АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
АВТОМЕТРИЯ

№ 1 1970

УДК 62—506.2.001

М. С. ХАЙРЕТДИНОВ

(Новосибирск)

ОБ ЭФФЕКТИВНОСТИ КОРРЕЛИРОВАННЫХ ПРИЗНАКОВ В ЗАДАЧАХ КЛАССИФИКАЦИИ СОСТОЯНИЙ ОБЪЕКТОВ

В практике измерений часто возникает задача классификации определенных состояний объектов статистической природы на основании измерения совокупности их параметров — признаков; при выборе признаков возникает вопрос о влиянии степени коррелированности их между собой на качество классификации, характеризуемое вероятностью ошибочной классификации. Определенное состояние объекта — класс — в многомерном пространстве признаков описывается случайным вектором, имеющим плотность распределения вероятностей с определенными параметрами. Вопрос о влиянии степени коррелированности признаков на ошибку классификации реализации двумерного случайного вектора — образа, принадлежащего к одному из двух классов с нормальными плотностями распределения с разными средними значениями и одинаковыми ковариационными матрицами, рассматривался в [1].

Целью данной работы в предположении, что классификация осуществляется с помощью адаптивного классификатора [2], является рассмотрение зависимости от степени коррелированности признаков вероятности ошибочной классификации для случаев, когда классы характеризуются произвольными плотностями распределения вероятностей появления образов с разными средними значениями и неравными ковариационными матрицами и для вынесения решения используется линейная дискриминантная функция (л. д. ф.) в первом случае и квадратичная дискриминантная функция (к. д. ф.) во втором.

Метод определения действительного значения вероятности ошибочной классификации образов с произвольными плотностями распределения вероятностей пока не разработан, поэтому в дальнейшем используется минимизированное значение верхней границы вероятности ошибки, полученное из неравенства Чебышева для линейного классификатора образов в случае двух классов [3]

$$\alpha = \frac{1}{\overline{P}^T \sum^{-1} \overline{P}}.$$

Здесь

$$\overline{P} = \begin{pmatrix} q_1 \bar{\mu}_1 - q_2 \mu_2 \\ q_1 - q_2 \end{pmatrix};$$

$$\sum = \begin{pmatrix} q_1 \sum_1 + q_2 \sum_2 + q_1 q_2 (\bar\mu_1 + \bar\mu_2)(\bar\mu_1 + \bar\mu_2)^T & 2 q_1 q_2 (\bar\mu_1 + \bar\mu_2) \\ 2 q_1 q_2 (\bar\mu_1 + \bar\mu_2)^T & 4 q_1 q_2 \end{pmatrix}$$

После раскрытия квадратичной формы $\bar P^T \sum^{-1} \bar P$ с учетом матриц $\bar P$ и Σ и предполагая $q_1 = q_2 = \frac{1}{2}$, получим

$$\alpha \leqslant \frac{1}{\frac{1}{2}(\bar\mu_1 - \bar\mu_2)^T \sum^{-1} (\bar\mu_1 - \bar\mu_2)} = \frac{1}{\frac{1}{2} a}. \tag{1}$$

В приведенных выражениях q_1, $\bar\mu_1$, Σ_1 и q_2, $\bar\mu_2$, Σ_2 — соответственно априорная вероятность появления образов, вектор среднего значения и ковариационная матрица, относящиеся к 1-му и 2-му классам соответственно; a — величина, характеризующая расстояние между классами.

Следует отметить, что в случае нормального распределения вероятности появления образов верхняя граница вероятности ошибки приближается к действительной и имеет вид [3] $\alpha \leqslant \frac{2}{9} \frac{1}{a}$. Наконец, из сравнения полученного соотношения с теоретическими и экспериментальными данными (отчасти взятыми из работы [4]) при $a > 1$ значение верхней границы, определяемое соотношением

$$\alpha \leqslant \frac{2}{9} \frac{1}{a+1}, \tag{2}$$

приближается к действительной (байесовой) вероятности ошибки с погрешностью не более 10%.

