УДК 519.7

А.В. Анисимов, А.А. Галкин

ИССЛЕДОВАНИЕ АСИМПТОТИЧЕСКИХ СВОЙСТВ НЕПАРАМЕТРИЧЕСКИХ КЛАССИФИКАТОРОВ НА ОСНОВЕ ФУНКЦИЙ ГЛУБИНЫ

Введение. Когда границы классов достаточно сложные, линейные классификаторы могут быть нецелесообразны для использования. В таком случае для различения конкурирующих классов необходима зависимость от нелинейных разделительных поверхностей. Для построения таких поверхностей можно спроектировать данные $z_i = (z_{i1}, z_{i2}, ..., z_{id})$ в пространство более высокой размерности для получения нового вектора измеряемых величин $t_i = (\omega_1(z_i), \omega_2(z_i), ..., \omega_r(z_i))$ и выполнить линейную классификацию в l-мерном пространстве. Если данные проектируются в пространство квадратичных функций, это может рассматриваться как линейная классификация с $l = r + \binom{r}{2}$ измеряемыми величинами, которые в конечном итоге приводят к квадратичному разделению в начальном r-мерном пространстве. Величины

$$\Phi_m(\varphi) = \frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \Lambda\{\varphi'(t_{1i} - t_{2j}) > 0\} \qquad (1)$$

и

$$\Psi_m(\varphi, \gamma) = \frac{p_1}{m_1} \sum_{i=1}^{m_1} \Lambda\{\varphi' t_{1i} + \gamma < 0\} + \frac{p_2}{m_2} \sum_{i=1}^{m_2} \Lambda\{\varphi' t_{2i} + \gamma > 0\} \qquad (2)$$

могут быть оптимизированы для получения соответствующих оценок φ и γ, которые используются для формирования разделительной поверхности.

Постановка задачи. Предположим, имеется задача с двумя классами, а $z_{11}, z_{12}, ..., z_{1m_1}$ и $z_{21}, z_{22}, ..., z_{2m_2}$ являются двумя независимыми множествами r-мерных независимых одинаково распределенных случайных данных из двух r-мерных конкурирующих множеств. Пусть $t_{11}, t_{12}, ..., t_{1m_1}$ и $t_{21}, t_{22}, ..., t_{2m_2}$ — их трансформации в l-мерное пространство, а $\overline{\varphi}_{hsd}$ — максимизирующая оценка $\Phi_m(\varphi)$, в то время как $\overline{\varphi}_{rgr}$ и $\overline{\gamma}_{rgr}$ — минимизирующие оценки $\Psi_m(\varphi, \gamma)$.

Теорема 1. Предположим, что, при $M = m_1 + m_2 \to \infty$, $m_1 / M \to \alpha$ $(0 < \alpha < 1)$. Определим

$$\Phi(\varphi) = \text{prob}\{\varphi'(t_{1i} - t_{2j}) > 0\} \qquad (3)$$

и

$$\Psi(\varphi, \gamma) = p_1 \operatorname{prob}\{\varphi' t_{1i} + \gamma < 0\} + p_2 \operatorname{prob}\{\varphi' t_{2j} + \gamma > 0\}. \tag{4}$$

Тогда при $M \to \infty$ имеем следующие случаи:

(а) $\left|\Phi_m(\overline{\varphi}_{hsd}) - \max_{\varphi} \Phi(\varphi)\right| \to 0,\ \left|\Phi(\overline{\varphi}_{hsd}) - \max_{\varphi} \Phi(\varphi)\right| \to 0;$

(б) $\left|\Psi_m(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr}) - \min_{\varphi,\gamma} \Psi(\varphi, \gamma)\right| \to 0,\ \left|\Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr}) - \min_{\varphi,\gamma} \Psi(\varphi, \gamma)\right| \to 0.$

Когда существуют особые оптимизаторы φ^*_{hsd} и $(\varphi^*_{rgr}, \gamma^*_{rgr})$ для $\Phi(\varphi)$ и $\Psi(\varphi, \gamma)$ соответственно, а Φ и Ψ являются непрерывными функциями своих аргументов [1], $\overline{\varphi}_{hsd}$ сходится к φ^*_{hsd}, а $(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr})$ — почти всюду к $(\varphi^*_{rgr}, \gamma^*_{rgr})$ при $M \to \infty$.

