ДОПОВІДІ
НАЦІОНАЛЬНОЇ
АКАДЕМІЇ НАУК
УКРАЇНИ

6 • 2016

http://dx.doi.org/10.15407/dopovidi2016.06.025

УДК 519.7

О. А. Галкін

Київський національний університет ім. Тараса Шевченка

E-mail: galkin.o.a@gmail.com

Асимптотичні властивості Σ-класифікатора для багатокласових задач розпізнавання з нееліптичним розподілом даних

(Представлено членом-кореспондентом НАН України А. В. Анісімовим)

Досліджуються асимптотичні властивості Σ-класифікатора, що не вимагає попередньої інформації про розподіл або форму розділової кривої. Побудовано математичний апарат для розв'язання багатокласових задач розпізнавання з нееліптичним розподілом даних на основі методу мажоритарного голосування. Досліджено процедуру визначення форм розділових кривих Σ-класифікатора по геометричній структурі даних, що лежать в основі Σ-схеми. Визначено умови, при яких Σ-класифікатор є асимптотично еквівалентним правилу Байеса.

Ключові слова: правило Байеса, Σ-класифікатор, асимптотична збіжність.

Постановка задачі. Дана робота присвячена дослідженню результатів узгодженості Σ-класифікатора, який описано в [1], а також визначенню умов, при яких запропоноване правило класифікації буде асимптотично еквівалентним правилу Байєса. Враховуючи той факт, що досліджувані функції глибини є обмеженими, будемо вважати, що вони обмежені. Визначимо $V_\tau\ (r_1, r_2) = 1$, якщо $r_2 > d_\tau(r_1)$ та $V_\tau(r_1, r_2) = 0$, якщо $r_2 \leqslant d_\tau(r_1)$ при $r_1, r_2 \in [0, 1]$ та $\tau \in \mathbb{R}^{\Delta_0}$. Вирази $V_\tau(E_H(z), E_U(z))$ та $V_{\overline{\tau}_M}(E_{H_n}(z), E_{U_m}(z))$ відповідають Σ-класифікатору та його емпіричному аналогу, що має таку форму при $\overline{\tau}_M = \arg\min\{\overline{\Psi}_M(\tau)\}$: а) якщо $E_{U_m}(z) > d_{\overline{\tau}_M}(E_{H_n}(z))$, тоді $z \in U$; б) якщо $E_{U_m}(z) < d_{\overline{\tau}_M}(E_{H_n}(z))$, тоді $z \in H$.

Коефіцієнти помилкової класифікації для $V_\tau(E_H(z), E_U(z))$ та $V_{\overline{\tau}_M}(E_{H_n}(z), E_{U_m}(z))$ можна визначити таким чином:

$$\Psi(V_\tau) = p_1 P_H\{x\colon V_\tau(E_H(x), E_U(x)) = 1\} + p_2 P_U\{x\colon V_\tau(E_H(x), E_U(x)) = 0\} \tag{1}$$

та

$$\Psi_M(\overline{V}_M) = p_1 P_H\{x\colon V_{\overline{\tau}_M}(E_{H_n}(x), E_{U_m}(x)) = 1\} + \\ + p_2 P_U\{x\colon V_{\overline{\tau}_M}(E_{H_n}(x), E_{U_m}(x)) = 0\}, \qquad (2)$$

де $\Psi_M(\overline{V}_M)$ є умовною ймовірністю помилкової класифікації, коли $E_{U_m}(X)$ та $d_{\overline{\tau}_M}(E_{H_n}(X))$ використовуються для класифікації елемента X.

Властивості Σ-класифікатора. Розглянемо функцію $V_\omega(r_1, r_2) = \omega(r_1)$, де змінну ω будемо використовувати у якості показника скінченного об'єднання інтервалів [2]. Після застосування даної функції до $(E_H(x), E_U(x))$, класифікація точки x відбуватиметься виключно на основі її глибини відносно H. Крім того, пов'язана із функцією ω помилка невірної класифікації може бути визначена таким чином:

$$\Psi(V_\omega) = p_1 P_H\{x\colon V_\omega(E_H(x), E_U(x)) = 1\} + p_2 P_U\{x\colon V_\omega(E_H(x), E_U(x)) = 0\}. \qquad (3)$$

Далі припустимо, що для $a \in \mathbb{R}$ та $\forall \tau \in \mathbb{R}^{\Delta_0}$

$$P_H\{x\colon E_U(x) = d_\tau(E_H(x))\} = P_U\{x\colon E_U(x) = d_\tau(E_H(x))\} = 0, \qquad (4)$$

$$P_H\{x\colon E_H(x) = a\} = P_U\{x\colon E_H(x) = a\} = 0. \qquad (5)$$

Твердження 1 . *$\exists V_0 \in G$, що $\Psi(V_0) = \inf\limits_{V\in G} \Psi(V)$ при умові, якщо H та U задовольняють* (4) *та* (5).

