Математика и информатика

Е.Е. ЖУК, Е.В. СЕРИКОВА

СТАТИСТИЧЕСКАЯ КЛАССИФИКАЦИЯ ПРИ ПРЯМОМ ВЫБОРЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ И ЕЕ РИСК

The problem of statistical classification of multivariate normal (Gaussian) observations in the subspace of informative features is studied. The iterative step by step method for noninformative feature rejection is used and efficiency of transition to the selected features is analytically investigated.

Пусть в пространстве R^N зарегистрировано случайное наблюдение $x \in R^N$, принадлежащее к одному из $L \geq 2$ классов $\{\Omega_1, \ldots, \Omega_L\}$. Введем в рассмотрение дискретную случайную величину d^0 со значениями из множества номеров классов $S=\{1, \ldots, L\}$: $d^0 \in S$ и распределением вероятностей:

$$P\{d^0 = i\} = \pi_i > 0,\ i \in S\ \ (\pi_1 + \ldots + \pi_L = 1), \tag{1}$$

где $\{\pi_i\}_{i \in S}$ – априорные вероятности классов $\{\Omega_i\}_{i \in S}$ [1–5]. Значение d^0 случайной величины является ненаблюдаемым истинным номером класса, к которому принадлежит наблюдение x. При фиксированном $d^0=i (i \in S)$ наблюдение $x \in R^N$ описывается условной плотностью распределения вероятностей:

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

Классы $\{\Omega_i\}_{i \in S}$ полностью определяются своими характеристиками $\{\pi_i, p_i(\cdot)\}_{i \in S}$, по которым можно принимать решение о принадлежности наблюдения к тому или иному классу: например, при помощи оптимального (байесовского) решающего правила [1, 3, 4], минимизирующего риск (вероятность ошибочной классификации).

Однако на практике пространство признаков R^N зачастую избыточно (его размерность N велика) [1, 2, 4], и необходимо выбрать из исходного пространства N признаков некоторое подмножество информативных признаков [2].

Будем рассматривать часто встречающуюся в приложениях модель Фишера [1, 3, 4] смеси многомерных нормальных (гауссовских) распределений:

$$p_i(x) = n_N(x \mid \mu_i,\ \Sigma),\ x \in R^N,\ i \in S, \tag{2}$$

где $n_N(x \mid \mu_i,\ \Sigma) = (2\pi)^{-N/2}(\det(\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i)\right)$ – плотность N-мерного гауссовского распределения вероятностей с вектором математического ожидания $\mu_i=E\{x|d^0=i\} \in R^N$ (так называемый центр [1, 5] i-го класса) и невырожденной ковариационной $(N \times N)$-матрицей $\Sigma=E\{(x-\mu_i)\times$

$\times(x-\mu_i)^T|d^0=i\}$ $(\det(\Sigma)\neq 0)$, общей для всех классов, где "T" – символ транспонирования.

В качестве метода выбора информативных признаков будем использовать пошаговый переход из исходного пространства R^N в подпространство информативных признаков. На каждом шаге из исходного пространства признаков исключается один неинформативный признак. Не ограничивая общности, будем считать, что этот признак последний, и наблюдение $x\in R^N$ допускает представление:

$$x=(x_1,\ x_2,\ ...,\ x_{N-1},\ x_N)^T=((x^{N-1})^T \vdots x_N)^T\in R^N.$$

Такое разбиение порождает разбиение векторов условного математического ожидания:

$$\mu_i=(\mu_{i,1},\ \mu_{i,2},\ ...,\ \mu_{i,N-1},\ \mu_{i,N})^T=((\mu_i^{N-1})^T \vdots \mu_{i,N})^T\in R^N$$

