УДК 519.85

Н.Г. ЖУРБЕНКО

ЛИНЕЙНЫЙ КЛАССИФИКАТОР И ПРОЕКЦИЯ НА ПОЛИТОП[1]

Аннотация. Рассмотрен алгоритм построения линейных бинарных классификаторов. Объекты распознавания представляются точками *n*-мерного евклидова пространства. Алгоритм основан на решении задачи проектирования нуля на выпуклую оболочку конечного числа точек евклидова пространства

Ключевые слова: линейный классификатор, выпуклая оболочка, проекция на политоп, штрафные функции.

ВВЕДЕНИЕ

В настоящее время теория и практика тематики распознавания образов (теория обучения машин) настолько развита, что кратко характеризовать достижения в этой области чрезвычайно сложно. Достаточно полное современное состояние и актуальные проблемы теории и практики этой интенсивно развивающейся области компьютерной математики изложены, например, в работах [1, 2].

В данной статье мы рассматриваем линейный классификатор, который соответствует одной из первых и простейших моделей классификации двух классов объектов. Известно, что построение линейного классификатора связано с задачей определения ненулевого решения однородной системы линейных неравенств, которая, в свою очередь, связана с задачей проектирования нуля на выпуклую оболочку конечного числа точек в арифметическом пространстве R^n (с задачей проектирования на политоп). Если эта проекция отлична от нуля (т.е. политоп не содержит нуля), то она определяет линейный классификатор. Однако такой классификатор не является оптимальным. В настоящей работе приводится такая постановка задачи проектирования на политоп, когда ее решение определяет оптимальный классификатор со сколь угодно большой точностью.

1. ЛИНЕЙНЫЙ БИНАРНЫЙ КЛАССИФИКАТОР

Рассмотрим задачу построения линейных классификаторов для двух классов объектов: K_1 и K_2. Объект определяется точкой в евклидовом пространстве R^n. Пусть в результате серии экспериментов определена обучающая выборка, которая состоит из двух конечных множеств объектов: $G_1 = \{a^j \in R^n, \quad j \in J_1 = \overline{\{1, m_1\}}\}$ и $G_2 = \{b^j \in R^n, j \in J_2 = \overline{\{1, m_2\}}\}$ $(m_1 \geq 1$, $m_2 \geq 1)$, принадлежащих классам K_1 и K_2 соответственно.

Линейный классификатор $L(\eta, c)$ задается ненулевым вектором $\eta \in R^n$, $\eta \neq 0$, и константой $c \in R^1$. Классификатор $L(\eta, c)$ определяет открытые полупространства в R^n : $P^+ = \{x \mid (x, \eta) + c > 0\}$, $P^- = \{x \mid (x, \eta) + c < 0\}$. Множества G_1 и G_2 должны содержаться в разных полупространствах, например $G_1 \subset P^+$, $G_2 \subset P^-$. Плоскость $P(\eta, c) = \{x \mid (x, \eta) + c = 0\}$ называют разделяющей плоскостью для множеств G_1 и G_2.

Если плоскость разделяет множества G_1 и G_2 , то она разделяет и множества $\operatorname{conv}\{G_1\}$ и $\operatorname{conv}\{G_2\}$ ($\operatorname{conv}\{\cdot\}$ — выпуклая оболочка множества). Критерий существования разделяющей плоскости для множеств G_1, G_2: $\operatorname{conv}\{G_1\} \cap \operatorname{conv}\{G_2\} = \varnothing$.

---

[1] Работа выполнена при поддержке Volkswagen Foundation (грант № 90 306).

