УДК 519.7

А.А. ГАЛКИН

АФИННО-ИНВАРИАНТНЫЙ КЛАССИФИКАТОР ЭКСТРАПОЛЯЦИОННОЙ ГЛУБИНЫ НА ОСНОВЕ МНОГОУРОВНЕВОЙ СТРУКТУРЫ СГЛАЖИВАНИЯ

Аннотация. Предложен и исследован непараметрический аффинно-инвариантный классификатор экстраполяционной глубины, устойчивый к выбросам и экстремальным значениям. Предложена многоуровневая структура сглаживания, позволяющая получать глобальные свойства функций плотности и границ класса при соответствующих условиях регулярности. Описанный классификатор использует ядерные оценки плотности для эффективной классификации многомерных данных на различных уровнях сглаживания.

Ключевые слова: ядерная оценка плотности, уровень сглаживания, функция глубины.

ВВЕДЕНИЕ

Использование классификаторов максимальной экстраполяционной глубины позволяет получать относительно низкие коэффициенты ошибочной классификации в случае, когда априорные вероятности множеств данных равны, а их распределения отличаются только параметрами расположения. Однако на практике распределения множеств данных зачастую имеют различные матрицы разброса и формы, а также априорные вероятности. Описанные особенности обусловливают актуальность задачи разработки развитых версий классификаторов максимальной глубины. Существующие в настоящее время версии классификатора экстраполяционной глубины позволяют решать прикладные задачи при монотонном отношении между глубинными функциями и функциями плотности, а также при условии, что множества данных имеют различные матрицы разброса [1].

КЛАССИФИКАТОР ЭКСТРАПОЛЯЦИОННОЙ ГЛУБИНЫ НА ОСНОВЕ ЭЛЛИПТИЧЕСКОЙ СИММЕТРИИ РАСПРЕДЕЛЕНИЙ

Рассмотрим случай, когда распределения множеств данных эллиптические. Если $E(z, H_l)$ является глубиной z относительно H_l, то байесовский классификатор задается как

$$\mathfrak{F}_{\mathrm{B}}(z) = \arg \max_{1 \le l \le L} p_l o_l \{E(z, H_l)\},$$

где o_l — функциия преобразования, причем она монотонно убывающая, а также равна для всех групп множеств данных, если функции h_l являются унимодальными, а распределения множеств данных отличаются только параметрами расположения [2]. Кроме того, байесовский классификатор эквивалентен классификатору максимальной глубины, если p_l равны. Однако, если не выполняется хотя бы одно из приведенных условий, возникает необходимость в получении информации относительно функциональных форм o_l.

Лемма 1. Если $\xi_l(\cdot)$ является функцией плотности от $F_e(z, H_l)$, а функции $h_1, h_2, \ldots, h_L$ эллиптически-симметричные, то байесовский классификатор задается как

$$\mathfrak{F}_{\mathrm{B}}(z) = \arg \max_{l \in \{1, \ldots, L\}} \alpha_l \xi_l \{F_e(z, H_l)\} \{F_e(z, H_l)\}^{r-3} / \{1 - F_e(z, H_l)\}^{r-1},$$

где α_l — постоянная величина.

Доказательство. Учитывая, что функция h_l эллиптически-симметричная, имеем

$$h_l(z) = \mathrm{I}(r/2)(2p)^{-r/2} |\Xi_l|^{-1/2}\ c_l(C(z, H_l)) / C(z, H_l)^{r-1},$$

где c_l — вероятностная функция плотности от $C(z, H_l) = \{(z - \varepsilon_l)' \Xi_l^{-1} (z - \varepsilon_l)\}^{1/2}$, а ε_l и Ξ_l — соответственно параметры расположения и масштаба для h_l.

Следовательно, можно утверждать, что

$$\mathfrak{F}_{\mathrm{B}}(z) = \arg \max_{1 \le l \le L} p_l h_l(z) = \arg \max_{1 \le l \le L} \alpha_l \lambda_l \{\mathrm{I}(z, H_l)\} / \{\mathrm{I}(z, H_l)\}^{r-1},$$

где постоянная величина α_l зависит от H_l и p_l, а λ_l является функцией плотности от $\mathrm{I}(z, H_l)$. Поскольку $F_e(z, H_l) = \{1 + \mathrm{I}(z, H_l)\}^{-1}$, доказательство следует из свойств выборочного распределения.

Лемма доказана.