Заметим, что если X — ограниченная случайная переменная с $E(X) = \varepsilon$ и $P(0 \le X \le 1) = 1$, тогда выполняется следующее условие:

$$E\{e^{w(X-\varepsilon)}\} \le e^{w^2/8} \tag{5}$$

для $\forall w > 0$.

Доказательство. (а) Пусть задана ограниченная функция ядра

$$u(\varphi' t_1, \varphi' t_2) = \Lambda\{\varphi' t_1 > \varphi' t_2\} \tag{6}$$

для обобщенной несмещенной статистики $\Phi_m(\varphi)$, где $0 \le u \le 1$.

Предположим, что $m_1 \le m_2$, и определим

$$Q(i_1, i_2, \ldots, i_{m_1}) = m_1^{-1} \sum_{j=1}^{m_1} u(\varphi' t_{1j}, \varphi' t_{2i_j}) \tag{7}$$

для определенных перестановок $(i_1, i_2, \ldots, i_{m_1})$ со множества $\{1, 2, \ldots, m_2\}$. Кроме того, $\Phi_m(\varphi)$ можно выразить следующим образом:

$$\Phi_m(\varphi) = \frac{(m_2 - m_1)!}{m_2!} \sum_{(i_1, i_2, \ldots, i_{m_1}) \in T} Q(i_1, i_2, \ldots, i_{m_1}), \tag{8}$$

где T является множеством всех возможных перестановок $(i_1, i_2, \ldots, i_{m_1})$ элементов множества $\{1, 2, \ldots, m_2\}$.

Теперь, поскольку выражения в сумме, определяющие Q, являются независимыми и одинаково распределенными и используется неравенство Йенсена на выпуклой функции e^z, получаем следующее неравенство:

$$e^{w\Phi_m(\varphi)} \le \frac{(m_2 - m_1)!}{m_2!} \sum_{(i_1, i_2, \ldots, i_{m_1}) \in T} e^{wQ(i_1, i_2, \ldots, i_{m_1})} \tag{9}$$

для каждого $w > 0$. Из этого следует, что

$$E\{e^{w\Phi_m(\varphi)}\} \le E\{e^{wQ(i_1, i_2, \ldots, i_{m_1})}\} = [E\{e^{wu(\varphi' t_{11}, \varphi' t_{21})/m_1}\}]^{m_1}, \tag{10}$$

а также

$$E\{e^{w[\Phi_m(\varphi) - \Phi(\varphi)]}\} \le [E\{e^{w[u(\varphi' t_{11}, \varphi' t_{21}) - \Phi(\varphi)]/m_1}\}]^{m_1} = \{\Psi_u(w/m_1)\}^{m_1}. \tag{11}$$

Поскольку

$$E\{\Phi_m(\varphi)\} = E\{Q(i_1, i_2, \ldots, i_{m_1})\} = E\{u(\varphi' t_1, \varphi' t_2)\} = P\{\varphi' t_{11} > \varphi' t_{21}\} = \Phi(\varphi) \tag{12}$$

 ISSN 0572-2691

и применяется условие (5), получим следующее неравенство:

$$P\{\Phi_m(\varphi)-\Phi(\varphi)\geq\kappa\}\leq E\{e^{w[\Phi_m(\varphi)-\Phi(\varphi)-\kappa]}\}\leq e^{-w\kappa}\{\Psi_u(w/m_1)\}^{m_1}\leq e^{-w\kappa+\frac{w^2}{wm_1}} \quad (13)$$

для любого $\kappa>0$. Минимизируя данное выражение относительно w, получаем такой результат [2]:

$$P\{\Phi_m(\varphi)-\Phi(\varphi)\geq\kappa\}\leq e^{-2m_1\kappa^2}. \quad (14)$$

Кроме того, можно показать, что

$$P\{\Phi_m(\varphi)-\Phi(\varphi)\leq-\kappa\}\leq e^{-2m_1\kappa^2} \quad (15)$$

для любого положительного κ. Объединяя эти два результата, получаем

$$P\{\left|\Phi_m(\varphi)-\Phi(\varphi)\right|\geq\kappa\}\leq 2e^{-2m_1\kappa^2} \quad (16)$$

для каждого $\kappa>0$.