Определение классификатора $L(\eta, c)$ сводится к определению решения $(\eta, c) \in R^{n+1}$ следующей системы линейных неравенств:

$$(a^j, \eta) + c > 0, \; a^j \in G_1, \; j = \overline{1, m_1}, \tag{1}$$

$$(-b^j, \eta) - c > 0, \; b^j \in G_2, \; j = \overline{1, m_2}. \tag{2}$$

Относительно определения классификатора $L(\eta, c)$ заметим следующее. Очевидно, что $L(\eta, c+\delta)$ для достаточно малых δ также является классификатором. Поэтому из всех классификаторов с данным вектором η выделяется классификатор $L(\eta, c^*)$, для которого разделяющая плоскость P находится на одинаковом расстоянии от множеств G_1 и G_2. Определение классификатора $L(\eta, c^*)$ не представляет трудностей и состоит в следующем. Пусть для $k = 1, 2$ $h_k = \max\{(a, \gamma_k \eta) | a \in G_k\}$, $G_k^* = \{a \in G_k | (a, \gamma_k \eta) = h_k\}$, где $\gamma_1 = -1$, $\gamma_2 = +1$ (отметим, что $h_k = H_k(\gamma_k \eta)$, где $H_k(u) = \max\{(x, u) | x \in \operatorname{conv}\{G_k\}\}$ — опорная функция множества $\operatorname{conv}\{G_k\}$ [3]). Для $\forall a^* \in G_k^*$ плоскость $P_k(\gamma_k \eta, -h_k)$ является опорной к множеству $\operatorname{conv}\{G_k\}$ (с внешней нормалью $\gamma_k \eta$) и проходит через все точки a, принадлежащие G_k^* Множества G_k^* назовем множествами активных объектов.

Несмотря на некоторую громоздкость приведенных формальных определений, геометрический смысл введенных плоскостей достаточно прост и заключается в следующем. Параллельные плоскости P_1 и P_2 определяют полосу, разделяющую множества G_1 и G_2. Пусть a^*, b^* – некоторые (любые) активные объекты из G_1, G_2 ($a^* \in G_1^*$, $b^* \in G_2^*$), $c^* = -(\eta, (a^* + b^*)/2)$. Тогда $P(\eta, c^*)$ — плоскость, проходящая на одинаковом расстоянии от G_1 и G_2. Легко видеть, что ширина разделяющей полосы h классификатора $L(\eta, c^*)$ определяется формулой $h = ((a^* - b^*), \eta) / \|\eta\|$. Величину h называют зазором классификатора (для данного вектора нормали η).

Вектор η можно нормировать, что обычно и делается, т.е. рассматривать классификатор $L(\eta / \|\eta\|, c^* / \|\eta\|)$ с отделяющей плоскостью $P(\eta / \|\eta\|, c^* / \|\eta\|)$.

2. СИСТЕМА СТРОГИХ ЛИНЕЙНЫХ НЕРАВЕНСТВ

В 1921 г. В. Карвер сформулировал следующий математически изящный критерий совместности системы строгих линейных неравенств [4] (см. также [5, § 37]) : система линейных неравенств $(a^j, x) > 0$, $j = \overline{1, m}$, совместна тогда и только тогда, когда $0 \notin \operatorname{conv}\{a^j\}$.

Таким образом, критерий совместности однородной системы строгих линейных неравенств сводится к задаче определения наименьшего по норме вектора p, $p \in \operatorname{conv}\{a^j\}$ (здесь используется стандартная евклидова норма). Для такого вектора принято следующее обозначение: $p = Nr\{a^j, j = \overline{1, m}\}$; p — проекция нуля на политоп $\operatorname{conv}\{a^j\}$.

Необходимое и достаточное условие того, что $p = Nr\{a^j, j = \overline{1, m}\}$, заключается в следующем:

$$p \in \operatorname{conv}\{a^j\}, \; (a^j - p, p) \geq 0, \; j = \overline{1, m}. \tag{3}$$

Если $p \neq 0$, то $(a^j, p) \geq |p|^2 > 0$, $j = \overline{1, m}$. Таким образом, p является решением системы $(a^j, x) > 0$, $j = \overline{1, m}$.

