УДК [519.245+519.214]: 519.237.8

М.А. ИВАНЧУК, И.В. МАЛЫК

АЛГОРИТМ ПОСТРОЕНИЯ РАЗДЕЛИМЫХ ε-СЕТЕЙ ДВУХ МНОЖЕСТВ

Аннотация. Предложен новый метод решения задачи классификации, основанный на разделении двух множеств в пространстве R^d. Доказаны необходимые и достаточные условия ε-разделимости. Сформулирован алгоритм построения разделимых ε-сетей двух множеств размера $[2d/\varepsilon]$. Рассмотрен пример использования данного алгоритма для двух множеств, сгенерированных из нормально распределенных совокупностей. Результаты классификации предложенного метода сравниваются с результатами классификации по методу опорных векторов.

Ключевые слова: ε-сети, размерность Вапника–Червоненкиса, разделение множеств, разделяющая плоскость.

ВВЕДЕНИЕ

В настоящей статье предложен новый метод решения задачи классификации, основанный на разделении двух множеств в пространстве R^d путем построения и разделения ε-сетей [1, 2] этих множеств в ранжированном пространстве относительно гиперплоскостей. Данная работа является логическим продолжением статьи [3], где было введено понятие области разделения — тех значений ε, при которых возможно разделить множества. В [3] были приведены примеры области разделения для случайных величин с наиболее часто используемыми законами распределения и сформулирована теорема о сходимости, для доказательства которой было введено понятие совокупности всех возможных ε-сетей произвольного множества, доказаны некоторые свойства этой совокупности, а также рассматривались одномерные случайные величины. Однако все утверждения несложно обобщались для многомерного случая. В настоящей статье изучаются выборки из векторных распределений.

ОБЛАСТЬ РАЗДЕЛЕНИЯ ДЛЯ МНОГОМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Рассмотрим многомерные случайные величины $\xi=(\xi_1,\xi_2,\dots,\xi_d)$ и $\eta=(\eta_1,\eta_2,\dots,\eta_d)$, генерирующие генеральные совокупности A и B. Пусть известны их функции распределения

$$F_\xi(x_1,x_2,\dots,x_d)=P(\xi_1<x_1,\ \xi_2<x_2,\dots,\ \xi_d<x_d)$$

и

$$F_\eta(x_1,x_2,\dots,x_d)=P(\eta_1<x_1,\ \eta_2<x_2,\dots,\ \eta_d<x_d)\,.$$

Пусть h — гиперплоскость в пространстве R^d; h^+,h^- — полупространства, которые порождаются гиперплоскостью h.

Определение 1. Множество D_l,

$$D_l:=\{(x,\ y)\in(0,1)^2:\exists h\in R^d,\ P\{\xi\in h_+\}\le x,\ P\{\eta\in h_-\}\le y\}\,,$$

называется областью разделения.

Рассмотрим примеры области разделения для двухмерных случайных величин.

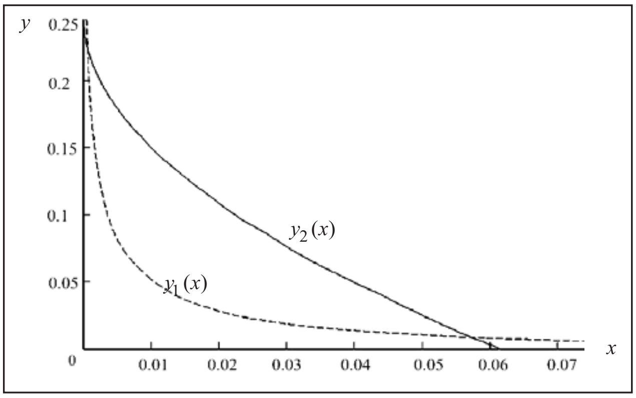

Рис. 1. Нижняя граница области разделения для нормального $y_1(x)$ и равномерного $y_2(x)$ распределений

Пример 1. Пусть независимые векторные случайные величины ξ и η распределены по нормальному закону с параметрами $\mu_A = (3;5)$, $\sigma^A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $\mu_B = (9;9)$, $\sigma^B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. Тогда область разделения ограничена кривыми $x = 1$, $y = 1$, $y = y_1(x)$. Нижняя граница области разделения изображена на рис. 1.