Отметим, что пока α_l изменяются в зависимости от выбора одномерных мер расположения и масштаба, лемма 1 справедлива для любого определения функций экстраполяционной глубины.

ЯДЕРНЫЕ ОЦЕНКИ ПЛОТНОСТИ

Для построения развитой версии классификатора экстраполяционной глубины используется метод ядерных оценок плотности для оценки ξ_l, а также выборочная форма $F_e(z, H_{lm_l})$ для оценки $F_e(z, H_l)$. В данном случае оценивается одномерная плотность независимо от размерности пространства измерений, что позволяет избежать проблемы так называемого «проклятия размерности», которая часто имеет место в многомерных непараметрических оценках плотности [3].

Заметим, что выбор полосы пропускания a_l обязателен для оценки ξ_l, $1 \le l \le L$. Данная оценка плотности задается как

$$\bar{\xi}_{la_l}(\omega) = (m_l a_l)^{-1} \sum_{i=1}^{m_l} \Theta\{a_l^{-1}(\omega - \overline{\omega}_{m_l}^{(l)}(z_{li}))\},$$

где Θ является функцией ядра, а $\overline{\omega}_{m_l}^{(l)}(z) = F_e(z, H_{lm_l})$.

Лемма 2. Пусть имеют место следующие предположения:

а) функция $h_l(z) > 0\ \ \forall\, z \in \mathrm{R}^r$ и $l = 1, 2$;

б) для $l = 1, 2$ функция $H_{\beta, l}(\bar{z}) = P(\beta(Z) \le \bar{z})$ является равномерно непрерывной в $\bar{z}$, где $\beta(z) = d^{(2)}(z) / d^{(1)}(z)$, $d^{(l)}(z) = \omega_l(\xi^{(l)}(z))(\xi^{(l)}(z))^{r-3} / /(1 - \xi^{(l)}(z))^{r-1}$, $\xi^{(l)}(z) = F_{np}(z, H_l)$, а Z принадлежит l-му классу;

в) для $l = 1, 2$ величина $a_l \to 0$ и $m_l a_l^4 \to \infty$ при $m_l \to \infty$.

Предположим также, что h_1 и h_2 являются эллиптически-симметричными функциями. Если искомая оценка Δ выбрана путем минимизации оценки перекрестной проверки частоты ошибок, то коэффициент ошибочной классификации классификатора экстраполяционной глубины $\mathfrak{F}_2(\cdot)$ сходится к байесовскому риску при $\min\{m_1, m_2\} \to \infty$.

Доказательство. Очевидно, что

$$|\Psi(r_2) - \Psi_V| \le \sum_{l=1}^{2} \int \left| \prod_{i=1, i \ne l}^{2} \Lambda \left\{ \frac{d_{m_l, a_l}^{(l)}(z)}{d_{m_i, a_i}^{(i)}(z)} \ge v_m \right\} - \prod_{i=1, i \ne l}^{2} \Lambda \left\{ \frac{d^{(l)}(z)}{d^{(i)}(z)} \ge v \right\} \right| h_l(z) dz.$$

Можно утверждать, что $|\Psi(r_2)-\Psi_V|$ вероятностно сходится к нулю согласно теореме Лебега о мажорируемой сходимости, где показатели ограничены соответствующими функциями. Результат можно получить, используя выборочное ожидание и повторно применяя теорему Лебега о мажорируемой сходимости.

Лемма доказана.

В проведенном исследовании используем гауссовское ядро, предполагая, что ядро Θ имеет ограниченную первую производную. Рассматривая двухклассовую задачу, в которой $d^{(l)}_{m_l,a_l}(z)=\bar{\xi}_{la_l}(\overline{\omega}^{(l)}_{m_l}(z))(\overline{\omega}^{(l)}_{m_l}(z))^{r-3}/(1-\overline{\omega}^{(l)}_{m_l}(z))^{r-1}$ для $l=1,2$, а $\Delta=\log(\alpha_2/\alpha_1)$, можно утверждать, что результирующий классификатор $\mathfrak{F}_2(z)=1$, если $\log[d^{(1)}_{m_1,a_1}(z)]-\log[d^{(2)}_{m_2,a_2}(z)]>\Delta$, и $\mathfrak{F}_2(z)=2$ в противном случае. Очевидно, что выбор a_1, a_2 и Δ влияет на производительность классификатора $\mathfrak{F}_2(\cdot)$. Поэтому при увеличении размера выборки, а также согласно предположениям а) и б) частота ошибок развитой версии классификатора экстраполяционной глубины $\mathfrak{F}_2(\cdot)$ сходится к байесовскому риску при условии, что a_1 и a_2 удовлетворяют предположению в), а Δ выбирается путем минимизации оценки перекрестной проверки частоты ошибок [4].