и ковариационной матрицы:

$$\begin{array}{cc} \begin{array}{cc} N-1 & 1 \end{array} & \\ \Sigma=\left(\begin{array}{c|c} \Sigma_{N-1,\,N-1} & \sigma_{N-1} \\ \hline \sigma_{N-1}^T & \sigma_{N,\,N} \end{array}\right) & \begin{array}{c} N-1 \\ 1 \end{array} \end{array}.$$

Под неинформативностью будем понимать сильную коррелированность исключаемого признака с остальными, т. е. близость множественного коэффициента корреляции [6] $\rho_N=\sqrt{\sigma_{N-1}^T\Sigma_{N-1,\,N-1}^{-1}\sigma_{N-1}/\sigma_{N,\,N}}$ к единице. В качестве критерия эффективности перехода из исходного пространства будем использовать риск (вероятность ошибочной классификации) [1, 4, 5].

Воспользуемся тем фактом [1, 3, 4], что байесовское решающее правило (БРП), минимизирующее риск для модели Фишера (1), (2), в исходном пространстве допускает представление ($x\in R^N$):

$$d_0^N(x)=\arg\max_{i\in S}\{2\ln\pi_i-(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)\},$$

а его риск задается соотношением:

$$r_0^N=P\{d_0^N(x)\neq d^0\}=\sum_{i\in S}\pi_i P\left\{d_0^N(x)\neq i\,\middle|\,d^0=i\right\}= \tag{3}$$

$$=\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\int_{R^N}\prod_{\substack{k\in S\\ k\neq j}}I\left((x-\mu_j)^T\Sigma^{-1}(\mu_j-\mu_k)+\frac{{}_N\Delta_{jk}^2}{2}-\ln\frac{\pi_k}{\pi_j}\right)n_N(x\,|\,\mu_i,\Sigma)dx,$$

где $I(z)=\{1,$ если $z\geq 0$; $0,$ если $z<0\}$ – единичная функция Хевисайда; ${}_N\Delta_{jk}=\sqrt{(\mu_j-\mu_k)^T\Sigma^{-1}(\mu_j-\mu_k)}$ – расстояние Махаланобиса [1, 4] между классами Ω_j и Ω_k в исходном пространстве ($k\neq j\in S$).

В пространстве отобранных признаков ($x^{N-1}\in R^{N-1}$), которые также описываются моделью Фишера, но с параметрами $\left\{\mu_i^{N-1}\right\}_{i\in S}$ и $\Sigma_{N-1,\,N-1}$, БРП принимает вид:

$$d_0^{N-1}(x^{N-1})=\arg\max_{i\in S}\{2\ln\pi_i-(x^{N-1}-\mu_i^{N-1})^T\Sigma_{N-1,N-1}^{-1}(x^{N-1}-\mu_i^{N-1})\},$$

а его риск $r_0^{N-1}=P\{d_0^{N-1}(x^{N-1})\neq d^0\}$:

$$r_0^{N-1}=\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\int_{R^{N-1}}\prod_{\substack{k\in S\\ k\neq j}}I\left((x^{N-1}-\mu_j^{N-1})^T\Sigma_{N-1,\,N-1}^{-1}(\mu_j^{N-1}-\mu_k^{N-1})+\right. \tag{4}$$

$$\left.+\frac{{}_{N-1}\Delta_{jk}^2}{2}-\ln\frac{\pi_k}{\pi_j}\right)n_{N-1}(x^{N-1}\,|\,\mu_i^{N-1},\ \Sigma_{N-1,\,N-1})dx^{N-1},$$

где ${}_{N-1}\Delta_{jk}=\sqrt{(\mu_j^{N-1}-\mu_k^{N-1})^T\Sigma_{N-1,\,N-1}^{-1}(\mu_j^{N-1}-\mu_k^{N-1})}$ – межклассовое расстояние Махаланобиса в пространстве отобранных признаков ($k\neq j\in S$).