Пример 2. Пусть векторная случайная величина ξ распределена равномерно на квадрате $\{(x, y) \in (1;5)^2\}$, а векторная случайная величина η распределена равномерно на квадрате $\{(x, y) \in (4;8)^2\}$, причем случайные величины независимы. Тогда область разделения ограничена кривыми $x = 1$, $y = 1$, $y = y_2(x)$. Нижняя граница области разделения изображена на рис. 1.

НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ ε-РАЗДЕЛИМОСТИ

Рассмотрим ранжированное пространство (R^d, H^d) и множество $A \subset R^d$. Для ε-сетей множества A справедлива следующая лемма.

Лемма 1. Пусть задано ранжированное пространство (R^d, H^d) и конечное множество $A \subset R^d$, $|A| = n_A$; $N = \{x_1, \ldots, x_{n_N}\}$ — ε-сеть множества A в заданном ранжированном пространстве такая, что

1) $n_N \geq \dfrac{8\delta}{\varepsilon}$, где $\delta = VC(R^d, H^d)$,

2) $|A \setminus A_1| < n_A \varepsilon \left(1 - \dfrac{1}{8\delta}\right)$, где $A_1 = A \cap \mathrm{conv}_N$.

Тогда существует такое $i \in [1, n_N]$, что множество $N_1 = N \setminus \{x_i\}$ является ε-сетью множества A.

Доказательство. Докажем от противного. Пусть не существует такого $i \in [1, n_N]$, при котором множество $N_1 = N \setminus \{x_i\}$ является ε-сетью множества A. Это означает, что для любой точки $x_i \in N$ справедливо выражение

$$\exists r_i, |r_i \cap A| \geq \varepsilon n_A : r_i \cap N = \{x_i\}.$$

Очевидно, что это выражение не выполняется для внутренних точек conv_N, поэтому для любой внутренней точки x_i множество $N_1 = N \setminus \{x_i\}$ является ε-сетью множества A. Рассмотрим случай, когда множество N содержит только вершины многогранника conv_N. Обозначим $X_i = \{x_i^1, x_i^2, \ldots, x_i^{n_i}\}$ вершины conv_N, которые имеют общие ребра с произвольной вершиной x_i выпуклой оболочки множества N. Согласно предположению существует такое полупространство r_i, для которого $|r_i \cap A| \geq \varepsilon n_A$, при этом ни одна точка из X_i не принадлежит r_i. Обозначим m_{in} точки из $r \cap A$, принадлежащие conv_N, и m_{ex} — точки из $r \cap A$, не принадлежащие conv_N. Тогда $m_{ex} + m_{in} \geq \varepsilon n_A$.

Используем условие 2 леммы 1:

$$m_{in} \geq \varepsilon n_A - m_{ex} > \varepsilon n_A - \varepsilon n_A \left(1 - \frac{1}{8\delta}\right) = \frac{\varepsilon n_A}{8\delta}.$$

Согласно предположению неравенство (1) выполняется для всех точек из N. Тогда согласно условию 1 леммы 1

$$|A| \geq \frac{8\delta}{\varepsilon} m_{in} > n_A.$$

Полученное противоречие доказывает лемму 1.

Рассмотрим два ε-разделимых множества A и B из R^d. Согласно определению 1 из [3] существуют $A_1 \subset A$, $B_1 \subset B$, для которых справедливы соотношения

$$\operatorname{conv}(A \setminus A_1) \cap \operatorname{conv}(B \setminus B_1) = \varnothing \tag{1}$$

и

$$|A_1| + |B_1| < \varepsilon(n_A + n_B). \tag{2}$$

Для множеств A и B справедлива следующая теорема.