1. Решение на основе л. д. ф. Рассмотрим решение задачи в первом случае. Пусть вектор признаков представляет собой двумерный случайный вектор с ковариационными матрицами для обоих классов соответственно

$$\sum_1 = \begin{pmatrix} \sigma_{11}^{(1)} & \sigma_{12}^{(1)} \\ \sigma_{21}^{(1)} & \sigma_{22}^{(1)} \end{pmatrix}; \quad \sum_2 = \begin{pmatrix} \sigma_{11}^{(2)} & \sigma_{12}^{(2)} \\ \sigma_{21}^{(2)} & \sigma_{22}^{(2)} \end{pmatrix},$$

где $\sigma_{11}^{(1)} = \sigma_1^{(1)^2}$ — дисперсия 1-го признака 1-го класса; $\sigma_{22}^{(1)} = \sigma_2^{(1)^2}$ — дисперсия 2-го признака 1-го класса; $\sigma_{12}^{(1)} = \sigma_{21}^{(1)} = \sigma_1^{(1)} \sigma_2^{(1)} \rho_1$; ρ_1 — коэффициент корреляции между признаками 1-го класса. Соответственно $\sigma_{11}^{(2)} = \sigma_1^{(2)^2}$ — дисперсия 1-го признака 2-го класса; $\sigma_{22}^{(2)} = \sigma_2^{(2)^2}$ — дисперсия 2-го признака 2-го класса; $\sigma_{12}^{(2)} = \sigma_{21}^{(2)} = \sigma_1^{(2)} \sigma_2^{(2)} \rho_2$; ρ_2 — коэффициент корреляции между признаками, относящимися ко 2-му классу.

Положим, что $\sigma_1^{(1)} = c_1 \sigma_1^{(2)}$, $\sigma_2^{(1)} = c_2 \sigma_2^{(2)}$. Введя обозначения m_{11}, m_{21} и m_{12}, m_{22}, соответствующие средним значениям признаков 1-го и 2-го классов, и расписав квадратичную форму, для a получим

$$a = \frac{1}{(c_1^2+1)(c_2^2+1) - (\rho_1 c_1 c_2 + \rho_2)^2} \left[\frac{(m_{11} - m_{12})^2 (c_2^2+1)}{\sigma_1^{(2)^2}} - \right.$$
$$\left. - \frac{2(\rho_1 c_1 c_2 + \rho_2)(m_{11} - m_{12})(m_{21} - m_{22})}{\sigma_1^{(2)} \sigma_2^{(2)}} + \frac{(m_{21} - m_{22})^2 (c_1^2+1)}{\sigma_2^{(2)^2}} \right]. \tag{3}$$

Обозначим $\frac{(m_{11}-m_{12})^2}{\sigma_1^{(2)2}} = A_1^2$, $\frac{(m_{21}-m_{22})^2}{\sigma_2^{(2)2}} = A_2^2$; тогда соотношение для a можно представить в виде

$$a = \frac{1}{(c_1^2+1)(c_2^2+1)-(c_1c_2\rho_1+\rho_2)^2}\,[A_1^2(c_2^2+1)-2(c_1c_2\rho_1+\rho_2)A_1A_2+ + A_2^2(c_1^2+1)]. \tag{4}$$

Проанализируем функциональную зависимость a от переменных ρ_1 и ρ_2.

1. При определении экстремума a из условий $\frac{da}{d\rho_1}=0$, $\frac{da}{d\rho_2}=0$ получим, что соответствующие значения ρ_1 и ρ_2 связаны соотношением

$$x=c_1c_2\rho_1+\rho_2, \tag{5}$$

в котором x удовлетворяет одному из двух условий:

$$\text{а) } x_1=\frac{A_1}{A_2}(c_2^2+1);\quad \text{б) } x_2=\frac{A_2}{A_1}(c_1^2+1).$$