Отметим, что множество гиперплоскостей $S=\{x:\varphi'x=0\}$ в R^r, проходящих через начало координат, имеет размерность Вапника–Червоненкиса l. Поэтому множество $\{x:\varphi'x>0\}$ имеет полиномиальную разделимость, где l является степенью многочлена [3].

В результате имеем, что

$$P\left\{\sup_{\varphi}\left|\Phi_m(\varphi)-\Phi(\varphi)\right|>\kappa\right\}<2(m_1m_2)^l e^{-2m_1\kappa^2} \quad (17)$$

для каждого $\kappa>0$ с использованием вероятностных неравенств на соответствующих множествах.

Поскольку $m_1/M\to\alpha$ $(0<\alpha<1)$, а $M\to\infty$ и $\sum_{M\geq1}M^{2l}e^{-VM}<\infty$ для любого $V>0$, получаем следующую сходимость:

$$\sup_{\varphi}\left|\Phi_m(\varphi)-\Phi(\varphi)\right|\to 0 \text{ почти всюду при } M\to\infty. \quad (18)$$

Данный результат следует из леммы Бореля–Кантелли.

Предположим, что $\overline{\varphi}_{hsd}$ — максимизирующая оценка $\Phi_m(\varphi)$, а φ^*_{hsd} — максимизирующая оценка $\Phi(\varphi)$, которая не обязательно уникальна. Поэтому

$$\left|\Phi_m(\overline{\varphi}_{hsd})-\Phi(\overline{\varphi}_{hsd})\right|\to 0 \quad (19)$$

и

$$\left|\Phi_m(\varphi^*_{hsd})-\Phi(\varphi^*_{hsd})\right|\to 0 \quad (20)$$

при $M\to\infty$.

Из определения $\overline{\varphi}_{hsd}$ и φ^*_{hsd} следует, что

$$\Phi(\varphi^*_{hsd})\geq\Phi(\overline{\varphi}_{hsd}) \text{ и } \Phi_m(\overline{\varphi}_{hsd})\geq\Phi(\varphi^*_{hsd}) \quad (21)$$

для каждого m.

Итак,

$$\left|\Phi_m(\overline{\varphi}_{hsd})-\max_{\varphi}\Phi(\varphi)\right|=\left|\Phi_m(\overline{\varphi}_{hsd})-\Phi(\varphi^*_{hsd})\right|\to 0 \text{ при } M\to\infty. \quad (22)$$

Поэтому

$$\left|\Phi(\overline{\varphi}_{hsd})-\max_{\varphi}\Phi(\varphi)\right|\to 0 \text{ при } M\to\infty. \quad (23)$$

(б) Отметим, что

$$\frac{1}{m_1}\sum_{i=1}^{m_1}\Lambda\{\varphi' t_{1i}+\gamma<0\}$$

являются средними независимыми одинаково распределенными и ограниченными случайными величинами для некоторых фиксированных φ и γ. Поэтому, используя неравенство Хевдинга, имеем следующее неравенство:

$$P\left\{\left|\frac{1}{m_1}\sum_{i=1}^{m_1}\Lambda\{\varphi' t_{1i}+\gamma<0\}-P\{\varphi' t_{11}+\gamma<0\}\right|>\phi/2\right\}<2e^{-m_1\phi^2/2} \quad (24)$$

для каждого $\phi>0$.