Опорная плоскость к $\mathrm{conv}\{a^j\}$ определяется как

$$P=\{(x-p,\ p)=0\}. \tag{4}$$

Определим в R^{n+1} точки $\widetilde{a}^j(t)=(a^j,t)^{\mathrm{T}}, a^j\in G_1$; $\widetilde{b}^j(t)=(-b^j,-t)^{\mathrm{T}}$, $b^j\in G_2$. Здесь параметр $t>0$. Пусть $\widetilde{G}_1(t)=\{\widetilde{a}^j(t),\ j=\overline{1,m_1}\}$, $\widetilde{G}_2(t)=$ $=\{\widetilde{b}^j(t),\ j=\overline{1,m_2}\}$. Множества $\widetilde{G}_1(t)$, $\widetilde{G}_2(t)$ можно интерпретировать как множества $G_1(1)$, $G_2(1)$, преобразованные оператором растяжения по $(n+1)$-й координате: $(n+1)$-я координата точек умножается на положительное число t.

Определение линейного классификатора (т.е. решение системы неравенств (1), (2) с учетом указанного преобразования) сводится к определению кратчайшего вектора из множества $\widetilde{G}(t)\in R^{n+1}$, $\widetilde{G}(t)=\mathrm{conv}\{\cup\widetilde{G}_k(t),\ k=1,2\}$.

3. ОПТИМАЛЬНЫЙ КЛАССИФИКАТОР

В теории распознавания образов особое внимание уделяется классификатору с максимальным зазором (оптимальному классификатору). Существуют различные формулировки определения оптимального классификатора [1, 2]. Компактная и геометрически естественная формулировка оптимального классификатора представляется следующей задачей квадратичного программирования.

$$h_{\max}^2=\min\{\|x-y\|^2:\ x\in\mathrm{conv}\{G_1\};\ y\in\mathrm{conv}\{G_2\}\}. \tag{5}$$

Пусть x^*, y^* — решение задачи (5), $\eta^*=x^*-y^*\neq 0$. Тогда отделяющая плоскость оптимального классификатора определяется равенством $(\eta^*, x-(x^*+y^*)/2)=0$. Параметр c^* (величина порога) оптимального классификатора $L(\eta^*, c^*)$ равен $(\eta^*, -(x^*+y^*)/2)=-(\|x^*\|^2+\|y^*\|^2)/2$. Естественно, вектор η^* можно нормировать: $L(\eta^*/\|\eta^*\|,\ c^*/\|\eta^*\|)\equiv L(\eta^*, c^*)$.

Величина зазора оптимального классификатора определяется формулой $h_{\max}=\|x^*-y^*\|$.

Строгая выпуклость евклидовой нормы обеспечивает единственность вектора $\eta^*=x^*-y^*$ и, следовательно, единственность оптимального классификатора (однако непосредственно векторы x^*, y^* могут определяться неоднозначно).

Пусть классификатор определяется на основе проекции нуля на множество $\widetilde{G}(t)$: $(\eta,\eta_{n+1})^{\mathrm{T}}=Nr\{\widetilde{G}(t)\}$. Эта задача также имеет единственное решение (для стандартной евклидовой нормы). Тогда возникает вопрос: является ли этот классификатор классификатором с максимальным зазором? Нетрудно заметить, что в общем случае это не так. Достаточно рассмотреть простейшие примеры даже для одномерного случая ($n=1$). Суть введения параметра t состоит в следующем: проекция нуля на множество $\widetilde{G}(t)$ определяет классификатор, величина зазора которого монотонно возрастает с увеличением параметра t и стремится к величине зазора оптимального классификатора.

Запишем задачу определения вектора $p(t)=Nr\{\widetilde{G}(t)\}$ в явном виде (задача $L(t)$):

$$p(t)=(\eta^*(t),\eta_{n+1}^*(t))^{\mathrm{T}}=\arg\min\{4(\|\eta\|^2+\eta_{n+1}^2)\}, \tag{6}$$

$$\eta=\sum_1^{m_1}\lambda_j a^j-\sum_1^{m_2}\mu_j b^j, \tag{7}$$

$$\eta_{n+1} = t\left(\sum_1^{m_1} \lambda_j - \sum_1^{m_2} \mu_j\right), \tag{8}$$

$$\sum_1^{m_1} \lambda_j + \sum_1^{m_2} \mu_j = 1, \tag{9}$$

$$\lambda_j \ge 0,\ j = \overline{1, m_1};\ \mu_j \ge 0,\ j = \overline{1, m_2}. \tag{10}$$

Далее предположим, что $p(t) \ne \theta$. Тогда, как отмечалось выше, множества G_1, G_2 разделимы (в противном случае G_1, G_2 неразделимы).