Теорема 1. Пусть $\overline{\beta}_m(z)=d^{(2)}_{m_2,a_2}(z)/d^{(1)}_{m_1,a_1}(z)$ и $\beta(z)=d^{(2)}(z)/d^{(1)}(z)$. Тогда $\exists G_\mu$, что $P(G_\mu)>1-\mu$ и $\sup_{z\in W_\mu}|\overline{\beta}_m(z)-\beta(z)|\xrightarrow{P}0$ при $\min\{m_1,m_2\}\to\infty$, где $i=1,2$, а Z принадлежит i-му классу.

Доказательство. Определим $\bar{\xi}_{ia_i}$ и $\bar{\xi}^*_{ia_i}(\omega)=\dfrac{1}{m_i a_i}\sum\limits_{l=1}^{m_i}\Theta\left\{\dfrac{\omega-\omega^{(i)}(z_{il})}{a_i}\right\}$ для $i=1,2$. Также отметим, что

$$\sup_z|\bar{\xi}_{ia_i}(\overline{\omega}^{(i)}_{m_i}(z))-\xi_i(\omega^{(i)}(z))|\le\sup_z|\bar{\xi}_{ia_i}(\overline{\omega}^{(i)}_{m_i}(z))-\bar{\xi}_{ia_i}(\omega^{(i)}(z))|+$$

$$+\sup_z|\bar{\xi}_{ia_i}(\omega^{(i)}(z))-\xi^*_{ia_i}(\omega^{(i)}(z))|+\sup_z|\xi^*_{ia_i}(\omega^{(i)}(z))-\xi_i(\omega^{(i)}(z))|.$$

Используя предположение в) леммы 2 и принимая, что $N_\Theta=\sup_\kappa|\Theta'(\kappa)|<\infty$, имеем

$$\sup_z|\overline{\omega}_{ia_i}(\bar{\xi}^{(i)}_{m_i}(z))-\overline{\omega}_{ia_i}(\xi^{(i)}(z))|\le N_\Theta\sup_z|\bar{\xi}^{(i)}_{m_i}(z)-\xi^{(i)}(z)|/a_1^2\xrightarrow{P}0 \qquad (1)$$

при $m_i\to\infty$. Заметим, что неравенство (1) основано на том, что $\sup_z|\bar{\xi}^{(i)}_{m_i}(z)-\xi^{(i)}(z)|=O_P(m_i^{-1/2})$ и функция h_i эллиптически-симметричная. Итак, отсюда имеем

$$\sup_z|\overline{\omega}_{ia_i}(\xi^{(i)}(z))-\omega^*_{ia_i}(\xi^{(i)}(z))|\xrightarrow{P}0 \qquad (2)$$

при $m_i\to\infty$.

В результате, используя свойства равномерной непрерывности функции экстраполяционной глубины, которые следуют из эллиптической симметричности h_i, получаем

$$\sup_z|\omega^*_{ia_i}(\xi^{(i)}(z))-\omega_i(\xi^{(i)}(z))|\xrightarrow{P}0 \qquad (3)$$

при $m_i\to\infty$. Заметим, что сходимость (3) выполняется, если осуществляется предположение в) леммы 2 и используются свойства ядерных оценок плотности [5].

В итоге, объединив сходимости (2) и (3), получим $\sup_z |\overline{\omega}_{ia_i}(\overline{\xi}^{(i)}_{m_i}(z)) - -\omega_i(\xi^{(i)}(z))| \xrightarrow{P} 0$ при $m_i \to \infty$.

Для всех $\mu > 0$ можно найти такое $\theta = \theta(\mu) > 0$, что множество $G_\mu = \{z: \theta \le \xi^{(1)}(z), \xi^{(2)}(z) \le 1 - \theta\}$ будет иметь более высокую вероятность, чем $1 - \mu$ относительно распределения вероятностей двух классов.