Теорема 1. *Нехай $E_H(\cdot)$ та $E_U(\cdot)$ задовольняють асимптотичній збіжності $\sup\limits_z |E_{H_n}(z) - E_H(z)| \to 0$ та $\sup\limits_z |E_{U_m}(z) - E_U(z)| \to 0$ при $\min(n, m) \to \infty$, а також є неперервними. Крім того припустимо, що H та U задовольняють* (4) *та* (5). *Тоді $\overline{\Psi}_M(V)$ асимптотично сходиться до $\Psi(V)$при$\min(n, m) \to \infty$для$\forall\ V \in G$.*

Доведення. Розглядаючи майже напевну збіжність послідовностей, ми використаємо теорему Скорохода в сенсі розподільної збіжності для побудови відповідного представлення емпіричних розподілів [3]. Припустимо, що випадкові вибірки $\{Z_n\}$ та $\{X_m\}$ визначені на ймовірнісному просторі (E, η, ε), а випадковість введених величин залежить від $e \in E$. Таким чином, ми приймаємо $\{Z_n(e)\}$ та $\{X_m(e)\}$ для послідовностей, а такожH_n^e та U_m^e для емпіричних розподілів, що базуються на $\{Z_1(e), \ldots, Z_n(e)\}$ та $\{X_1(e), \ldots, X_m(e)\}$ відповідно. Для спрощення запису введемо позначення $\overline{\Psi}_M^e$ та $\overline{V}_M^e$.

Можна стверджувати, що $\exists$ множина $E_0 \in \eta$, яка $\varepsilon(E_0) = 1$ та для $\forall \varepsilon \in E_0$ послідовності функцій розподілу $\{H_n^e\}$ та $\{U_m^e\}$ сходяться до H та U відповідно. Даний висновок слідує з r-вимірної теореми Гливенко–Кантеллі, оскільки $\min(n, m) \to \infty$. Тоді $\exists$ такий ймовірнісний простір $(\mathrm{I}, \eta_\mathrm{I}, \varepsilon_\mathrm{I})$, що якщо $e \in E_0$, тоді $\exists$ такі послідовності випадкових величин $\{X_n^{e,H}\}_{n\geqslant 0}$ та $\{X_m^{e,U}\}_{m\geqslant 0}$, що

а) розподілом $X_n^{e,H}$ є H_n^e, а розподілом $X_m^{e,U}$–U_m^e для $\forall n$, $m = 1, \ldots$. Також розподілом $X_0^{e,H}$ є H, а розподілом $X_0^{e,U}$–U.

б) послідовності $\{X_n^{e,H}\}_{n\geqslant 1}$ та $\{X_m^{e,U}\}_{m\geqslant 1}$ сходяться майже напевно до випадкових величин $X_0^{e,H}$ та $X_0^{e,U}$ відповідно.

Наступні результати мають місце при застосуванні доданих послідовностей теореми Скорохода.

Зазначимо, що множина $E^* = E_0 \bigcap E_1$ має ймовірність 1, оскільки E_1 є множиною одиничної ймовірності, в якій збіжності $\sup\limits_z |E_{H_n}(z) - E_H(z)| \to 0$ та $\sup\limits_z |E_{U_m}(z) - E_U(z)| \to 0$ задоволені. Отже ми маємо

$$\begin{aligned} \overline{\Psi}^e_M(V) &= p_1 P_{H^e_n}\{z : V(E_{H^e_n}(z), E_{U^e_m}(z)) = 1\} + p_2 P_{U^e_m}\{z : V(E_{H^e_n}(z), E_{U^e_m}(z)) = 0\} = \\ &= p_1 \varepsilon_{\mathrm{I}}\{V(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n)) = 1\} + p_2 \varepsilon_{\mathrm{I}}\{V(E_{H^e_n}(X^{e,U}_m), E_{U^e_m}(X^{e,U}_m)) = 0\} = \\ &= W^e_n + Q^e_m. \end{aligned} \tag{6}$$

де $e \in E^*$.