В случае двух классов (L=2) выражения для риска (3), (4) упрощаются [1, 4]:

$$r_0^N=\pi_1\Phi\left(-\frac{\Delta_N}{2}-\frac{\ln h}{\Delta_N}\right)+\pi_2\Phi\left(-\frac{\Delta_N}{2}+\frac{\ln h}{\Delta_N}\right),\quad \Delta_N={}_N\Delta_{12},\ h=\frac{\pi_1}{\pi_2};\qquad(5)$$

$$r_0^{N-1}=\pi_1\Phi\left(-\frac{\Delta_{N-1}}{2}-\frac{\ln h}{\Delta_{N-1}}\right)+\pi_2\Phi\left(-\frac{\Delta_{N-1}}{2}+\frac{\ln h}{\Delta_{N-1}}\right),\ \Delta_{N-1}={}_{N-1}\Delta_{12},\qquad(6)$$

где $\Phi(z)=\int_{-\infty}^{z}\varphi(\omega)d\omega$ – функция распределения вероятностей стандартного нормального закона $N_1(0,\,1)$ с плотностью $\varphi(\omega)=n_1(\omega|0,1)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\omega^2}{2}\right)$.

Выясним, как связаны риски (3) и (4) (риски (5) и (6) в случае двух классов), тем самым будет исследована эффективность перехода от исходного пространства R^N к подпространству отобранных признаков R^{N-1}.

Лемма 1. *В условиях модели Фишера в пространстве исходных признаков R^N для функционала риска* (3) *справедлива следующая оценка:*

$$r_0^N\le\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\Phi\left(-\frac{{}_N\Delta_{ij}}{2}-\frac{\ln\pi_i/\pi_j}{{}_N\Delta_{ij}}\right).\qquad(7)$$

Д о к а з а т е л ь с т в о. Воспользуемся соотношением (3) и оценим функционал риска сверху:

$$r_0^N\le\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\int_{R^N}I\left((x-\mu_j)^T\Sigma^{-1}(\mu_j-\mu_i)+\frac{{}_N\Delta_{ij}^2}{2}-\ln\frac{\pi_i}{\pi_j}\right)n_N(x|\mu_i,\Sigma)dx.$$

Выполнив последовательно замены

$$y=(x-\mu_i)^T\Sigma^{-1}(\mu_j-\mu_i)+\frac{{}_N\Delta_{ij}^2}{2}-\ln\frac{\pi_i}{\pi_j}\quad\text{и}\quad z=\frac{y}{{}_N\Delta_{ij}}+\frac{{}_N\Delta_{ij}}{2}+\frac{\ln\pi_i/\pi_j}{{}_N\Delta_{ij}},$$

получим (7). Лемма доказана.

Следствие. В условиях модели Фишера в пространстве отобранных признаков R^{N-1} для функционала риска (4) *будет иметь место аналогичная оценка:*

$$r_0^{N-1}\le\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\Phi\left(-\frac{{}_{N-1}\Delta_{ij}}{2}-\frac{\ln\pi_i/\pi_j}{{}_{N-1}\Delta_{ij}}\right).\qquad(8)$$

Замечание. В случае двух классов (L=2) *выражения* (7) *и* (8) *являются точными представлениями для функционалов риска и совпадают с* (5) *и* (6).

Лемма 2. *В пространстве отобранных признаков R^{N-1} расстояние Махаланобиса ${}_{N-1}\Delta_{ij}$ связано с соответствующим расстоянием ${}_N\Delta_{ij}$ в исходном пространстве R^N следующим соотношением* ($k\neq j\in S$):

$${}_{N-1}\Delta_{ij}^2={}_N\Delta_{ij}^2-(1-\rho_v^2)^{-1}\sigma_{N,N}^{-1}\left((\mu_{i,N}-\mu_{j,N})-\sigma_{N-1}^T\Sigma_{N-1,N-1}^{-1}(\mu_i^{N-1}-\mu_j^{N-1})\right)^2.\qquad(9)$$

