию отступа x_i относительно w:

$$p_i(w) = y_i w^T x_i.$$

Для решения задачи нужно найти вектор w_* такой, что:

$$\min_i p_i(w_*) = \max_w \min_i p_i(w) = \max_w \varphi(w).$$

Суперградиент

Определение

[c] g - *суперградиент вогнутой функции* $f(x)$ *в точке* x $\forall y \in R^n$, *если*

$$f(y) \leq f(x) + \langle g, y - x \rangle.$$

[c] Поляк Б. Т. Введение в оптимизацию. - Москва: Наука, 1983.

Итерация алгоритма

Сначала выбирается начальное приближение w_0. Например, для этого можно взять разность между "центрами" A и B:

$$w_0 = \frac{\frac{1}{N}\sum_i x_i - \frac{1}{M}\sum_j x_j}{\|\frac{1}{N}\sum_i x_i - \frac{1}{M}\sum_j x_j\|}.$$

Тогда он будет направлен к области, где с большей вероятностью будут разделены множества, что поспособствует поиску оптимального решения за меньшее число итераций.

Шаг алгоритма

На каждом шаге ищутся точки, в которых достигается минимум:

$$\varphi(w_k) = \min_i \{y_i \langle w_k, x_i \rangle\}. \tag{1}$$

Далее вычисляется суперградиент g_k.

1. Если минимум функции (1) достигается в одной точке, то $g_k = -y_j x_j$.
2. Если минимум достигается в нескольких точках, то g_k будет равен выпуклой комбинации этих точек, имеющей минимальную длину. Для того, чтобы ее найти, из этих точек строится симплекс, затем между ним и 0 с помощью классического алгоритма Б. Н. Козинца вычисляется кратчайшее расстояние. Получим отрезок $[0, x_{g,k}]$. Тогда $g_k = x_{g,k}$.

Итерация алгоритма

Следующее приближение строится по формуле:

$$v_k = w_k + \gamma_k g_k, \quad w_{k+1} = P_B(v_k). \tag{2}$$

γ_k - коэффициент усиления.
P_B - ортогональная проекция на единичный шар:

$$P_B(v_k) = \begin{cases} v_k, & \|v_k\| \le 1, \\ \frac{v_k}{\|v_k\|}, & \|v_k\| > 1. \end{cases}$$

Для функции φ выполняется следующее свойство:

$$\varphi(tw) = t\varphi(w), \quad t \ge 0.$$

Из однородности следует, что максимум будет достигаться на границе $\|w\| = 1$.

Определение

[d] *Функция $f(\gamma)$ называется унимодальной, если она имеет ровно один локальный максимум (минимум) на интервале $[a, b]$, и она монотонна с обеих сторон точки минимума.*

[d] Гончаров В. А. Методы оптимизации. - Москва: Физматлит, 2003.

Метод золотого сечения

Для такой функции $f(\gamma)$ на отрезке $[a, b]$ требуется найти:

$$\gamma^* = \arg\max_{\gamma \in [a,b]} f(\gamma).$$

Метод использует свойство унимодальности так: если максимум один, то можно сужать интервал, не теряя максимум, сравнивая значения функции в двух точках, реализующих "золотое сечение" отрезка:

$$\varphi = \frac{1 + \sqrt{5}}{2} \approx 1.618, \quad \frac{1}{\varphi} \approx 0.618.$$

На каждом шаге выбираются две точки внутри интервала:

$$c = b - \frac{b - a}{\varphi}, \quad d = a + \frac{b - a}{\varphi}.$$

Метод золотого сечения

1. Вычисляются значения функции в точках c и d: $f(c), f(d)$.
2. Поскольку функция унимодальна, можно определить, в какой части интервала находится максимум. Если $f(c) < f(d)$, то максимум лежит в правой части $[c, b]$, и интервал сужается: $a \leftarrow c$. Иначе (если $f(c) \geq f(d)$), максимум лежит в левой части $[a, d]$, и интервал меняется: $b \leftarrow d$.
3. На каждом шаге длина интервала сокращается:

$$|b - a| \leftarrow \frac{1}{\varphi} \cdot |b - a|.$$

Метод золотого сечения

В рассматриваемой модификации алгоритма Козинца метод золотого сечения используется для нахождения максимума функции:

$$f(\gamma) = \varphi(w_k + \gamma g_k).$$

Функция φ является вогнутой как минимум конечного набора линейных функций. Тогда функция $f(\gamma)$ тоже является вогнутой (по γ), т. к. это свойство сохраняется при линейном преобразовании аргумента, следовательно, она является унимодальной на отрезке. Значит, она подходит для применения метода золотого сечения.

