УДК 519.21

О. Кубайчук, канд.фіз.-мат.наук
katrissa@voliacable.com

АСИМПТОТИКА ОЦІНКИ ДЛЯ БАЄСОВОГО ПОРОГУ

Розглянуто асимптотику оцінки для баєсового порогу, побудованої методом мінімізації емпіричного ризику і методом емпірично-баєсової класифікації для порогових класифікаторів для вибірки із суміші зі змінними концентраціями.

The asymptotic of the estimator for Bayesian border that is constructed by the Empirical Risk Minimization method and the method of the Empirical-Bayesian Classification for border classifiers for the sample from mixture with varying concentrations is considered.

1. Вступ

Розглядається задача класифікації деякого об'єкту O за спостереженням його числової характеристики $\xi = \xi(O)$. Вважаємо, що об'єкт може належати лише одному з двох класів. Розглядаємо порогові класифікатори вигляду

$$g^1_{t_1,t_2}(\xi) = \begin{cases} 1, \textit{при } \xi \in [t_1, t_2], \\ 2, \textit{при } \xi \notin [t_1, t_2], \end{cases} \qquad (27)$$

тобто об'єкт відносять до першого класу, якщо його характеристика потрапляє в проміжок $[t_1, t_2]$ і до другого класу в іншому випадку. Приклад такої класифікації – визначення людини (об'єкт) як хворої (другий клас), якщо її температура (характеристика ξ) перевищує 37° (або рівень гемоглобіну в крові перевищує 84 одиниці) (поріг t_2) і є меншою за 36° (рівень гемоглобіну є меншим за 72 одиниці) (поріг t_1). Можливий також варіант

$$g^2_{t_1,t_2}(\xi) = \begin{cases} 1, \textit{при } \xi \notin [t_1, t_2], \\ 2, \textit{при } \xi \in [t_1, t_2]. \end{cases} \qquad (28)$$

Найкращим (баєсовим) вважають поріг $\overline{t}^B = (t_1^B, t_2^B)$, при якому $g^i_{t_1,t_2}$, $i = 1,2$ має найменшу ймовірність помилки. Об'єкт O, у якого спостерігається деяка числова характеристика $\xi = \xi(O)$, може належати одному з двох класів; невідомий номер класу, якому належить O позначимо $ind(O)$. Вважаються відомими апріорні ймовірності $p_i = P(ind(O) = i)$, $i = 1,2$. Характеристика ξ – випадкова, її розподіл залежить від $ind(O)$: $P(\xi(O) < x \mid ind(O) = i) = H_i(x)$. Розподіли H_i невідомі, але будемо вважати, що вони мають неперервні щільності відносно міри Лебега – h_i.

2. Методи оцінки і основні результати

Множину всіх класифікаторів позначимо G. Ймовірність помилки класифікатора:

$$L(g^1_{\overline{t}}) = L^1(\overline{t}) = P\left\{g^1_{\overline{t}}(\xi(O)) \ne ind(O)\right\} = \sum_{i=1}^{2} P\left\{ind(O) = i\right\} P\left\{g^1_{\overline{t}}(\xi(O)) = 3 - i \mid ind(O) = i\right\} =$$

$$= p_1(H_1(t_1) + 1 - H_1(t_2)) + p_2(H_2(t_2) - H_2(t_1)),$$

$$L(g^2_{\overline{t}}) = L^2(\overline{t}) = P\left\{g^2_{\overline{t}}(\xi(O)) \ne ind(O)\right\} = \sum_{i=1}^{2} P\left\{ind(O) = i\right\} P\left\{g^2_{\overline{t}}(\xi(O)) = 3 - i \mid ind(O) = i\right\} =$$

$$= p_1(H_1(t_2) - H_1(t_1)) + p_2(H_2(t_1) + 1 - H_2(t_2)).$$

Баєсовим класифікатором у класі G називають класифікатор $g^B \in G$, на якому досягається мінімум $L(g)$:

$$g^B = \arg\min_{g \in G} L(g_{\overline{t}}).$$

Поріг $\overline{t}^B$ баєсового класифікатора є баєсовим порогом: $\overline{t}^B = \arg\min_{\overline{t} \in R^2} L(\overline{t})$.