Утверждение 1. Пусть $p(t) \ne \theta$. Тогда $\eta^*(t) \ne \theta$.

Доказательство. Допустим, что $\eta^*(t) = \theta$. Пусть $a \in \widetilde{G}_1,\ b \in \widetilde{G}_2$. Тогда в соответствии с (3), $t\eta^*_{n+1}(t) \ge 0, -t\eta^*_{n+1}(t) \ge 0$. Так как $t > 0$, то $\eta^*_{n+1}(t) = 0$. Но тогда $p(t) = \theta$. Имеем противоречие. □

Утверждение 2. Для $\forall t > 0$ решение задачи (6)–(10) определяет классификатор $L(\eta^*(t), t\eta^*_{n+1})$. Отделяющая плоскость этого классификатора задается уравнением $(x \in R^n)$:

$$(x \in R^n, \eta^*(t)) - t\eta^*_{n+1}(t) = 0.$$

Доказательство. В соответствии с (3) (применительно к $\widetilde{G}_1(t), \widetilde{G}_2(t)$) имеем

$$(a^j, \eta^*(t)) + t\,\eta^*_{n+1}(t) \ge \| p(t) \| > 0,\ a^j \in G_1,$$

$$(-b^j, \eta^*(t)) - t\eta^*_{n+1}(t) \ge \| p(t) \| > 0,\ b^j \in G_2.$$

□

Утверждение 3. Если $\eta^*_{n+1}(t) = 0$, то решение задачи (6)–(10) определяет оптимальный классификатор.

Доказательство. Из условия утверждения 3 и уравнений (8), (9) следует, что $\sum_1^{m_1} \lambda^*_j = \sum_1^{m_2} \mu^*_j = 1/2$. Поэтому задача (6)–(9) эквивалентна следующей задаче:

$$\arg\min\ 4\| \eta \|^2, \tag{11}$$

$$\eta = \frac{1}{2}\left(\sum_1^{m_1} (2\lambda_j)\, a^j - \sum_1^{m_2} (2\mu_j)\, b^j\right),$$

$$\sum_1^{m_1} (2\lambda_j) = 1;\ \lambda_j \ge 0,\ j = \overline{1, m_1};\ \sum_1^{m_2} (2\mu_j) = 1;\ \ \mu_j \ge 0,\ j = \overline{1, m_2}.$$

Если ввести новые переменные $\widetilde{\lambda}_j = 2\lambda_j,\ \widetilde{\mu}_j = 2\mu_j$, то становится очевидным, что задача (11) эквивалентна задаче (5). □

Утверждение 4. Пусть $\eta^*(t),\ \eta^*_{n+1}(t),\ p(t)$ — решение задачи (6)–(9). Тогда

$$\lim_{t \to \infty} \| p(t) \| \to h_{\max}; \qquad \lim_{t \to \infty} \eta^*_{n+1}(t) \to 0;$$

вектор $\eta^* = \lim\limits_{t \to \infty} \eta^*(t)$ определяет оптимальный классификатор.

Доказательство. Дополним задачу $L(t)$ (задачу (6)–(10)) следующим ограничением, не нарушающим совместности ее ограничений (задача L^*):

$$\sum_1^{m_1} \lambda_j - \sum_1^{m_2} \mu_j = 0. \tag{12}$$

Из утверждения 3 следует, что решение задачи L^* определяет оптимальный классификатор.

Легко видеть, что задача $L(t)$ является задачей метода штрафных функций применительно к задаче L^* относительно одного ограничения — (12).

Приведем теоретические результаты по методу штрафных функций для задачи минимизации с ограничениями типа равенств [6, гл. 8, §2, п. 5].