Д о к а з а т е л ь с т в о. Согласно формулам Фробениуса [6], обратная ковариационная матрица Σ^{-1} может быть представлена в следующем виде:

$$\Sigma^{-1} = \begin{array}{c} \\ \left(\begin{array}{c|c} \Psi^{N-1,\, N-1} & \Psi_{N-1} \\ \hline \Psi_{N-1}^{T} & \Psi_{N,\, N} \end{array} \right) \end{array} \begin{array}{c} N-1 \\ 1 \end{array},$$

где

$$\Psi_{N,\, N} = (\sigma_{N,\, N} - \sigma_{N-1}^{T}\Sigma_{N-1,\, N-1}^{-1}\sigma_{N-1})^{-1} \in R^{1}, \qquad \Psi_{N-1} = -\Sigma_{N-1,\, N-1}^{-1}\sigma_{N-1}\Psi_{N,\, N} \in R^{N-1},$$

а $\Psi^{N-1,\, N-1} = \Sigma_{N-1,\, N-1}^{-1} + \Psi_{N-1}\Psi_{N-1}^{T}\Psi_{N,\, N}^{-1}$ – матрица порядка $(N-1)\times(N-1)$.

Преобразуем квадрат расстояния Махаланобиса в исходном пространстве:

$$_N\Delta_{ij}^2 = (\mu_i - \mu_j)^T\Sigma^{-1}(\mu_i - \mu_j) = (\mu_i^{N-1} - \mu_j^{N-1})^T\Psi^{N-1,\, N-1}(\mu_i^{N-1} - \mu_j^{N-1}) +$$

$$+2(\mu_{i,\, N} - \mu_{j,\, N})\Psi_{N-1}^{T}(\mu_i^{N-1} - \mu_j^{N-1}) + (\mu_{i,\, N} - \mu_{j,\, N})\Psi_{N,\, N}(\mu_{i,\, N} - \mu_{j,\, N}).$$

Заметим, что $\Psi_{N,\, N} = (1-\rho_N^2)^{-1}\sigma_{N,\, N}^{-1}$, и продолжим преобразования, обозначив через $\bar{\Psi}_{N-1} = \Sigma_{N-1,\, N-1}^{-1}\sigma_{N-1}$:

$$_N\Delta_{ij}^2 = (\mu_i^{N-1} - \mu_j^{N-1})^T\Sigma_{N-1,\, N-1}^{-1}(\mu_i^{N-1} - \mu_j^{N-1}) + \left(\bar{\Psi}_{N-1}^{T}(\mu_i^{N-1} - \mu_j^{N-1})\right)^2\Psi_{N,\, N} -$$

$$-2(\mu_{i,\, N} - \mu_{j,\, N})\bar{\Psi}_{N-1}^{T}(\mu_i^{N-1} - \mu_j^{N-1})\Psi_{N,\, N} + (\mu_{i,\, N} - \mu_{j,\, N})^2\Psi_{N,\, N} =$$

$$= {}_{N-1}\Delta_{ij}^2 + \left((\mu_{i,\, N} - \mu_{j,\, N}) - \sigma_{N-1}^{T}\Sigma_{N-1,\, N-1}^{-1}(\mu_i^{N-1} - \mu_j^{N-1})\right)^2(1-\rho_N^2)^{-1}\sigma_{N,\, N}^{-1},$$

что совпадает с (9). Лемма доказана.