Для $g^1_{\overline{t}}$ маємо: $\overline{t}^{1B} = \arg\min_{t_1 \in R, t_2 \in R} L^1(t_1, t_2) = (\arg\min_{t_1 \in R} L^1(t_1, t_2), \arg\min_{t_2 \in R} L^1(t_1, t_2)) = (\arg\min_{t_1 \in R} L^1_1(t_1), \arg\min_{t_2 \in R} L^1_2(t_2))$

де $L^1_1(t_1) = p_1 H_1(t_1) - p_2 H_2(t_1)$, $L^1_2(t_2) = p_1(1 - H_1(t_2)) + p_2 H_2(t_2)$.

Для $g^2_{\overline{t}}$ маємо: $\overline{t}^{2B} = \arg\min_{t_1 \in R, t_2 \in R} L^2(t_1, t_2) = (\arg\min_{t_1 \in R} L^2(t_1, t_2), \arg\min_{t_2 \in R} L^2(t_1, t_2)) = (\arg\min_{t_1 \in R} L^2_1(t_1), \arg\min_{t_2 \in R} L^2_2(t_2))$

де $L^2_1(t_1) = -p_1 H_1(t_1) + p_2 H_2(t_1)$, $L^2_2(t_2) = p_1 H_1(t_2) + p_2(1 - H_2(t_2))$.

Найкращим (баєсовим) вважають поріг $\overline{t}^B = (t_1^B, t_2^B)$, при якому g_{t_1,t_2} має найменшу ймовірність помилки. При цьому виникає проблема вибору (оцінки) порогу на основі навчаючої вибірки. Найбільш поширеними методами оцінювання $\overline{t}^B$ за повністю класифікованою вибіркою є емпірично-баєсова класифікація (ЕБК) [3;5] та метод мінімізації емпіричного ризику (МЕР) [2;7].

Розглянемо перший метод. Вважаємо, що навчаюча вибірка отримана із суміші зі змінними концентраціями. Будемо досліджувати асимптотичну поведінку цього методу.

Розглянемо випадок (27). Функції H_i (а, значить, і h_i) вважаються невідомими. Їх можна оцінити за даними, що являють собою вибірку із суміші зі змінними концентраціями: $\left\{\xi_{j:N}\right\}_{j=1}^{N}$, $\xi_{j:N}$ – незалежні між собою при фіксованому N і $P\left\{\xi_{j:N} < x\right\} = w_{j:N}H_1(x) + (1 - w_{j:N})H_2(x)$, де $w_{j:N}$ – відома концентрація об'єктів першого класу у суміші в момент j-го спостереження [6]. Для оцінки функції розподілу H_i використовують зважені емпіричні функції розподілу

$$\widehat{H}_i^N(x) = \frac{1}{N}\sum_{j=1}^{N} a_{j:N}^i 1\left\{\xi_j < x\right\},$$

де $1\{A\}$ – індикатор події A, $a_{j:N}^i$ – вагові коефіцієнти:

$$a_{j:N}^1 = \frac{1}{\Delta_N}\left((1 - S_N^1)w_{j:N} + (S_N^2 - S_N^1)\right),\quad a_{j:N}^2 = \frac{1}{\Delta_N}\left(S_N^2 - S_N^1 w_{j:N}\right),\quad S_N^k = \frac{1}{N}\sum_{j=1}^{N}(w_{j:N})^k,\quad \Delta_N = S_N^2 - (S_N^1)^2$$ (див. [6]).

Для оцінки щільностей розподілів h_i можна скористатися ядерними оцінками

$$\widehat{h}_i^N(x) = \frac{1}{Nk_N}\sum_{j=1}^{N} a_{j:N}^i K\left(\frac{x - \xi_{j:N}}{k_N}\right),$$

де K – ядро (щільність деякого ймовірнісного розподілу), k_N – параметр згладжування [1;4].

Оцінка ЕБК будується наступним чином: знаходиться множина T_N всіх розв'язків рівняння $p_1\widehat{h}_1^N(t) - p_2\widehat{h}_2^N(t) = 0$ і на роль оцінки використовується $\overline{\widehat{t}}_N^{EBC} = \arg\min_{t_1,t_2\in T_N} L_N^1(t_1,t_2)$, де

$$L_N^1(t_1,t_2) = p_1\left(\widehat{H}_1^N(t_1) + 1 - \widehat{H}_1^N(t_2)\right) + p_2\left(\widehat{H}_2^N(t_2) - \widehat{H}_2^N(t_1)\right) = L_{N_1}^1(t_1) + L_{N_2}^1(t_2),$$

$$L_{N_1}^1(t_1) = p_1\widehat{H}_1^N(t_1) - p_2\widehat{H}_2^N(t_1),\quad L_{N_2}^1(t_2) = p_1(1 - \widehat{H}_1^N(t_2)) + p_2\widehat{H}_2^N(t_2);$$

$$\widehat{t}_{N_1}^{EBC} = \arg\min_{t_1\in T_N} L_{N_1}^1(t_1),\quad \widehat{t}_{N_2}^{EBC} = \arg\min_{t_2\in T_N} L_{N_2}^1(t_2).$$

