УДК 519.7

А.А. Галкин

ПРИМЕНЕНИЕ МЕР СГЛАЖИВАНИЯ В НЕПАРАМЕТРИЧЕСКИХ ЯДЕРНЫХ КЛАССИФИКАТОРАХ С ИСПОЛЬЗОВАНИЕМ НОРМАЛЬНОЙ АППРОКСИМАЦИИ ВЕРОЯТНОСТЕЙ

Введение

На основе результатов современных исследований в области распознавания образов, а также в результате значительного сдвига ядерные оценки плотности распределения становятся достаточно масштабными для больших полос пропускания среднеквадратических интегрированных ошибок [1–3]. Кроме того, коэффициент ошибочной классификации ядерного классификатора $\Psi(a)$ достигает минимума и становится почти гладким для широкого диапазона больших значений полосы пропускания a. В данной работе предполагается, что исследуются m элементов данных выборки каждого множества, а общая полоса пропускания a используется для различных оценок плотности множества данных. Переменные $\Omega\{\overline{h}_{la}(z)\}$ и $\overline{h}_{la}(z)$ могут рассматриваться соответственно как теоретические и эмпирические масштабно-пространственные функции от l-го множества данных для различных вариаций сглаживающего параметра a. Отметим, что теоретические масштабно-пространственные функции $\Omega\{\overline{h}_{la}(z)\}$ являются свертками истинных функций плотности $h_l(z)$ с ядром Θ и полосой пропускания a. Кроме того, дисперсия ядерной оценки плотности распределения, что является средним значением множества независимых и одинаково распределенных случайных величин, уменьшается, и, как следствие, для некоторой фиксированной полосы пропускания a распределение $\overline{h}_{la}(z)$ имеет тенденцию к почти вырожденности в $\Omega\{\overline{h}_{la}(z)\}$. Данный результат имеет место, когда размер выборки достаточно большой и имеет тенденцию к росту. Отметим, что классификатор на основе ядерной оценки плотности распределения будет относить элемент данных к классу, имеющему наибольшее значение для теоретической функции масштабируемого пространства, когда размер выборки m стремится к бесконечности, а апостериорные вероятности для различных множеств данных равны при фиксированном значении a. Указанные результаты также имеют место для случая свертки, когда h и ядро Θ являются сферически-симметричными и строго нисходящими функциями расстояния от их центров симметрии. В данном случае функции масштабируемого пространства сохраняют порядок среди начальных плотностей, когда они удовлетворяют модели локализации смещения для всех значений a.

Постановка задачи

В данной работе предполагается, что функция Θ является плотностью с модой в 0 и ограничена третьей производной, а функция $h(z)$ такова, что $\int \| z \|^6 h(z)dz < \infty.$ Математическое ожидание и дисперсия $\overline{h}_a(z)$ задаются таким образом:

$$\Omega\{\overline{h}_a(x)\} = a^{-r}[\Theta(0) + (1/2a^2)\Omega_h\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a^{-3})] \tag{1}$$

и

$$D\{\overline{h}_a(z)\} = (4ma^{2r+4})^{-1}[D_h\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a^{-1})] \tag{2}$$

при $a \to \infty.$ В данном случае математическое ожидание и дисперсия $\overline{h}_a(z)$ могут быть записаны как

$$\Omega\{\overline{h}_a(z)\} = a^{-r}\Omega_h[\Theta\{(z-Z)/a\}] \tag{3}$$

и

$$D\{\overline{h}_a(z)\} = m^{-1}a^{-2r}D_h[\Theta\{(z-Z)/a\}]. \tag{4}$$

Выражение $\Theta\{(z-Z)/a\}$ может быть записано как

$$\Theta\{(z-Z)/a\} = \Theta(0) + (1/2a^2)\{(a-Z)'K^2\Theta(0)(z-Z)\} + (1/6a^3)\sum_{i,l,k} X_{i,l,k} \tag{5}$$

