Эффективность решающего правила максимального правдоподобия при классификации многомерных наблюдений в условиях неравновероятных классов

Жук Е.Е.; Шумилин С.Н.
ММАД, ФПМИ
БГУ
г. Минск, Республика Беларусь
e-mail: zhukee@mail.ru

***Аннотация* — Рассматривается задача статистической классификации в случае, когда информация об априорных вероятностях классов частично или полностью отсутствует.**

Ключевые слова: байесовское решающее правило; априорные вероятности классов; модель Фишера

I. Введение

На практике при решении задачи статистической классификации зачастую априорные вероятности классов адекватно оценить невозможно, в этом случае вместо оптимального (в смысле риска) байесовского решающего правила используется решающее правило максимального правдоподобия.

Аналитически исследовано, насколько решающее правило максимального правдоподобия хуже по риску, чем байесовское решающее правило.

II. Статистическая классификация и проблема задания априорных вероятностей классов

Пусть в пространстве R^N $(N \geq 1)$ регистрируются случайные наблюдения $x \in R^N$ из $L \geq 2$ классов $\{\Omega_1, \ldots, \Omega_L\}$. Наблюдение x принадлежит к классу со случайным номером $d^o \in S$ $(S = \{1, \ldots, L\}$ – множество номеров классов):

$$\mathrm{P}\{d^o = i\} = \pi_i^o > 0,\ i \in S;\ \sum_{i \in S} \pi_i^o = 1, \qquad (1)$$

где $\{\pi_i^o\}_{i \in S}$ – так называемые априорные вероятности классов $\{\Omega_i\}_{i \in S}$ [1]. При фиксированном номере $d^o = i$ наблюдения из класса Ω_i описываются условной плотностью распределения вероятностей [1]:

$$p_i(x) \geq 0,\ x \in R^N:\quad \int_{R^N} p_i(x)dx = 1,\ i \in S, \qquad (2)$$

где плотности $\{p_i(\cdot)\}_{i \in S}$ в (2) могут быть параметризованы [1]: $p_i(x) = p(x; \theta_i^o)$, $\theta_i^o \in \Theta \subseteq R^m$ $(m \geq 1)$, $i \in S$.

Задача статистической классификации заключается в построении решающего правила (РП) [1]: $d = d(x)\colon R^N \to S$, являющегося статистической оценкой для d^o и относящего наблюдение x к одному из классов $\{\Omega_i\}_{i \in S}$ $(d \in S)$.

В качестве критерия эффективности РП $d = d(x) \in S$, $x \in R^N$, используется риск, имеющий смысл средних (ожидаемых) потерь при классификации [1]: $r = r(d) = \mathrm{E}\{w_{d^o, d(x)}\}$, где $W = (w_{ij})_{i,j \in S}$ – матрица потерь. Далее будем рассматривать случай классификации данных с $(0-1)$-матрицей потерь $W = (w_{ij})_{i,j \in S}$, $w_{ij} = 1 - \delta_{ij}$. Известно [1], что наименьшие потери гарантирует байесовское решающее правило (БРП):

$$d_o(x) = d(x; \pi^o) = \arg\max_{i \in S}\{\pi_i p_i(x)\},\quad x \in R^N, \qquad (3)$$

обладающее минимально возможным значением риска:

$$r_o = r(d_o) = P\{d(x; \pi^0) \neq d^0\} = \int_{R^N} \max_{i \in S}\{\pi_i p_i(x)\}dx. \qquad (4)$$

Пусть у нас есть предполагаемые значения априорных вероятностей $\{\pi_i\}_{i \in S}$:

$$\pi_i = \pi_i^o + \varepsilon_i,\ i \in S;\ \sum_{i \in S} \pi_i = 1, \qquad (5)$$

где $\{\varepsilon_i\}_{i \in S}$ – их отклонения от соответствующих истинных значений из (1):

$$-\pi_i^o < \varepsilon_i < 1 - \pi_i^o,\ i \in S;\ \sum_{i \in S} \varepsilon_i = 0. \qquad (6)$$

Степень этих отклонений будем характеризовать величиной:

$$\varepsilon_+ = \max_{i \in S} |\varepsilon_i|, \qquad (7)$$

называемой уровнем отклонения [2].

