Теорія ймовір. та матем. статист.
Вип. 72, 2005, стор. 140–149

Teor. Ĭmovir. ta Matem. Statyst.
No. 72, 2005, pp. 140–149

РОЗПІЗНАВАННЯ КОМПОНЕНТІВ СУМІШІ

УДК 519.21

О. В. СУГАКОВА

Анотація. Рассматривается задача классификации объектов, выбранных из смеси нескольких компонент с разными вероятностными распределениями. Для построения классификатора используются ядерные оценки плотности компонент смесей для одномерной случайной величины $S_j^N(b) = \sum_{i=1}^d b_i \xi_j^{N,i}$, которая является проекцией вектора наблюдений $\xi_j^N = (\xi_j^{N,1}, \xi_j^{N,2}, \dots, \xi_j^{N,d})$ на неслучайное направление $b = (b_1, b_2, \dots, b_d)$. Получена оценка $\hat{b}$ для наилучшего возможного направления b. Доказано, що вероятность ошибки классификатора, использующего $S(\hat{b})$, сходится к наименьшей из возможных для таких классификаторов вероятности ошибки.

1. Вступ і огляд літератури

Однією з основних задач статистичного аналізу є класифікація: визначення класу, до якого належить об'єкт на основі його спостережуваних характеристик ξ [1]. Розглянемо наступну постановку задачі.

Нехай ми маємо вибірку з N об'єктів, які можуть належати одному з M класів. Номер класу, якому належить j-й об'єкт, позначимо $\operatorname{ind}(j)$. Цей номер невідомий, але відомі ймовірності $w_{j,N}^k = \mathsf{P}\{\operatorname{ind}(j) = k\}$ (концентрація k-ї компоненти у суміші під час j-го спостереження). Спостерігається векторна характеристика $\xi_j^N = (\xi_j^{N,1}, \xi_j^{N,2}, \dots, \xi_j^{N,d}) \in \mathbf{R}^d$ об'єкта з номером j, причому щільність розподілу ξ_j^N при умові, що $\operatorname{ind}(j) = k$, є $h_k(x_1, x_2, \dots, x_d)$. Таким чином, безумовна щільність ξ_j^N є

$$f_{\xi_j^N}(x_1, \dots, x_d) = \sum_{k=1}^M w_{j,N}^k h_k(x_1, \dots, x_d).$$

(Величини $w_{j,N}^k$ вважаємо відомими, а $h_k(x_1, x_2, \dots, x_d)$ — невідомими.) Характеристики ξ_j^N вважаємо незалежними між собою при фіксованих N.

Нам потрібно, маючи вибірку $X = (\xi_j^N, j = 1, \dots, N)$, побудувати класифікатор $g\colon \mathbf{R}^d \to \{1, \dots, M\}$, який на основі характеристики ξ нового об'єкта O (незалежно від X) оцінював би номер $\operatorname{ind}(O)$ класу, якому цей об'єкт належить.

Як відомо, при класичному байєсовому підході [2] найкращим вважається класифікатор, який мінімізує ймовірність помилки (байєсів класифікатор). Для побудови такого класифікатора потрібно знати щільності розподілів характеристик об'єктів, що належать різним класам. Часто статистик не знає цих щільностей, але має розкласифіковану вибірку з об'єктів різних класів (навчаючу вибірку). У такому випадку за навчаючою вибіркою будують оцінки невідомих щільностей і підставляють їх у формулу для байєсового класифікатора. У результаті отримують так званий емпірично-байєсів класифікатор [2]. Для гауссових спостережень ця процедура

2000 *Mathematics Subject Classification.* Primary 62H30; Secondary 62G07.

Ключові слова і фрази. Ядерні оцінки щільності, емпірично-байєсів класифікатор, оцінки компонент суміші.

приводить до звичайного дискримінантного аналізу [3]: класифікація зводиться до порівняння кількох "дискримінантних" функцій, які є лінійними комбінаціями спостережуваних характеристик.

У випадку непараметричного оцінювання щільностей ймовірність помилки емпірично-байєсових класифікаторів прямує до помилки байєсового класифікатора [2]. Але для багатовимірних спостережень емпірично-байєсові класифікатори мають дуже складну, непрозору форму, їх важко інтерпретувати. Тому на практиці часто використовують наступний підхід. Багатовимірні характеристики аналізованих об'єктів проектують на деякий напрямок, тобто обирають певну лінійну комбінацію всіх спостережуваних змінних, що характеризують об'єкт. У економетриці та соціальній статистиці такі комбінації звуть "сумарними індексами", у психометриці — "шкалами". Класифікація об'єктів проводиться на основі подібного загального індекса.

