ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

9 • 2015

ІНФОРМАТИКА ТА КІБЕРНЕТИКА

УДК 519.7

Член-кореспондент НАН України **А. В. Анісімов, О. А. Галкін**

Побудова класифікаторів на основі ядерних оцінок щільності з використанням апостеріорних ймовірностей конкуруючих множин

Запропоновано підхід до побудови класифікаторів на основі ядерних оцінок щільності для розв'язання задач розпізнавання образів. Підхід грунтується на використанні апостеріорної ймовірності та розділової міри типу π-значення для ефективного розділення конкуруючих множин. Для кожної оцінки щільності класу застосовано сімейство оцінок щільності для кожної множини в широкому діапазоні смуг пропускання. Запропоновано та адаптовано процедуру об'єднання результатів класифікації на різних рівнях згладжування, що забезпечило гнучке використання різних смуг пропускання для різних пар конкуруючих класів. Статистичні невизначеності обчислено на основі приблизно оцінених ймовірностей помилкової класифікації.

Ключові слова: оцінка щільності, вагова функція, правило класифікації.

Постановка задачі. Для обчислення ядерних оцінок щільності в задачах розпізнавання необхідно нормалізувати всі дані в класі, використовуючи оцінку дисперсійної матриці класу для набуття даними сферичної форми, що дозволяє більш ефективно використовувати загальні смуги пропускання a_l для всіх координатних змінних. Припустимо, що z_{l1}, $z_{l2}, \ldots, z_{lm_l}$ є навчальною вибіркою даних з l-го класу, де $1 \leqslant l \leqslant L$. Для класифікації елемента z в один з L класів необхідно отримати оцінки щільності $\overline{h}_{la_l}(z)$ в точці z для всіх $l = 1, 2, \ldots, L$. У разі, коли випадкові вектори проходять лінійне перетворення, оцінки щільності для вихідних векторів даних можуть бути отримані з нормалізованих векторів даних з використанням формули перетворення для ймовірнісної функції щільності [1].

Для фіксованої пари смуг пропускання a_1 та a_2 для двох оцінок щільності класу та пари конкуруючих класів існує порядок між функціями $p_1\overline{h}_{1a_1}(z)$ та $p_2\overline{h}_{2a_2}(z)$, що визначає, який з двох класів є більш вірогідним. Апостеріорна ймовірність для першої множини даних у двокласовій задачі для заданого елемента даних z та заданої пари смуг пропускання a_1 та a_2 може бути задана як

$$\breve{P}_{a_1,a_2}(1|z) = \frac{p_1\overline{h}_{1a_1}(z)}{p_1\overline{h}_{1a_1}(z) + p_2\overline{h}_{2a_2}(z)}. \tag{1}$$

Для обчислення цих апостеріорних ймовірностей можна використовувати широкий спектр значень для a_1 та a_2. У двокласових задачах із застосуванням ядерних методів статистичного аналізу [2] елемент z класифікується в першу множину даних, якщо $p_1\overline{h}_{1a_1}(z) > p_2\overline{h}_{2a_2}(z)$.

Для заданого елемента z розглянемо ймовірність

$$P_{a_1,a_2}(z) = P\{p_1\overline{h}_{1a}(z) > p_2\overline{h}_{2a}(z)|z\}, \tag{2}$$

великі та малі значення якої дають розв'язки для першої та другої множин даних відповідно. Зауважимо, що оцінки щільності є середніми значеннями незалежних та однаково розподілених випадкових величин, а оцінки щільності для різних множин грунтуються на незалежних множинах даних. Тому нормальне наближення може використовуватися для оцінки вказаної ймовірності з високим ступенем точності для дуже великих розмірів вибірки для фіксованих a_1 та a_2. В результаті, використовуючи нормальне наближення з оціненими середніми значеннями та дисперсіями, отримуємо таку ймовірність:

$$P_{a_1,a_2}(z) = F\left(\frac{p_1\Omega[\overline{h}_{1a_1}(z)|z] - p_2\Omega[\overline{h}_{2a_2}(z)|z]}{\sqrt{p_1^2 D[\overline{h}_{1a_1}(z)|z] + p_2^2 D[\overline{h}_{2a_2}(z)|z]}}\right) = F\left(\frac{p_1\overline{h}_{1a_1}(z) - p_2\overline{h}_{2a_2}(z)}{\sqrt{p_1^2\overline{w}_{1a_1}^2(z) + p_2^2\overline{w}_{2a_1}^2(z)}}\right), \tag{3}$$