Аналогично

$$P\left\{\left|\frac{1}{m_2}\sum_{i=1}^{m_2}\Lambda\{\varphi' t_{2i}+\gamma>0\}-P\{\varphi' t_{21}+\gamma>0\}\right|>\phi/2\right\}<2e^{-m_2\phi^2/2} \quad (25)$$

для каждого $\phi>0$. Отсюда следует, что

$$P\{\left|\Psi_m(\varphi,\gamma)-\Psi(\varphi,\gamma)\right|>\phi\}<P\left\{\left|\frac{1}{m_1}\sum_{i=1}^{m_1}\Lambda\{\varphi' t_{1i}+\gamma<0\}-P\{\varphi' t_{11}+\gamma<0\}\right|>\phi/2\right\}+$$

$$+P\left\{\left|\frac{1}{m_2}\sum_{i=1}^{m_2}\Lambda\{\varphi' t_{2i}+\gamma>0\}-P\{\varphi' t_{21}+\gamma>0\}\right|>\phi/2\right\}<2(e^{-m_1\phi^2/2}+e^{-m_2\phi^2/2}). \quad (26)$$

Применяя аналогичный вывод по размерности Вапника–Червоненкиса для гиперплоскостей в R^l, а также используя свойства полиномиальной разделимости [4], получаем, что

$$P\left\{\sup_{\varphi,\gamma}\left|\Psi_m(\varphi,\gamma)-\Psi(\varphi,\gamma)\right|>\phi\right\}<2(m_1+m_2)^{l+1}(e^{-m_1\phi^2/2}+e^{-m_2\phi^2/2}). \quad (27)$$

Поскольку $\sum_{M\geq 1}M^{l+1}e^{-VM}<\infty$ для любого $\Psi>0$, имеем, что

$$\sup_{\varphi,\gamma}\left|\Psi_m(\varphi,\gamma)-\Psi(\varphi,\gamma)\right|\to 0$$

почти всюду при $M\to\infty$. Данный результат следует из леммы Бореля–Кантелли.

Далее проводится верификация того, что

$$\left|\Psi(\overline{\varphi}_{rgr},\overline{\gamma}_{rgr})-\min_{\varphi,\gamma}\Psi(\varphi,\gamma)\right|\to 0 \text{ и } \left|\Psi_m(\overline{\varphi}_{rgr},\overline{\gamma}_{rgr})-\min_{\varphi,\gamma}\Psi(\varphi,\gamma)\right|\to 0$$

почти всюду при $M\to\infty$.

Предположим, что максимизирующая оценка φ^*_{hsd} для $\Phi(\varphi)$ уникальна. Поскольку уже продемонстрировано, что на вероятностном множестве $\Phi(\overline{\varphi}_{hsd})$ сходится к $\Phi(\varphi^*_{hsd})$ при $M\to\infty$, установлено, что на том же множестве $\overline{\varphi}_{hsd}$ должно сходиться к φ^*_{hsd} из-за уникальности φ^*_{hsd} и непрерывности функции $\Phi(\varphi)$. Итак, любая подпоследовательность последовательности данной оценки будет иметь сходящуюся подпоследовательность, которая на вероятностном множестве сходится к φ^*_{hsd}, поскольку $\overline{\varphi}_{hsd}$ всегда находится на компактной поверхности единичного шара в R^n. В результате $\overline{\varphi}_{hsd}$ должно сходиться к φ^*_{hsd} почти всюду.

На основе аналогичных соображений, если $(\varphi^*_{rgr}, \gamma^*_{rgr})$ является уникальной минимизируемой оценкой для $\Psi(\varphi, \gamma)$, $(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr}) \to (\varphi^*_{rgr}, \gamma^*_{rgr})$ при $M \to \infty$, поскольку $\Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr})$ сходится к $\Psi(\varphi^*_{rgr}, \gamma^*_{rgr})$ почти всюду.

Теорема доказана.

Переменная $\Phi(\varphi)$ является мерой линейной или нелинейной разделимости между двумя конкурирующими многомерными распределениями вдоль направления φ, а $\max_{\varphi} \Phi(\varphi)$ определяет максимальную линейную или нелинейную разделимость между двумя многомерными множествами данных. Отметим также, что $\Psi(\varphi, \gamma)$ является средней вероятностью ошибочной классификации, когда поверхность $\varphi' t + \gamma = 0$ применяется для разделения между двумя конкурирующими множествами данных, а $\min_{\varphi,\gamma} \Psi(\varphi, \gamma)$ является наиболее подходящей средней вероятностью ошибочной классификации, которая достигается с использованием линейных или нелинейных классификаторов. Здесь будет уместно отметить, что $\Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr})$ может рассматриваться как условная средняя вероятность ошибочной классификации на учебной выборке, когда поверхность $\overline{\varphi}'_{rgr} t + \overline{\gamma}_{rgr} = 0$ используется для классификации данных, поступающих с одного из двух конкурирующих множеств [5].