Очевидно, что

$$\sup_{z \in G_\mu} \left| \frac{(\overline{\xi}^{(i)}_{m_i}(z))^{r-3}}{(1 - \overline{\xi}^{(i)}_{m_i}(z))^{r-1}} - \frac{(\xi^{(i)}(z))^{r-3}}{(1 - \xi^{(i)}(z))^{r-1}} \right| \xrightarrow{P} 0$$

для $i = 1, 2$. Отсюда следует, что

$$\sup_{z \in G_\mu} | d^{(i)}_{m_i, a_i}(z) - d^{(i)}(z) | \xrightarrow{P} 0$$

при $m_i \to \infty$. Итак, поскольку $\inf_{z \in G_\mu} d^{(i)}(z) > 0$ для $i = 1, 2$, получаем желаемый результат.

Теорема доказана.

Теорема 2. Пусть $v_m = \arg\min_\Delta \Psi^{VB}_m(\Delta)$, $v = \arg\min_\Delta \Psi(\Delta)$ и имеют место предположения леммы 2, а также

$$\Psi^{VB}_m(\Delta) = \sum_{i=1, l \ne i}^{2} \frac{p_i}{m_i} \sum_{j=1}^{m_i} \Lambda \left\{ \frac{d^{(l)}_{m_l, a_l}(z_{ij})}{d^{(i)}_{m_i, a_i}(z_{ij})} \ge \Delta_i \right\},$$

$$\Psi(\Delta) = \sum_{i=1, l \ne i}^{2} p_i P \left\{ \frac{d^{(l)}(Z)}{d^{(i)}(Z)} \ge \Delta_i \right\},$$

где $m = (m_1, m_2)$, $\Delta_1 = 1/\Delta$, $\Delta_2 = \Delta$, а Z принадлежит i-му классу. Тогда $v_m \xrightarrow{P} v$ при $\min(m_1, m_2) \to \infty$ с условием, что v является уникальным.

Доказательство. Покажем, что $\sup_\Delta |\Psi^{VB}_m(\Delta) - \Psi(\Delta)| \xrightarrow{P} 0$ при $\min(m_1, m_2) \to \to \infty$. Отметим, что имеет место сходимость $v_m \xrightarrow{P} v$, которая следует из $\sup_\Delta |\Psi^{VB}_m(\Delta) - \Psi(\Delta)| \xrightarrow{P} 0$, поскольку $\Psi(\cdot)$ является уникальным минимумом.

Отметим, что

$$|\Psi^{VB}_m(\Delta) - \Psi(\Delta)| \le \sum_{i=1, l \ne i}^{2} \frac{p_i}{m_i} \sum_{j=1}^{m_i} \left| \Lambda \left\{ \frac{d^{(l)}_{m_l, a_l}(z_{ij})}{d^{(i)}_{m_i, a_i}(z_{ij})} \ge \Delta_i \right\} - P \left\{ \frac{d^{(l)}(Z)}{d^{(i)}(Z)} \ge \Delta_i \right\} \right| \le$$

$$\le \sum_{i=1, l \ne i}^{2} \frac{p_i}{m_i} \sum_{j=1}^{m_i} \left| \Lambda \left\{ \frac{d^{(l)}_{m_l, a_l}(z_{ij})}{d^{(i)}_{m_i, a_i}(z_{ij})} \ge \Delta_i \right\} - \Lambda \left\{ \frac{d^{(l)}(z_{ij})}{d^{(i)}(z_{ij})} \ge \Delta_i \right\} \right| +$$

$$+\sum_{i=1, j\neq i}^{2}\frac{p_i}{m_i}\sum_{j=1}^{m_i}\left|\Lambda\left\{\frac{d^{(l)}(z_{ij})}{d^{(i)}(z_{ij})}\geq\Delta_i\right\}-P\left\{\frac{d^{(l)}(Z)}{d^{(i)}(Z)}\geq\Delta_i\right\}\right|,$$

где Z принадлежит i-му классу.

Далее определим величины

$$G_m(\Delta_1)=\frac{1}{m_1}\sum_{i=1}^{m_1}|\Lambda\{\beta(z_{1i})\geq\Delta_1\}-P\{\beta(Z)\geq\Delta_1\}|,$$

$$V_m(\Delta_1)=\frac{1}{m_1}\sum_{i=1}^{m_1}|\Lambda\{\overline{\beta}_m(z_{1i})\geq\Delta_1\}-\Lambda\{\beta(z_{1i})\geq\Delta_1\}|,$$

где Z принадлежит l-му классу. Можно показать, что $\sup_{\Delta_1}|G_m(\Delta_1)|\xrightarrow{ac}0$, используя лемму Гливенко–Кантелли.