Теорема 1. Для того чтобы множества $A, B \subset R^d$ были ε-разделимыми, необходимо и достаточно, чтобы в ранжированном пространстве (R^d, H^d) существовали их ε-сети $N_{\varepsilon_A}^A$, $N_{\varepsilon_B}^B$, для которых выполнялись следующие условия:

1) $\operatorname{conv} N_{\varepsilon_A}^A \cap \operatorname{conv} N_{\varepsilon_B}^B = \varnothing$;

2) $\varepsilon_A n_A + \varepsilon_B n_B < \varepsilon(n_A + n_B)$;

3) $|N_{\varepsilon_A}^A| \leq \dfrac{8\delta}{\varepsilon_A}$, $|N_{\varepsilon_B}^B| \leq \dfrac{8\delta}{\varepsilon_B}$, где $\delta = VC(R^d, H^d)$.

Доказательство. Необходимость. Предположим, что множества A и B — ε-разделимые. Обозначим $|A_1| = n_{A_1}$, $|B_1| = n_{B_1}$. Пусть $\varepsilon_A = \dfrac{n_{A_1}}{n_A} + \theta$, $\varepsilon_B = \dfrac{n_{B_1}}{n_B} + \theta$, где θ следует из условия $(n_{A_1} + n_{B_1}) + 2\theta < \varepsilon(n_A + n_B)$.

Примем в качестве сетей множеств $N_{\varepsilon_A}^A = A \setminus A_1$, $N_{\varepsilon_B}^B = B \setminus B_1$.

Согласно определению ε-разделимости для множеств A и B выполняются соотношения (1) и (2). Тогда условие 1 теоремы 1 выполняется согласно (1), а условие 2 — согласно (2).

Покажем, что условие 3 выполняется. Рассмотрим сеть $N_{\varepsilon_A}^A$. Предположим, что $|N_{\varepsilon_A}^A| > \dfrac{8\delta}{\varepsilon}$. Тогда согласно лемме 1 размер сети $N_{\varepsilon_A}^A$ можно уменьшить на единицу. По индукции существует такое множество точек $X = \{x_1, x_2, \ldots x_p\}$, $X \subset N_{\varepsilon_A}^A$, $p \geq |N_{\varepsilon_A}^A| - \dfrac{8\delta}{\varepsilon}$, для которого множество $N_{\varepsilon_A}^A \setminus X$ является ε-сетью множества A размера не более $\dfrac{8\delta}{\varepsilon}$. По аналогии для сети $N_{\varepsilon_B}^B$ условие 3 также выполняется.

Достаточность. Предположим, что выполняются условия 1–3 теоремы 1. Докажем, что множества A и B — ε-разделимые. Предположим, что множества A, B не являются ε-разделимыми. Это означает, что (1) может выполняться только в случае, когда

$$n_{A_1} + n_{B_1} \geq \varepsilon(n_A + n_B). \tag{3}$$

Поскольку выполняется условие 2, то существует гиперплоскость L, относительно которой выпуклые оболочки $\operatorname{conv} N_{\varepsilon_A}^A$ и $\operatorname{conv} N_{\varepsilon_B}^B$ лежат в разных полупространствах. Предположим, что conv $N_{\varepsilon_A}^{A} \in L_+$ и conv $N_{\varepsilon_B}^{B} \in L_-$. Введем обозначения $A^N = \{x \in A: x \in L_+\}$, $B^N = \{x \in B: x \in L_-\}$.

Множества $A \setminus A^N$ и $B \setminus B^N$ согласно определению ε-сетей удовлетворяют соотношениям $|A \setminus A^N| < \varepsilon_A n_A$, $|B \setminus B^N| < \varepsilon_B n_B$.

Рассмотрим исключаемые множества $A_1 = A \setminus A^N$, $B_1 = B \setminus B^N$. Для этих множеств согласно условию 2 справедливо $n_{A_1} + n_{B_1} < \varepsilon(n_A + n_B)$. Таким образом, неравенство (3) для данных множеств не выполняется. Получили противоречие, которое доказывает достаточность.

Теорема 1 доказана.