В случае, когда размер учебной выборки стремится к бесконечности, средняя вероятность ошибочной классификации линейного или нелинейного классификатора на основе регрессионной глубины асимптотически сходится к наиболее подходящему возможному среднему коэффициенту ошибочной классификации, который может быть получен с использованием линейного или нелинейного классификатора. Линейный или нелинейный классификатор на основе регрессионной глубины сходится почти всюду к оптимальной разделительной гиперплоскости или нелинейной поверхности, когда линейный или нелинейный классификатор уникален.

Выражение $\left| \Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr}) - \min_{\varphi,\gamma} \Psi(\varphi, \gamma) \right| \to 0$ почти всюду при $M \to \infty$, что было доказано в теореме 1, где $\Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr})$ — условная средняя вероятность ошибочной классификации для данных из учебной выборки. При использовании Ψ как функции, ограниченной от 0 до 1, а также теоремы Лебега о мажорируемой сходимости данный вывод очевиден, если принять ожидание $\Psi(\overline{\varphi}_{rgr}, \overline{\gamma}_{rgr})$ на учебной выборке.

Теорема 2. Пусть функции плотности h_1 и h_2 двух конкурирующих классов являются эллиптически-симметричными с общей матрицей рассеивания Ξ. Предположим, что $h_i(z) = c(z - \varepsilon_i)$, $i = (1, 2)$, для некоторых параметров локализации ε_1 и ε_2 и общей эллиптически-симметричной функции плотности c, что удовлетворяет условию $c(kz) \geq c(z)$ для каждого z и $0 < k < 1$. При условиях, которые предусмотрены в теореме 1, средние вероятности ошибочной классификации линейного классификатора на основе регрессионной глубины сходятся к оптимальной байесовской ошибке, когда размер учебной выборки стремится к бесконечности при условии, что априорные вероятности двух классов равны. Кроме того, в случае равных априорных вероятностей, если байесовский классификатор является оптимальным, а $\Phi(\varphi)$ имеет уникальную максимизирующую оценку, то же самое имеет место для классификатора на основе полупространственной глубины, где оба этих глубинных классификатора сходятся почти всюду к байесовскому классификатору. Когда априорные вероятности являются неравными, приведенные выше результаты конвергенции для глубинных линейных классификаторов остаются верными для нормально распределенных множеств данных с общей дисперсионной матрицей, но различными векторами средних значений.

Доказательство. При соответствующих условиях линейный классификатор с $\varphi = \Xi^{-1}(\varepsilon_1 - \varepsilon_2)$ и $\gamma = (\varepsilon_2' \Xi^{-1} \varepsilon_2 - \varepsilon_1' \Xi^{-1} \varepsilon_1)/2$ является байесовским. В результате средняя ошибка ошибочной классификации линейного классификатора на основе регрессионной глубины сходится к оптимальному байесовскому риску, что следует из результатов, полученных с применением теоремы Лебега о мажорируемой сходимости.

Кроме того, когда байесовский классификатор уникален, линейный классификатор на основе регрессионной глубины сходится почти всюду к байесовскому классификатору, что следует из доказательства (б) теоремы 1 [6].

Из теоремы 1 следует, что $\overline{\varphi}_{hsd}$ сходится почти всюду к φ^* при $M \to \infty$ в случае, когда $\Phi(\varphi)$ имеет уникальную максимизирующую оценку

$$\varphi_{hsd}^* = \Xi^{-1}(\varepsilon_1 - \varepsilon_2).$$

Рассмотрим два независимых случайных вектора $Z_1 \approx h_1$ и $Z_2 \approx h_2$, которые полностью независимы на учебной выборке. Определим

$$X_{1,m} = \overline{\varphi}_{hsd}' Z_1, \quad X_{2,m} = \overline{\varphi}_{hsd}' Z_2, \quad X_1 = \varphi^{*'} Z_1 \text{ и } X_2 = \varphi^{*'} Z_2,$$

используя данные случайные векторы. Можно утверждать, что $(X_{1,m}, X_{2,m}) \overset{J}{\to}$ $\overset{J}{\to} (X_1, X_2)$ почти всюду при $M \to \infty$, учитывая сходимость почти всюду $\overline{\varphi}_{hsd}$ к φ^*.