Сходимость алгоритма

Теорема

Пусть функция φ имеет вид

$$\varphi(w) = \min_i y_i \langle w, x_i \rangle,$$

где $y_i \in \{+1, -1\}, x_i \in R^n$. *Обозначим множество оптимальных решений:*

$$W^* = \arg \max_{\|w\| \leq 1} \varphi(w).$$

Пусть $w^* \in W^*$. *Данные ограничены, т. е.* $\exists R > 0 : \|x_i\| \leq R \quad \forall i$. *На каждом шаге выбирается суперградиент* $g_k \in \partial\varphi(w_k)$, *а новое приближение строится по формуле (2). Тогда последовательность* $\{w_k\}$ *имеет непустое множество предельных точек, любые предельные точки принадлежат* W^* *и*

$$\lim_{k \to \infty} \varphi(w_k) = \varphi(w_*) = \max_{\|w\| \leq 1} \varphi(w).$$

Сходимость алгоритма

Доказательство (1)

Для $\varphi(w) = \min_i y_i\langle w, x_i\rangle$ *имеем для любого* w*:*

$$\partial\varphi(w) = \mathrm{conv}\{y_i x_i : \ i \in I(w)\}, \quad I(w) = \{i : \ y_i\langle w, x_i\rangle = \varphi(w)\}.$$

$\partial\varphi(w)$ *- субдифференциал, выпуклая оболочка градиентов активных линейных функций,* $I(w)$ *- множество индексов точек, в которых достигается минимум.*
В частности, для любого $g \in \partial\varphi(w)$ *выполняются*

$$\langle g, w\rangle = \varphi(w), \quad \|g\| \le G.$$

Доказательство (2)

$\langle g, w\rangle = \varphi(w)$ *выполняется потому, что*

$$g = \sum_{i \in I(w)} \alpha_i y_i x_i \quad (\alpha_i \geq 0, \sum \alpha_i = 1),$$

$$\langle g, w\rangle = \sum \alpha_i y_i \langle x_i, w\rangle = \sum \alpha_i \varphi(w) = \varphi(w).$$

$\|g\| \leq G$ *потому, что*

$$\exists G > 0 : \forall i \|y_i x_i\| \leq G,$$

поскольку данные ограничены.

Сходимость алгоритма

Доказательство (3)

Для любого w^* *и* $g_k \in \partial\varphi(w_k)$ *по определению суперградиента*

$$\varphi(w^*) \leq \varphi(w_k) + \langle g_k, w^* - w_k \rangle, \tag{3}$$

откуда следует

$$\langle g_k, w^* \rangle \geq \varphi(w^*).$$

Сходимость алгоритма

Доказательство (4)

Оценим расстояние до оптимального решения. Так как B - замкнутое выпуклое множество (в данном случае единичный шар), ортогональная проекция P_B нерасширяющаяся:

$$\|P_B(u) - P_B(v)\| \leq \|u - v\|.$$

В частности, для любого оптимального $w^* \in B$:

$$\|w_{k+1} - w^*\| \leq \|v_k - w^*\|.$$

Квадратичная оценка:

$$\|w_{k+1} - w^*\|^2 = \|w_k - w^*\|^2 - 2\gamma_k \langle g_k, w^* - w_k \rangle + \gamma_k^2 \|g_k\|^2.$$

Доказательство (5)

Применим неравенство (3) и получим:

$$\|w_{k+1} - w^*\|^2 \leq \|w_k - w^*\|^2 - 2\gamma_k(\varphi_1(w^*) - \varphi_1(w_k)) + \gamma_k^2\|g_k\|^2.$$

Вынесем γ_k за скобку:

$$-2\gamma_k(\varphi_1(w^*) - \varphi_1(w_k)) + \gamma_k^2\|g_k\|^2 = -2\gamma_k(\varphi_1(w^*) - \varphi_1(w_k) - \frac{\gamma_k}{2}\|g_k\|^2).$$