Исходя из определения области разделения (см. [3]) $(\varepsilon_A, \varepsilon_B) \in D_{A,B}$. Согласно многомерному обобщению теоремы 2 из [3] имеет место сходимость $\lim_{n_A, n_B \to \infty} D(n_A, n_B) = D_l$. Поскольку функции распределения случайных величин ξ, η известны, согласно лемме 1 из [3] можно построить функцию

$$y(x) = \min(F_\eta(F_\xi^{-1}(1-x)), 1 - F_\eta(F_\xi^{-1}(x))), \tag{4}$$

которая будет разделять множества D_l и $\overline{D_l}$. Таким образом, чтобы ε-сети $N_A^{\varepsilon_A}$, $N_B^{\varepsilon_B}$ были разделимыми, необходимо выбирать $\varepsilon_A, \varepsilon_B$ так, чтобы удовлетворялось неравенство

$$\varepsilon_B \geq y(\varepsilon_A). \tag{5}$$

АЛГОРИТМ ПОСТРОЕНИЯ РАЗДЕЛИМЫХ ε-СЕТЕЙ ДВУХ ε-РАЗДЕЛИМЫХ МНОЖЕСТВ

Пусть из генеральных совокупностей, сгенерированных случайными величинами ξ и η, получены выборки множеств A и B объемами n_A, n_B. Задача состоит в нахождении разделяющей гиперплоскости L, для которой справедливо соотношение

$$P\{\xi \in L^+, \eta \in L^-\} = \sup_{l \in H^d} P\{\xi \in l^+, \eta \in l^-\}.$$

Выберем ε_A и ε_B, удовлетворяющие (5), и построим для множеств A и B ε-сети $N_A^{\varepsilon_A}$, $N_B^{\varepsilon_B}$ в ранжированном пространстве (R^d, H^d). Для удобства восприятия приведем алгоритм для двумерного случая. Для d-мерного случая алгоритм аналогичный.

Предположим, что множество A содержит точку с наименьшей ординатой. Обозначим $a_{\min}^1$ точку множества A с минимальной абсциссой и $a_{\max}^1$ — точку с максимальной абсциссой. Проведем $k = [1/\varepsilon_A] + 1$ вертикальных линий от $a_{\min}$ до $a_{\max}$ так, чтобы в каждую из $[1/\varepsilon_A]$ полученных полос размещалось одинаковое $(\varepsilon_A n_A)$ число точек. Искомые вертикальные линии, разделяющие полосы, описываются уравнениями $x = C_i$, $i = \overline{1, k}$. Рассмотрим функцию распределения случайной величины ξ_1 : $F_{\xi_1}(x) = P\{\xi_1 < x\}$. В первую полосу должны попасть $\varepsilon_A n_A$ точек. Константу C_1 находим из уравнения $F_{\xi_1}(C_1) = P\{\xi_1 < C_1\} = \dfrac{\varepsilon_A n_A}{n_A} = \varepsilon_A$. Тогда во вторую полосу попадает $\varepsilon_A n_A$ точек, если C_2 удовлетворяет уравнению $F_{\xi_1}(C_2) = P\{\xi_1 < C_2\} = 2\varepsilon_A$. Таким образом, $\forall i = \overline{1, k}$ находим константу C_i из уравнения $F_{\xi_1}(C_i) = i\varepsilon_A$.

Для каждой *i*-й полосы, $1 \le i \le [1/\varepsilon_A]$, введем обозначения: A^i — множество, содержащее точки множества A, попавшие в *i*-ю полосу; B^i— множество (возможно, пустое), содержащее точки множества B, попавшие в *i*-ю полосу; $a^2_{\min}(i)$, $a^2_{\max}(i)$ — точки множества A^i соответственно с наименьшей и наибольшей ординатами; $b^2_{\min}(i)$, $b^2_{\max}(i)$ — точки множества B^i соответственно с наименьшей и наибольшей ординатами.

Обозначим $N_A^{\varepsilon_A}$ множество точек, которые будем выбирать в *ε*-сеть множества *A*. Из *i*-й полосы множества *A* в $N_A^{\varepsilon_A}$ отбираем две точки. Первая из них — это точка $a^2_{\min}(i)$. Вторую точку из A^i в $N_A^{\varepsilon_A}$ отбираем по следующему алгоритму.