Поэтому $\sup\limits_{\gamma} \left| \Psi(\overline{\varphi}_{hsd}, \gamma) - \Psi(\varphi^*, \gamma) \right| \to 0$ почти всюду при $M \to \infty$, поскольку слабая сходимость к непрерывному распределению обеспечивает равномерную сходимость, а X_1 и X_2 являются непрерывно распределенными. Однако заметим, что из доказательства (б) теоремы 1 очевидно, что $\sup\limits_{\gamma} \left| \Psi_m(\overline{\varphi}_{hsd}, \gamma) - \Psi(\overline{\varphi}_{hsd}, \gamma) \right| \to 0$ почти всюду при $M \to \infty$. Поэтому $\sup\limits_{\gamma} \left| \Psi_m(\overline{\varphi}_{hsd}, \gamma) - \Psi(\varphi^*, \gamma) \right| \to 0$ почти всюду при $M \to \infty$.

Аналогично доказательству теоремы 1

$$\left| \Psi_m(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd}) - \min_{\gamma} \Psi(\varphi^*, \gamma) \right| = \left| \Psi_m(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd}) - \min_{\varphi, \gamma} \Psi(\varphi, \gamma) \right| \to 0 \qquad (28)$$

почти всюду при $M \to \infty$. Кроме того, необходимо, чтобы $\left| \Psi_m(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd}) - \right.$ $\left. - \Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd}) \right| \to 0$ почти всюду при $M \to \infty$. Итак, $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ сходится почти всюду к $\min\limits_{\varphi, \gamma} \Psi(\varphi, \gamma)$, что является байесовским риском [7].

Поскольку $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ является условной средней вероятностью ошибочной классификации для исследуемых данных из учебной выборки, а также учитывая ожидание $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ на учебной выборке данных, получаем безусловную среднюю вероятность ошибочной классификации линейного классификатора на основе полупространственной глубины. Применением теоремы Лебега о мажорируемой сходимости, а также тем, что Ψ является функцией, ограниченной от 0 до 1, завершается доказательство сходимости.

Далее необходимо показать, что $\overline{\gamma}_{hsd}$ сходится почти всюду к постоянной величине для доказательства сходимости почти всюду линейного классификатора на основе полупространственной глубины. Можно проверить, что оптимальный байесовский риск строго меньше 0,5, если два множества данных не являются статистически неразделимыми при наличии идентичных распределений в случае равных априорных вероятностей с двумя конкурирующими множествами данных [8].

Поскольку $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ сходится к байесовскому риску, а $\overline{\varphi}_{hsd}$ — к φ^* при $M \to \infty$ на множестве с единичной вероятностью, на данном множестве значение $\overline{\gamma}_{hsd}$ должно оставаться ограниченным. Это является причиной того, что, из-за сходимости $\overline{\varphi}_{hsd}$ к φ^*, $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ будет сходиться к значению 0,5 по подпоследовательности, вдоль которой $\left|\overline{\gamma}_{hsd}\right| \to \infty$ при $M \to \infty$.

Однако каждый раз, когда $\overline{\gamma}_{hsd}$ сходится к действительному числу γ, $\Psi(\overline{\varphi}_{hsd}, \overline{\gamma}_{hsd})$ должно сходиться к $\Psi(\varphi^*, \gamma)$ на данном множестве с единичной вероятностью из-за непрерывности Ψ. Учитывая то, что любая ограниченная последовательность должна иметь сходящуюся подпоследовательность, $\overline{\gamma}_{hsd}$ должно сходиться к γ^*, где $\Xi(\varphi^*, \gamma^*) = \min\limits_{\varphi,\gamma} \Psi(\varphi, \gamma)$. Заметим, что полученный результат аналогичен байесовскому риску [9].