Рассмотрим задачу A:

$$\min f(x), \quad x \in R^n,$$

$$g_i(x) = 0, \; i = 1, \ldots, m.$$

Метод штрафных функций состоит в следующем:

$$x^k = \arg\min_{x \in Q} f_k(x), \quad f_k(x) = f(x) + \frac{1}{2} K_k \| g(x) \|^2, \; K_k \to \infty. \tag{13}$$

1. Пусть задача A имеет решение. Множество ее решений обозначим X^*. Пусть f и g_i непрерывны, Q ограничено и замкнуто, $Q \cap X^* \neq \theta$. Тогда любая предельная точка метода (13) (здесь x^k — глобальный минимум $f_k(x)$ на Q) является глобальным минимумом для задачи A.

2 Пусть x^* — точка невырожденного минимума и $\nabla^2 g_i(x)$ удовлетворяет условию Липшица в окрестности x^*. Тогда $x^k \to x^*$ при $K_k \to \infty$, причем

$$\| x^k - x^* \| = O(1/K_k), \tag{14}$$

$$K_k g(x^k) \to y^*,$$

где y^* — оптимальные множители Лагранжа ограничений задачи A (двойственные переменные) [7].

Применительно к задаче (6)–(10) имеем:

- ограничениям задачи A соответствует одно ограничение (12): $\sum_1^{m_1} \lambda_j - \sum_1^{m_2} \mu_j = 0$;
- множество Q определяется ограничениями (7)–(10);
- параметру K_k штрафной функции соответствует $8t^2$;
- $\frac{1}{2} K_k \| g(x) \|^2$ соответствует $4\eta_{n+1}^2(t)$ (отсюда следует, что $\eta_{n+1}^*(t) \to 0$);
- $K_k g_1(x^k)$ соответствует $8t\eta_{n+1}(t)$ (отсюда следует, что $t\eta_{n+1}^*(t) \to \frac{1}{8} y^*$).

Для задачи L^* (простейшей задачи квадратичного программирования) все условия приведенных утверждений 1, 2 по методу штрафных функций выполняются. Отсюда с учетом утверждения 3 следует утверждение 4. □

4. К ЧИСЛЕННОЙ РЕАЛИЗАЦИИ АЛГОРИТМА

Настройка численной реализации описанного выше подхода построения оптимального классификатора достаточно очевидна и в кратком изложении она состоит в следующем.

1. Выполняем масштабирование обучающей выборки.

Простейший вариант масштабирования: умножение всех векторов a^j, b^j на $\nu = 1/\max \| a^j \|, \| b^j \|$. Тогда все векторы будут принадлежать единичному шару в R^n. При этом величина максимального зазора не больше двух.

2. Определяем значение параметра t. При указанном масштабировании параметр можно положить равным ≈ 10. Решаем задачу (6)–(10).

3. Выполняем решение ряда задач с бóльшими значениями параметра t (например, $t := 10t$).

Прекращение работы алгоритма определяется на основании следующих факторов (условие останова алгоритма):

— значение величины $\left|\sum_1^{m_1} \lambda_j - \sum_1^{m_2} \mu_j\right| \approx 0$ (т.е. $\sum_1^{m_1} \lambda_j \approx 1/2$, $\sum_1^{m_2} \mu_j \approx 1/2$);

— стабилизация с ростом параметра t величин $\| p(t) \|$ и $t\eta_{n+1}^*(t)$ (оценка оптимального значения двойственной переменной).

4. Классификатор определяется на основе вектора $\eta^*(t)$, полученного в результате решения задачи (6)–(10) при останове алгоритма. Процедура определения такого классификатора для заданного вектора нормали приведена в разд. 1.

5. ЗАМЕЧАНИЯ

1. Для любого значения параметра $t > 0$ алгоритм определяет некоторый классификатор. С ростом параметра t величина зазора такого классификатора монотонно увеличивается, приближаясь к величине зазора оптимального классификатора. О скорости приближения к оптимальному значению можно судить по оценке (14)

2. Имеется возможность точного определения оптимального классификатора на основании решения задачи при достаточно большом значении параметра t. Таким образом, рассматриваемая задача является примером задачи, для которой точное решение можно определить на основании использования гладкой штрафной функции. Поясним кратко такую возможность.