Если $B^i = \varnothing$ (т.е. в *i*-й полосе нет точек множества *B*), включаем в множество $N_A^{\varepsilon_A}$ точку $ay^i_{\max}$, иначе если $a^2_{\max}(i) < b^2_{\min}(i)$ (т.е. в *i*-й полосе выпуклые оболочки множеств не пересекаются), включаем в множество $N_A^{\varepsilon_A}$ точку $a^2_{\max}(i)$, иначе включаем в множество $N_A^{\varepsilon_A}$ точку $a(i)$ из множества A^i, которая является ближайшим соседом точки $b^2_{\min}(i)$. Точку $a(i)$ будем называть базисной точкой множества *A*.

Лемма 2. Множество точек в $N_A^{\varepsilon_A}$ является *ε*-сетью множества *A*.

Доказательство. Докажем от противного. Пусть $N_A^{\varepsilon_A}$ не является *ε*-сетью множества A. Это означает, что существует такое полупространство *H* в R^2, которому принадлежит не менее $\varepsilon_A n_A$ точек множества *A*, но ни одна из них не принадлежит множеству $N_A^{\varepsilon_A}$. Обозначим *Z* множество точек из *A*, принадлежащих полупространству H, причем $|Z| \ge \varepsilon_A n_A$. Рассмотрим произвольную точку $z \in Z$. Эта точка принадлежит одной из горизонтальных полос и одной из вертикальных. Согласно построению вместе с точкой *z* в полупространство *H* попадает хотя бы одна из крайних точек горизонтальной или вертикальной полосы либо базисная точка. Поскольку согласно построению множество $N_A^{\varepsilon_A}$ состоит из крайних или базисных точек множества A, то $Z \cap N_A^{\varepsilon_A} \ne \varnothing$, что противоречит предположению.

Лемма 2 доказана.

Согласно построению размер *ε*-сети множества *A* равен $[2d/\varepsilon_A]$; *ε*-сеть множества *B* строится по аналогичному алгоритму.

Разделение *ε*-сетей. Разделим множества $N_A^{\varepsilon_A}$ и $N_B^{\varepsilon_B}$ по методу линейного разделения выпуклых оболочек, изложенным в работах [4–6]. Представим краткое описание алгоритма данного метода.

1. Построить выпуклые оболочки conv $N_A^{\varepsilon_A}$ и conv $N_B^{\varepsilon_B}$.

2. Найти точки-промахи, если таковые существуют. Для уменьшения вычислительной сложности алгоритма будем искать промахи только среди базисных точек. Промахом множества $N_A^{\varepsilon_A}$ будем считать такую точку $x \in NB_A$, для которой $x \in \operatorname{conv} N_B^{\varepsilon_B}$. Множество точек-промахов множества $N_A^{\varepsilon_A}$ обозначим $P_A^{\varepsilon_A}$.

3. Исключить промахи из множества $N_A^{\varepsilon_A}$ и построить выпуклую оболочку множества $N'_A = N_A^{\varepsilon_A} \setminus P_A^{\varepsilon_A}$.

4. Среди ребер многоугольников conv N'_A и conv $N_B^{\varepsilon_B}$ найти ребро такое, чтобы точки множества N'_A и множества $N_B^{\varepsilon_B}$ были размещены в разных полупространствах относительно прямой l, содержащей это ребро.

5. Прямую l считать *ε*-разделяющей прямой множеств $N_A^{\varepsilon_A}$ и $N_B^{\varepsilon_B}$.