У даній статті розглядається задача побудови такого "лінійного" індекса, який забезпечував би асимптотично мінімальну помилку класифікації. Для цього спостережувані дані проектуються на різні напрямки, за кожною проекцією будується емпірично-байєсів класифікатор, оцінюється ймовірність його помилки, і на роль "найкращого" індекса беруть той, для якого ця оцінка є найменшою. Показано, що помилка емпірично-байєсової класифікації на основі індексів, вибраних таким способом, прямує до найменшої можливої помилки класифікації за лінійними індексами. При цьому на роль навчаючої вибірки беруть вибірку з суміші зі змінними концентраціями [4, 5]. Класифікація по сумішах зі змінними концентраціями розглядалась також у [6].

2. Постановка задачі

Ми будемо розглядати "лінійні індекси" вигляду

$$S_j^N(b) = \sum_{i=1}^{d} b_i \xi_j^{N,i}, \tag{1}$$

де $b = (b_1, b_2, \dots, b_d) \in \mathbf{R}^d$; довжина невипадкового вектора b дорівнює одиниці:

$$b_1^2 + b_2^2 + \cdots + b_d^2 = 1. \tag{2}$$

Індекс $S(b)$ є, по суті, проекцією ξ на напрямок b. Ми будемо мати справу лише з класифікаторами вигляду $\bar g_b(\xi) = g(S(b))$, де $g\colon \mathbf{R} \to \{1, \dots, M\}$ — довільна вимірна функція. Ймовірність помилки класифікатора $\bar g_b$

$$L_{\bar g_b} = \mathsf{P}\{\bar g_b(\xi) \neq \operatorname{ind}(O)\}.$$

Як відомо [2], мінімум $L_{\bar g_b}$ по всіх g при фіксованому b досягається на байєсовому класифікаторі

$$g^*(x) = \arg\max_i \left(p_i u_i(x)\right),$$

де $p_i = \mathsf{P}\{\operatorname{ind}(O) = i\}$ — концентрації компонентів суміші під час спостереження O; $u_i(x)$ — щільності в.в. $\eta_i(b) = \sum_{j=1}^d \eta_i^j b_j$, де $\eta_i = (\eta_i^1, \eta_i^2, \dots \eta_i^d)$ має щільність h_i. Ймовірність помилки цього класифікатора має вигляд

$$L_g^* = L^*(b) = 1 - \int_{-\infty}^{\infty} \max_{1 \le k \le M} (p_k u_k(x))\, dx.$$

Найкращим для побудови класифікатора є індекс, отриманий проекцією ξ на напрямок

$$b^* = \arg\min_b L^*(b) = \arg\max \int_{-\infty}^{\infty} \max_{1 \le k \le M} (p_k u_k(x))\, dx.$$

Оскільки справжні значення u_k невідомі, замінимо їх ядерними оцінками, побудованими за вибіркою $(S_j(b))_{j=1}^N$. Ці оцінки мають вигляд [4]:

$$\hat{u}_i^N(b,x) = \frac{1}{N\sigma_N}\sum_{j=1}^N a_{j,N}^i K\left(\frac{x - S_j(b)}{\sigma_N}\right), \tag{3}$$

де

$$a_{j,N}^i = \frac{1}{\det \Gamma_N}\sum_{k=1}^M (-1)^{i+k}\gamma_{ik}^N w_{j,N}^k; \qquad i,j = 1,\dots,M;$$

тут у нас $\Gamma_N = (\langle w_\cdot^k, w_\cdot^l\rangle)_{k,l=1}^M$ — матриця Грама системи функцій $w_{j,N}^k$ зі скалярним добутком

$$\langle w_\cdot^k, w_\cdot^l\rangle_N = \frac{1}{N}\sum_{j=1}^N w_{j,N}^k w_{j,N}^l,$$

а γ_{jk}^N — j,k -й мінор матриці Γ_N. При цьому $a_{j,N}^i$ задовольняють умову

$$\langle a_\cdot^i, w_\cdot^k\rangle_N = \frac{1}{N}\sum_{j=1}^N a_{j,N}^i w_{j,N}^k = \chi\{k = i\}. \tag{4}$$

Тут $\chi\{A\}$ — індикатор події A. Вважаємо, що σ_N є параметром згладжування; $\sigma_N \to 0$, при $N \to \infty$, але $N\sigma_N \to \infty$; K — ядро оцінки — функція, що являє собою щільність деякої в.в.η. У [7] доведена асимптотична незсуненість та конзистентність оцінки типу (3).