с использованием разложения Тейлора в окрестности 0, где $K\Theta(0) = 0,$ а

$$X_{i,l,k} = (z_i - Z_i)(z_l - Z_l)(z_k - Z_k)\frac{\partial^3\Theta(\kappa)}{\partial\kappa_i\,\partial\kappa_l\,\partial\kappa_k}\Big|_{\kappa=\rho} \tag{6}$$

для некоторого промежуточного вектора ρ между 0 и $(z-Z)/a.$

В результате получаем такие равенства:

$$\Omega_h[\Theta\{(z-Z)/a\}] = \Theta(0) + (1/2a^2)\Omega_h\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a^{-3}) \tag{7}$$

и

$$\begin{aligned} D_h[\Theta\{(z-Z)/a\}] &= D_h[(1/2a^2)\{(z-Z)'K^2\Theta(0)(z-Z)\} + (1/6a^3)\sum_{i,l,k} X_{i,l,k}] = \\ &= (1/4a^4)D_h\{(z-Z)'K^2\Theta(0)(z-Z)\} + O(a^{-5}), \end{aligned} \tag{8}$$

поскольку функция Θ ограничена третьей производной, а $\int \| z \|^6 h(z)dz < \infty.$

Теорема 1. Пусть $\overline{h}_{1a}$ и $\overline{h}_{2a}$ — ядерные оценки плотности распределения для двух функций плотности h_1 и h_2 соответственно, где функции $h_1,$ h_2 и Θ удовлетворяют условиям (1) и (2). Тогда для $\overline{h}_{1a}$ и $\overline{h}_{2a}$ имеют место следующие результаты сходимости:

(а) $P\{\overline{h}_{1a}(z) < \overline{h}_{2a}(z)\} \to 0$ при а $\to \infty,$ если $p_1 > p_2;$

(б) $P\{\overline{h}_{1a}(z) < \overline{h}_{2a}(z)\} \to 0\,(\to 1)$ при $m, a \to \infty$ в зависимости от $z'K^2\Theta(0) \times$ $\times \{\Omega_{h_2}(Z) - \Omega_{h_1}(Z)\} > (<)\,(1/2)[\Omega_{h_2}\{Z'K^2\Theta(0)Z\} - \Omega_{h_1}\{Z'K^2\Theta(0)Z\}]$ при условии $p_1 = p_2 = 1/2.$

ISSN 0572-2691

Доказательство. Пусть $X_a(z) = p_1\bar{h}_{1a}(z) - p_2\bar{h}_{2a}(z), \varepsilon_a(z) = \Omega\{X_a(z)\}$ и $w_a^2(z) = D\{X_a(z)\}$.

(а) $w_a^2(z)/\varepsilon_a^2(z) \to 0,$ а $\varepsilon_a(z)$ остается положительным, когда $p_1 > p_2$ и $a \to \infty.$ Поэтому из неравенства Чебышева имеем $\lim\limits_{a\to\infty} P\{X_a(z) \le 0\} = 0.$

(б) В случае, когда $p_1 = p_2$, из (1) и (2) очевидно следующее:

- знаки для $\varepsilon_a(z)$ и для

$$z'K^2\Theta(0)\{\Omega_{h_2}(Z) - \Omega_{h_1}(Z)\} - (1/2)[\Omega_{h_2}\{Z'K^2\Theta(0)Z\} - \Omega_{h_1}\{X'K^2\Theta(0)Z\}]$$

одинаковы, когда $a \to \infty$;

- $w_a^2(z)/\varepsilon_a^2(z) \to 0$ при $m,\ a \to \infty.$

Из неравенства Чебышева следует, что

$$P\{Z_a(z) \le 0\} \ge \frac{\varepsilon_a^2(z)}{\varepsilon_a^2(z) + w_a^2(z)}, \text{ когда } \varepsilon_a(z) \le 0, \tag{9}$$

и

$$PZ_a(z) \le 0 \le \frac{w_a^2(z)}{\varepsilon_a^2(z) + w_a^2(z)}, \text{ когда } \varepsilon_a(z) > 0. \tag{10}$$

В неравенствах (9) и (10) правые части стремятся к 1 и 0 соответственно при $m, a \to \infty$. В результате $P\{X_a(z) \le 0\}$ также стремится к 1 и 0.