Расширим разделяющую полосу оптимального классификатора на достаточно малую величину. Тогда в силу конечности числа точек множеств G_1 и G_2 расширенная полоса будет содержать только активные объекты G_1^*, G_2^* оптимального классификатора. Естественно, что расширенная на малую величину полоса классификатора, определенного на основании $\eta^*(t)$ (решение задачи (6)–(10)), будет содержать только точки из множеств G_1^* и G_2^*. Но тогда определение оптимального вектора η^* сводится к задаче проектирования точки на аффинную оболочку конечного числа точек.

3. Сведение задачи определения оптимального классификатора к задаче проектирования на политоп обусловлено тем, что операция проектирования не инвариантна линейному преобразованию пространства. Естественно, использование линейного преобразования пространства можно интерпретировать как изменение метрики пространства. В частности, задачу (6)–(10) можно рассматривать как задачу проектирования нуля на множество $\widetilde{G}(1) = \operatorname{conv}\{\cup\{\widetilde{G}_k(1), k = 1, 2\}\}$ в метрике $\| x \|' = \sqrt{\sum_{i=1}^{n} x_i^2 + (tx_{n+1})^2}$.

Отметим, что вместо используемой нами стандартной евклидовой нормы рассмотрим норму в R^{n+1}: $\| x \|_* = \sqrt{\sum_1^n x_i^2} + |x_{n+1}|$. Тогда задача проектирования нуля на $\widetilde{G}(t)$ (задача (6)–(10)) в норме $\| \cdot \|_*$ состоит в решении задачи (6)–(10), в которой целевая функция $4(\| \eta \|^2 + \eta_{n+1}^2)$ заменяется функцией $4(\| \eta \|^2 + |\eta_{n+1}|)$.

Но тогда такую задачу можно рассматривать как задачу минимизации негладкой штрафной функции относительно ограничения равенства (12). Поэтому оптимальный классификатор определяется для конечных (достаточно больших) значений параметра t. В связи с этим отметим, что известный метод опорных векторов [2] можно интерпретировать как метод с использованием точных негладких штрафных функций.

4. Разработка универсального, «наилучшего» алгоритма для решения даже такого достаточно узкого класса задач оптимизации, как определение оптимального линейного классификатора, является чрезвычайно сложной проблемой. Однако для конкретной области приложений отдельный алгоритм может быть вполне приемлемым для решения реальных задач этой прикладной области. Но для искусственно построенных примеров этот алгоритм может вообще не решить задачу за приемлемое время. Так, например, скорость сходимости известного алгоритма Козинца [1, 8] не зависит от размерности задачи и равна $O(1/\sqrt{k})$. Она соответствует неулучшаемой теоретически возможной скорости алгоритмов для класса выпуклых функций [9] (для алгоритмов с равномерной по размерности пространства переменных скоростью сходимости). Таким образом, алгоритм Козинца не выделяет класс задач определения оптимального классификатора из всего класса выпуклых задач (как класс более «простых» задач). Однако для конкретной прикладной области приложения алгоритм Козинца может быть вполне приемлемым.

В работе предполагалась линейная разделимость множеств обучающей выборки. В случае линейной неразделимости множеств (т.е. политоп содержит нуль) в результате решения задачи проектирования на политоп находят коэффициенты представления нуля выпуклой комбинацией точек политопа. Эти коэффициенты определяют полезную информацию для построения классификатора и в случае линейной неразделимости множеств. Так, в работе [10] рассматривается задача построения классификатора на основе удаления объектов обучающей выборки для обеспечения разделимости множеств. В этой задаче значения коэффициентов разложения нуля можно использовать для установления приоритета относительно удаления объектов: чем бóльший коэффициент, тем бóльший приоритет на удаление. Аналогично эти значения можно использовать также в методе опорных векторов [2].