За оцінку ймовірності помилкової класифікації $L^*(b)$ беремо

$$\hat{L}^N(b) = 1 - \int_{-\infty}^{\infty} \max_{1\le k\le M}\left(p_k \hat{u}_k^N(b,x)\right) dx,$$

а за оцінку b^* —

$$\hat{b} = \arg\min \hat{L}^N(b).$$

"Квазибайесів класифікатор" для спостереження ξ будуємо за формулою

$$\hat{g}(\hat{b},x) = \arg\max_{1\le i\le M}(p_i \hat{u}_i^N(\hat{b},x)).$$

Для оцінки якості класифікатора служить умовна ймовірність помилки при фіксованій навчаючій вибірці [2]:

$$\mathsf{P}\{\hat{g}(\hat{b},x) \ne \mathrm{ind}(O) \mid X\} = L_{\hat{g}}(\hat{b}) = 1 - \int_{-\infty}^{\infty} p_{\hat{g}(\hat{b},x)} u_{\hat{g}(\hat{b},x)}\, dx.$$

3. Основні результати

Припустимо, що виконуються наступні умови

$$\text{(i)} \qquad K(x) \le a, \qquad \int_{-\infty}^{\infty} |x| K(x)\, dx < \infty,$$

де a — деяка константа.

Сукупність множин $S = \{A\}$, де

$$\text{(ii)} \qquad A = \left\{y\colon K\left(\frac{x - \sum_{i=1}^d b_i y_i}{\sigma_N}\right) \ge c\right\}$$

(c —деяка стала), утворює VC-клас [1, розд. X], [5, с. 55].

Усі частинні похідні щільностей $h_i(x_1,\dots,x_d)$ для усіх $i = 1,\dots,M$ існують і обмежені:

$$\text{(iii)} \qquad \left|\frac{\partial h_i}{\partial x_s}(x_1,\dots,x_d)\right| \le c_{is}, \qquad s = 1,\dots,d.$$

Вагові коефіцієнти $a^i_{j,N}$ задовольняють умові

$$\text{(iv)} \qquad \left|a^i_{j,N}\right| \le \bar{a},$$

де $\bar{a}$ — деяка стала.

$$\text{(v)} \qquad \frac{1}{\sigma_N} = o(N^{1/4}), \qquad N \to \infty.$$

Теорема 3.1. *При виконанні умов* (i)–(v)

$$\mathsf{E}\, L_{\hat{g}}(\hat{b}) \to L^*_g(b^*) \quad \textit{при } N \to \infty.$$

Наслідок 3.1. *В умовах теореми*

$$L_{\hat{g}}(\hat{b}) \to L^*_g(b^*) \quad \textit{при } N \to \infty$$

за ймовірністю.

Доведемо наступні леми.

Лема 3.1. *За умов* (i)–(v) *для усіх* $k = 1, \dots, M$

$$\mathsf{E} \sup_{b,x} \left|\hat{u}^N_k(b,x) - u_k(b,x)\right| \to 0 \quad \textit{при } N \to \infty.$$

Доведення леми 3.1. Зробимо наступні оцінки

$$\begin{aligned}
&\sup_{b,x} \left|\hat{u}^N_k(b,x) - u_k(b,x)\right| \\
&\quad \le \sup_{b,x} \left| \frac{1}{N\sigma_N} \sum_{j=1}^N a^k_{j,N} K\left(\frac{x - S^N_j(b)}{\sigma_N}\right) - \frac{1}{N\sigma_N} \mathsf{E} \sum_{j=1}^N a^k_{j,N} K\left(\frac{x - S^N_j(b)}{\sigma_N}\right) \right| \\
&\qquad + \sup_{b,x} \left| \frac{1}{N\sigma_N} \mathsf{E} \sum_{j=1}^N a^k_{j,N} K\left(\frac{x - S^N_j(b)}{\sigma_N}\right) - u_k(b,x) \right| \\
&\quad = S^N_1 + S^N_2.
\end{aligned} \tag{5}$$