Теорема доказана.

Средняя вероятность ошибочной классификации при выполнении классификатора на основе ядерной оценки плотности распределения приближается к результатам линейного классификатора, который задается таким образом:

$$\Im_{lin.}(z) = \arg\min_l [z'K^2\Theta(0)\Omega_{h_l}(Z) - (1/2)\Omega_{h_l}\{Z'K^2\Theta(0)Z\}]. \tag{11}$$

Данный результат имеет место, когда априорные вероятности равны при $m, a \to \infty$, а h_l являются функциями плотности, удовлетворяющими условию $\int \|z\|^6 h_l(z)\,dz < \infty$ для всех $l = 1, 2, ..., L.$ В данном случае ядро Θ является функцией плотности с модой в 0 и ограничено третьей производной.

Для проверки полученных результатов примем, что G_{il}^a — событие, где $\bar{h}_{ia}(z) - \bar{h}_{la}(z) > 0$ для $1 \le i \ne l \le L$. Очевидно, что $P(G_{il}^a) + P(G_{li}^a) = 1$ и $\sum\limits_{l:l\ne i} P(G_{il}^a) - (L-2) \le P\left\{\bigcap\limits_{l:l\ne i} G_{il}^a\right\} \le \min\limits_{l\ne i} P(G_{il}^a).$ Поэтому, для $\forall i, P\left\{\bigcap\limits_{l:l\ne i} G_{il}^a\right\} \to 1$ тогда и только тогда, когда $P(G_{il}^a) \to 1$ для всех $l \ne i.$ Ограниченный линейный классификатор почти эквивалентен классификатору, который относит элемент данных z к классу l_0, что максимизирует $z'\Omega_{h_l}(Z) - (1/2)\,\Omega_{h_l}(ZZ'),$ поскольку $K^2\Theta(0)$ является отрицательно-определенной функцией, или минимизирует $\Omega_{h_l}(\|z - Z\|^2)$ для $1 \le l \le L.$ Данный вывод имеет место, когда ядро Θ является сферически-симметричной и строго убывающей функцией нормы ее аргумента [4].

Далее исследуем поведение средней вероятности ошибочной классификации на основе оценки плотности, когда $m, a \to \infty.$ Предполагается, что функции плотности h_l удовлетворяют модели сдвига таким образом, что $h_l(z) = c(z - \varepsilon_l)$ для общей функции плотности c с нулевым средним значением и параметром локализации ε_l для всех $l = 1, 2, ..., L$.

Средняя вероятность ошибочной классификации ядерного классификатора эквивалентна линейному классификатору для задачи n-класса, если существуют n-максимумы среди априорных вероятностей, а $p_{l_1} = p_{l_2} = ... = p_{l_n} > p_l$ для всех $l \notin \{l_1, l_2, ..., l_n\}$. Это возможно, когда классы имеют максимальную априорную вероятность. Кроме того, средний показатель ошибочной классификации приближается к элементарному классификатору, который относит все данные к множеству l_0, если существует такое l_0, что $p_{l_0} > p_l$ для всех $l \neq l_0$. Средний показатель ошибочной классификации также демонстрирует поведение такого линейного классификатора, как

$$\Im_{lin.}(z) = \arg\min_l [z'K^2\Theta(0)\,\varepsilon_l - \{\varepsilon_l' K^2\Theta(0)\,\varepsilon_l\}/2], \tag{12}$$

если $p_1 = p_2 = ... = p_l$. Итак, для $i \neq j$, $\Omega_{h_i}(Z) - \Omega_{h_l}(Z) = \varepsilon_i - \varepsilon_l$, а $\Omega_{h_i}(Z'K^2 \times \times \Theta(0)Z) - \Omega_{h_l}(Z'K^2\Theta(0)Z) = \varepsilon_i' K^2\Theta(0)\varepsilon_i - \varepsilon_l' K^2\Theta(0)\varepsilon_l$, поскольку функции плотности множеств данных удовлетворяют модели локального сдвига с параметром локализации ε_l для l-класса.