Если $\frac{A_1}{A_2}\leqslant\sqrt{\frac{c_1^2+1}{c_2^2+1}}$, то x определяется из условия а). Если $\frac{A_1}{A_2}\leqslant\sqrt{\frac{c_2^2+1}{c_1^2+1}}$, то x удовлетворяет условию б). При любом из этих значений x функция a достигает своего экстремума — минимума. Обоснование этих выводов приведено в приложении 1. Минимальное значение a, например при $x=x_1$, равно, как следует из (4),

$$a_{\min}=\frac{A_2^2(c_1^2+1)-A_1^2(c_2^2+1)}{(c_1^2+1)(c_2^2+1)-\frac{A_1^2}{A_2^2}(c_2^2+1)}.$$

2. При $|\rho_1|\leqslant 1$ и $|\rho_2|\leqslant 1$ максимальное значение a получается при $\rho_1=\rho_2=1$ или $\rho_1=\rho_2=-1$, ибо, как показано в приложении 1, при $x\neq x_1$ $a=f(x)$ является монотонно возрастающей функцией, а $|x|$, в свою очередь, на основании (5) имеет максимальное значение при $\rho_1=\rho_2=1$ или $\rho_1=\rho_2=-1$. При $\rho_1=\rho_2=1$, как следует из соотношения (4),

$$a_{1\max}=\frac{(A_1c_2-A_2c_1)^2+(A_1-A_2)^2}{(c_1-c_2)^2}.$$

При $\rho_1=\rho_2=-1$

$$a_{2\max}=\frac{(A_1c_2+A_2c_1)^2+(A_1+A_2)^2}{(c_1-c_2)^2}.$$

В случае $c_1=c_2$ $a_{1\max}$ и $a_{2\max}\to\infty$. Если $A_1>0$, $A_2>0$, $c_1>0$, $c_2>0$, график функции $a=f(x)$ будет иметь вид, представленный на рис. 1.

2. Решение на основе к. д. ф. В данном разделе рассмотрим вопрос о влиянии коррелированных признаков на качество классификации в случае принятия решения с использованием к. д. ф. Адаптивный классификатор рассматривается как линейный классификатор, на входы которого поступают четыре признака: x_1, x_1^2, x_2, x_2^2. Будем считать, что распределение плотностей вероятности значений признаков x_1 и x_2 являются нормальными с параметрами для обоих классов соответственно: $(m_{11}, \sigma_1^{(1)})$, $(m_{21}, \sigma_2^{(1)})$ и $(m_{12}, \sigma_1^{(2)})$, $(m_{22}, \sigma_2^{(2)})$.

Без потери общности полученных результатов в дальнейшем для упрощения выводов положим средние значения признаков обоих классов равными $m_{11}=m_{12}=m_{21}=0$ и $m_{22}\neq 0$. Сумма ковариационных мат-

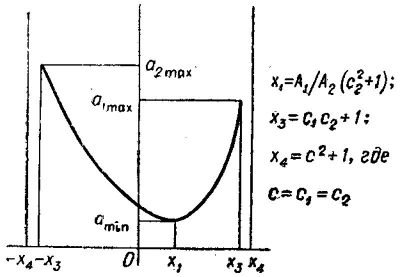

Рис. 1.

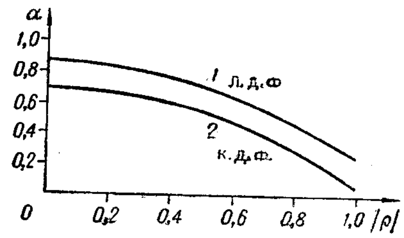

Рис. 2.

риц двух 4-мерных случайных векторов, относящихся к 1-му и 2-му классам, будет иметь вид

$$\sum=\begin{pmatrix} \sigma_{11}^{(1)}+\sigma_{11}^{(2)} & \sigma_{12}^{(1)}+\sigma_{12}^{(2)} & \sigma_{13}^{(1)}+\sigma_{13}^{(2)} & \sigma_{14}^{(1)}+\sigma_{14}^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{41}^{(1)}+\sigma_{41}^{(2)} & \sigma_{42}^{(1)}+\sigma_{42}^{(2)} & \sigma_{43}^{(1)}+\sigma_{43}^{(2)} & \sigma_{44}^{(1)}+\sigma_{44}^{(2)} \end{pmatrix}$$