Розглянемо S^N_1. Визначимо зважену емпіричну міру співвідношенням

$$\hat{\mu}_N(A) = \frac{1}{N} \sum_{j=1}^N a^k_{j,N} \chi\{\xi^N_j \in A\}, \qquad \bar{\mu}_N(A) = \mathsf{E}\, \hat{\mu}_N(A). \tag{6}$$

Нехай $\mathsf{P}(dy)$ — ймовірнісна міра. Тоді, використовуючи доведення з [1, с. 287], можемо написати

$$\begin{aligned}
S^N_1 = \sup_{b,x} \Bigg| &\frac{1}{\sigma_N} \int_{\mathbf{R}^d} K\left(\frac{x - \sum_{i=1}^d b_i y_i}{\sigma_N}\right) \mathsf{P}(dy) \\
&- \frac{1}{\sigma_N} \int_{\mathbf{R}^d} K\left(\frac{x - \sum_{i=1}^d b_i y_i}{\sigma_N}\right) \hat{\mu}_N(dy) \Bigg| \\
\le \frac{a}{\sigma_N} &\sup_{A \in S} \left|\mathsf{P}(A) - \hat{\mu}_N(A)\right|,
\end{aligned} \tag{7}$$

де S — система множин з умови (ii). Тут ми скористалися умовою (i). Далі, відзначимо, що

$$\sup_{A\in S} \left|\mathsf{P}(A) - \hat{\mu}_N(A)\right| \le \sup_{1\le j\le N} \left|a^i_{j,N}\right| + 1 \le \bar{a} + 1,$$

а також введемо позначення

$$\tilde{C}^k_N = \sup_{1\le j\le N} \left|a^i_{j,N}\right| + \sup_{1\le j\le N} a^i_{j,N} - \inf_{1\le j\le N} a^i_{j,N}.$$

Застосуємо наступну теорему.

Теорема 3.2 ([5, с. 60]). *Нехай $a_{j,N}$ — довільний масив вагових коефіцієнтів; S є VC-класом, функцію росту якого позначимо $g^G(N)$. $\hat{\mu}_N(A)$, $\bar{\mu}_N(A)$ визначені рівностями* (6). *Тоді для усіх $\lambda > 2M/N$*

$$\mathsf{P}\left\{\sup \frac{|\hat{\mu}_N(A) - \bar{\mu}_N(A)|}{\bar{C}_N^k} \geq \lambda\right\} \leq M\left(6Ng^G(2N)\exp\left(-\frac{\lambda^2 N}{32M^2}\right) + 2\exp\left(-\frac{\lambda^2 N}{8M^2}\right)\right).$$

Зробимо наступну оцінку:

$$\mathsf{E}\sup_{A\in S}\left|\frac{\mathsf{P}(A) - \hat{\mu}_N(A)}{\bar{C}_N^k}\right| \leq \int_0^{\frac{\bar{a}+1}{\bar{C}_N^k}} \mathsf{P}\left\{\sup_{A\in S}\left|\frac{\mathsf{P}(A) - \hat{\mu}_N(A)}{\bar{C}_N^k}\right| \geq \lambda\right\} d\lambda.$$

Оскільки $\bar{C}_N^k$ — обмежена знизу величина, то, не обмежуючи загальності, можемо вважати, що

$$\frac{2M}{\sqrt[4]{N}} < \frac{\bar{a}+1}{\bar{C}_N^k}.$$

Розіб'ємо інтеграл на дві частини, і до оцінки другої застосуємо теорему 2:

$$\begin{aligned}
&\mathsf{E}\sup_{A\in S}\left|\frac{\mathsf{P}(A) - \hat{\mu}_N(A)}{\bar{C}_N^k}\right| \\
&\quad\leq \int_0^{\frac{2M}{\sqrt[4]{N}}} d\lambda + \int_{\frac{2M}{\sqrt[4]{N}}}^{\frac{\bar{a}+1}{\bar{C}_N^k}} \left(6MNg^S(2N)\exp\left(-\frac{\lambda^2 N}{32M^2}\right) + 2\exp\left(-\frac{\lambda^2 N}{8M^2}\right)\right) d\lambda \qquad (8)\\
&\quad\leq 2MN^{-1/4} \\
&\qquad + \left(6MNg^S(2N)\exp\left(-\frac{1}{8}\sqrt{N}\right) + 2\exp\left(-\frac{1}{2}\sqrt{N}\right)\right)\left(\frac{\bar{a}+1}{\bar{C}_N^k} - \frac{2M}{\sqrt[4]{N}}\right).
\end{aligned}$$