Крім того, має місце така рівність:

$$\begin{aligned} |E_{H^e_n}(X^{e,H}_n) - E_H(X^{e,H}_0)| &= |E_{H^e_n}(X^{e,H}_n) - E_H(X^{e,H}_n)| + |E_H(X^{e,H}_n) - E_H(X^{e,H}_0)| = \\ &= W^e_{n,1} + W^e_{n,2}. \end{aligned} \tag{7}$$

Оскільки точка e розглядається як така, що належить до E_1, то $\{W^e_{n,1}\}_n \to 0$. Однак, оскільки $e \in E_0$, $\varepsilon_{\mathrm{I}} \overset{\text{ac.}}{\to} X^{e,H}_0$, де $X^{e,H}_n$ є випадковими величинами, що визначені на I.

Отже, $W^e_{n,2}\varepsilon_{\mathrm{I}} \overset{\text{ac.}}{\to} 0$, оскільки E_H є неперервною функцією. Використовуючи аналогічний підхід до $E_{U^e_m}(X^{e,H}_n)$, ми отримуємо

$$(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n)) \overset{\text{ac.}}{\to} (E_H(X^{e,H}_0), E_U(X^{e,H}_0))$$

відносно ймовірності ε_{I}. У даному випадку межею множини $\{(t,b)\colon V(t,b) = 0\}$ є $\{(t,b)\colon t = d(b)\}$, коли $V = V_d$ (d є многочленом) та множиною вигляду $\{\{a_i\} \times [0,1]\colon i = 1, \ldots, f+1\}$, коли V є показником об'єднання f інтервалів.

Тому, як слідує з (4) та (5),

$$W^e_n \to p_1 \varepsilon_{\mathrm{I}}\{V(E_H(X^{e,H}_0), E_U(X^{e,H}_0)) = 1\} = p_1 P_H(z\colon V(E_H(z), E_U(z)) = 1\}, \tag{8}$$

оскільки розподілом $X^{e,H}_0$ є H. Отже, результат має місце, оскільки (8) виконується для $\forall e \in E^*$, що є множиною одиничної ймовірності. Теорему доведено.

Далі наведено лему 1, з якої слідує, що $\overline{V}_M \overset{\text{ac.}}{\to} V_0$, де $\overline{V}_M$ мінімізує емпіричний коефіцієнт помилкової класифікації, а V_0 — множинний коефіцієнт помилкової класифікації.

Лема 1. *Нехай* $\overline{V}_M = \arg\min\limits_{V \in G}\{\overline{\Psi}_M(V)\}$ *та* $V_0 = \arg\min\limits_{V \in G}\{\Psi(V)\}$. *Крім того, нехай асимптотична збіжність розглядається, як*

$$P_H\{x\colon \overline{V}_M(E_H(x), E_U(x)) \to V_0(E_H(x), E_U(x))\} = 1 \tag{9}$$

та

$$P_U\{x\colon \overline{V}_M(E_H(x), E_U(x)) \to V_0(E_H(x), E_U(x))\} = 1. \tag{10}$$

Тоді, якщо V_0 *є унікальним, то* $\overline{V}_M \overset{\text{ac.}}{\to} V_0$ *при* $\min(n,m) \to \infty$.

Доведення. Як слідує з теореми 1, $\exists$ така множина одиничної ймовірності E^*, що $\overline{\Psi}^e_M(V_0) \overset{\text{ac.}}{\to} \Psi(V_0)$ для $\forall e \in E^*$. Крім того, E^* містить множину E_0, в якій виконується асимптотична збіжність $\sup\limits_z |E_{H_n}(z) - E_H(z)| \to 0$ та $\sup\limits_z |E_{U_m}(z) - E_U(z)| \to 0$, а також має місце теорема Скорохода [4].

Розглянемо послідовність $\{\overline{V}^e_M\}_M$, де $e \in E^*$ є фіксованою точкою. Для доведення даної леми необхідно показати, що вся послідовність сходиться до E_0. Дана збіжність буде мати місце, якщо кожна підпослідовність $\{\overline{V}^e_M\}_M$ буде містити додаткову підпослідовність, що сходиться до E_0. Отже, розглянемо підпослідовність $\{\overline{V}^e_M\}_M$. Зазначимо, що дана підпослідовність містить асимптотично збіжну підпослідовність. Позначимо границю збіжної підпослідовності послідовності $\{\overline{V}^e_M\}_M$, як V^e.