III. Исследование риска классификации при неточно заданных априорных вероятностях классов

Отметим, что БРП (3) в условиях равновероятности классов ($\pi_i = 1/L; i \in S$) есть решающее правило максимального правдоподобия (РПМП):

$$d_0^+(x) = \arg\max_{i \in S} p_i(x),\ x \in R^N. \qquad (8)$$

Пусть в БРП $d(\cdot; \pi^o)$ из (3) вместо истинных значений априорных вероятностей $\pi^o = (\pi_1^o, \ldots, \pi_L^o)^{\mathrm{T}}$ из (1) используются отличные от них априорные вероятности $\pi = (\pi_1, \ldots, \pi_L)^{\mathrm{T}}$, определяемые (5), (6):

$$d_*(x) = d(x; \pi) = \arg\max_{i \in S}\{\pi_i p_i(x)\},\ x \in R^N. \qquad (9)$$

Введем в рассмотрение следующие риски классификации: $r_* = R(\pi, \pi^0)$ - риск решающего правила (9):

$$r_* = 1 - \sum_{i \in S} \pi_i^o \int_{R^N} \prod_{\substack{k \in S \\ k \neq i}} U\big(p_i(x) - p_k(x)\big) p_i(x)dx; \qquad (10)$$

причем если $\pi_i = 1/L, i \in S$, то $d_* = d_*(x)$ совпадает с РПМП из (11);

r_0^+ - риск БРП из (3) при $\pi_i^0 = 1/L, i \in S$:

$$r_0^+ = 1 - \frac{1}{L}\sum_{i\in S}\int_{R^N}\prod_{\substack{k\in S\\k\neq i}} U\big(p_i(x) - p_k(x)\big)p_i(x)dx \tag{11}$$

Теперь в качестве предполагаемых значений априорных вероятностей классов из (5) возьмем следующие:

$$\pi = \{\pi_i\}_{i\in S};\ \pi_i = \frac{1}{L}, i \in S, \tag{12}$$

и рассмотрим решающее правило (9).

Получена формула точного представления риска РП (9), с вектором предполагаемых значений априорных вероятностей из (12), через риск БРП, совпадающего с РПМП (8) ($r_* = r_0^+ + O(\varepsilon)$):

$$r_* = r_0^+ + \sum_{i\in S}\varepsilon_i\int_{R^N}\prod_{\substack{k\in S\\k\neq i}} U\big(p_i(x) - p_k(x)\big)p_i(x)dx \tag{13}$$

Условным риском классификации решающего правила $d(\cdot)$ в классе с номером i назовем следующую величину:

$$r_i = 1 - P\{d(x) = i \mid d^0 = i\}. \tag{14}$$

Предположим, что $r_i = r_j = r_., \forall i, j \in S$, тогда учитывая это в (13):

$$r_* = r_0^+ + (1 - r_.)\sum_{i\in S}\varepsilon_i = r_0^+. \tag{15}$$

Т. е. значение риска правила $d_* = d_*(x)$ не зависит от предполагаемых значений $\{\pi_i\}_{i\in S}$ априорных вероятностей из (14).

IV. Слуйчай двух классов. Модель Фишера

Рассмотрим часто встречающийся на практике случай двух классов (L=2), в частности модель Фишера [1], когда условные плотности $\{p_i(\cdot)\}_{i\in S}$ из (2), описывающие классы $\{\Omega_i\}_{i\in S}$, многомерные гауссовские, с общей для обоих классов невырожденной ковариационной матрицей.