З оцінок (7), (8) випливає, що

$$\begin{aligned}
\mathsf{E}\,S_1^N &\leq O\left(\sigma_N^{-1}N^{-1/4}\right) + O\left(\sigma_N^{-1}Ng^S(2N)\exp\left(-\frac{1}{8}\sqrt{N}\right)\right) \\
&\quad + O\left(\sigma_N^{-1}\exp\left(-\frac{1}{2}\sqrt{N}\right)\right).
\end{aligned} \qquad (9)$$

При $N\to\infty$ права частина (9) прямує до нуля: перший доданок через умову (v); другий і третій завдяки тому, що $g^S(N)$, як функція росту VC класу S, зростає як степенева функція, σ_N^{-1} теж, а експоненційні множники прямують до нуля швидше, ніж степеневі. Отже,

$$\mathsf{E}\,S_1^N = o(1) \quad \text{при } N\to\infty. \qquad (10)$$

Тепер оцінимо S_2^N:

$$\begin{aligned}
\frac{1}{N\sigma_N}\mathsf{E}\sum_{j=1}^N a_{j,N}^k K\left(\frac{x - S_j^N(b)}{\sigma_N}\right) &= \frac{1}{N\sigma_N}\sum_{j=1}^N a_{j,N}^k\,\mathsf{E}\,K\left(\frac{x - S_j^N(b)}{\sigma_N}\right) \\
&= \frac{1}{N\sigma_N}\sum_{j=1}^N a_{j,N}^k \sum_{k=1}^M w_{j,N}^k \int_{-\infty}^{\infty} K\left(\frac{x-y}{\sigma_N}\right) u_k(b,y)\,dy \\
&= \int_{-\infty}^{\infty} K(z)u_k(b, x - \sigma_N z)\,dz.
\end{aligned}$$

Тут ми скористалися рівністю (4). Отже, S_2^N можна представити наступним чином

$$S_2^N = |\mathsf{E}\,u_k(b, x - \sigma_N\eta) - u_k(b, X)|,$$

де η — в.в. зі щільністю $K(x)$. Внаслідок умови (2) у вектора $b = (b_1, b_2, \ldots, b_d)$ хоча б одна компонента ненульова, нехай $b_d \neq 0$. Тоді щільність випадкової величини μ_k можна представити так:

$$u_k(b, x) = \frac{1}{b_d} \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} h_k \left(x_1, \ldots, x_{d-1}, \frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i \right) \right) dx_{d-1}.$$

Для спрощення далі будемо писати $h_k(x_d)$ замість $h_k(x_1, \ldots, x_{d-1}, x_d)$. Виберемо послідовність $t = t_N \to \infty$, $N \to \infty$. Зробимо наступні оцінки

$$\begin{aligned} S_2^N &= \frac{1}{|b_d|} \mathsf{E} \int_{-\infty}^{\infty} dx_1 \ldots \int_{-\infty}^{\infty} \left| h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i - \sigma_N \eta \right) \right) \right. \\ &\qquad\qquad\qquad\qquad \left. - h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i \right) \right) \right| dx_{d-1} \\ &\le \frac{1}{|b_d|} \mathsf{E} \int\limits_{x_1^2 + \cdots + x_{d-1}^2 \le t^2} dx_1 \ldots dx_{d-1} \left| h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i - \sigma_N \eta \right) \right) \right. \\ &\qquad\qquad\qquad\qquad \left. - h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i \right) \right) \right| \\ &\quad + \frac{1}{|b_d|} \mathsf{E} \int_{x_1^2 + \cdots + x_{d-1}^2 > t^2} dx_1 \ldots dx_{d-1} h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i - \sigma_N \eta \right) \right) \\ &\quad + \frac{1}{|b_d|} \int_{x_1^2 + \cdots + x_{d-1}^2 > t^2} dx_1 \ldots dx_{d-1} h_k \left(\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i \right) \right) \\ &= A + B + C. \end{aligned} \tag{11}$$

Для оцінки A застосуємо теорему Лагранжа:

$$A \le \mathsf{E} \frac{\sigma_N |\eta|}{b_d^2} \int_{x_1^2 + \cdots + x_{d-1}^2 \le t^2} dx_1 \ldots dx_{d-1} \left| \frac{\partial h_k}{\partial x_d} \left(x_1, \ldots, x_{d-1}, \frac{s - \sum_{i=1}^{d-1} b_i x_i}{b_d} \right) \right|. \tag{12}$$