Теорема 1. *Пусть в условиях модели Фишера выполняется:* $\delta_{ij} = \left|(\mu_{i,N} - \mu_{j,N}) - \sigma_{N-1}^{T}\Sigma_{N-1,N-1}^{-1}(\mu_i^{N-1} - \mu_j^{N-1})\right| \to 0$, $i \neq j \in S$, *тогда для приращения риска* $\Delta r_0 = r_0^{N-1} - r_0^{N} \geq 0$ *справедлива оценка сверху*:

$$\Delta r_0 \leq \frac{1}{2}\left(1-\rho_N^2\right)^{-1}\sigma_{N,\, N}^{-1}\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j>i}}\varphi\left(-\frac{_N\Delta_{ij}}{2} - \frac{\ln \pi_i/\pi_j}{_N\Delta_{ij}}\right)\frac{1}{_N\Delta_{ij}}\delta_{ij}^2 + O\left((\max_{i,\, j\in S}\delta_{ij})^4\right), \quad (10)$$

а в случае двух классов (L=2) *справедливо соотношение*:

$$r_0^{N-1} = r_0^{N} + \frac{1}{2}(1-\rho_N^2)^{-1}\sigma_{N,\, N}^{-1}\frac{\pi_1}{\Delta_N}\varphi\left(-\frac{\Delta_N}{2} - \frac{\ln h}{\Delta_N}\right)\delta_{12}^2 + O(\delta_{12}^4). \quad (11)$$

Доказательство. Очевидно, что при переходе в пространство информативных признаков вероятность ошибочной классификации может только увеличиться, следовательно, $\Delta r_0 \geq 0$. Воспользуемся оценками для риска (7) и (8), соотношением (9) для расстояния Махаланобиса, малостью величин $\delta_{ij} = \left|(\mu_{i,\, N} - \mu_{j,\, N}) - \sigma_{N-1}^{T}\Sigma_{N-1,\, N-1}^{-1}(\mu_i^{N-1} - \mu_j^{N-1})\right|$ и с учетом соотношения $_{N-1}\Delta_{ij} = \sqrt{_N\Delta_{ij}^2 - \delta_{ij}^2}$ разложим оценку для приращения риска в ряд Тейлора:

$$\Delta r_0 \leq \left(1-\rho_N^2\right)^{-1}\sigma_{N,\, N}^{-1}\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j\neq i}}\left(\frac{1}{4\,_N\Delta_{ij}} - \frac{\ln \pi_i/\pi_j}{2\,_N\Delta_{ij}^3}\right)\varphi\left(-\frac{_N\Delta_{ij}}{2} - \frac{\ln \pi_i/\pi_j}{_N\Delta_{ij}}\right)\delta_{ij}^2 + O\left((\max_{i,\, j\in S}\delta_{ij})^4\right),$$

учтем, что $\pi_i\varphi\left(-\frac{_N\Delta_{ij}}{2} - \frac{\ln \pi_i/\pi_j}{_N\Delta_{ij}}\right) = \pi_j\varphi\left(-\frac{_N\Delta_{ij}}{2} + \frac{\ln \pi_i/\pi_j}{_N\Delta_{ij}}\right)$, и получим соотношение (10).

В случае двух классов выражение (11) доказывается аналогичным образом с учетом замечания ко второй лемме. Теорема доказана.

Отметим, что требование малости величин $\{\delta_{ij}\}_{i \neq j \in s}$ в условиях теоремы 1 обусловлено предполагаемой сильной коррелированностью (а следовательно, "почти линейной" зависимостью [6]) исключаемого признака x_N от остальных $x^{N-1} \in R^{N-1}$.

Практическая ценность соотношения (11) и приведенной выше оценки сверху для приращения риска (10) состоит в том, что они позволяют на каждом шаге выбирать исключаемый неинформативный признак таким образом, чтобы проигрыш по риску не превосходил некоторой наперед заданной величины.

Основным недостатком полученных соотношений является невозможность непосредственно учитывать влияние исключенных на предыдущих шагах признаков. Для решения этой проблемы получим соотношение, напрямую связывающее функционалы риска в исходном пространстве наблюдений R^N и подпространстве информативных признаков R^m, сформированном на $(N-m)$-м шаге процедуры исключения.