де m_1 та m_2 — розміри вибірки для двох класів; F — стандартна функція нормального розподілу; $\overline{w}_{la_l}^2(z)$ — оцінена дисперсія $\overline{h}_{la_l}(z)$ $(l = 1, 2)$, отримана із вибірки з використанням вибіркової дисперсії від

$$W_l^{-r}\Theta\{W_l^{-1}(z_{l1} - z)\}, \quad W_l^{-r}\Theta\{W_l^{-1}(z_{l2} - z)\}, \quad \ldots, \quad W_l^{-r}\Theta\{W_l^{-1}(z_{lm_l} - z)\}.$$

Оцінка ймовірності з високим ступенем точності. Нормальне наближення $P_{a_1,a_2}(z)$ може бути задане таким чином. Для заданого елемента z та пари смуг пропускання a_1 та a_2 візьмемо пару гіпотез H_0: $p_1\Omega\{\overline{h}_{1a}(z)\} \geqslant p_2\Omega\{\overline{h}_{2a}(z)\}$ та H_A: $p_1\Omega\{\overline{h}_{1a}(z)\} < p_2\Omega\{\overline{h}_{2a}(z)\}$. Якщо навчальна вибірка використовується для тестування даних гіпотез з застосуванням ядерних оцінок щільності, то нормальне наближення може бути визначене, як одностороннє π-значення [3].

Лема 1. *Припустимо, що* $\Omega[\Theta^2\{W_i^{-1}(z - z_{i1})\}|z] < \infty$ *для* $i = 1, 2$, *а також визначимо* $\varepsilon_{ia_i}(z) = \Omega\{\overline{h}_{ia_i}(z)\}$ *для* $i = 1, 2$. *За умови, якщо* $m_1/M \to \alpha (0 < \alpha < 1)$ *при* $M = m_1 + m_2 \to \infty$, *мають місце такі випадки:*

$$а) \quad \left|P_{a_1,a_2}(1|z) - \frac{p_1\varepsilon_{1a_1}(z)}{p_1\varepsilon_{1a_1}(z) + p_2\varepsilon_{2a_2}(z)}\right| = O_P(M^{-1/2}), \tag{4}$$

$$б) \quad |P_{a_1,a_2}(z) - \Lambda\{p_1\varepsilon_{1a_1}(z) > p_2\varepsilon_{2a_2}(z)\}| = O_P(M^{-1/2}e^{-VM}), \tag{5}$$

де $\Lambda\{\cdot\}$ *— характеристична функція;* $V > 0$.

Доведення. 1. Визначимо $\breve{M}_i = p_i\overline{h}_{ia_i}(z)$ для $i = 1, 2$. Оскільки $\breve{M}_i$ є середнім значенням незалежних та однаково розподілених випадкових величин, $\breve{M}_i$ сходиться до нормального розподілу із середнім значенням $\vartheta_i = p_i\phi_{ia_i}(z)$ та дисперсією $\mu_i = p_i^2 w_{ia_i}^2(z)$, що має порядок $O(m_i^{-1})$ при допустимій умові моменту. Дане твердження випливає з центральної граничної теореми. Далі визначимо $\mathrm{O}(\breve{M}_1, \breve{M}_2) = \breve{M}_1/(\breve{M}_1 + \breve{M}_2)$. Функція O є неперервно-диференційовною в $\breve{M}_1$ та $\breve{M}_2$, які є незалежними та додатно визначеними випадковими величинами.