Тут s — деяка в.в., що знаходиться між величинами

$$\frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i - \sigma_N \eta \right) \quad \text{та} \quad \frac{1}{b_d} \left(x - \sum_{i=1}^{d-1} b_i x_i \right).$$

У інтегралі правої частини (12) робимо заміну $x_1 = v b_d^2$, а потім використовуємо умову обмеженості частинних похідних (iii). Маємо

$$\begin{aligned} A &\le \sigma_N \mathsf{E} |\eta| \int_{v^2/b_d^2 + \cdots + x_{d-1}^2 \le t^2} dv \ldots dx_{d-1} \\ &\qquad\qquad \times \left| \frac{\partial h_k}{\partial x_d} \left(v b_d^2, \ldots, x_{d-1}, \frac{1}{b_d} \left(s - \sum_{i=1}^{d-1} b_i x_i \right) \right) \right| \\ &\le \sigma_N c_{kd} \mathsf{E} |\eta| \operatorname{mes} \left\{ (v, x_2, \ldots, x_{d-1}) \in \mathbf{R}^{d-1} \colon \frac{v^2}{b_d^2} + x_2^2 + \cdots + x_{d-1}^2 \le t^2 \right\} \\ &\le \sigma_N c_{kd} \mathsf{E} |\eta| \operatorname{mes} \left\{ (v, x_2, \ldots, x_{d-1}) \in \mathbf{R}^{d-1} \colon v^2 + x_2^2 + \cdots + x_{d-1}^2 \le t^2 \right\} \\ &\le \sigma_N c_{kd} \mathsf{E} |\eta| (2t)^{d-1}. \end{aligned} \tag{13}$$

Тепер оцінимо B:

$$B = \frac{1}{|b_d|} \mathsf{E} \int_{x_1^2+\cdots+x_{d-1}^2>t^2;\, x_d = x - \sum_{i=1}^{d-1} b_i x_i - \sigma_N \eta} dx_1 \ldots dx_{d-1} h_k \left(x_1, \ldots, x_{d-1}, \frac{x_d}{b_d} \right).$$

Оскільки підінтегральна функція невід'ємна, то, збільшуючи область інтегрування, ми не зменшимо величину інтеграла.

$$B \le \frac{1}{|b_d|} \int_{x_1^2+\cdots+x_{d-1}^2>t^2;\, -\infty<x_d<\infty} dx_1 \ldots dx_d h_k \left(x_1, \ldots, x_{d-1}, \frac{x_d}{b_d} \right).$$

Зробимо заміну змінної $v_d = x_d/b_d$ в інтегралі:

$$B \le \int_{x_1^2+\cdots+x_{d-1}^2>t^2;\, -\infty<v_d<\infty} dx_1 \ldots dv_d h_k (x_1, \ldots, v_d). \tag{14}$$

Доданок C оцінюється аналогічно B:

$$C \le \int_{x_1^2+\cdots+x_{d-1}^2>t^2;\, -\infty<v_d<\infty} dx_1 \ldots dv_d h_k (x_1, \ldots, v_d). \tag{15}$$

Отже, враховуючи (11)—(15), можемо написати

$$S_2^N \le c_{kd} \mathsf{E} |\eta| \sigma_N (2t)^{d-1} + 2 \int_{x_1^2+\cdots+x_{d-1}^2>t^2;\, -\infty<v_d<\infty} dx_1 \ldots dv_d h_k (x_1, \ldots, v_d). \tag{16}$$

Зрозуміло, що можна вибрати $t = t_N$ таким чином, щоб

$$\sigma_N (2t)^{d-1} \to 0 \quad \text{при } N \to \infty.$$

При такому виборі права частина (16) буде прямувати до 0 при $N \to \infty$ і отже,

$$S_2^N \to 0 \quad \text{при } N \to \infty \tag{17}$$

рівномірно по x та b. Збираючи докупи твердження (5), (10) та (17), завершуємо доведення леми 3.1. □

Зауваження 3.1. Якщо $K(x)$ — кусково-монотонна функція зі скінченною кількістю інтервалів монотонності, то система подій S, визначена в (ii), являє собою VC-клас.