Теорема 2. Пусть ядро Θ является симметричным в окрестности 0 с $\int \|x\|^3\Theta(x)dx < \infty$, а функция плотности h ограничена третьей производной. Тогда, при $a \to \infty$, математическое ожидание и дисперсия $h_a(z)$ задаются таким образом:

$$\Omega\{\bar{h}_a(z)\} = h(z) + O(a^2), \tag{13}$$

$$D\{\bar{h}_a(z)\} = (ma^r)^{-1}\{\gamma h(z) + O(a^2)\} \tag{14}$$

при $a \to \infty$ и $\gamma = \int \Theta^2(x)\,dx$.

Доказательство. Для некоторого промежуточного вектора ρ между z и $z - ax$ имеет место следующее равенство:

$$\begin{aligned} \Omega\{h_a(z)\} &= \Omega_h[a^{-d}\Theta\{(z-Z)/a\}] = a^{-r}\int \Theta\{(z-Z)/a\}h(Z)dZ = \int \Theta(x)\,h(z- \\ &- ax)\,dx = \int \Theta(x)[h(z) - a\{x'Kh(z)\} + (a^2/2)\{x'K^2h(z)x\} + (a^3/3!)\sum_{i,l,k} X_{i,l,k}]dx, \end{aligned} \tag{15}$$

где $X_{i,l,k} = x_i x_l x_k \dfrac{\partial^3 h(\kappa)}{\partial\kappa_i \partial\kappa_l \partial\kappa_k}|_{t=\xi}$.

Поэтому

$$\Omega\{\bar{h}_a(z)\} = h(z) + \frac{a^2}{2}\int \{x'K^2h(z)x\}dx + o(a^2) = h(z) + O(a^2), \tag{16}$$

а также, учитывая, что $\gamma = \int \Theta^2(x)\,dx$, получаем следующее равенство:

$$\begin{aligned} &\Omega_h[a^{-2r}\Theta^2\{(z-Z)/a\}] = a^{-r}\int \Theta^2(x)[h(z) - a\{x'Kh(z)\} + \\ &+ (a^2/2)\{x'K^2h(z)x\} + (a^3/3)\sum_{i,l,k} X_{i,l,k}]dx = a^{-r}[\gamma h(z) + O(a^2)]. \end{aligned} \tag{17}$$

В результате из приведенных выше соображений следует, что

$$D\{\bar{h}_a(z)\} = m^{-1}D_h\{a^{-r}\Theta((z-Z)/a)\} = m^{-1}a^{-r}\{\gamma h(z) + O(a^2)\}. \tag{18}$$

Теорема доказана.

Теорема 3. Пусть функции плотности $h_1, h_2, ..., h_j$ и ядро Θ удовлетворяют условиям, указанным в теореме 2. Для каждого $l \in \{1, 2, ..., L\}$ и $M \to \infty$ предполагается, что $a_{m_l} \to 0$ и $a_{m_l} / a_{m_i} \to V_{li} > 0$ для всех i, $m_l a_{m_l}^r \to \infty$, а $m_l / M \to \alpha_l$ такое, что $0 < \alpha_i < 1$. Определим $Z_1, Z_2, ... Z_L$, которые являются независимыми нормально распределенными величинами, где

$$\Omega(Z_i) = p_i \varepsilon_{ia_{m_i}} = \Omega\{p_i \bar{h}_{ia_{m_i}}(z)\} \tag{19}$$