Для всех $\mu>0$ получаем такое $\xi_\mu>0$, что

$$\sup_{\Delta_1}|H_{\beta,1}(\Delta_1+\xi_\mu/2)-H_{\beta,1}(\Delta_1-\xi_\mu/2)|<\mu$$

согласно предположению б) леммы 2. Кроме того, используя теорему 1, получаем такое G_μ, что $P(G_\mu)>1-\mu$ для $l=1,2$ и Z принадлежит l-му классу.

Далее определяем множество $W_\mu=\{z:|\beta(z)-\Delta_1|>\xi_\mu/2\}\cap\{z: z\in G_\mu\}$, используя ξ_μ и G_μ.

Имеем, что

$$V_m(\Delta_1)=\frac{1}{m_1}\sum_{\{i:z_{1i}\notin W_\mu\}}|\Lambda\{\overline{\beta}_m(z_{1i})<\Delta_1\}-\Lambda\{\beta(z_{1i})<\Delta_1\}|+$$

$$+\frac{1}{m_1}\sum_{\{i:z_{1i}\in W_\mu\}}|\Lambda\{\overline{\beta}_m(z_{1i})<\Delta_1\}-\Lambda\{\beta(z_{1i})<\Delta_1\}|\leq$$

$$\leq\frac{1}{m_1}\sum_{i=1}^{m_1}\Lambda\{z_{1i}\notin W_\mu\}+\frac{1}{m_1}\sum_{\{i:z_{1i}\in W_\mu\}}|\Lambda\{\overline{\beta}_m(z_{1i})<\Delta_1\}-\Lambda\{\beta(z_{1i})<\Delta_1\}|.$$

Согласно предположению б) леммы 2 и теореме 1 имеет место асимптотическая сходимость

$$\frac{1}{m_1}\sum_{i=1}^{m_1}\Lambda\{z_{1i}\notin W_\mu\}\to P(Z_1\notin W_\mu)\leq P(|(\beta(Z_1)-\Delta_1|\leq\xi_\mu/2)+P(Z_1\notin G_\mu)<2\mu$$

при $\min\{m_1,m_2\}\to\infty$.

Как следует из $|\beta(z)-\Delta_1|>\xi_\mu/2$, $\exists M_0\geq 1$, что для всех $m=(m_1,m_2)$, где $\min\{m_1,m_2\}\geq M_0$, имеем $|\overline{\beta}_m(z)-\Delta_1|>\xi_\mu/2$.

Итак,

$$\frac{1}{m_1}\sum_{\{i:\,W(z_{1i})\}}|\Lambda\{\overline{\beta}_m(z_{1i})<\Delta_1\}-\Lambda\{\beta(z_{1i})<\Delta_1\}|=0,$$

откуда следует, что $V_m(\Delta_1)\leq 2\mu$.

В результате доказательство данной теоремы имеет место с применением аналогичных рассуждений для случая, когда $i=2$.

Теорема доказана.

Заметим, что такой же классификатор на основе полупространственной глубины является сложным и недостаточно эффективным при работе с нулевыми глубинами. Экспериментальная модификация функции полупространственной глубины принимает только дискретные значения, что приводит к потере информации для непрерывных распределений. В результате получаем неточные оценки плотности с пиками в окрестности этих дискретных значений. Кроме того, существенной проблемой являются неравенства в хвосте исходной оценки плотности $\bar{h}_l$, что вызвано наличием объектов с нулевыми глубинами. Отметим, что экспериментальная модификация функции экстраполяционной глубины не имеет таких проблем и является непрерывной в z. Поэтому развитая версия классификатора экстраполяционной глубины часто превосходит классификатор полупространственной глубины.

