УДК [519.245+519.214]: 519.237.8

ПОБУДОВА ВІДОКРЕМЛЮВАНИХ ε-СІТОК ДВОХ МНОЖИН

***М.А. Іванчук**, асистент*
Буковинський державний медичний університет
mgracia@ukr.net
***І.В. Малик,** к.ф.-м.н., доцент*
Чернівецький національний університет ім.Ю.Федьковича
malyk.igor.v@gmail.com

В статті розглядається спосіб відокремлення двох множин в просторі R^n шляхом побудови та відокремлення ε-сіток цих множин в ранжованому просторі відносно гіперплощин.

Ivanchuk M.A., Malyk I.V. Building the separated ε-nets for two sets. In the article are discussed the method of the separation of two sets in the space R^n, which is based on building and separation ε-nets of these sets in the range space w.r.t. hyperplanes.

Ключові слова: ε-СІТКИ, ВІДОКРЕМЛЕННЯ МНОЖИН.
Keywords: ε-NETS, SETS' SEPARATION.

В даній роботі розглядається задача класифікації двох множин, що генеруються незалежними випадковими величинами. Нехай з генеральних сукупностей, що генеруються випадковими величинами ξ та η, отримані вибірки A та B об'ємами n_A, n_B. Задача полягає в знаходженні відокремлюючої гіперплощини L, для якої справедливе співвідношення

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

В [1] були доведені необхідні та достатні умови існування відокремлюваних ε-сіток [3] цих множин.

Теорема 1. Нехай множини A, B розміру n_A, n_B відповідно генеруються незалежними неперервними випадковими величинами ξ, η; 2) множини $D_{A,B} = \left\{(\varepsilon_1, \varepsilon_2) \in (0,1)^2 : \exists N_A^{\varepsilon_1}, N_B^{\varepsilon_2}, convN_A^{\varepsilon_1} \cap convN_B^{\varepsilon_2} = \varnothing\right\}$ та $\overline{D}_{A,B} := (0,1)^2 \setminus D_{A,B}$ розділені кривою $y_{A,B}(x)$. Тоді $\forall x \in (0,1)$ має місце співвідношення $\lim\limits_{n_A, n_B \to \infty} y_{A,B}(x) = y(x)$, де $y(x) = \min\left(F_\eta\left((F_\xi)^{-1}(1-x)\right), 1 - F_\eta\left((F_\xi)^{-1}(x)\right)\right)$.

Алгоритм побудови відокремлюваних ε-сіток двох множин

Виберемо $\varepsilon_A, \varepsilon_B$, що задовольняють нерівності $\varepsilon_B \geq y(\varepsilon_A)$ та побудуємо для множин A та B ε-сітки $N_A^{\varepsilon_A}$, $N_B^{\varepsilon_B}$ в ранжованому просторі (R^d, H^d). Для зручності наведемо алгоритм для двовимірного випадку. Для d-вимірного випадку алгоритм аналогічний.

Припустимо, що множина A містить точку з найменшою ординатою. Позначимо $a_{\min}^1$ - точка з множини A з мінімальною абсцисою та $a_{\max}^1$ - точка з максимальною абсцисою. Проведемо $k = \left[\frac{1}{\varepsilon_A}\right] + 1$ вертикальних ліній від $a_{\min}$ до $a_{\max}$ так, щоб в кожну з $\left[\frac{1}{\varepsilon_A}\right]$ отриманих смуг попала однакова ($\varepsilon_A n_A$) кількість точок. Шукані вертикальні лінії, що відокремлюють смуги, описуються рівняннями $x = C_i, i = \overline{1,k}$, де константу C_i, $\forall i = \overline{1,k}$ шукаємо з рівняння $F(C_i) = i\varepsilon_A$

Для кожної i-ї смуги, $1 \leq i \leq \left[\frac{1}{\varepsilon}\right]$ введемо позначення:

A^i - множина, що містить точки з множини A, що попали в i-ту смугу; B^i - множина (можливо, порожня), що містить точки з множини B, що попали в i-ту смугу; $a_{\min}^2(i)$, $a_{\max}^2(i)$ - точки множини A^i з найменшою та найбільшою ординатами; $b_{\min}^2(i)$,

$b_{\max}^2(i)$ - точки множини B^i з найменшою та найбільшою ординатами. Позначимо $N_A^{\varepsilon_A}$ - множина точок, які ми обиратимемо в ε-сітку множини A.

З i-ї смуги в $N_A^{\varepsilon_A}$ відбираємо дві точки. Перша – це точка $a_{\min}^2(i)$. Другу точку з A^i в $N_A^{\varepsilon_A}$ відбираємо за наступним алгоритмом.

Якщо $B^i = \varnothing$, включаємо в множину $N_A^{\varepsilon_A}$ точку $ay_{\max}^i$

інакше якщо $a_{\max}^2(i) < b_{\min}^2(i)$, включаємо в множину $N_A^{\varepsilon_A}$ точку $a_{\max}^2(i)$

інакше включаємо в множину $N_A^{\varepsilon_A}$ точку $a(i)$ з множини A^i, що є найближчим сусідом до точки $b_{\min}^2(i)$. Точку $a(i)$ називатимемо ***базисною точкою*** множини A.

Лема 2. Множина точок $N_A^{\varepsilon_A}$ є ε-сіткою множини A.

Будемо відокремлювати множини $N_A^{\varepsilon_A}$ та $N_B^{\varepsilon_B}$ за методом лінійного відокремлення опуклих оболонок (ЛВОО), який детально був розглянутий у [3]. Згідно теореми 1, відокремлююча гіперплощина для ε-сіток є ε-відокремлюючою для заданих множин.

В доповіді наведено новий метод розв'язання задачі класифікації, що полягає у побудові та відокремленні ε-сіток двох множин в ранжованому просторі $\left(R^d, H^d\right)$.

Література

1. Ivanchuk M. A. Using ε-Nets for Linear Separation of Two Sets in a Euclidean Space R^d / M. A. Ivanchuk, I. V. Malyk // Cybernetics and Systems Analysis. - Volume 51. - Issue 6 (2015). – P. 965-968.
2. Haussler D. Epsilon-nets and simplex range queries/ D. Haussler and E. Welzl// Discrete Comput. Geom. –1987. – №2. – 127–151
3. Іванчук М.А. Класифікація множин методом лінійного відокремлення їх опуклих оболонок // Математичне та

комп'ютерне моделювання - Випуск 12. – 2015 р. – С. 113-120