Не нарушая общности, будем считать, что отброшены последние $N-m$ признаков, и наблюдение $x \in R^N$ допускает представление:

$$x = ((x^m)^T \vdots (x^{N-m})^T)^T \in R^N, \quad x^m \in R^m, \quad x^{N-m} \in R^{N-m}.$$

Такое разбиение порождает соответствующее разбиение "центров" классов ($i \in S$):

$$\mu_i = \left((\mu_i^m)^T \vdots (\mu_i^{N-m})^T\right)^T \in R^N, \ \mu_i^m \in R^m, \ \mu_i^{N-m} \in R^{N-m},$$

и ковариационной матрицы:

$$\Sigma = \begin{array}{cc} \begin{array}{cc} \quad m & \quad N-m \end{array} & \\ \left(\begin{array}{c|c} \Sigma_{m,\,m} & \Sigma_{m,\,N-m} \\ \hline \Sigma_{m,\,N-m}^T & \Sigma_{N-m,\,N-m} \end{array}\right) & \begin{array}{c} m \\ N-m \end{array} \end{array}.$$

В пространстве информативных признаков ($x^m \in R^m$), которые описываются моделью Фишера с параметрами $\{\mu_i^m\}_{i \in S}$ и $\Sigma_{m,\,m}$, БРП принимает вид:

$$d_0^m(x^m) = \arg\max_{i \in S}\{2\ln\pi_i - (x^m - \mu_i^m)^T \Sigma_{m,\,m}^{-1}(x^m - \mu_i^m)\},$$

а его риск $r_0^m = P\{d_0^m(x^m) \neq d^0\}$:

$$r_0^m = \sum_{i \in S}\pi_i \sum_{\substack{j \in S \\ j \neq i}} \int_{R^m} \prod_{\substack{k \in S \\ k \neq j}} I\left((x^m - \mu_j^m)^T \Sigma_{m,\,m}^{-1}(\mu_j^m - \mu_k^m) + \right.$$
$$\left. + \frac{{}_m\Delta_{jk}^2}{2} - \ln\frac{\pi_k}{\pi_j}\right) n_m(x^m \mid \mu_i^m, \Sigma_{m,\,m})dx^m, \tag{12}$$

где ${}_m\Delta_{jk} = \sqrt{(\mu_j^m - \mu_k^m)^T \Sigma_{m,\,m}^{-1}(\mu_j^m - \mu_k^m)}$ – межклассовое расстояние Махаланобиса в пространстве m информативных признаков ($k \neq j \in S$).

В случае двух классов ($L = 2$) для риска (12) имеем:

$$r_0^m = \pi_1\Phi\left(-\frac{\Delta_m}{2} - \frac{\ln h}{\Delta_m}\right) + \pi_2\Phi\left(-\frac{\Delta_m}{2} + \frac{\ln h}{\Delta_m}\right), \quad \Delta_m = {}_m\Delta_{12}, \ h = \frac{\pi_1}{\pi_2}. \tag{13}$$

Лемма 3. *В пространстве информативных признаков R^m расстояние Махаланобиса ${}_m\Delta_{ij}$ связано с соответствующим расстоянием ${}_N\Delta_{ij}$ в исходном пространстве R^N следующим соотношением ($i \neq j \in S$):*

$${}_m\Delta_{ij}^2 = {}_N\Delta_{ij}^2 - \tag{14}$$

$$-\left((\mu_i^{N-m}-\mu_j^{N-m})-B(\mu_i^m-\mu_j^m)\right)^T\Sigma^{-1}_{N-m,\ N-m|m}\left((\mu_i^{N-m}-\mu_j^{N-m})-B(\mu_i^m-\mu_j^m)\right),$$

где $\Sigma_{N-m,\ N-m|m}=\Sigma_{N-m,\ N-m}-\Sigma^T_{m,\ N-m}\Sigma^{-1}_{m,\ m}\Sigma_{m,\ N-m}$ *– условная матрица ковариаций случайного вектора* $x^{N-m}\in R^{N-m}$ *при фиксированном векторе* $x^m\in R^m$, *а* $B=\Sigma^T_{m,\ N-m}\Sigma^{-1}_{m,\ m}$ *– матрица коэффициентов многомерной регрессии случайного вектора* $x^{N-m}\in R^{N-m}$ *на вектор* $x^m\in R^m$ [6].