Відзначимо, що з асимптотичного розкладу Тейлора відомо, що

$$\frac{\{\mathrm{O}(\breve{M}_1, \breve{M}_2) - \mathrm{O}(\vartheta_1, \vartheta_2)\}}{\mu} \xrightarrow{J} \text{нормаль } (0,1), \tag{6}$$

де

$$\mu = \left\{ \sum_{i=1}^{2} \mu_i \left(\frac{\partial \mathrm{O}}{\partial \breve{M}_i} \right)^2_{\breve{M}_1=\vartheta_1, \breve{M}_2=\vartheta_2} \right\}^{1/2}.$$

В результаті, оскільки $m_1/M \to \alpha$, а $m_2/M \to 1-\alpha (0 < \alpha < 1)$ при $M \to \infty$, одержуємо

$$|\mathrm{O}(\breve{M}_1, \breve{M}_2) - \mathrm{O}(\vartheta_1, \vartheta_2)| = O_P(M^{-1/2}). \tag{7}$$

2. Нехай $\Lambda\{p_1\phi_{1a_1}(z) > p_2\phi_{2a_2}(z)\} = 1$ при $\vartheta_1 > \vartheta_2$. Для фіксованих a_1, a_2 та z з частини 1 даної леми випливає, що

$$\frac{1}{\sqrt{\mu_1 + \mu_2}}[(\breve{M}_1 - \breve{M}_2) - (\vartheta_1 - \vartheta_2)] \xrightarrow{J} \text{нормаль}(0,1) \tag{8}$$

при $M \to \infty$.

Далі визначимо $T_{a_1,a_2}(z) = (\breve{M}_1 - \breve{M}_2)/\sqrt{\mu_1 + \mu_2}$. Зазначимо, що $T_{a_1,a_2}(z)/M$ ймовірнісно сходиться до постійного значення V, а $T_{a_1,a_2}(z) = O_P(M^{1/2})$, що можна визначити з порядку μ_1 та μ_2. Оскільки $f(z)/z < 1 - F(z) < (1/z - 1/z^3)f(z)$, має місце рівність

$$1 - P_{a_1,a_2}(z) = 1 - F(T_{a_1,a_2}(z)) = O_p(M^{-1/2}e^{-VM}), \tag{9}$$

де $z > 0$, а $f(\cdot)$ та $F(\cdot)$ означають ймовірнісну та кумулятивну функції щільності стандартного нормального розподілу відповідно. Лему доведено.

Асимптотична поведінка розділових мір. Для пари смуг пропускання (a_1, a_2) та заданого z оцінена апостеріорна ймовірність $P_{a_1,a_2}(z)$ сходиться до $p_1\varepsilon_{1a_1}(z)/[p_1\varepsilon_{1a_1}(z) + p_2\varepsilon_{2a_2}(z)]$ зі швидкістю $O(M^{-1/2})$. Проте $P_{a_1,a_2}(z)$ експоненціально сходиться до 0 або 1, залежно від $\varepsilon_{1a_1}(z)$, $\varepsilon_{2a_2}(z)$ та апостеріорних ймовірностей $P_{a_1,a_2}(1|z)$. Якщо $p_1\varepsilon_{1a_1}(z) < p_2\varepsilon_{2a_2}(z)$, $P_{a_1,a_2}(1|z)$ матиме швидкість збіжності $\sqrt{M}$ до значення, що є меншим 0,5, відповідні міри типу π-значення експоненціально сходяться до нуля набагато швидше. Тому $P_{a_1,a_2}(z)$ завжди даватиме більш переконливі результати, ніж $P_{a_1,a_2}(1|z)$ за чи проти першої множини даних для заданих (a_1, a_2) та при зростанні об'єму вибірки [4].

Для отримання рішення щодо класифікації елемента z необхідно знати, які значення пари смуг пропускання (a_1, a_2) приведуть до статистично більш достовірного результату класифікації. Для двокласової задачі розглянемо середню ймовірність помилкової класифікації, а саме:

$$\Psi(a_1, a_2) = p_1 \int\limits_{z \in \mathrm{Re}^q_{a_1,a_2}} h_1(z)\,dz + p_2 \int\limits_{z \in \mathrm{Re}_{a_1,a_2}} h_2(z)\,dz. \tag{10}$$

При фіксованому виборі (a_1, a_2) Re_{a_1,a_2} є множиною всіх z, що класифіковані до першої множини даних, а $\mathrm{Re}^q_{a_1,a_2}$ — множиною доповнення.

Для об'єднання результатів, отриманих на різних рівнях згладжування при досягненні остаточного рішення, необхідно формувати відповідні середньозважені апостеріорні ймовірності, обчислені для різних варіантів вибору (a_1, a_2). Для цього необхідно використовувати відповідну вагову функцію, яка повинна набувати більш великих значень для даної пари смуг пропускання, що призводить до нижчих показників помилкової класифікації.