и

$$D(Z_i) = p_i^2 w_{ia_{m_i}}^2 = D\{p_i \bar{h}_{ia_{m_i}}(z)\} \tag{20}$$

для некоторого $z \in R^r$. Тогда для произвольного значения z получаем

$$\lim_{M\to\infty} \left| P(Z_1) > Z_i, \ (\text{для всех } i \neq 1) - \prod_{i\neq 1} \Phi\left(\frac{p_1 \varepsilon_{1a_{m_1}} - p_i \varepsilon_{ia_{m_i}}}{p_i w_{ia_{m_i}}}\right)\right| = 0. \tag{21}$$

Доказательство. Для всех $i \neq 1$

$$P\{Z_i < Z_1\} = \int P\{Z_i < z\}\, c(z)dz = \int \prod_{i\neq 1} F\left(\frac{z - p_i \varepsilon_{ia_{m_i}}}{p_i w_{ia_{m_i}}}\right) c(z)dz = \Omega_c\left\{\prod_{i\neq 1} F\left(\frac{z - p_i \varepsilon_{ia_{m_i}}}{p_i w_{ia_{m_i}}}\right)\right\}, \tag{22}$$

где $c(\cdot)$ — вероятностная функция плотности от Z_1. Если принять, что $\lambda_M(z) = \prod_{i\neq 1}\left(\frac{z - p_i \varepsilon_{ia_{m_i}}}{p_i w_{ia_{m_i}}}\right)$, а также использовать разложение Тейлора в окрестности $p_1 \varepsilon_{1a_{m_1}}$, выражение $\lambda_M(z)$ может быть определено как $\lambda_M(z) = \lambda_M(p_1 \varepsilon_{1a_{m_1}}) + (z - \; - p_1 \varepsilon_{1a_{m_1}})\lambda'_M(\rho)$ для некоторого значения ρ, что находится между $p_1 \varepsilon_{1a_{m_1}}$ и z.

Итак,

$$(z - p_1 \varepsilon_{1a_{m_1}})\lambda'_M(\rho) = \sum_{l=2}^{L} \frac{z - p_1 \varepsilon_{1a_{m_1}}}{p_l w_{la_{m_l}}} \gamma_{la_{m_l}}(\rho), \tag{23}$$

где $\gamma_{la_{m_l}}(\rho) = f\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right) \prod_{k:\, k\neq 1,\, l} F\left(\frac{\rho - p_k \varepsilon_{ka_{m_k}}}{p_k w_{ka_{m_k}}}\right)$.

В результате получаем

$$\Omega\left(\left|\frac{z - p_1 \varepsilon_{1a_{m_1}}}{p_l w_{la_{m_1}}} \gamma_{la_{m_l}}(\rho)\right|\right) \le \Omega\left\{\left|\frac{z - p_1 \varepsilon_{1a_{m_1}}}{p_l w_{la_{m_l}}} f\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right)\right|\right\} \le \Omega^{1/2} \times$$
$$\times \left(\frac{z - p_1 \varepsilon_{1a_{m_1}}}{p_l w_{la_{m_l}}}\right)^2 \Omega^{1/2}\left\{f^2\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right)\right\} = \frac{p_1^2}{p_l^2}\frac{w_{1a_{m_1}}^2}{w_{la_{ml}}^2}\Omega^{1/2}\left\{f^2\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right)\right\}, \tag{24}$$

где $|F(\cdot)| \le 1$.

Как следует из теоремы 2, $W^2_{1a_{m_1}}$ и $w^2_{la_{m_l}}$ сходятся к 0 при $M \to \infty$, но $w^2_{1a_{m_1}} / w^2_{la_{m_l}}$ сходится к положительной постоянной величине [5].