Доказательство основано на использовании формул Фробениуса [6] для обратной матрицы Σ^{-1} и аналогично доказательству леммы 2.

Ковариационную матрицу $\Sigma_{N-m,N-m}$ можно представить в виде [6]:

$$\Sigma_{N-m,\ N-m}=\left(\Sigma^{1/2}_{N-m,\ N-m}\right)^T\Sigma^{1/2}_{N-m,\ N-m},$$

где $\left(\Sigma^{-1/2}_{N-m,\ N-m}\right)^T\Sigma_{N-m,\ N-m}\Sigma^{-1/2}_{N-m,\ N-m}=I_{N-m}$ – единичная матрица порядка $(N-m)\times(N-m)$. Введем следующие обозначения:

$$\overline{\rho}^2_{N-m}=\left(\Sigma^{-1/2}_{N-m,\ N-m}\right)^T\Sigma^T_{m,\ N-m}\Sigma^{-1}_{m,\ m}\Sigma_{m,\ N-m}\Sigma^{-1/2}_{N-m,\ N-m},$$

$$\delta^*_{ij}=\left(\mu_i^{N-m}-\mu_j^{N-m}\right)-B\left(\mu_i^m-\mu_j^m\right)\in R^{N-m},$$

тогда выражение (14) перепишется в виде:

$$_m\Delta^2_{ij}={}_N\Delta^2_{ij}-\left(\delta^*_{ij}\right)^T\left(\Sigma^{-1/2}_{N-m,\ N-m}\right)^T\left(I_{N-m}-\overline{\rho}^2_{N-m}\right)^{-1}\Sigma^{-1/2}_{N-m,\ N-m}\delta^*_{ij},\qquad(15)$$

что представляет собой аналог соответствующего выражения (9), полученного в лемме 2, где $\overline{\rho}_{N-m}$ – матрица порядка $(N-m)\times(N-m)$, являющаяся аналогом множественного коэффициента корреляции $\rho_N\in R^1$.

Выясним, как связаны риски (3) и (12) (риски (5) и (13) в случае двух классов), тем самым будет исследована эффективность перехода от исходного пространства R^N к подпространству информативных признаков R^m.

Теорема 2. *Пусть в условиях модели Фишера выполняется:* $\overline{\delta}_{ij}=\left|\delta^*_{ij}\right|\to 0$, *где* $\left|\delta^*_{ij}\right|=\sqrt{\left(\delta^*_{ij}\right)^T\delta^*_{ij}}$ *– норма Евклида, тогда для приращения риска* $\overline{\Delta}r_0=r_0^m-r_0^N\geq 0$ *справедлива оценка сверху*:

$$\overline{\Delta}r_0\leq\frac{1}{2}\sum_{i\in S}\pi_i\sum_{\substack{j\in S\\ j>i}}\frac{1}{_N\Delta_{ij}}\varphi\left(-\frac{_N\Delta_{ij}}{2}-\frac{\ln\pi_i/\pi_j}{_N\Delta_{ij}}\right)\times\qquad(16)$$

$$\times\left(\delta^*_{ij}\right)^T\left(\Sigma^{-1/2}_{N-m,\ N-m}\right)^T\left(I_{N-m}-\overline{\rho}^2_{N-m}\right)^{-1}\Sigma^{-1/2}_{N-m,\ N-m}\delta^*_{ij}+O\left((\max_{i,\ j\in S}\overline{\delta}_{ij})^4\right),$$