Будемо вважати, що виконуються наступні умови:

(А). $\overline{t}^B$ існує і є єдиною точкою глобального мінімуму $L^1(\overline{t})$ (t_1^B є точкою глобального мінімуму $L_1^1(t_1)$, t_2^B – $L_2^1(t_2)$).

(B_k). Існують границі $S^i = \lim_{n\to\infty} S_N^i$, $i = 1,2,...k$ і $\Delta = S^2 - (S^i)^2 > 0$.

Теорема 1. Нехай виконуються (А), (B_2), існують і є неперервними щільності h_i, $k_N \to 0$, $N_{k_N} \to \infty$, K – неперервна функція $d^2 \overset{def}{=} \int_{-\infty}^{\infty} K^2(t)dt < \infty$.

Тоді $\overline{\widehat{t}}_N^{EBC} \to \overline{t}^B$ ($\widehat{t}_{N_1}^{EBC} \to t_1^B$, $\widehat{t}_{N_2}^{EBC} \to t_2^B$) за ймовірністю при $N \to \infty$.

Доведення. Згідно з теоремою 1 у [1], в наших умовах, $\widehat{h}_i^N(x) \to h_i(x)$ за ймовірністю в кожній точці $x \in R$. Отже $u_N(x) \square\, p_2\widehat{h}_2^N(x) - p_1\widehat{h}_2^N(x) \to u(x) \square\, p_2h_2(x) - p_1h_1(x)$ за ймовірністю. Для $\delta_i > 0$ позначимо

$$A_N(\delta_i) = \left\{\exists t_i : \left|t_i - t_i^B\right| \le \delta_i,\, u_N(t_i) = 0\right\}.$$

Покажемо, що

$$P\left(A_N(\delta_i)\right) \to,\; N \to \infty. \tag{29}$$

Оскільки t_1^B – точка мінімуму $L_1^1(t)$, t_2^B – точка мінімуму $L_2^1(t)$, а $-L_1^1(t) = +L_2^1(t) = u(t)$ – неперервна функція, то $u(t)$ повинна змінювати знак в околі точки t_i^B, тобто існують такі t_i^-, t_i^+, що $t_i^B - \delta_i < t_i^- < t_i^B < t_i^+ < t_i^B + \delta_i$ і $u(t_i^-)u(t_i^+) < 0$. Отже, $P\left(u_N(t_i^-)u_N(t_i^+) < 0\right) \to 1$. Але, оскільки u_N – неперервна, то $\left\{u_N(t_i^-)u_N(t_i^+) < 0\right\} \subseteq A_N(\delta_i)$. Отже, (29) доведено для $i = 1,2$.

Фіксуємо δ_i, $i = 1,2$. Оскільки, L_1^1, L_2^1 – неперервні на R функції, $L_1^1(-\infty) = 0$, $L_1^1(+\infty) = p_1 - p_2$, $L_2^1(-\infty) = p_1$, $L_2^1(+\infty) = p_2$ і виконана умова (A), то $\forall\delta_i > 0\;\; \exists\varepsilon_i$ таке, що для всіх t_i, для яких $\left|t_i - t_i^B\right| > \delta_i$ має місце нерівність $L^1(t_i) > L^1(t_i^B) + \varepsilon_i$, $i = 1,2$. Виберемо $0 < \delta_i^{'} < \delta_i$ так, щоб для всіх $t \in \left[t_i^B - \delta_i^{'},\, t_i^B + \delta_i^{'}\right]$ виконувалось $L^1(t_i) < L^1(t_i^B) + \frac{\varepsilon}{4}$. Позначимо $B_{N_i} = \left\{\inf_{t\notin\left[t_i^B - \delta_i^{'},\, t_i^B + \delta_i^{'}\right]} L_{N_i}^1(t) > L^1(t_i^B) + \frac{\varepsilon}{2} > \inf_{t\in\left[t_i^B - \delta_i^{'},\, t_i^B + \delta_i^{'}\right]} L_{N_i}^1(t)\right\}$.