Выпишем элементы матрицы:

$$\sigma_{11}^{(1)}+\sigma_{11}^{(2)}=\sigma_1^{(1)^2}+\sigma_1^{(2)^2};\quad \sigma_{13}^{(1)}+\sigma_{13}^{(2)}=\rho_1\sigma_1^{(1)}\sigma_2^{(1)}+\rho_2\sigma_1^{(2)}\sigma_2^{(2)};$$
$$\sigma_{33}^{(1)}+\sigma_{33}^{(2)}=\sigma_2^{(1)^2}+\sigma_2^{(2)^2}.$$

С учетом принятых средних значений признаков вид остальных элементов матрицы на основании выводов, приведенных в приложении 2, будет таков:

$$\sigma_{12}^{(1)}+\sigma_{12}^{(2)}=0;\quad \sigma_{14}^{(1)}+\sigma_{14}^{(2)}=2m_{22}\rho_2\sigma_1^{(2)}\sigma_2^{(2)};\quad \sigma_{22}^{(1)}+\sigma_{22}^{(2)}=2\sigma_1^{(1)^4}+2\sigma_1^{(2)^4};$$
$$\sigma_{23}^{(1)}+\sigma_{23}^{(2)}=0;\quad \sigma_{24}^{(1)}+\sigma_{24}^{(2)}=2\rho_2^2\sigma_1^{(2)^2}\sigma_2^{(2)^2}+2\rho_1^{(2)}\sigma_1^{(1)^2}\sigma_2^{(1)^2};$$
$$\sigma_{34}^{(1)}+\sigma_{34}^{(2)}=2m_{22}\sigma_2^{(2)^2};\quad \sigma_{44}^{(1)}+\sigma_{44}^{(2)}=2\sigma_2^{(1)^4}+4m_{22}^{(2)}\sigma_2^{(2)^2}+2\sigma_2^{(2)^4}.$$

Расстояние $a=(\bar{\mu}_1-\bar{\mu}_2)^T\sum^{-1}(\bar{\mu}_1-\bar{\mu}_2)$ между классами в рассматриваемом случае будет равно

$$a=\frac{1}{|A|}(p_1^2A_{11}+2p_1p_2A_{12}+2p_1p_3A_{13}+2p_1p_4A_{14}+p_2^2A_{22}+$$
$$+2p_2p_3A_{23}+2p_2p_4A_{24}+2p_3p_4A_{34}+p_3^2A_{33}+p_4^2A_{44}).$$

Здесь $A_{11}, A_{12}, \ldots, A_{44}$ — элементы обратной ковариационной матрицы Σ^{-1}; $|A|$ — определитель матрицы Σ; расстояния между средними значениями признаков следующие:

$$p_1=M^{(1)}[x_1]-M^{(2)}[x_1]=m_{11}-m_{12}=0;$$

$$p_2=M^{(1)}[x_1^2]-M^{(2)}[x_1^2]=m_{11}^2+\sigma_1^{(1)2}-m_{12}^2-\sigma_1^{(2)^2}=\sigma_1^{(1)^2}-\sigma_1^{(2)^2},$$

$$p_3 = M^{(1)}[x_2] - M^{(2)}[x_2] = m_{21} - m_{22} = -m_{22};$$

$$p_4 = M^{(1)}[x_2^2] - M^{(2)}[x_2^2] = m_{21}^2 + \sigma_2^{(1)^2} - m_{22}^2 - \sigma_2^{(2)^2} = \sigma_2^{(1)^2} - m_{22}^2 - \sigma_2^{(2)^2}.$$

Символами $M^{(1)}[x_1]$, $M^{(2)}[x_1]$, $M^{(1)}[x_2]$, $M^{(2)}[x_2]$, $M^{(1)}[x_1^2]$, $M^{(2)}[x_1^2]$, $M^{(1)}[x_2^2]$, $M^{(2)}[x_2^2]$ обозначены соответственно каждому классу средние значения признаков и их квадратов. Вывод соотношений для средних значений квадратов признаков приведен в приложении 2. С учетом $p_1 = 0$

$$a = \frac{1}{|A|}\left(p_2^2 A_{22} + p_3^2 A_{33} + p_4^2 A_{44} + 2p_2 p_3 A_{23} + 2p_2 p_4 A_{24} + 2p_3 p_4 A_{34}\right).$$