Скориговані вагові функції. Визначимо $\Psi_0 = \min\limits_{a_1,a_2} \overline{\Psi}(a_1, a_2)$ та розглянемо вагові функції $d(a_1, a_2)$, які є спадними функціями від $\overline{\Psi}(a_1, a_2)$, що еквівалентно $\overline{\Psi}(a_1, a_2) - \Psi_0$.

Зазначимо, що оскільки продуктивність класифікатора буде нижчою порівняно з тривіальним класифікатором, який класифікує всі елементи до класу, що має найвищу апріорну ймовірність, $d(a_1, a_2)$ повинні зникати кожного разу, коли відповідне значення $\overline{\Psi}(a_1, a_2)$ перевищує будь-яку з двох апріорних ймовірностей. Для класифікації елемента z необхідно включати у вагах відповідну міру $P_{a_1,a_2}(z)$ типу π-значення, а також використовувати такі пари смуг пропускання, які призводять до більш переконливих розв'язків для одного або двох класів та відповідно налаштувати значення вагової функції. Налаштовані ваги будуть залежати як від загально оцінених ймовірностей помилкової класифікації, так і від характерного елемента даних [5].

У даній роботі використано скориговану вагову функцію

$$d_z(a_1, a_2) = d(a_1, a_2)|P_{a1,a2}(z) - 0{,}5|, \tag{11}$$

де

$$d(a_1, a_2) = \exp\left\{-\frac{1}{2}\frac{(\overline{\Psi}(a_1, a_2) - \Psi_0)^2}{\Psi_0(1 - \Psi_0)/M}\right\}, \tag{12}$$

якщо $\dfrac{\overline{\Psi}(a_1, a_2) - \Psi_0}{[\Psi_0(1 - \Psi_0)/M]^{1/2}} \leqslant \vartheta$ і $\overline{\Psi}(a_1, a_2) < \min\{p_1, p_2\}$, та $d(a_1, a_2) = 0$ — в іншому випадку. Для $M = m_1 + m_2$, коли найбільш ефективний класифікатор на основі ядерних оцінок щільності використовується для класифікації M незалежних елементів даних, Ψ_0 та $\Psi_0(1 - \Psi_0)/M$ можуть розглядатися як оцінки для середнього значення та дисперсії емпіричного показника помилкової класифікації. Постійна величина ϑ визначає максимальне число відхилень від мінімального оціненого показника помилкової класифікації в нормалізованій шкалі, за якою вагова схема ігнорує пару смуг пропускання (a_1, a_2), поставивши на них нульову вагу. У даному випадку $\vartheta = 0$ відповідає прикріпленню всіх ваг лише на пару смуг пропускання (a_1, a_2), для якої $\overline{\Psi}(a_1, a_2) = \Psi_0$.

Зауважимо, що відповідно до вибору вагової функції гауссівського типу, немає потреби розглядати значення ϑ, що є більшими трьох. Вибір скоригованої вагової функції, використаної у даній роботі, є дещо суб'єктивним, тому можна використовувати інші відповідні функції. Встановлено, що якщо застосовується прийнятна вагова функція, кінцевий результат не є надто чутливим до вагової процедури [6].

При наявності більше двох класів, наслідком чого є зростання обчислювальної складності, досить непросто знайти оптимальні смуги пропускання, що мінімізують оцінки загальної середньої ймовірності помилкової класифікації $\Psi(a_1, a_2, \ldots, a_L)$. У даному випадку необхідно розкласти багатокласові задачі в ряд задач бінарної класифікації, результати яких комбінуються для отримання кінцевого правила рішення. Найбільш ефективним інструментом для об'єднання результатів попарної класифікації є застосування методу мажоритарного голосування, який в задачі з J класами після $\dbinom{J}{2}$ попарних порівнянь класифікує елемент даних до класу, що має максимальне число голосів [7]. Однак застосування даного методу може призвести до появи областей нерозв'язності, де більше, ніж один клас може мати максимальне число голосів. Дану проблему можна вирішити, застосовуючи метод попарного з'єднання, який об'єднує оцінені апостеріорні ймовірності для різних попарних класифікацій для визначення кінцевих апостеріорних ймовірностей для конкуруючих класів [8].