Для двух конкурирующих нормально распределенных множеств данных с параметрами (ε_1, Ξ) и (ε_2, Ξ), а также для не обязательно равных априорных вероятностей p_1 и p_2 имеет место следующее неравенство:

$$p_1 h_1(z) > p_2 h_2(z) \Leftrightarrow p_1 \left|\Xi\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(z-\varepsilon_1)'\Xi^{-1}(z-\varepsilon_1)} > p_2 \left|\Xi\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(z-\varepsilon_2)'\Xi^{-1}(z-\varepsilon_2)} \Leftrightarrow$$

$$\Leftrightarrow (z-\varepsilon_2)'\Xi^{-1}(z-\varepsilon_2) - (z-\varepsilon_1)'\Xi^{-1}(z-\varepsilon_1) > V \Leftrightarrow$$

$$\Leftrightarrow 2z'\Xi^{-1}(\varepsilon_1 - \varepsilon_2) > \{\varepsilon_1'\Xi^{-1}\varepsilon_1 - \varepsilon_2'\Xi^{-1}\varepsilon_2\} + V, \tag{29}$$

где $V = 2\log(p_2 / p_1)$.

В результате оптимальное байесовское правило имеет свойство уникальности и линейности, поэтому доказательство теоремы является завершенным, поскольку Φ и Ψ — непрерывные функции в рамках многомерного нормального распределения.

Теорема доказана.

Отметим, что исследованные асимптотические результаты сходимости определены для случая, когда размерность проекционного пространства l не меняется при изменении размера выборки M. Однако в некоторых непараметрических методах статистического анализа для задач распознавания, а именно методах на основе опорно векторной машины или нейронных сетей, размерность проекционного пространства возрастает с размером выборки данных [10, 11]. Указанное свойство гибкости по выбору разделительной поверхности можно получить для глубинных методов. Установлено, что если l растет с M таким образом, что для всех положительных значений V имеем $\sum\limits_{M \ge 1} M^{2l} e^{-VM} < \infty$, результаты сходимости в (а) и (б) теоремы 1 остаются в силе.

Другими словами, если l растет со скоростью M^{κ} для любого $0<\kappa<1$, данные результаты сходимости имеют место.

Как уже было отмечено, максимизация $\Phi_m(\varphi)$ относительно φ требует нахождения начала координат функции полупространственной глубины в множестве данных, определенных l-мерными векторами разниц $t_{1i}-t_{2j}$ $(i=1,2,...,m_1;$ $j=1,2,...,m_2)$. Имеет место конечная задача максимизации, однако проведение максимизации по всем пересчетам даст вычислительную сложность порядка $O(m_{\circ}^{2l})$, где $m_{\circ}=\max\{m_1,m_2\}$. Максимизация $\Phi_m(\varphi)$ может быть сведена к φ с $\|\varphi\|=1$, а минимизация $\Psi(\varphi,\gamma)$ — к (φ,γ) с $\|(\varphi,\gamma)\|=1$. Учитывая, что порядок вычислительной сложности быстро растет с размерностью l, точная оптимизация $\Phi_m(\varphi)$ и $\Psi_m(\varphi,\gamma)$ не может быть применена к задачам большой размерности, поэтому можно выполнять только приблизительную оптимизацию.

Заключение. В настоящей работе предложен подход, где индикаторные функции, фигурирующие в выражениях для Φ_m и Ψ_m, аппроксимируются с помощью соответственно определенных функций сглаживания. С помощью указанной аппроксимации использовались производные для определения направления наиболее быстрого подъема или спуска целевой функции, которую необходимо оптимизировать. Логистическая функция $1/(1+e^{-\kappa z})$ с большим положительным κ является аппроксимацией для индикаторной функции Λ $(z>0)$. Понятно, что недостаточно большое значение κ приведет к неточной аппроксимации. Однако очень большое значение κ приведет к достаточно точной аппроксимации, но числовая оптимизация на основе наиболее быстрого подъема или спуска будет численно неустойчивой. Установлено, что высокий коэффициент численной устойчивости в оптимизационном алгоритме может быть получен даже для очень большого значения t. Это стало возможным, когда все измеряемые величины были нормализованы к процессу аппроксимации. На основе практических экспериментов было установлено, что если используется значение $5\le\kappa\le10$ после процесса нормализации измеряемых величин, средние коэффициенты ошибочной классификации для полученных алгоритмов остаются почти одинаковыми и достаточно малыми. В результате получены наиболее соответствующие значения в этом диапазоне. В двумерном случае для линейных методов статистического анализа для задач распознавания, где точное определение $\Phi_m(\varphi)$ и $\Psi_m(\varphi,\gamma)$ является достаточно простым, проведен сравнительный анализ эффективности точных и приближенных глубинных методов классификации, а также установлено, что они достигают практически одинаковых средних коэффициентов ошибочной классификации. Начиная с разных случайных начальных данных, приближенные реализации оптимизационных алгоритмов выполнялись несколько раз для решения проблемы возможного наличия нескольких локальных минимумов.