Можно показать, что после вычисления определителя $|A|$ и выражения в круглых скобках функция $a = f(\rho_1, \rho_2)$ является положительной и монотонно возрастающей при $|\rho_1| \to 1$ и $|\rho_2| \to 1$. При $|\rho_1| = 1$ и $|\rho_2| = 1$ определитель $|A| = 4\sigma_1^{(2)^4}\sigma_2^{(2)^4}(c_2 - c_1)^4(c_2 + c_1)^2$; отсюда при $c_2 = c_1$ $a \to \infty$. Не приводя громоздких выкладок, обосновывающих приведенные выводы, проиллюстрируем сказанное на примере, из которого будет видна качественная зависимость верхней границы вероятности ошибки от коэффициента корреляции между признаками. Положим $m_{11} = m_{12} = m_{21} = 0$, $m_{22} = 4$, $\sigma_1^{(1)^2} = 4$, $\sigma_1^{(2)^2} = 1$, $\sigma_2^{(1)^2} = 2$, $\sigma_2^{(2)^2} = 5$. Пусть $\rho_1 = \rho_2 = \rho$. Зависимость верхней границы вероятности ошибки — $\alpha = f(\rho)$ — при выбранных параметрах распределений дана на рис. 2. Здесь же при тех же параметрах для сравнения приведена зависимость $\alpha = f(\rho)$ в случае л. д. ф.

ВЫВОДЫ

При решении задач классификации образов с использованием л. д. ф. для классов с произвольными распределениями плотностей вероятности появления образов минимальным расстояние между классами в зависимости от коэффициентов корреляции ρ_1, ρ_2 между признаками оказывается при $c_1 c_2 \rho_1 + \rho_2 = \frac{A_1}{A_2}(c_2^2 + 1)$ или $c_1 c_2 \rho_1 + \rho_2 = \frac{A_2}{A_1}(c_1^2 + 1)$. Получаемому при этом значению $a_{\min}$ соответствует максимальное значение верхней границы вероятности ошибки $\alpha \leqslant \frac{1}{1/2\, a_{\min}}$ или $\alpha \leqslant \frac{2}{9}\frac{1}{a_{\min} + 1}$ для классов с нормальным распределением вероятностей появления образов.

И для к. д. ф. и для л. д. ф. при значениях $|\rho_1| \to 1$ и $|\rho_2| \to 1$ расстояние между классами a монотонно растет. В диапазоне $|\rho_1| \leqslant 1$ и $|\rho_2| \leqslant 1$ максимальное значение $a = a_{\max}$ достигается при $\rho_1 = \rho_2 = 1$ или $\rho_1 = \rho_2 = -1$. В случае, если соотношения между дисперсиями первых и вторых признаков обоих классов равны между собой ($c_1 = c_2$), то $a_{\max} \to \infty$, а значение вероятности ошибки стремится к нулю. Качественная зависимость верхней границы α от ρ (при $\rho_1 = \rho_2 = \rho$) и для к. д. ф. и л. д. ф. одинакова, за исключением значений ρ, близких к 1. Для этих значений ρ значение верхней границы α в случае к. д. ф. уменьшается быстрее, нежели в случае л. д. ф.