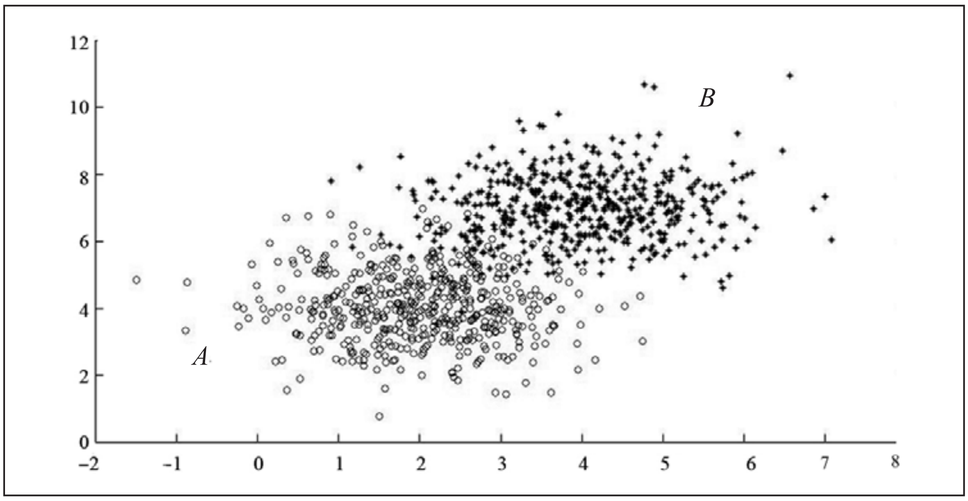

Рис. 2. Точки множеств A и B

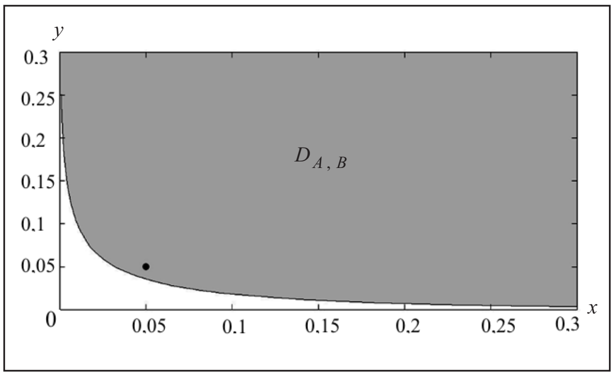

Рис. 3. Область разделения для множеств A и B и выбранная на ней точка (0.05; 0.05)

Оценка сложности алгоритма. Оценим сложность алгоритма предложенного выше метода разделения двух множеств с использованием их ε-сетей.

— Сложность алгоритма нахождения точек, попавших в соответствующую полосу, линейная. Таким образом, распределение точек обоих множеств по полосам имеет сложность $O(n)$, где $n = \max(|A|, |B|)$.

— Нахождение минимальных и максимальных точек в каждой из $[1/\varepsilon]$ горизонтальных и из $[1/\varepsilon]$ вертикальных полос можно оценить, как сложность алгоритма упорядочивания массива, состоящего из εn точек, т.е. $O(\varepsilon n \ln \varepsilon n)$, где $\varepsilon = \max(\varepsilon_A, \varepsilon_B)$.

— Поиск базисных точек также можно оценить как $O(\varepsilon n \ln \varepsilon n)$.

— Наиболее сложный шаг в методе линейного разделения выпуклых оболочек — построение выпуклых оболочек, сложность алгоритма которого с учетом $|N_A| = [2d/\varepsilon_A]$ оценивается как $O\left(\frac{2d}{\varepsilon}\ln\frac{2d}{\varepsilon}\right)$.

Таким образом, сложность алгоритма разделения множеств с использованием ε-сетей можно оценить как $O(n \ln n)$, что не больше, чем сложность алгоритма наиболее популярного на данный момент метода классификации — метода опорных векторов [7, 8].

Пример 3. Рассмотрим множества, имеющие двумерное нормальное распределение с параметрами $\mu^A = (2; 4)$, $\mu^B = (4; 7)$, $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. На рис. 2 изображены точки множеств A и B; каждое множество состоит из 500 точек.

Выпуклые оболочки множеств пересекаются, причем

$$\frac{|A \cap \text{conv } B| + |B \cap \text{conv } A|}{n_A + n_B} = 0.083\,.$$