а в случае двух классов ($L=2$, $\Delta_N={}_N\Delta_{12}$, $h=\dfrac{\pi_1}{\pi_2}$):

$$\overline{\Delta}r_0=\frac{1}{2}\frac{\pi_1}{\Delta_N}\varphi\left(-\frac{\Delta_N}{2}-\frac{\ln h}{\Delta_N}\right)\times\qquad(17)$$

$$\times\left(\delta^*_{12}\right)^T\left(\Sigma^{-1/2}_{N-m,\ N-m}\right)^T\left(I_{N-m}-\overline{\rho}^2_{N-m}\right)^{-1}\Sigma^{-1/2}_{N-m,\ N-m}\delta^*_{12}+O\left(\overline{\delta}^4_{12}\right).$$

Доказательство основано на соотношении (5) и аналогично доказательству теоремы 1.

Соотношения (16) и (17) для приращения риска позволяют на каждом шаге процедуры исключения следить за тем, чтобы общий проигрыш по риску не превосходил некоторой наперед заданной величины, а полученные ранее выражения (10) и (11) позволяют контролировать локальный проигрыш по риску на конкретном шаге для одного признака.

1. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М., 1989.
2. Методы, критерии и алгоритмы, используемые при преобразовании, выделении и выборе признаков в анализе данных. Вильнюс, 1988.
3. Миленький А.В. Классификация сигналов в условиях неопределенности. М., 1975.
4. Фукунага К. Введение в статистическую теорию распознавания образов. М., 1979.
5. Жук Е.Е., Харин Ю.С. Устойчивость в кластер-анализе многомерных наблюдений. Мн., 1998.
6. Андерсон Т. Введение в многомерный статистический анализ. М., 1963.

Поступила в редакцию 27.03.2002.

Евгений Евгеньевич Жук – доктор физико-математических наук, доцент кафедры ММАД.
Екатерина Викторовна Серикова – студентка 5-го курса ФПМИ.

УДК 519.24

Н.Н. ДЕМЕШ, Т.В. СОБОЛЕВА

ПОСТРОЕНИЕ СОСТОЯТЕЛЬНОЙ ОЦЕНКИ СПЕКТРАЛЬНОЙ ПЛОТНОСТИ ДЕЙСТВИТЕЛЬНОГО УСТОЙЧИВОГО ПРОЦЕССА

The consistent estimate of a spectral density of a discrete real stationary random process by 2π-periodical spectral windows has been constructed.

Будем рассматривать действительный дискретный симметричный α-устойчивый стационарный случайный процесс $X(t)$, $t\in Z=\{0, \pm 1, \ldots\}$, $0<\alpha<2$ (см. [1, 2, 6]), имеющий спектральное представление

$$X(t)=\int_{-\pi}^{\pi}\cos(\lambda t)d\xi(\lambda), \qquad (1)$$

где $\xi(\lambda)$ – действительный α-устойчивый с независимыми приращениями процесс такой, что

$$\left\{E\left|d\xi(\lambda)\right|^{p}\right\}^{\alpha/p}=C(p,\alpha)\varphi(\lambda)d\lambda$$

для всех $0<p<\alpha$, где $C(p, \alpha)$ зависит от α, p и не зависит от $\xi(\lambda)$, а $\varphi(\lambda)$, $\lambda\in\Pi=[-\pi, \pi]$, – неотрицательная, интегрируемая, 2π-периодическая функция, которую будем называть спектральной плотностью дискретного устойчивого стационарного случайного процесса (1). При $\alpha=2$ процесс $X(t)$, $t\in Z$, является гауссовским, а функция $\varphi(\lambda)$ – обычная спектральная плотность и, следовательно, в этом случае для исследования $X(t)$, $t\in Z$, применим традиционный спектральный анализ. При $0<\alpha<2$ функция $\varphi(\lambda)$ не является спектральной плотностью в обычном смысле, но при решении задач линейного предсказания и фильтрации играет ту же роль, что и спектральная плотность процессов второго порядка.

В качестве оценки спектральной плотности рассмотрим следующую статистику