Поэтому, используя теорему Лебега о мажорируемой сходимости, получаем

$$f^2\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right) \to 0 \Rightarrow \Omega\left\{ f^2\left(\frac{\rho - p_l \varepsilon_{la_{m_l}}}{p_l w_{la_{m_l}}}\right)\right\} \to 0 \Rightarrow \Omega\left|(z - p_1 \varepsilon_{1a_{m_1}})\tilde{\lambda}'_M(\rho)\right| \to 0 \quad (25)$$

при $M \to \infty$.

Теорема доказана.

Для работы с большими выборками необходима нормальная аппроксимация распределения ядерных оценок плотности. Поскольку ядерные оценки плотности распределения являются средним арифметическим независимых и одинаково распределенных случайных величин, в такой аппроксимации не будет значительной потери точности. Если принять среднее значение и дисперсию $\varepsilon_{la_l}(z)$ и $w^2_{la_l}(z)$ для $\bar{h}_{la_l}(z)$, $l = 1, 2, ..., L$, соответственно, средняя вероятность ошибочной классификации может быть аппроксимирована следующей функцией:

$$\begin{aligned} o(a_1, a_2, ..., a_L) = 1 - \sum_{l=1}^{L} p_l \int \Bigg[\int \prod_{i \neq l} F\left\{\frac{g - p_i \varepsilon_{ia_i}(z)}{p_i w_{ia_i}(z)}\right\} \times \\ \times f\{g, p_l \varepsilon_{la_l}(z), p_l w_{la_l}(z)\}\, dg \Bigg] h_l(z)\, dz, \end{aligned} \quad (26)$$

где $f(\cdot, \varepsilon, w)$ — функция плотности вероятностей нормального распределения со средним значением ε и стандартным отклонением w, а $F(\cdot)$ — кумулятивная функция стандартного нормального распределения [6].

Теорема 4. Пусть функции плотности $h_1, h_2, ..., h_L$ имеют ограниченные третьи производные, а ядро Θ является ограниченным и симметричным относительно 0, что удовлетворяет условию $\int \|x\|^3 \Theta^2(x) dx < \infty$. Для каждого значения $l \in \{1, 2, ..., L\}$ предполагается, что $a_{m_l} \to 0$ и $a_{m_l} / a_{m_i} \to V_{li} > 0$ для всех i, $m_l a^r_{m_l} \to \infty$ и $m_l / M \to \alpha_l$ такое, что $0 < \alpha_l < 1$ $(M \to \infty)$.

Тогда

$$\left| \Psi(a_{1_{m_1}}, a_{2_{m_2}}, ..., a_{L_{m_L}}) - o(a_{1_{m_1}}, a_{2_{m_2}}, ..., a_{L_{m_L}}) \right| \to 0, \quad (27)$$

а $\Psi(a_{1_{m_1}}, a_{2_{m_2}}, ..., a_{L_{m_L}})$ и $o(a_{1_{m_1}}, a_{2_{m_2}}, ..., a_{L_{m_L}})$ демонстрируют поведение, соответствующее оптимальному байесовскому риску при $M \to \infty$.

Доказательство. Поскольку $p_l \varepsilon_{la_{m_l}}(z)$ и $p^2_l w^2_{la_{m_l}}(z)$ — среднее значение и дисперсия для $p_l \bar{h}_{ia_{m_i}}(z)$, выражение $o(a_{m1}, a_{m2}, ..., a_{mL})$ задается таким образом:

$$\begin{aligned} o(a_{m1}, a_{m2}, ..., a_{mL}) = 1 - \sum_{l=1}^{L} p_l \int \Bigg[\int \prod_{i \neq l} F\left\{\frac{g - p_i \varepsilon_{ia_{m_i}}(z)}{p_i w_{ia_{m_i}}(z)}\right\} \times \\ \times f\{g, p_l \varepsilon_{la_{m_l}}(z), p_l w_{la_{m_l}}(z)\} dg \Bigg] h_l(z) dz. \end{aligned} \quad (28)$$