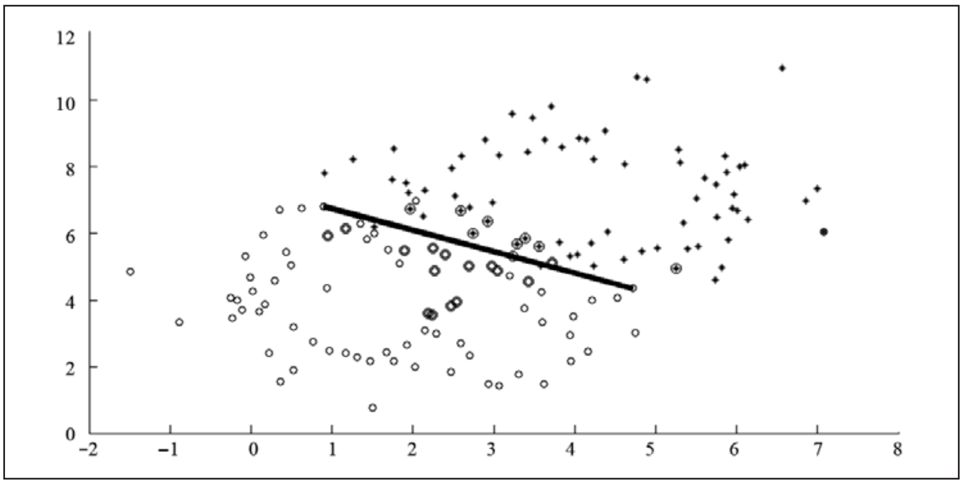

Рис. 4. ε-сети множеств A и B и разделяющая их прямая

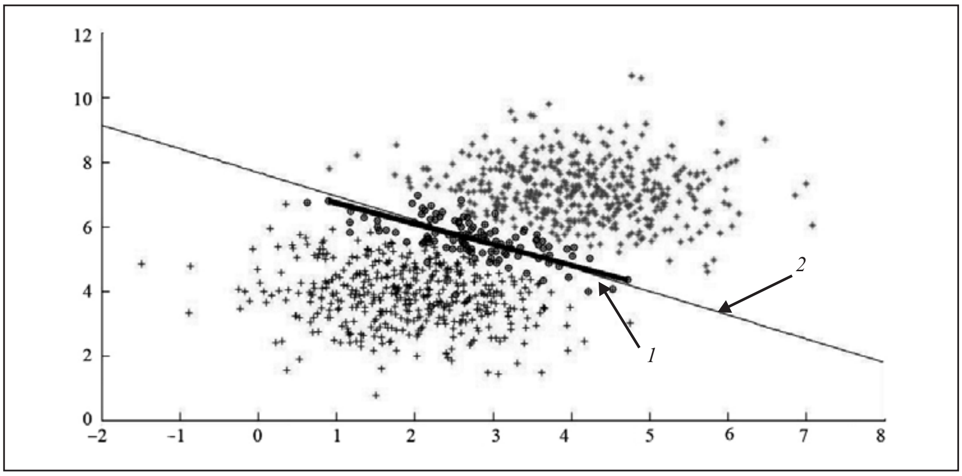

Рис. 5. Разделяющая прямая (*1*), построенная с использованием ε-сетей; разделяющая прямая (*2*), найденная по методу опорных векторов

Построим область $D_{A,B}$ и выберем точку $(\varepsilon_A, \varepsilon_B) \in D_{A,B}$ так, чтобы выполнялось неравенство

$$\frac{\varepsilon_A n_A + \varepsilon_B n_B}{n_A + n_B} \le 0.083 .$$

Пусть $\varepsilon_A = 0.05$; $\varepsilon_B = 0.05$ (см. рис. 3). На рис. 4 изображены полученные ε-сети $N_A^{\varepsilon_A}$, $N_B^{\varepsilon_B}$ и разделяющая их прямая. Эта же прямая является ε-разделяющей для множеств A и B. На рис. 5 изображены разделяющая прямая, построенная с использованием ε-сетей (*1*), и разделяющая прямая, найденная по методу опорных векторов (*2*).

Результаты классификации с использованием ε-сетей сравнивались с результатами классификации по методу опорных векторов. Для оценки качества классификации использовали метод Монте-Карло [9]. Было выполнено 1000 итераций для множеств, каждое из которых состояло из 10000 точек. Среднее количество ошибок классификации собственным методом — 4,30%, методом опорных векторов — 3,57%.