На практике на множестве данных необходимо оценивать a_l, оптимальный асимптотический порядок которых обоснован в лемме 2, где использован пропускной метод перекрестной проверки для выбора a_1, a_2 и Δ. Для снижения вычислительной стоимости выбрано $a_1 = (w_1 / w_2) a_2$, поскольку полосы пропускания должны быть пропорциональны дисперсиям множеств, где w_l $(l = 1, 2)$ является дисперсионной мерой оценочных функций глубины $\{\bar{\xi}_{m_l}^{(l)}(z_{l1}), \bar{\xi}_{m_l}^{(l)}(z_{l2}), \ldots$
$\ldots, \bar{\xi}_{m_l}^{(l)}(z_{lm_l})\}$. Далее вычислим

$$d_{m_i, a_i}^{(i)}(z_{lj}) = \bar{\varpi}_{ia_i}^{*}(\bar{\xi}_{m_i}^{(i)}(z_{lj}))(\bar{\xi}_{m_i}^{(i)}(z_{lj}))^{r-3} / (1 - \bar{\xi}_{m_i}^{(i)}(z_{lj}))^{r-1}$$

для a_2, $a_1 = (w_1 / w_2) a_2$, $i, l = 1, 2$ и $j = 1, \ldots, m_l$, где $\bar{\varpi}^{*}$ соответствует пропускной ядерной оценке плотности для $l = i$. Постоянная величина Δ, которая зависит от a_2, найдена на порядковых статистиках $\log[d_{m_1, a_1}^{(1)}(z_{lj})] - \log[d_{m_2, a_2}^{(2)}(z_{lj})]$, $l = 1, 2$, $j = 1, 2, \ldots, m_l$, для минимизации частоты ошибок перекрестной проверки. Отметим, что выбор a_2 в диапазоне значений обусловлен получением низкого коэффициента ошибок перекрестной проверки. Кроме того, выбран максимальный оптимизатор из множества минимизаторов, имеющих место вследствие ступенчатой природы частоты ошибок перекрестной проверки [6].

Данные результаты также можно получить для глубинной классификации с использованием глубины Махаланобиса. Поскольку v_H является постоянной величиной, которая зависит от исходного распределения H, оценка минимального ковариантного определителя матрицы разброса $\Xi_H \to v_H \Xi_H$. Однако независимо от значения v_H форма байесовского классификатора аналогична лемме 1. Данный метод классификации можно адаптировать к разработке развитой версии классификатора на основе глубины Махаланобиса, а его асимптотическую оптимальность доказать на основе леммы 2.

Для многоклассовой классификации $a_1, a_2, \ldots, a_L$ и $\alpha_1, \alpha_2, \ldots, \alpha_L$ выбираются аналогично, однако на практике вычислительно сложно минимизировать частоту ошибок перекрестной проверки относительно нескольких параметров. Таким образом, выполняем $\binom{L}{2}$ бинарных классификаций, рассматривая пару классов, где результаты всех попарных классификаций объединяются с помощью метода мажоритарного голосования. Заметим, что при соответствующих условиях регулярности можно доказать согласованность байесовского риска развитой версии классификатора экстраполяционной глубины для многоклассовых задач на основе леммы 2.

МНОГОУРОВНЕВОЕ СГЛАЖИВАНИЕ ДЛЯ КЛАССИФИКАЦИИ МНОГОМЕРНЫХ ДАННЫХ

Оценка параметра сглаживания в ядерных оценках плотности проведена с помощью метода перекрестной проверки для развитой версии глубинного классификатора. Однако при решении практических задач классификации достаточно часто имеет место неопределенность модели при использовании одной пары полос пропускания (a_1, a_2). Наряду с проблемой выборочной зависимости существенен выбор параметра сглаживания, который зависит от характерного объекта классификации. В данном случае определенный уровень сглаживания может определять различное поведение в различных областях пространства измерений. Следовательно, актуальна задача исследования результатов классификации для различных масштабов сглаживания вместо использования фиксированной пары (a_1, a_2) в определенном диапазоне. Объединять данные, которые индексированы по полосам пропускания, можно с помощью принятия взвешенного среднего значения оцененных апостериорных вероятностей [7].

Отметим, что $e^{\rho_{m,a_1,a_2}(z)}$ дает оценку $p_1 h_1(z) / p_2 h_2(z)$, поскольку элемент z относится к первому классу, если $\rho_{m,a_1,a_2}(z) = \log[d^{(1)}_{m_1,a_1}(z)] - \log[d^{(2)}_{m_2,a_2}(z)] - \Delta > 0$, где Δ выбирается путем минимизации ошибки перекрестной проверки для фиксированных (a_1, a_2). Итак, имеем $\overline{\pi}_{m,a_1,a_2}(1|z) = e^{\rho_{m,a_1,a_2}(z)} / (1 + e^{\rho_{m,a_1,a_2}(z)})$, что является оцененной апостериорной вероятностью класса.