Розглядаючи послідовність $\{\overline{\Psi}^e_M(\overline{V}^e_M)\}_M$, ми маємо

$$\overline{\Psi}^e_M(\overline{V}^e_M) = p_1\varepsilon_{\mathrm{I}}\{\overline{V}^e_M(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n)) = 1\} + \\ + p_2\varepsilon_{\mathrm{I}}\{\overline{V}^e_M(E_{H^e_n}(X^{e,U}_m), E_{U^e_m}(X^{e,U}_m)) = 0\}, \tag{11}$$

як і в доведенні теореми 1.

Аналогічним чином ми отримуємо, що

$$(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n)) \overset{\text{ас.}}{\to} (E_H(X^{e,H}_0), E_U(X^{e,H}_0)). \tag{12}$$

Зазначимо, що збіжність $\overline{V}^e_M \to V^e$ означає, що $\overline{V}^e_M(E_H(z), E_U(z)) = 1$ при умові, якщо $V^e(E_H(z), E_U(z)) = 1$, а $(E_H(z), E_U(z))$ не належить межі множини $\{V^e = 1\}$.

Розглядаючи асимптотичну збіжність відносно ймовірності ε_{I}, ми отримуємо такі асимптотичні збіжності:

$$\Lambda_{\{\overline{V}^e_M(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n))=1\}} \to \Lambda_{\{\overline{V}^e(E_H(X^{e,H}_0), E_U(X^{e,H}_0))=1\}} \tag{13}$$

та

$$\Lambda_{\{\overline{V}^e_M(E_{H^e_n}(X^{e,H}_n), E_{U^e_m}(X^{e,H}_n))=0\}} \to \Lambda_{\{\overline{V}^e(E_H(X^{e,H}_0), E_U(X^{e,H}_0))=0\}}, \tag{14}$$

оскільки ймовірність того, що $X^{e,H}_0$ належить межі $\{V^e = 1\}$ дорівнює нулю.

З рівності (11) та теореми Лебега про мажоровану збіжність слідує, що $\overline{\Psi}^e_M(\overline{V}^e_M) \to \to \Psi(V^e)$ для точки e, оскільки всі випадкові величини є додатними та обмеженими 1 [5]. Крім того, в даній точці e має місце збіжність $\overline{\Psi}^e_M(V_0) \overset{\text{ас.}}{\to} \Psi(V_0)$, та як $e \in E^*$.

Таким чином, $\Psi(V^e) = \Psi(V_0)$, оскільки

$$\Psi(V_0) = \lim \overline{\Psi}^e_M(V_0) \geqslant \lim \overline{\Psi}^e_M(\overline{V}^e_M) = \Psi(V^e) \geqslant \Psi(V_0), \tag{15}$$

що слідує з визначення V_0 та $\overline{V}^e_M$. Отже, $V^e = V_0$, враховуючи унікальність V_0. Лему доведено.

Далі наведемо основний результат роботи.

Теорема 2. *Припустимо, що* $p_1 = p_2$, *а функції щільності* $h_1(\cdot)$ *та* $h_2(\cdot)$ *з* H *та* U *відповідно мають вигляд* $h_i(z) = v_i|\Xi_i|^{-1/2} f_i((z-\varepsilon_i)'\Xi_i^{-1}(z-\varepsilon_i))$ *з* $h_1(\cdot) = h_2(\cdot)$ *та* $\Xi_1 = \Xi_2$, *де* $i = 1, 2$. *Якщо* $\tau_1 = (1, 0, \ldots, 0)$, *а функція глибини, що використовується в алгоритмі класифікації є глибиною Махаланобіса, напівпросторовою глибиною, симпліціальною глибиною, або проекційною глибиною, тоді ми маємо, що* $\Omega(\Psi_M(\overline{V}_M)) \to \Psi(V_{\tau_1})$ *при* $\min(n, m) \to \infty$.

Доведення. З леми 1 слідує, що $\overline{V}_M \overset{\text{ас.}}{\to} V_{\tau_1}$ при $\min(n, m) \to \infty$. Крім того, $V_0 = V_{\tau_1}$.