Дійсно, в цьому випадку існує скінченна кількість G відрізків $[d_l, d_{l+1}]$ таких, що

$$\left\{ y \colon K \left(\frac{x - \sum_{i=1}^d b_i y_i}{\sigma_N} \right) \ge c \right\} = \cup_{l=1}^G \left\{ x - \sigma_N d_{l+1} \le \sum_{i=1}^d b_i y_i \le x - \sigma_N d_l \right\}.$$

Кожна така множина — прошарок між двома гіперплощинами; об'єднання скінченної кількості таких множин є VC-класом [1, с. 232, 233].

Лема 3.2. *В умовах* (i)—(iv)

$$\mathsf{E} \sup_b \left| \hat{L}_g^N(b) - L_g^*(b) \right| \to 0 \quad \textit{при } N \to \infty.$$

Доведення леми 3.2. Скористаємося наступною нерівністю Дьорфі, доведеної в [2].

Теорема 3.3 ([2, с. 154]).

$$\hat{L}_g^N - L_g^* \le 2 \sum_{k=1}^M \int_{-\infty}^{\infty} \left| u_k(x) - \hat{u}_k^N(x) \right| dx. \tag{18}$$

Виберемо послідовність $D_N \to \infty$, $N \to \infty$, та зробимо наступні оцінки

$$\begin{aligned}\mathsf{E}\int_{-\infty}^{\infty}\left|\hat{u}_k^N(b,x)-u_k(b,x)\right|dx &\le \mathsf{E}\int_{-D_N}^{D_N}\left|\hat{u}_k^N(b,x)-u_k(b,x)\right|dx\\ &\quad+\mathsf{E}\left(\int_{D_N}^{\infty}\hat{u}_k^N(b,x)\,dx+\int_{-\infty}^{-D_N}\hat{u}_k^N(b,x)\,dx\right)\\ &\quad+\left(\int_{D_N}^{\infty}u_k(b,x)\,dx+\int_{-\infty}^{-D_N}u_k(b,x)\,dx\right)\\ &=A+B+C.\end{aligned} \tag{19}$$

Оцінимо кожний доданок у правій частині (19). Маємо

$$A \le 2D_N\,\mathsf{E}\sup_{b,x}\left|\hat{u}_k^N(b,x)-u_k(b,x)\right|. \tag{20}$$

За лемою 1,

$$\mathsf{E}\sup_{b,x}\left|\hat{u}_k^N(b,x)-u_k(b,x)\right| \to 0, \qquad N\to\infty,$$

і отже, можна вибрати таку послідовність $D_N \to \infty$, щоб права частина (20) прямувала до нуля.

Тепер оцінимо доданок C. За нерівністю Коші–Буняковського та (2)

$$-\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2} \le \sum_{i=1}^{d} b_i\eta_k^i \le \sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2}. \tag{21}$$

Позначимо функцію розподілу в.в. μ_k як $F(x)$ та при оцінюванні скористаємося нерівністю (21):

$$\begin{aligned}\int_{D_N}^{\infty}u_k(b,x)\,dx &= 1-F(D_N) \le \mathsf{P}\left\{\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2} \ge D_N\right\};\\ \int_{-\infty}^{-D_N}u_k(b,x)\,dx &= F(-D_N) \le \mathsf{P}\left\{-\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2} \le -D_N\right\}\\ &= \mathsf{P}\left\{\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2} > D_N\right\}.\end{aligned}$$

З цих оцінок випливає, що

$$C \le 2\,\mathsf{P}\left\{\sum_{i=1}^{d}\left(\eta_k^i\right)^2 \ge D_N^2\right\} \to 0 \quad \text{при } N\to\infty. \tag{22}$$

Аналогічно оцінимо доданок B. Не обмежуючи загальності, вважаємо, що $\sigma_N \le 1$. За нерівністю Коші–Буняковського

$$\left(\sum_{i=1}^{d} b_i\eta_k^i+\sigma_N\eta\right)^2 \le \left(1+\sigma_N^2\right)\left(\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2\right).$$