Відомо, що $\overline{h}_{la_l}(z)$ є середнім значенням незалежних та однаково розподілених випадкових величин для фіксованого a_l. У випадку, коли розмір вибірки m_l збільшується, її дисперсія прямує до нуля та ймовірнісно сходиться до $\varepsilon_{la_l}(z) = \Omega\{\overline{h}_{la_l}(z)\} = \Theta_{a_l} * h_l(z)$, що є згорткою функції щільності h_l з ядром Θ та смугою пропускання a_l. Порядок $p_1\varepsilon_{1a_1}(z)$ та $p_2\varepsilon_{2a_2}(z)$ визначає асимптотичне правило розв'язання у двокласовій задачі. Тому $\varepsilon_{la}(z)$ $(l = 1, 2)$ зберігає порядок фактичних щільностей для всіх значень a, якщо розподіли множини даних задовольняють модель зсуву розташування, тобто $h_l(z) = c(z-\varepsilon_l)$ для загальної щільності c та параметрів розташування ε_l, а загальна смуга пропускання a використовується для обох множин даних. Крім того, у випадку рівних апріоних ймовірностей для всіх додатних значень $a = a_1 = a_2$ відповідна ймовірність помилкової класифікації асимптотично стає оптимальним байєсівським ризиком [9]. Тому, оскільки дисперсія прямує до нуля досить швидко, дана збіжність є значно швидшою для великих значень a. Зазначимо, що функції щільності множин даних не завжди можуть задовольняти умову симетрії. Для дослідження поведінки класифікатора на основі ядерних оцінок щільності для великих розмірів вибірок та великих смуг пропускання наведемо таку лему.

Лема 2. *Припустимо, що ядро Θ обмежене третіми похідними та є r-вимірною функцією щільності з модою в 0. Нехай функції h_1 та h_2 є такими, що $\int \|z\|^6 h_i(z)\,dz < \infty$. Також визначимо постійну величину $V_p = p_2/p_1$ та припустимо, що a_1, a_2 змінюються таким чином, що $a_2/a_1 = V_a$ та є постійною величиною. Отже, $\Psi(a_1, a_2)$ демонструє таку асимптотичну поведінку при $a_1 \to \infty$. Якщо $V_p = V_a^r$ при a_1, $a_2 \to \infty$, $\Psi(a_1, a_2)$ прямує до ймовірності помилкової класифікації квадратичного правила класифікації, що задається $\mathrm{Im}_{кв}(z) = 1$, якщо*

$$V_a^2 \Omega_{h_1}\left\{(z-Z)' K^2 \Theta(0)(z-Z)\right\} > \Omega_{h_2}\left\{(z-Z)' K^2 \Theta(0)(z-Z)\right\},$$

та $\mathrm{Im}_{кв}(z) = 2$ — в іншому випадку. Якщо $V_p > V_a^r$ при $a_1, a_2 \to \infty$, то $\Psi(a_1, a_2) \to p_1$. А також, якщо $V_p < V_a^r$ при $a_1, a_2 \to \infty$, то $\Psi(a_1, a_2) \to p_2$.

Доведення. На основі визначення $\overline{h}_{la_l}(z)$ $(l = 1, 2)$ встановлено, що

$$\Omega_{h_l}\{\overline{h}_{a_l}(z)\} = a_l^{-r} \Omega_{h_l}\left[\Theta\left\{\frac{z-Z}{a_l}\right\}\right] \tag{13}$$

та

$$D_{h_l}\{\overline{h}_{a_l}(z)\} = m_l^{-1} a_l^{-2r} D_{h_l}\left[\Theta\left\{\frac{z-Z}{a_l}\right\}\right] \tag{14}$$

з урахуванням того факту, що

$$\Psi(a_1, a_2) = p_1 \Omega_{h_1}\{\Lambda(p_1\overline{h}_{1a_1} < p_2\overline{h}_{2a_2})\} + p_2 \Omega_{h_2}\{\Lambda(p_1\overline{h}_{1a_1} > p_2\overline{h}_{2a_2})\}. \tag{15}$$

Використовуючи розклад Тейлора в 0, $\Theta\{(z-Z)/a_l\}$ може бути виражено як

$$\Theta\left\{\frac{z-Z}{a_l}\right\} = \Theta(0) + \frac{1}{2a_l^2}\{(z-Z)'\} K^2 \Theta(0)(z-Z) + \frac{1}{6a_l^3}\sum_{i,k,n} X_{i,k,n}, \tag{16}$$