Получаем итоговую оценку для расстояния:

$$\boxed{\|w_{k+1} - w^*\|^2 - \|w_k - w^*\|^2 \leq -2\gamma_k(\varphi(w^*) - \varphi(w_k) - \frac{\gamma_k}{2}\|g_k\|^2).}$$

Доказательство (6)

Поскольку w^* *- точка, в которой достигается максимум функции* $\varphi(w)$, *то выполняется*

$$\forall \varepsilon \quad \exists \delta > 0 : \varphi(w^*) - \varphi(w) > \delta, \|w^* - w\| > \varepsilon.$$

Значит, если алгоритм не сходится, то неравенство

$$\varphi(w^*) - \varphi(w_k) > \delta$$

выполняется при $k = k_s$ *бесконечно много раз. Тогда из ограниченности нормы суперградиента* $\|g_k\| \leq G$ *следует, что для этих* k_s *сумма*

$$2\gamma_0 \sum_{s=1}^{\infty} (\delta - \frac{\gamma_{k_s}}{2} G^2)$$

сколь угодно велика, если $0 < \gamma_- < \gamma_k < \gamma_+ < \frac{2\delta}{G^2}$.

Сходимость алгоритма

Доказательство (7)

Это противоречит неравенству:

$$\|w^* - w_k\|^2 \geq 0.$$

Таким образом, получили оценку для γ_k. *Длина шага* γ_k *может быть любой, но отделена от 0 и* $\frac{2\delta}{G^2}$*:*

$$0 < \gamma_- < \gamma_k < \gamma_+ < \frac{2\delta}{G^2}. \tag{4}$$

Проверим, удовлетворяют ли γ_k, *полученные методом золотого сечения, оценке (4). При подборе коэффициентов* γ_k *в методе золотого сечения задается интервал* $[a, b]$, *на котором производятся вычисления. Значит,* $a \leq \gamma_k \leq b$. *Поэтому если сделать отрезок* $[\gamma_-, \gamma_+]$, *то* γ_k *будут удовлетворять условию (4), а отпимальное решение будет достигаться.*

Сходимость алгоритма

Доказательство (8)

Остается проверить ограниченность g_k. Для этого воспользуемся следующей леммой:

Лемма

[c] *Суперградиенты вогнутой функции $f(x)$ ограничены на всяком ограниченном множестве.*

[c] Поляк Б. Т. Введение в оптимизацию. - Москва: Наука, 1983.

Т.к. множество $\{w : \|w\| \leq 1\}$ ограничено, суперградиент g_k тоже ограничен, и все условия для сходимости выполнены.

Реализация и тестирование алгоритма

Для численного эксперимента использовались бинарные выборки в пространстве признаков $\mathbb{R}^{20}$. Каждый датасет содержит $N = 800$ объектов с метками $y_i \in \{-1, +1\}$, распределённых поровну между классами.

Векторы признаков x_i генерируются из многомерного нормального распределения с единичной ковариационной матрицей:

$$x_i \sim \mathcal{N}(\mu_{y_i}, I).$$

μ_+ и μ_- - центры классов.

Степень перекрытия классов определяется расстоянием между центрами:

$$d = \|\mu_+ - \mu_-\|.$$

Таким образом, подбором параметра d можно регулировать степень перекрытия классов.

Реализация и тестирование алгоритма

Сравнение алгоритмов по точности (оценивается количество правильно классифицированных точек)

Modified Kozinets	SVM
0.735	0.775
0.740	0.760
0.745	0.725
0.690	0.700
0.765	0.775
0.710	0.765
0.745	0.785
0.725	0.790
0.685	0.715
0.690	0.775

Сравнение алгоритмов по времени счета

Modified Kozinets	SVM
1.048	0.113
1.285	0.207
1.289	0.205
1.689	0.207
1.291	0.208
0.987	0.207
1.187	0.202
1.294	0.209
1.185	0.112
1.485	0.204

Заключение

1. В работе была представлена модификация алгоритма Б. Н. Козинца, решающая задачу мягкого разделения пересекающихся классов.

2. Предоставлено доказательство сходимости предложенного алгоритма.

3. Алгоритм был реализован и проверен на различных примерах, полученные результаты были сравнены с результатами алгоритма SVM.

Спасибо за внимание!