Отже,

$$-\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2}\le\sum_{i=1}^{d}b_i\eta_k^i+\sigma_N\eta\le\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2},$$

$$\begin{aligned}
\mathsf{E}\int_{D_N}^{\infty}\hat{u}_k^N(b,x)\,dx&=\frac{1}{N\sigma_N}\,\mathsf{E}\int_{D_N}^{\infty}\sum_{j=1}^{N}a_{j,N}^kK\left(\frac{x-S_j^N(b)}{\sigma_N}\right)dx\\
&=\int_{D_N}^{\infty}\mathsf{E}\,u_k(b,x-\sigma_N\eta)\,dx=\mathsf{E}(1-F(D_N-\sigma_N\eta))\\
&=\mathsf{P}\left\{\sum_{i=1}^{d}b_i\eta_k^i+\sigma_N\eta\ge D_N\right\}\\
&\le\mathsf{P}\left\{2\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2}\ge D_N\right\},
\end{aligned}\tag{23}$$

$$\begin{aligned}
\mathsf{E}\int_{-\infty}^{-D_N}\hat{u}_k^N(b,x)\,dx&=\mathsf{E}\,F(-D_N-\sigma_N\eta)\le\mathsf{P}\left\{-2\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2}<-D_N\right\}\\
&=\mathsf{P}\left\{2\sqrt{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2}>D_N\right\}.
\end{aligned}$$

З оцінок (23) випливає

$$B\le 2\,\mathsf{P}\left\{\sum_{i=1}^{d}\left(\eta_k^i\right)^2+\eta^2\ge\frac{D_N^2}{4}\right\}\to 0\quad\text{при } N\to\infty.\tag{24}$$

Збираючи докупи граничні твердження (20), (22), (24) та оцінку (19), за нерівністю Дьорфі отримуємо твердження леми 3.2.

Доведення теореми 3.1. Зробимо такі перетворення

$$L_{\hat g}(\hat b)-L_g^*(b^*)=\left(L_{\hat g}(\hat b)-\hat L_g^N(\hat b)\right)+\left(\hat L_g^N(\hat b)-\hat L_g^N(b^*)\right)+\hat L_g^N(b^*)-L_g^*(b^*).\tag{25}$$

Оскільки

$$\hat b=\arg\min_b \hat L_g^N(b),$$

а класифікатор побудований так, що

$$L_{\hat g}(\hat b)\le\hat L_g^N(\hat b),$$

то перші два доданки у (25) є від'ємними.

З (25) випливає:

$$0\le L_{\hat g}(\hat b)-L_g^*(b^*)\le\left|\hat L_g^N(b^*)-L_g^*(b^*)\right|\le\sup_b\left|\hat L_g^N(b)-L_g^*(b)\right|.\tag{26}$$

За лемою 2 математичне сподівання правої частини (26) прямує до 0 при $N\to\infty$. Теорема 3.1 доведена. □

Твердження наслідку випливає з теореми 1 та нерівності Чебишова.

4. Висновки

Отже, нами побудований квазибайєсів класифікатор на основі "лінійних індексів" $S(b)$ і знайдена консистентна оцінка $\hat{b}$ для напрямку b: ймовірність помилки класифікації при використанні $S(\hat{b})$ збігається до найменшої можливої. Подальші зусилля треба зосередити на дослідженні швидкості збіжності отриманих оцінок.

Література

1. В. Н. Вапник, А. Я. Червоненкис, *Теория распознавания образов*, "Мир", Москва, 1974.
2. Л. Деврой, Л. Дьерфи, *Непараметрическое оценивание плотности*, "Мир", Москва, 1988.
3. А. Афифи, С. Эйзен, *Статистический анализ: Подход с использованием ЭВМ*, "Мир", Москва, 1982.
4. Р. Є. Майборода, *Непараметрична статистика неоднорідних спостережень*, Докт. дис., Київ, 1994.
5. Р. Є. Майборода, *Статистичний аналіз сумішей. Курс лекцій*, "Київ. ун-т.", Київ, 2004.
6. Ю. О. Іванько, Р. Є. Майборода, *Експоненціальні оцінки емпірично-баєсового ризику при класифікації суміші зі змінними концентраціями*, Укр. мат. ж. **54** (2002), № 10, 1421–1428.
7. О. В. Сугакова, *Асимптотика ядерної оцінки щільності розподілу по спостереженнях суміші зі змінними концентраціями*, Теор. ймовірност. матем. статист. **59** (1998), 156–166.

Кафедра математики та теоретичної радіофізики, радіофізичний факультет, Національний університет ім. Тараса Шевченка, проспект Глушкова, 2, корп. 5, Київ 03127, Україна

Адреса електронної пошти: `sugak@univ.kiev.ua`

Надійшла 02/04/2004