Приведенные приближенные (1) и уточненные (2) соотношения могут быть использованы при решении вопроса по выбору информативных признаков.

$$-2\,(c_1 c_2 \rho_1 + \rho_2)\, A_1 A_2 + A_2^2\,(c_1^2 + 1)]\,(c_1 c_2 \rho_1 + \rho_2)\}.$$

После введения обозначений

$$(c_1^2 + 1)(c_2^2 + 1) = L, \quad (c_1 c_2 \rho_1 + \rho_2) = x, \quad A_1^2 (c_2^2 + 1) + A_2^2 (c_1^2 + 1) = B$$

полученное соотношение можно представить в виде

$$\frac{da}{d\rho_1} = 2\, c_1 c_2\, (-A_1 A_2 x^2 + Bx - A_1 A_2 L). \tag{1}$$

По аналогии

$$\frac{da}{d\rho_2} = 2\,(-A_1 A_2 x^2 + Bx - A_1 A_2 L). \tag{2}$$

Значения x, при которых достигается экстремум a, определяются уравнением

$$\frac{da}{d\rho_1} = \frac{da}{d\rho_2} = A_1 A_2 x^2 - Bx + A_1 A_2 L = 0.$$

В результате решения получим:

$$x_1 = \frac{A_1}{A_2}\,(c_2^2 + 1); \tag{3}$$

$$x_2 = \frac{A_2}{A_1}\,(c_1^2 + 1). \tag{4}$$

Рассмотрим, какой из корней подходит при $|\rho_1| \leqslant 1$, $|\rho_2| \leqslant 1$. Прежде всего покажем, что $x^2 \leqslant L$. Действительно, это эквивалентно

$$(c_1 c_2 \rho_1 + \rho_2)^2 \leqslant (c_1^2 + 1)\,(c_1^2 + 1).$$

Положим $\rho_1 = \rho_2 = 1$. Тогда $(c_1 c_2 + 1)^2 \leqslant (c_1^2 + 1)\,(c_2^2 + 1)$, что, в свою очередь, эквивалентно неравенству $c_1^2 + c_2^2 - 2 c_1 c_2 \geqslant 0$ или $(c_1 - c_2)^2 \geqslant 0$. Аналогичное неравенство получим при $\rho_1 = \rho_2 = -1$. Итак, условие, что $x^2 \leqslant L$, доказано. Отсюда должно выполняться условие $x_1^2 \leqslant L$, что эквивалентно неравенству $\frac{A_1}{A_2} \leqslant \sqrt{\frac{c_1^2+1}{c_2^2+1}} = k$. Выполнение условия $x_2^2 \leqslant L$ эквивалентно тому, что $\frac{A_2}{A_1} \leqslant \sqrt{\frac{c_2^2+1}{c_1^2+1}} = \frac{1}{k}$.

Итак, когда $A_1 \leqslant k\, A_2$, действителен корень x_1, в противном случае нужно использовать x_2. Из соотношений (1) и (2) и с учетом (3) или (4) нетрудно видеть, что при $x > x_1$ или $x > x_2$ $\frac{da}{d\rho_1} > 0$ и $\frac{da}{d\rho_2} > 0$. При $x < x_1$ или $x < x_2$ $\frac{da}{d\rho_1} < 0$ и $\frac{da}{d\rho_2} < 0$. Следовательно, при $x = x_1$ или $x = x_2$ функция a достигает своего минимума.

Приложение 2

Здесь будут рассмотрены выводы соотношений для элементов ковариационной матрицы Σ 4-мерного случайного вектора в пространстве признаков x_1, x_1^2, x_2, x_2^2. Найдем соотношение для $\sigma_{22}^{(1)} + \sigma_{22}^{(2)}$. По определению,

$$\sigma_{22}^{(1)} = M\,[(x_1^2 - M\,[x_1^2])^2] = M\,[x_1^4] - M^2\,[x_1^2].$$

Определим $M[x_1^2]$. Известно [5], что центрированная и нормированная по дисперсии величина $\overset{\circ}{x}_1 = \left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right)^2$ распределена по закону χ^2 с параметрами (1, 2). Используем это в дальнейшем. Предварительно, раскрыв скобки, найдем, что

$$x_1^2 = \sigma_1^{(1)2}\,\overset{\circ}{x}_1 + 2m_{11}x_1 - m_{11}^2, \tag{1}$$

отсюда

$$M[x_1^2] = \sigma_1^{(1)2} + m_{11}^2.$$

Определим $M[x_1^4]$. Из соотношения (1) следует:

$$x_1^4 = \sigma_1^{(1)4}\,\overset{\circ}{x}{}_1^2 + 4x_1^2 m_{11}^2 + m_{11}^4 + 4\sigma_1^{(1)2}\,\overset{\circ}{x}_1 x_1 m_{11} - 2m_{11}^2\sigma_1^{(1)2}\,\overset{\circ}{x}_1 - 4m_{11}^3 x_1;$$