ISSN 0572-2691

Для всех l

$$\left| \int \left[\prod_{i \neq l} F\left\{ \frac{g - p_i \varepsilon_{ia_{m_i}}(z)}{p_i w_{ia_{m_i}}(z)} \right\} f\{g,\, p_l \varepsilon_{la_{m_l}}(z),\, p_l w_{la_{m_l}}(z)\} \right] dg - \right.$$
$$\left. - \prod_{i \neq l} F\left\{ \frac{p_l \varepsilon_{la_{m_l}}(z) - p_i \varepsilon_{ia_{m_i}}(z)}{p_i w_{ia_{m_i}}(z)} \right\} \right| \to 0, \tag{29}$$

поскольку $M \to \infty$. Данный результат следует из теоремы 3. Также для каждого l

$$\prod_{i \neq l} F\left\{ \frac{p_l \varepsilon_{la_{m_l}}(z) - p_i \varepsilon_{ia_{m_i}}(z)}{p_i w_{ia_{m_i}}(z)} \right\} \to \begin{cases} 1, & \text{если } p_l h_l(z) > p_i h_i(z) \text{ для всех } i \neq l, \\ 0, & \text{в противном случае,} \end{cases} \tag{30}$$

поскольку $M \to \infty$. Данный результат следует из теоремы 2. Из теоремы Лебега о мажорируемой сходимости следует, что $o(a_{m1}, a_{m2}, ..., a_{mL}) \to 1 - \sum_{l=1}^{L} p_l \int_{h_l > h_i\ \forall i \neq l} h_l(z) dz,$ что является оптимальным байесовским риском [7].

Условие Линдеберга для многомерной центральной граничной теоремы имеет место для $\left\{ \frac{\bar{h}_{1a_{m1}}(z) - \varepsilon_{1a_{m1}}(z)}{w_{1a_{m1}}(z)}, \frac{\bar{h}_{2a_{m2}}(z) - \varepsilon_{2a_{m2}}(z)}{w_{2a_{m2}}(z)}, ..., \frac{\bar{h}_{La_{mL}}(z) - \varepsilon_{La_{mL}}(z)}{w_{La_{mL}}(z)} \right\},$ если учитывать асимптотические порядки $\Omega\{\bar{h}_{la_{m_l}}(z)\}$ и $D\{\bar{h}_{la_{m_l}}(z)\}$, полученные в теореме 2.

Таким образом, используя результаты по равномерной сходимости к многомерным нормальным вероятностям для выпуклых множеств с границами, которые имеют нулевую меру Лебега [8], получаем следующую сходимость для всех $i \neq l$:

$$\left| P\{p_l \bar{h}_{la_l}(z) > p_i \bar{h}_{ia_i}(z)\} - \int \prod_{i \neq l} F\left\{ \frac{g - p_i \varepsilon_{ia_{m_i}}(z)}{p_i w_{ia_{m_i}}(z)} \right\} f\{g,\, p_l \varepsilon_{la_{m_l}}(z),\, p_l w_{la_{m_l}}(z)\} dg \right| \to 0 \tag{31}$$

при $M \to \infty$. В результате, поскольку $M \to \infty$,

$$\left| \Psi(a_{m1}, a_{m2}, ..., a_{mL}) - o(a_{m1}, a_{m2}, ..., a_{mL}) \right| \to 0 \tag{32}$$

при использовании теоремы Лебега о мажорируемой сходимости.

Теорема доказана.

Таким образом, при соответствующих условиях регулярности, если проводится минимизация $o(a_1, a_2, ..., a_L)$ относительно $a_1, a_2, ..., a_L$, можно получить правило классификации на основе ядерной оценки плотности распределения с асимптотической средней вероятностью ошибочной классификации, которая эквивалентна оптимальному байесовскому риску.