Поскольку $\pi^*_m(l|z) = \sum\limits_{a_1,a_2 \in A} q_{a_1,a_2} \overline{\pi}_{m,a_1,a_2}(l|z)$, результирующий классификатор формируется путем объединения апостериорных оценок, полученных при различных значениях (a_1, a_2): $\mathfrak{I}_3(z) = \arg\max\limits_{l=1,2} \pi^*_m(l|z)$. Отметим, что q_{a_1,a_2} является весом, присвоенным классификатору, для которого a_1 и a_2 — полосы пропускания двух классов [8].

Объединение апостериорных оценок зависит от весовой функции q и диапазона полос пропускания $A = [a_1^j, a_1^g] \times [a_2^j, a_2^g]$. Однако независимо от выбора весовой функции частота ошибок $\mathfrak{I}_3(\cdot)$ асимптотически сходится к байесовскому риску, если верхняя и нижняя границы a_l^g и a_l^j от a_l удовлетворяют предположению в) леммы 2 для $l = 1, 2$.

Теорема 3. Предположим, что для $l = 1, 2$ величины h_1 и h_2 являются эллиптически-симметричными, где $h_l(z) > 0\ \forall z \in \mathrm{R}^r$, а $H_{\beta,l}(\bar{z}) = P(\beta(Z) \le \bar{z})$ — равномерно непрерывная функция в $\bar{z}$, где $\beta(z) = d^{(2)}(z) / d^{(1)}(z)$, $d^{(l)}(z) = \omega_l(\xi^{(l)}(z))(\xi^{(l)}(z))^{r-3} / (1 - \xi^{(l)}(z))^{r-1}$, $\xi^{(l)}(z) = F_{np}(z, H_l)$, а Z принадлежит l-му классу. Также предположим, что для a_l^g и a_l^j имеют место сходимости $a_l \to 0$ и $m_l a_l^4 \to \infty$ при $m_l \to \infty$. Тогда коэффициент ошибочной классификации многоуровневого классификатора экстраполяционной глубины $\mathfrak{I}_3(\cdot)$ сходится к байесовскому риску при $\min\{m_1, m_2\} \to \infty$.

Доказательство. Результат следует из теоремы Лебега о мажорируемой сходимости при условии, что для фиксированного z имеет место сходимость $\pi_m^*(1|z) \to \pi(1|z)$ при $\min\{m_1, m_2\} \to \infty$.

Предположим, что сходимость $\pi_m^*(1|z) \xrightarrow{P} \pi(1|z)$ не выполняется. Итак, $\exists \{m_\Delta = (m_{1\Delta}, m_{2\Delta}) : \Delta \geq 1\}$ и $\mu_0 > 0$, что $\forall \Delta \geq 1$, $|\pi_{m_\Delta}^*(1|z) - \pi(1|z)| > \mu_0$. Пусть $\{A_{m_\Delta}\}$, $\Delta \geq 1$, является соответствующей последовательностью диапазона полосы пропускания. Учитывая тот факт, что $\pi_{m_\Delta}^*(1|z)$ является взвешенным средним значением $\overline{\pi}_{m_\Delta, a_1, a_2}(1|z)$, можно получить такую подпоследовательность $\{(a_1^{m_\Delta}, a_2^{m_\Delta}) \in A_{m_\Delta}, \Delta \geq 1\}$, что $|\overline{\pi}_{m_\Delta, a_1^{m_\Delta}, a_2^{m_\Delta}}(1|z) - \pi(1|z)| > \mu_0 \quad \forall \Delta \geq 1$. Отсюда следует, что сходимость $\overline{\pi}_{m_\Delta, a_1^{m_\Delta}, a_2^{m_\Delta}}(1|z) \xrightarrow{P} \pi(1|z)$ не выполняется. В результате имеет место противоречие, поскольку последовательность полос пропускания удовлетворяет условию регулярности, при котором для $l = 1, 2$, имеет место сходимость $a_l \to 0$ и $m_l a_l^4 \to \infty$ при $m_l \to \infty$.

Теорема доказана.

На основе доказательства теоремы 3 можно утверждать, что выбор весовой функции q не имеет значительного влияния на выборочную производительность классификатора $\mathfrak{I}_3(\cdot)$. Однако выбор A и q является необходимым при использовании бесконечной выборки. Заметим, что вес должен постепенно уменьшаться при увеличении частоты ошибок с использованием более крупных весов для классификаторов, которые имеют более низкие частоты ошибок [9].