ЗАКЛЮЧЕНИЕ

Цель данной статьи заключается в следующем: на основе решения задачи проектирования на политоп получить классификатор с максимальным зазором. Решение этого вопроса имеет также некоторое практическое значение. Отметим, что решению классической задачи определения вектора минимальной длины из выпуклой оболочки конечного числа точек уделяется большое внимание [11–13]. Это связано с тем, что необходимость эффективного решения такой задачи возникает во многих алгоритмах оптимизации. Разумеется, задача максимального разделения двух множеств (5) не намного сложнее задачи проектирования на политоп. Она также является простейшим примером задачи квадратичного программирования. Однако разработке специализированных программных средств ее решения уделяется меньше внимания.

Практическое значение сведения задачи определения оптимального классификатора к задаче проектирования на политоп рассмотрим на следующем примере. Задача проектирования на политоп $\mathrm{conv}\{a^j : j=\overline{1,m}\}$ эквивалентна следующей задаче безусловной минимизации выпуклой (кусочно-сферической) функции: $\arg\min_x \{ \max_{j=\overline{1,m}} \{ \| x - a^j \|^2 - \| a^j \|^2 \}\}$. На основе субградиентных алгоритмов можно разработать специализированные алгоритмы решения этой задачи большой (супербольшой) размерности с учетом ее специфической структуры.

СПИСОК ЛІТЕРАТУРИ

1. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. Киев: Наук. думка, 2004. 545 с.
2. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). URL: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf.
3. Лейхтвейс К. Выпуклые множества. Москва: Наука, 1985. 336 с.
4. Carver W.B. Systems of linear inequalities. *Ann. Math.* 1921. T. 23, N 2. P. 212–220.
5. Чарин В.С. Линейные преобразования и выпуклые множества. Киев: Вища школа, 1978. 192 с.
6. Поляк Б.Т. Введение в оптимизацию. Москва: Наука, 1983. 384 с.
7. Поляк Б.Т. О скорости сходимости метода штрафных функций. *Журн. вычисл. математики и мат. физики*. 1971. Т. 11, № 1. С. 3–11.
8. Козинец Б.Н. Рекуррентный алгоритм разделения выпуклых оболочек двух множеств. Вапник В.Н. (ред.) *Алгоритмы обучения распознавания образов.* Москва: Сов. радио, 1973. 200 с.
9. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. Москва: Наука, 1979. 384 с.
10. Лаптин Ю.П., Журавлев Ю.И., Виноградов А.П. Минимизация эмпирического риска и задачи построения линейных классификаторов. *Кибернетика и системный анализ.* 2011. Т. 47, № 4. С. 155–164.
11. Wolfe P. Finding the nearest point in a polytope. *Math. Program*. 1976. Vol. 11, N 2. P. 128–149.
12. Нурминский Е.А. О сходимости метода подходящих аффинных подпространств для решения задачи о наименьшем расстоянии до симплекса. *Журн. вычисл. математики и мат. физики*. 2005. Т. 45, вып. 11. С. 1996–2004.
13. Журбенко Н.Г. Алгоритм проектирования на политоп. *Теорія оптимальних рішень.* 2008. № 7. С. 125–131.

Надійшла до редакції 05.03.2019

М.Г. Журбенко

ЛІНІЙНИЙ КЛАСИФІКАТОР І ПРОЄКЦІЯ НА ПОЛІТОП

Анотація. Запропоновано алгоритм побудови лінійних бінарних класифікаторів. Об'єкти розпізнавання представлено точками *n*-вимірного евклідового простору. Алгоритм ґрунтується на розв'язанні задачі проєктування нуля на опуклу оболонку кінцевого числа точок евклідового простору.

Ключові слова: лінійний класифікатор, опукла оболонка, проєкція на політоп, штрафні функції.

N.G. Zhurbenko

LINEAR CLASSIFIER AND PROJECTION ONTO A POLYTOPE

Abstract. An algorithm for constructing linear binary classifiers is proposed. Recognition objects are represented by points of *n*-dimensional Euclidean space. The algorithm is based on solving the problem of projecting zero onto the convex hull of a finite number of points of Euclidean space.

Keywords: linear classifier, convex hull, projection on polytope, penalty functions.

Журбенко Николай Георгиевич,
кандидат физ.-мат. наук, старший научный сотрудник Института кибернетики им. В.М. Глушкова НАН Украины, Киев, e-mail: zhurbnick@gmail.com.