Для модели Фишера в случае двух классов байесовский риск легко вычисляется аналитически [2]:

$$r_o = \pi_1^o\Phi\left(-\frac{\Delta}{2} - \frac{h}{\Delta}\right) + (1 - \pi_1^o)\Phi\left(-\frac{\Delta}{2} + \frac{h}{\Delta}\right), \tag{17}$$

$$h = \ln(H) = \ln\left(\frac{\pi_1^o}{1 - \pi_1^o}\right),$$

где $\Phi(z) = \int_{-\infty}^{z}\varphi(w)dw$, $z\in R$ – функция распределения вероятностей стандартного нормального закона с плотностью $\varphi(w) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{w^2}{2}\right)$, $w \in R$, а

$$\Delta = \sqrt{(\mu_1 - \mu_2)^T\Sigma^{-1}(\mu_1 - \mu_2)} \tag{18}$$

– межклассовое расстояние Махаланобиса [1]. Очевидно, что значение риска (11) $r_0^+ = \Phi\left(-\frac{\Delta}{2}\right)$.

Известно следующее соотношение для риска РП (9) в случае модели Фишера[2]:

$$r_* = r_o + \frac{\varphi\left(\frac{\Delta}{2} + \frac{h}{\Delta}\right)}{2\pi_1^o(1 - \pi_1^o)^2\Delta}\varepsilon^2 + O(|\varepsilon|^3)\cdot \tag{19}$$

Выражение для риска (19) позволяет на практике по заданному $\delta > 0$ определить так называемый δ-допустимый уровень отклонений [2]:

$$\varepsilon_+^* = \varepsilon_+^*(\delta) = (1 - \pi_1^o)\sqrt{\frac{2\pi_1^o\Delta\delta}{\varphi\left(\frac{\Delta}{2} + \frac{h}{\Delta}\right)}}, \tag{20}$$

гарантирующий, что при уровне отклонений, не превышающем ε_+^*: $\varepsilon_+ = |\varepsilon| \le \varepsilon_+^*$, приращение риска r_* относительно оптимального значения r_o не превзойдет δ: $r_* - r_o \le \delta$.

Положим далее, что значения априорных вероятностей классов совпадают и равны:

$$\pi_1 = \pi_1^o + \varepsilon = \frac{1}{2}, \pi_2 = \pi_2^o - \varepsilon = 1 - \pi_1^o - \varepsilon = \frac{1}{2},$$

(21) $\pi_1^0 = \pi_1 - \varepsilon = \frac{1}{2} - \varepsilon$.

Получено представление остатка r_* (10) через r_0 (11):

$$r_* = r_0 + \frac{1}{\sqrt{2\pi}}\frac{4e^{-\Delta^2/8}}{\Delta}\varepsilon_+^2 + o(\varepsilon_+^3) \tag{22}$$

Тогда, если

$$\delta \approx \frac{1}{\sqrt{2\pi}}\frac{4e^{-\Delta^2/8}}{\Delta}\varepsilon_+^2 \Rightarrow \varepsilon_+ = \sqrt{\frac{\delta\Delta\sqrt{2\pi}}{4e^{-\Delta^2/8}}}\cdot \tag{23}$$

Отметим также, что риск r_* из (10) в случае модели Фишера равен

$$r_* = \Phi\left(-\frac{\Delta}{2}\right)\cdot$$

Т. о., пользуясь РПМП в случае модели Фишера, при условии равенства всех вероятностей классов из (1), мы ничего не выигрываем в плане риска классификации. Также на точность классификации отклонения (6) от истинных значений априорных вероятностей не влияют при больших межклассовых расстояниях Δ из (18).

[1] Харин Ю. С., Жук Е. Е. Математическая и прикладная статистика. Мн., 2005.

[2] Жук Е. Е. Устойчивость байесовского решающего правила при неточно заданных априорных вероятностях классов //Вестн. Белорус. ун-та. Сер. 1. Физ. Мат. Информ. 2010. № 4. С. 14-20.