Относительно применения классификатора на основе полупространственной глубины, оценена величина γ из учебной выборки после оценки φ. Это стало возможным в результате пересчета порядковых статистик спроектированных данных $\overrightarrow{\varphi}_{hsd}t_{1i}$ и $\overrightarrow{\varphi}_{hsd}t_{2j}$ $(1\le i\le m_1;\ 1\le j\le m_2)$ вдоль оцененного направления $\overline{\varphi}_{hsd}$. В итоге получен результат, что вычислительная сложность при получении оценки $\overline{\gamma}_{hsd}$ не увеличивается с размерностью l из-за использования линейных проекций.

ISSN 0572-2691

А.В. Анісімов, О.А. Галкін

ДОСЛІДЖЕННЯ АСИМПТОТИЧНИХ ВЛАСТИВОСТЕЙ НЕПАРАМЕТРИЧНИХ КЛАСИФІКАТОРІВ НА ОСНОВІ ФУНКЦІЙ ГЛИБИНИ

Досліджено непараметричні методи класифікації та їх асимптотичні властивості на основі використання функцій напівпросторової та регресійної глибини, що використовуються для побудови лінійних та нелінійних розділових поверхонь серед конкуруючих множин даних. Запропоновано повністю незалежний від розподілу підхід, де для мінімізації коефіцієнтів помилкової класифікації застосовується розподільне розташування багатовимірних даних.

A.V. Anisimov, A.A. Galkin

THE RESEARCH OF THE ASYMPTOTIC PROPERTIES OF NONPARAMETRIC CLASSIFIERS BASED ON DEPTH FUNCTIONS

Nonparametric classification methods are considered as well as their asymptotic properties based on the use of half-space and regression depth functions that are used to build linear and nonlinear separating surfaces between competing sets of data. A fully distribution free approach is proposed where distributive location of multidimensional data is used to minimize misclassification coefficients.

1. *Mizera I.* On depth and deep points: a calculus // The Annals of Statistics. — 2002. — **30**. — P. 1681–1736.
2. *Mosler K.* Multivariate dispersions, central regions and depth. — New York : Springer-Verlag, 2002. — P. 1–10.
3. *Serfling R.* A depth function and a scale curve based on spatial quantiles // Statistics and data analysis based on L_1-norm and related methods. — Boston : Birkhäuser, 2002. — P. 25–38.
4. *Hall P.* Large sample optimality of least squares cross validations in density estimation // The Annals of Statistics. — 1983. — **11**. — P. 1156–1174.
5. *Vardi Y., Zhang C.H.* The multivariate on L_1-median and associated data depth // Proceedings of the National Academy of Sciences (USA). — 2000. — **97**. — P. 1423–1426.
6. *Silverman B.W.* Density estimation for statistics and data analysis. — London : Chapman and Hall, 1986. — P. 1–20.
7. *Lachenbruch P., Mickey M.* Estimation of error rates in discriminant analysis // Technometrics. — 1968. — **10**. — P. 1–11.
8. *Zuo Y., Serfling R.* Structural properties and convergence results for contours of sample statistical depth functions // The Annals of Statistics. — 2000. — **28**. — P. 483–499.
9. *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.
10. *Chaudhuri P., Marron J.* Scale space view of curve estimation // The Annals of Statistics. — 2000. — **28**. — P. 408–428.
11. *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.

Получено 23.04.2015

Статья представлена к публикации членом редколлегии чл.-корр. НАН Украины Чикрием А.А.