отсюда

$$M[x_1^4] = \sigma_1^{(1)4} M[\overset{\circ}{x}{}_1^2] + 4m_{11}^2 M[x_1^2] + m_{11}^4 +$$
$$+ 4\sigma_1^{(1)2} m_{11} M[\overset{\circ}{x}_1 x_1] - 2m_{11}^2\sigma_1^{(1)2} M[\overset{\circ}{x}_1] - 4m_{11}^3 M[x_1].$$

Найдем $M[\overset{\circ}{x}{}_1^2]$. Дисперсия $\overset{\circ}{x}{}_1^2$ равна Д $[\overset{\circ}{x}{}_1^2] = M[\overset{\circ}{x}{}_1^2] - M^2[\overset{\circ}{x}_1] = 2$.
Отсюда

$$M[\overset{\circ}{x}{}_1^2] = 2 + M^2[\overset{\circ}{x}_1] = 3.$$

Определим $M[\overset{\circ}{x}_1 x_1]$. Для этого рассмотрим произведение $x_1\left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right)^2$.

Представим его в виде

$$\left[\left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right) + \frac{m_{11}}{\sigma_1^{(1)}}\right]\sigma_1^{(1)}\left[\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right]^2.$$

Отсюда

$$M[\overset{\circ}{x}_1 x_1] = \sigma_1^{(1)} M\left[\left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right)^3\right] + m_{11} M\left[\left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right)^2\right].$$

Учитывая, что x_1 распределено нормально, имеем $M\left[\left(\frac{x_1 - m_{11}}{\sigma_1^{(1)}}\right)^3\right] = 0$. Окончательно $M[x_1\overset{\circ}{x}_1] = m_{11}$. Подставляя соотношения для $M[\overset{\circ}{x}{}_1^2]$ и $M[\overset{\circ}{x}_1, x_1]$, найдем, что $M[x_1^4] = m_{11}^4 + 6m_{11}^2\sigma_1^{(1)2} + 3\sigma_1^{(1)4}$. Итак, $\sigma_{22}^{(1)} = M[x_1^4] - M^2[x_1^2] = 4m_{11}^2\sigma_1^{(1)2} + 2\sigma_1^{(1)4}$. Соответственно $\sigma_{22}^{(2)} = 4m_{12}^2\sigma_1^{(2)2} + 2\sigma_1^{(2)4}$. По аналогии $\sigma_{44}^{(1)} = 4m_{21}^2\sigma_2^{(1)2} + 2\sigma_2^{(1)4}$, $\sigma_{44}^{(2)} = 4m_{22}^2\sigma_2^{(2)2} + 2\sigma_2^{(2)4}$.

Найдем $\left(\sigma_{14}^{(1)} + \sigma_{14}^{(2)}\right)$. По определению, $\sigma_{14}^{(1)} = M\left[(x_1 - M[x_1])\left(x_2^2 - M[x_2^2]\right)\right]$. Раскрыв скобки и используя соотношения $M[x_1] = m_{11}$, $M[x_2^2] = m_{21}^2 + \sigma_2^{(1)2}$, получим $\sigma_{14}^{(1)} = M\left[x_1x_2^{(2)}\right] - m_{11}\left(m_{21}^2 + \sigma_2^{(1)2}\right)$. Будем считать, что совместная плотность распределения $W(x_1, x_2, x_2)$ является нормальной; тогда [6] $M[(x_1 - m_{11})(x_2 - m_{21})(x_2 - m_{21})] = 0$. Раскрыв скобки и учитывая, что $M[x_1x_2] = \rho_1\sigma_1^{(1)}\sigma_2^{(1)} + m_{11}m_{21}$ найдем, что $M\left[x_1x_2^2\right] = 2m_{21}\rho_1\sigma_1^{(1)}\sigma_2^{(1)} + m_{11}\sigma_2^{(1)2} + m_{11}m_{21}^2$. Отсюда $\sigma_{14}^{(1)} = 2m_{21}\rho_1\sigma_1^{(1)}\sigma_2^{(1)}$. Аналогично $\sigma_{14}^{(2)} = 2m_{22}\rho_2\sigma_1^{(2)}\sigma_2^{(2)}$.