Фіксуємо довільне $\lambda_i > 0$. Використовуючи рівномірну збіжність L_N^1 до L^1, отримуємо, що для достатньо великих N, $P\left(B_{N_i}\right) > 1 - \frac{\lambda_i}{2}$. Згідно (29), при великих N, $P\left(A_N\left(\delta_i^{'}\right)\right) > 1 - \frac{\lambda_i}{2}$. Якщо виконано $A_N\left(\delta_i^{'}\right)$, то існують $t_i^* \in T_N \cap \left[t_i^B - \delta_i^{'}, t_i^B + \delta_i^{'}\right]$ і, при виконанні B_{N_i}, $L_{N_i}^1\left(t_i^*\right) < L_{N_i}^1\left(t_i\right)$ для всіх $t_i \notin \left[t_i^B - \delta_i, t_i^B + \delta_i\right]$. Тому

$$P\left\{\left|\hat{t}_i^{EBC} - t_i^B < \delta\right|\right\} \ge P\left(A_N\left(\delta_i^{'}\right) \cap B_{N_i}\right) \ge 1 - \frac{\lambda_i}{2} + 1 - \frac{\lambda_i}{2} = 1 - \lambda_i .$$

$$\left(P\left(A_N\left(\delta_i^{'}\right) \cap B_{N_i}\right) = P\left(A_N\left(\delta_i^{'}\right)\right) + P\left(B_{N_i}\right) - P\left(A_N\left(\delta_i^{'}\right) \cup B_{N_i}\right) \ge P\left(A_N\left(\delta_i^{'}\right)\right) + P\left(B_{N_i}\right) - 1 \right)$$

при великих N, $i = 1,2$. Враховуючи довільність λ_i, $i = 1,2$, отримаємо твердження теореми.

Далі оцінку порогу $\overline{t}^B$ будуємо методом мінімізації емпіричного ризику. Знову вважаємо, що навчаюча вибірка отримана із суміші зі змінними концентраціями. Припущення щодо оцінок для H_i і h_i такі ж як і раніше. Дослідимо асимптотичну поведінку цього методу. Оцінка МЕР визначається як $\hat{\overline{t}}_N^{MER} = \arg\min_{\overline{t} \in R^2} L^1{}_N(\overline{t})$, де

$$L^1{}_N(\overline{t}) = \left[p_1 \hat{H}_1^N(t_1) - p_2 \hat{H}_2^N(t_1)\right] + \left[p_1\left(1 - \hat{H}_1^N(t_2)\right) + p_2 \hat{H}_2^N(t_2)\right],$$

де $L^1{}_{N_1}(t_1) = p_1 \hat{H}_1^N(t_1) - p_2 \hat{H}_2^N(t_1)$, $\quad L^1{}_{N_2}(t_2) = p_1(1 - \hat{H}_1^N(t_2)) + p_2 \hat{H}_2^N(t_2)$.

Отже, $\hat{t}_{N_1}^{MER} = \arg\min_{t_1 \in R} L^1{}_{N_1}(t_1)$, $\quad \hat{t}_{N_2}^{MER} = \arg\min_{t_2 \in R} L^1{}_{N_2}(t_2)$. Будемо вважати, що виконуються наступні умови:

(A). $\overline{t}^B$ існує і є єдиною точкою глобального мінімуму $L^1(\overline{t})$.

(B_k). Існують границі $S^i = \lim_{n\to\infty} S_N^i$, $\quad i = 1,2,...k$ і $\Delta = S^2 - (S^i)^2 > 0$.

Теорема 2. Нехай виконуються (A), (B_2), H_i – неперервні функції на R. Тоді $\hat{\overline{t}}_N^{MER} \to \overline{t}^B$ ($\hat{t}_{N_1}^{MER} \to t_1^B$, $\hat{t}_{N_2}^{MER} \to t_2^B$) за ймовірністю при $N \to \infty$.