Зазначимо також, що

$$|\Psi_M(\overline{V}_M) - \Psi(V_{\tau_1})| \leqslant p_1 \int |\Lambda_{\{\overline{V}_M(E_{H_n}(x), E_{U_m}(x))=1\}} - \Lambda_{\{V_{\tau_1}(E_H(x), E_U(x))=1\}}| h_1(x)\, dx +$$

$$+ p_2 \int |\Lambda_{\{\overline{V}_M(E_{H_n}(x), E_{U_m}(x))=0\}} - \Lambda_{\{V_{\tau_1}(E_H(x), E_U(x))=0\}}| h_2(x)\, dx. \tag{16}$$

З теореми Лебега про мажоровану збіжність маємо

$$|\Psi_M(\overline{V}_M) - \Psi(V_{\tau_1})| \overset{\text{ac.}}{\to} 0 \tag{17}$$

при $\min(n, m) \to \infty$. Даний результат має місце при об'єднанні емпіричних функцій глибини з множинними функціями глибини з майже напевною поточковою збіжністю. Остаточний результат маємо при повторному застосуванні теореми Лебега про мажоровану збіжність. Теорему доведено.

Враховуючи припущення теореми 2, можна стверджувати, що правило Байєса еквівалентно такій схемі: 1) якщо $E_U(z) > E_H(z)$, тоді z належатиме U; 2) якщо $E_U(z) < E_H(z)$, тоді z належатиме H. Таким чином, $\Psi(V_{\tau_1})$ відповідає байєсівському ризику. Отже, на основі заданих припущень можна стверджувати, що запропонований Σ-класифікатор еквівалентний правилу Байєса.

Цитована література

1. *Галкін О. А.* Непараметричний метод класифікації для задач з нееліптичним розподілом даних на основі глибиннозалежної Σ-схеми // Вісн. Київ. нац. ун-ту ім. Тараса Шевченка. Сер. фіз.-мат. науки. – 2015. – № 3. – С. 60–65.
2. *Lange T., Mosler K, Mozharovskyi P.* Fast nonparametric classification based on data depth // Statist. Papers. – 2014. – **55**. – P. 53–67.
3. *Cuesta-Albertos J. A., Nieto-Reyes A.* The random Tukey depth // Computational Statistics & Data Analysis. – 2008. – **52**. – P. 4980–4987.
4. *Li J., ZuoR. Y.* New nonparametric tests of multivariate locations and scales using data depth // Statistical Sci. – 2004. – **19**. – P. 687–694.
5. *Zuo Y. J.* Projection-based depth functions and associated medians // The Annals of Statistics. – 2003. – **31**. – P. 1463–1484.

References

1. *Galkin O.* A. Bull. of Taras Shevchenko National University of Kiev. Ser. Phys. & Math., 2015, No 3: 60–65 (in Ukrainian).
2. *Lange T., Mosler K, Mozharovskyi P.* Statist. Papers, 2014, **55**: 53–67.
3. *Cuesta-Albertos J. A., Nieto-Reyes A.* Computational Statistics & Data Analysis, 2008, **52**: 4980–4987.
4. *Li J., Zuo R. Y.* Statistical Sci., 2004, **19**: 687–694.
5. *Zuo Y. J.* The Annals of Statistics, 2003, **31**: 1463–1484.

Надійшло до редакції 14.09.2015

А. А. Галкин

Киевский национальный университет им. Тараса Шевченко

E-mail: galkin.o.a@gmail.com

Асимптотические свойства Σ-классификатора для многоклассовых задач распознавания с неэллиптическим распределением данных

Исследуются асимптотические свойства Σ-классификатора, который не требует предварительной информации о распределении или форме разделительной кривой. Построен математический аппарат для решения многоклассовых задач распознавания с неэллиптиче-

ским распределением данных на основе метода мажоритарного голосования. Исследована процедура определения форм разделительных кривых Σ-классификатора по геометрической структуре данных, лежащих в основе Σ-схемы. Определены условия, при которых Σ-классификатор является асимптотически эквивалентным правилу Байеса.

Ключевые слова: правило Байеса, Σ-классификатор, асимптотическая сходимость.

O. A. Galkin

Taras Shevchenko National University of Kiev

E-mail: galkin.o.a@gmail.com

The asymptotic properties of a Σ-classifier for multiclass recognition problems with non-elliptic distribution of data

The asymptotic properties of a Σ-classifier that does not require a priori information about the distribution or shape of a dividing curve are studied. A mathematical apparatus is built to solve multiclass recognition problems with a non-elliptic distribution of data on the basis of the majority voting method. The procedure is studied to determine the shapes of dividing curves of a Σ-classifier by the geometric structure of data that are the basis of the Σ-scheme. The conditions, under which the Σ-classifier is asymptotically equivalent to the Bayes rule, are defined.

Keywords: Bayes rule, Σ-classifier, asymptotic convergence.