При $x_1 = x_2$ получим $\sigma_{12}^{(1)} = 2m_{11}\sigma_1^{(1)2}$ и $\sigma_{12}^{(2)} = 2m_{12}\sigma_1^{(2)2}$. По аналогии $\sigma_{34}^{(1)} = 2m_{21}\sigma_2^{(1)2}$ и $\sigma_{34}^{(2)} = 2m_{22}\sigma_2^{(2)2}$.

Найдем $\sigma_{24}^{(1)} + \sigma_{24}^{(2)}$. По определению, $\sigma_{24}^{(1)} = M\left[\left(x_1^2 - M[x_1^2]\right)\left(x_2^2 - M[x_2^2]\right)\right]$. Раскрыв скобки и используя значения $M[x_1^2]$ и $M[x_2^2]$, получим $\sigma_{24}^{(1)} = M\left[x_1^2x_2^2\right] - \left(m_{11}^2 + \sigma_1^{(1)2}\right)\left(m_{21}^2 + \sigma_2^{(1)2}\right)$.

Определим $M\left[x_1^2\, x_2^2\right]$. Предварительно примем допущение, что совместная плотность вероятности $W(x_1,\ x_2,\ x_1,\ x_2)$ является нормальной; тогда [6]

$$M\left[(x_1 - m_{11})(x_2 - m_{21})(x_1 - m_{11})(x_2 - m_{21})\right] = \sigma_1^{(1)^2}\sigma_2^{(1)^2} + 2\,\rho_1^2\,\sigma_1^{(1)^2}\sigma_2^{(1)^2}.$$

Раскрыв скобки в левой части равенства и подставляя значение $M\left[x_1 x_2^2\right]$, получим

$$M\left[x_1^2\, x_2^2\right] = \sigma_1^{(1)^2}\sigma_2^{(1)^2} + 2\,\rho_1^2\,\sigma_1^{(1)^2}\sigma_2^{(1)^2} + m_{11}^2\, m_{21}^2 +$$
$$+ m_{21}^2\,\sigma_1^{(1)^2} + 4\,m_{11}\,m_{21}\,\rho_1\,\sigma_1^{(1)}\sigma_2^{(1)} + m_{11}^2\,\sigma_2^{(1)^2}.$$

Окончательно $\sigma_{24}^{(1)} = 4\,m_{11}\,m_{21}\,\rho_1\,\sigma_1^{(1)}\sigma_2^{(1)} + 2\,\rho_1^2\,\sigma_1^{(1)^2}\sigma_2^{(1)^2}$. Соответственно $\sigma_{24}^{(2)} = 4\,m_{12}\,m_{22}\,\rho_2\,\sigma_1^{(2)}\sigma_2^{(2)} + 2\,\rho_2^2\,\sigma_1^{(2)^2}\sigma_2^{(2)^2}$.

ЛИТЕРАТУРА

1. А. Г. Француз. Некоторые вопросы статистической теории опознания образов.— В сб. «Бионика». М., «Наука», 1965.
2. Н. Нильсон. Обучающиеся машины. М., «Мир», 1967.
3. Яу, Лин. О верхней границе вероятности ошибки линейного классификатора вероятностных изображений.— Труды Института инженеров по электротехнике и радиоэлектронике, 1968, т. 56, № 3.
4. Patterson. An Adaptive Pattern Classification System.— IEEE Trans. on System Science and Cybernetics, 1966, v. SSC-2, № 1.
5. Г. Крамер. Математические методы статистики. М., Изд-во иностр. лит., 1948.
6. В. С. Пугачев. Теория случайных функций. М., «Наука», 1960.

Поступила в редакцию
18 июля 1969 г.