де $K\Theta(0) = 0$), а $X_{i,k,n} = (z_i - Z_i)(z_k - Z_k)(z_n - Z_n) \dfrac{\partial^3\Theta(\kappa)}{\partial\kappa_i \partial\kappa_k \partial\kappa_m}\Big|_{\kappa=\rho}$ — для деякого проміжного вектора ρ між 0 та $(z - Z)/a_l$. Оскільки $\int \|z\|^6 h(z)\,dz < \infty$, а функція Θ обмежена третіми похідними, отримуємо такі рівності:

$$\Omega_{h_l}\{\overline{h}_{la_l}(w)\} = a_l^{-r}\left[\Theta(0) + \frac{1}{2a_l^2}\Omega_{h_l}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_l^{-3})\right], \tag{17}$$

$$D_{h_l}\{\overline{h}_{la_l}(w)\} = (4m_l a_l^{2r+4})^{-1}\left[D_{h_l}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_l^{-3})\right]. \tag{18}$$

Класифікатор на основі ядерних оцінок щільності класифікує елемент z до першої множини даних тоді і тільки тоді, коли

$$\begin{aligned}
&p_1\Omega_{h_1}\{\overline{h}_{1a_1}(z)\} > p_2\Omega_{h_2}\{\overline{h}_{2a_2}(z)\} \Leftrightarrow \\
&\quad\Leftrightarrow p_1 a_1^{-r}\left[\Theta(0) + \frac{1}{2a_1^2}\Omega_{h_1}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_1^{-3})\right] > \\
&\quad > p_2 a_2^{-r}\left[\Theta(0) + \left[\Theta(0) + \frac{1}{2a_2^2}\Omega_{h_2}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_2^{-3})\right]\right] \Leftrightarrow \\
&\quad\Leftrightarrow V_p V_a^{-r}\left[\Theta(0) + \frac{1}{2a_1^2}\Omega_{h_1}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_1^{-3})\right] > \\
&\quad > \left[\Theta(0) + \frac{1}{2a_2^2}\Omega_{h_2}\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a_2^{-3})\right],
\end{aligned} \tag{19}$$

оскільки дисперсія ядерних оцінок щільності асимптотично прямує до нуля для елемента z та пари смуг пропускання (a_1, a_2).

Отже, якщо $V_p = V_a^r$, можна перевірити, що наведена вище нерівність має місце тоді і тільки тоді, коли

$$V_a^2\Omega_{h_1}\{(z-Z)'K^2\Theta(0)(z-Z)\} > \Omega_{h2}\{(z-Z)'K^2\Theta(0)(z-Z)\} \tag{20}$$

для великих значень a_1 та a_2. У випадку, коли $V_p > V_a^r$, вищенаведена нерівність має місце, незалежно від значення елемента z для великих значень a_1 та $a_2 = V_a a_1$. Також, якщо $V_p < V_a^r$, результуючий класифікатор завжди асимптотично класифікує кожен елемент z до другої множини даних. Лему доведено.

В результаті квадратичний класифікатор стає лінійним класифікатором форми

$$\mathrm{Im}_{\text{лін}}(z) = \arg\min_i\left[z'K^2\Theta(0)\Omega_{h_i}(Z) - \frac{1}{2}\Omega_{h_i}\{Z'K^2\Theta(0)Z\}\right] \tag{21}$$

у випадку, коли $p_1 = p_2$ або $V_p = V_a = 1$. Крім того, даний лінійний класифікатор можна виразити у спрощеному вигляді:

$$\mathrm{Im}_{\text{лін}}(z) = \arg\max_i\left\{z'\varepsilon_i - \frac{1}{2}\varepsilon_i'\varepsilon_i\right\}, \tag{22}$$

де ε_i — параметр розташування для i-ї множини даних ($i = 1, 2$). Цей висновок має місце на підставі таких умов: 1) ядерна функція Θ є сферичною; 2) $K^2\Theta(0)$ є від'ємно визначеною функцією; 3) мають місце припущення щодо зсуву розташування та сферичної симетрії. Отриманий лінійний класифікатор є оптимальним байєсівським класифікатором. Тому, у такому особливому випадку, ймовірність помилкової класифікації Ψ асимптотично сходиться до оптимального байєсівського ризику.