ЗАКЛЮЧЕНИЕ

В работе доказаны необходимые и достаточные условия ε-разделимости двух множеств в пространстве R^d, исходя из которых сформулирован алгоритм разделения двух ε-разделимых множеств путем разделения их ε-сетей в ранжированном пространстве (R^d, H^d). Предложенный алгоритм позволяет построить ε-сети размера $[2d/\varepsilon]$ и разделить их с вычислительной сложностью $O(n \ln n)$. Приведенный пример классификации двух нормально распределенных выборок показывает, что результат классификации с использованием ε-сетей аналогичен результату классификации по методу опорных векторов.

СПИСОК ЛИТЕРАТУРЫ

1. Haussler D., Welzl E. Epsilon-nets and simplex range queries // Discrete Comput. Geom. — 1987. — N 2. — P. 127–151.
2. Matousek J., Seidel R., Welzl E. How to net a lot with little: Small ε-nets for disks and halfspaces // Sixth Annual Symposium on Computational geometry, 1990. — P. 16–22.
3. Иванчук М.А., Малык И.В. Решение задач классификации с использованием ε-сетей // Кибернетика и системный анализ. — 2016. — **52,** № 4. — С. 134–144.
4. Іванчук М.А. Класифікація множин методом лінійного відокремлення їх опуклих оболонок // Математичне и комп'ютерне моделювання — 2015. — Вип. 12. — С. 113–120.
5. Иванчук М.А., Малык И.В. Разделение выпуклых оболочек как способ моделирования систем прогнозирования возникновения осложнений у больных // Проблемы управления и информатики. — 2015. — № 2. — С. 89–95.
6. Ivanchuk M.A., Malyk I.V. Separation of convex hulls as a way for modeling of systems of prediction of complications in patients // Journal of Automation and Information Sciences. — 2015. — **47,** Issue 4. — P. 78–84. — DOI: 10.1615/JAutomatInfScien.v47.i4.80.
7. Burges C.J.C. A tutorial on support vector machines for pattern recognition // Data Mining and Knowledge Discovery. — 1998. — **2.** — P. 121—167.
8. Tsang I.W., Kwok J.T., Cheung P.M. Core vector machines: Fast SVM training on very large data sets // Journal of Machine Learning Research. — 2005. — **6.** — P. 363–392.
9. Ермаков С.М. Метод Монте-Карло и смежные вопросы. — М.: Наука, 1971. — 328 с.

Надійшла до редакції 04.04.2016

М.А. Іванчук, І.В. Малик

АЛГОРИТМ ПОБУДОВИ ВІДОКРЕМЛЮВАНИХ ε-СІТОК ДВОХ МНОЖИН

Анотація. Запропоновано новий метод розв'язання задачі класифікації, що базується на відокремленні двох множин у просторі R^d. Доведено необхідні і достатні умови ε-відокремлюваності. Сформульований алгоритм побудови відокремлюваних ε-сіток двох множин розміром $[2d/\varepsilon]$. Розглянуто приклад використання цього алгоритму для двох множин, згенерованих з нормально розподілених сукупностей. Результати класифікації запропонованого методу порівняні з результатами класифікації за методом опорних векторів.

Ключові слова: ε-сітки, розмірність Вапніка–Червоненкіса, відокремлення множин, відокремлююча площина.

M.A. Ivanchuk, I.V. Malyk

AN ALGORITHM TO CONSTRUCT SEPARABLE ε-NETS OF TWO SETS

Abstract. The authors propose a new method to solve classification problem based on separation of two sets in space R^d. The necessary and sufficient conditions of ε-separability are proved. The algorithm of constructing two separable ε-nets of size $[2d/\varepsilon]$ is proposed. The paper contains an example of applying this algorithm to two sets generated from normally distributed sets. The classification results for the proposed method and for support vector machines are compared.

Keywords: ε-nets, VC-dimension, sets' separation, separating plane.

Иванчук Мария Анатольевна,
ассистент кафедры Буковинского государственного медицинского университета, Черновцы,
e-mail: mgracia@ukr.net.

Малык Игорь Владимирович,
кандидат физ.-мат. наук, доцент кафедры Черновицкого национального университета
имени Юрия Федьковича, e-mail: malyk.igor.v@gmail.com.