Доведення. Відмітимо, що з умови $\left(B_2\right)$ випливає рівномірна по N та j обмеженість вагових коефіцієнтів $a_{j:N}^i$. Тому за теоремою 2.4.2 з [3] $\sup_x \left|H_i^N\left(x\right) - H_i\left(x\right)\right| \to 0$ за ймовірністю при $N \to \infty$. Звідси випливає, що і $\sup_x \left|L_{N_1}^1\left(x\right) - L^1\left(x\right)\right| \to 0$ та $\sup_x \left|L_{N_2}^1\left(x\right) - L^1\left(x\right)\right| \to 0$ за ймовірністю. Фіксуємо довільні $\lambda_1, \lambda_2 > 0$, $\varepsilon_1, \varepsilon_2 > 0$. Нехай $A_{N_1} = \left\{\sup_x \left|L_{N_1}^1\left(x\right) - L^1\left(x\right)\right| < \frac{\varepsilon_1}{2}\right\}$, $A_{N_1} = \left\{\sup_x \left|L_{N_2}^1\left(x\right) - L^1\left(x\right)\right| < \frac{\varepsilon_2}{2}\right\}$.

При достатньо великих N, $\quad P\left(A_{N_1}\right) > 1 - \lambda_1$, $P\left(A_{N_2}\right) > 1 - \lambda_2$. Оскільки L_1^1, L_2^1 – неперервні функції на R, $L_1^1\left(-\infty\right) = 0$, $L_1^1\left(+\infty\right) = p_1 - p_2$, $L_2^1\left(-\infty\right) = p_1$, $L_2^1\left(+\infty\right) = p_2$ і виконана умова $\left(A\right)$, то $\forall \delta_i > 0$ $\exists \varepsilon_i$ таке, що для всіх t_i, для яких $\left|t_i - t_i^B\right| > \delta_i$ має місце нерівність $L^1\left(t_i\right) > L^1\left(t_i^B\right) + \varepsilon_i$, $i = 1,2$.

Нехай події A_{N_i} виконані . Тоді $\quad L^1\left(\hat{t}_{N_i}^{MER}\right) - \frac{\varepsilon_i}{2} \le L_{N_i}^1\left(\hat{t}_{N_i}^{MER}\right) \le L_{N_i}^1\left(t_i^B\right) \le L^1\left(t_i^B\right) + \frac{\varepsilon_i}{2}$, $i = 1,2$.

Отже, $L^1\left(\hat{t}_{N_i}^{MER}\right) \le L^1\left(t_i^B\right) + \varepsilon_i$ і $\left|\hat{t}_{N_i}^{MER} - t_i^B\right| \le \delta_i$, $i = 1,2$. Внаслідок довільності $\delta_1\left(\delta_2\right)$ та $\lambda_1\left(\lambda_2\right)$ теорема доведена.

Розглянемо випадок (28). В доведенні та формулюванні відбудуться незначні зміни, а саме, верхній індекс зміниться з 1 на 2, а також $L_1^1(-\infty) = 0$, $L_2^2(-\infty) = p_2$, $L_1^2(+\infty) = p_2 - p_1$, $L_2^2(+\infty) = p_1$, що не впливає на хід доведення .

3. Висновки

В даній роботі знайдено умови збіжності за ймовірністю оцінок для баєсового порогу, побудованих методом мінімізації емпіричного ризику і методом емпірично-баєсової класифікації для вибірки із суміші зі змінними концентраціями.

1. *Биллингсли П.* Сходимость вероятностных мер. – М., 1977. 2. Вапник В.Н. Индуктивные принципы поиска эмпирических закономерностей // Распознавание . Классификация. Прогноз, Вып.1. – М., 1989. 3. *Деврой Л., Дьерфи Л.* Непараметрическое оценивание плотности. – М., 1988. 4. *Іванько Ю.О.* Асимптотика ядерних оцінок щільностей та їх похідних, побудованих за спостереженнями із суміші зі змінними концентраціями // Вісник КНУ, сер. Математика. Механіка. – 2003. – №9. – С. 29–35. 5. *Іванько Ю.О., Майборода Р.Є.* Експоненціальні оцінки емпірично-баєсового ризику при класифікації суміші зі змінними концентраціями // Український математичний журнал. – 2002. – Т.54, №10. – С. 1421–1428. 6. *Майборода Р.Є.* Статистичний аналіз сумішей. – К., 2003. 7. *Vapnik V.N.* The nature of Statistical Learning Theory. – N. Y., 1996.

Надійшла до редколегії 27.09.07