Цитована література

1. *Godtliebsen F., Marron J. S., Chaudhuri P.* Significance in scale space for bivariate density estimation // J. of Computational and Graphical Statistics. – 2002. – **11**. – P. 3–21.
2. *Holmes C. C., Adams N. M.* A probabilistic nearest neighbor method for statistical pattern recognition // J. of the Royal Statistical Society. – 2002. – **64**. – P. 297–304.
3. *Hall P.* Large sample optimality of least squares cross validations in density estimation // The Annals of Statistics. – 1983. – **11**. – P. 1160–1173.
4. *Lachenbruch P., Mickey M.* Estimation of error rates in discriminant analysis // Technometrics. – 1968. – **10**. – P. 3–10.
5. *Silverman B. W.* Density estimation for Statistics and Data Analysis. – London: Chapman and Hall, 1986. – P. 1–7.
6. *Wand M., Jones M.* Kernel Smoothing. – London: Chapman and Hall, 1995. – P. 1–14.
7. *Ripley B.* Pattern Recognition and Neural Networks. – Cambridge: Cambridge University Press, 1996. – P. 1–17.
8. *Duda R., Hart P., Stork D.* Pattern Classification. – New York: Wiley, 2000. – P. 1–21.
9. *Chaudhuri P., Marron J.* Scale space view of curve estimation // The Annals of Statistics. – 2000. – **28**. – P. 410–427.

References

1. *Godtliebsen F., Marron J. S., Chaudhuri P.* J. of Computational and Graphical Statistics, 2002, **11**: 3–21.
2. *Holmes C. C., Adams N. M.* J. of the Royal Statistical Society, 2002, **64**: 297–304.
3. *Hall P.* The Annals of Statistics, 1983, **11**: 1160–1173.
4. *Lachenbruch P., Mickey M.* Technometrics, 1968, **10**: 3–10.
5. *Silverman B. W.* Density estimation for Statistics and Data Analysis, London: Chapman and Hall, 1986.
6. *Wand M., Jones M.* Kernel Smoothing, London: Chapman and Hall, 1995: 1–14.
7. *Ripley B.* Pattern Recognition and Neural Networks, Cambridge: Cambridge University Press, 1996: 1–17.
8. *Duda R., Hart P., Stork D.* Pattern Classification, New York: Wiley, 2000: 1–21.
9. *Chaudhuri P., Marron J.* The Annals of Statistics, 2000, **28**: 410–427.

Київський національний університет ім. Тараса Шевченка

Надійшло до редакції 12.05.2015

Член-корреспондент НАН Украины **А. В. Анисимов, А. А. Галкин**

Построение классификаторов на основе ядерных оценок плотности с использованием апостериорных вероятностей конкурирующих множеств

Киевский национальный университет им. Тараса Шевченко

Предложен подход к построению классификаторов на основе ядерных оценок плотности для решения задач распознавания образов. Подход основан на использовании апостериорной вероятности и разделительной меры типа π-значение для эффективного разделения конкурирующих множеств. Для каждой оценки плотности класса применено семейство оценок *плотности для каждого множества в широком диапазоне полос пропускания. Предложена и адаптирована процедура объединения результатов классификации на разных уровнях сглаживания, что обеспечило гибкое использование различных полос пропускания для различных пар конкурирующих классов. Статистические неопределенности вычислены на основе приближенно оцененных вероятностей ошибочной классификации.*

Ключевые слова: оценка плотности, весовая функция, правило классификации.

Corresponding Member of the NAS of Ukraine **A. V. Anisimov, O. A. Galkin**

Construction of classifiers based on kernel density estimations using the *a posteriori* probabilities of competing sets

Taras Shevchenko National University of Kiev

An approach is proposed to construct classifiers based on kernel density estimates for solving pattern recognition problems. The approach is based on the use of the a posteriori probability and a distributive π-type measure for the effective division of competing sets. The family of density estimates is applied to each set in a wide range of bandwidths for each estimate of the class density. A procedure is proposed and adapted to combine the classification results on different levels of smoothing that provides a flexible use of different bandwidths for different pairs of competing classes. Statistical uncertainties are calculated on the basis of approximate estimated probabilities of a misclassification.

Keywords: density estimate, weight function, classification rule.