Частота ошибок Ψ_{a_1, a_2} оценивается пропускным методом перекрестной проверки с помощью весовой функции

$$q_{a_1, a_2} = \exp\left[-\frac{1}{2} \frac{(\overline{\Psi}_{a_1, a_2} - \overline{\Psi}_0)^2}{\overline{\Psi}_0 (1 - \overline{\Psi}_0) / (m_1 + m_2)}\right] \Lambda[\overline{\Psi}_{a_1, a_2} \leq \min\{p_1, p_2\}],$$

где $\overline{\Psi}_0 = \min\limits_{a_1, a_2} \overline{\Psi}_{a_1, a_2}$.

В случае, когда развитая версия одноуровневого классификатора экстраполяционной глубины используется для классификации $(m_1 + m_2)$ объектов, в качестве оценок для среднего значения и дисперсии экспериментальной частоты ошибок можно рассматривать $\overline{\Psi}_0$ и $\overline{\Psi}_0 (1 - \overline{\Psi}_0) / (m_1 + m_2)$. Кроме того, частотой ошибок классификатора, который относит все объекты к классу с наибольшей априорной вероятностью, является $\min\{p_1, p_2\}$. Заметим, что весовая схема исследуемого классификатора демонстрирует нулевой вес, если классификатор с парой полос пропускания (a_1, a_2) менее эффективен по сравнению с тривиальным классификатором.

Использовался метод на основе квантилей парных расстояний для выбора A, а также определены 500 равноудаленных значений (a_1, a_2) в этом интервале, который удовлетворяет условию $a_1 = (w_1 / w_2) a_2$, где w_1 и w_2 одинаковы. Отметим, что в результате проведенных экспериментов получены высокие показатели вследствие удачного выбора диапазона полос пропускания и весовой функции.

ЗАКЛЮЧЕНИЕ

В данной работе предложен и исследован непараметрический аффинно-инвариантный классификатор на основе экстраполяционной глубины, который является устойчивым к выбросам и экстремальным значениям. Вследствие связи предложенного классификатора с расстоянием Махаланобиса, а также непрерывности его экспериментальной формы классификатор экстраполяционной глубины превосходит классификаторы полупространственной и ординальной глубины при решении широкого спектра практических задач классификации. Поскольку классификатор экстраполяционной глубины является легко модифицированным, его можно применять к глобальному классу параметрических моделей, тогда как линейные и квадратичные методы статистики и машинного обучения эффективно выполняются только при условии нормальности распределения. Кроме того, предложенный классификатор позволяет избавиться от «проклятия размерности», что касается экспоненциального роста необходимых экспериментальных данных в зависимости от размерности пространства при решении задач вероятностно-статистического распознавания образов и классификации. Следовательно, при работе с небольшими выборками в пространстве большой размерности классификатор на основе экстраполяционной глубины превосходит обычные непараметрические методы, когда множества данных являются почти эллиптическими. Отметим, что многоуровневая структура сглаживания позволяет исследовать глобальные свойства функций плотности и границ класса. В результате на практике предложенный многоуровневый метод является достаточно гибким вследствие агрегации результатов для различных масштабов сглаживания.

СПИСОК ЛИТЕРАТУРЫ

1. Chacón J.I., Duong T., Wand M.P. Asymptotics for general multivariate kernel density derivative estimators // Statistica Sinica. — 2011. — **21.** — P. 807–840.
2. Rousseeum P.J., Struyf A. Characterizing angular symmetry and regression symmetry // Statist. Plann. Inference. — 2004. — **122.** — P. 163–170.
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. Oja H., Paindaveine D. Optimal signed-rank tests based on hyperplanes // J. Statist. Plann. Inference. — 2005. — **135.** — P. 300–323.
5. Lange T., Mosler K., Mozharovskyi P. Fast nonparametric classification based on data depth // Statist. Papers. — 2014. — **55.** — P. 49–69.
6. 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.
7. Pollard D. Convergence of stochastic processes. — New York: Springer-Verlag, 1984. — P. 1–10.
8. Zuo Y., Serfling R. Structural properties and convergence results for contours of sample statistical depth functions // The Annals of Statistics. — 2000. — **28.** — P. 484–497.
9. Vardi Y., Zhang C.H. The multivariate on L_1-median and associated data depth // Proceedings of the National Academy of Sciences of the United States of America. — 2000. — **97.** — P. 1423–1426.

Поступила 06.10.2015