Заключение

В настоящей работе предложен метод выбора полос пропускания с использованием ядерных оценок плотности распределения, где вместо использования одной оптимальной полосы пропускания для каждой оценки плотности применены результаты различных мер сглаживания для ядерных оценок плотности. Разработан математический аппарат и проведен теоретический анализ проблемы сглаживания в непараметрических методах статистического анализа для решения задач распознавания образов. Один из основных выводов заключается в том, что в зависимости от того, являются ли априорные вероятности для различных множеств данных равными, средняя вероятность ошибочной классификации демонстрирует абсолютно разное поведение для различных вариантов выбора параметра полосы пропускания. Установлено, что когда все априорные вероятности равны, в зависимости от природы функций плотности и ядра, большие полосы пропускания могут приводить к почти оптимальным коэффициентам ошибочной классификации. Дополнительно установлено, что для больших полос пропускания средний коэффициент ошибочной классификации для классификатора на основе ядерной оценки плотности распределения асимптотически стремится к соответствующему коэффициенту ошибочной классификации для линейного правила решения и является оптимальным при некоторых предположениях относительно базовой плотности распределения. Вместо необработанных экспериментальных отношений, предложенный метод выбора полосы пропускания использует нормальную аппроксимацию определенных вероятностей, а его асимптотические свойства исследованы при соответствующих условиях регулярности.

О.А. Галкін

ЗАСТОСУВАННЯ МІР ЗГЛАДЖУВАННЯ В НЕПАРАМЕТРИЧНИХ ЯДЕРНИХ КЛАСИФІКАТОРАХ З ВИКОРИСТАННЯМ НОРМАЛЬНОЇ АПРОКСИМАЦІЇ ЙМОВІРНОСТЕЙ

Досліджено проблему вибору смуг пропускання з використанням ядерних оцінок щільності розподілу, де замість однієї оптимальної смуги пропускання для кожної оцінки щільності застосовано результати різних мір згладжування для ядерних оцінок щільності. Запропоновано підхід, де замість необроблених експериментальних відношень використовується нормальна апроксимація відповідних ймовірностей, а асимптотичні властивості досліджено при відповідних умовах регулярності.

A.A. Galkin

APPLICATION OF SMOOTHING MEASURES IN NONPARAMETRIC KERNEL CLASSIFIERS USING NORMAL APPROXIMATION OF PROBABILITIES

The problem of choosing the bandwidths is investigated using kernel density estimations of distribution, where the results of the various measures are applied for smoothing kernel density estimation instead of using a single optimum bandwidth for each of the density estimations. The approach is proposed where instead of the undeveloped experimental proportionality we use normal approximation of certain probabilities and also asymptotic properties are investigated under appropriate regularity conditions.

1. *Vardi Y., Zhang C.H.* The multivariate on L_1-median and associated data depth // Proc. of the National Academy of Sciences (USA). — 2000. — **97**. — P. 1423–1426.
2. *Godtliebsen F., Marron J.S., Chaudhuri P.* Significance in scale space for bivariate density estimation // Journal of Computational and Graphical Statistics. — 2002. — **11**. — P. 1–22.
3. *Holmes C.C., Adams N.M.* A probabilistic nearest neighbor method for statistical pattern recognition // Journal of the Royal Statistical Society. — 2002. — **64**. — P. 295–306.
4. *Hall P.* Large sample optimality of least squares cross validations in density estimation // The Annals of Statistics. — 1983. — **11**. — P. 1156–1174.
5. *Zuo Y., Serfling R.* Structural properties and convergence results for contours of sample statistical depth functions // Ibid. — 2000. — **28**. — P. 483–499.
6. *Mosler K.* Multivariate dispersions, central regions and depth. — New York: Springer-Verlag, 2002. — P. 1–14.
7. *Lachenbruch P., Mickey M.* Estimation of error rates in discriminant analysis // Technometrics. — 1968. — **10**. — P. 1–11.
8. *Silverman B.W.* Density estimation for statistics and data analysis. — London: Chapman and Hall, 1986. — P. 1–7.

Получено 22.05. 2015

Статья представлена к публикации чл.-корр. НАН Украины А.В. Анисимовым.