

ОБРАБОТКА ИЗОБРАЖЕНИЙ: МЕТОДЫ И ПРИКЛАДНЫЕ ЗАДАЧИ

О МОДИФИКАЦИЯХ МЕТОДА ПОСТРОЕНИЯ ЛИНЕЙНОЙ ДИСКРИМИНАНТНОЙ ФУНКЦИИ, ОСНОВАННОГО НА ПРОЦЕДУРЕ ПЕТЕРСОНА-МАТТСОНА

Мясников В.В.

Самарский государственный аэрокосмический университет,

Институт систем обработки изображений РАН

Аннотация

Рассматривается ряд новых модификаций известного метода построения линейной дискриминантной функции, минимизирующей вероятность ошибочной классификации, и основанного на процедуре Петерсона-Маттсона. Модификации метода касаются расширения его области применения для ряда типичных ситуаций, в которых известный метод не дает решений. В частности, на ситуацию построения классификатора по критерию Неймана-Пирсона или другим критериям, одновременно минимизирующим вероятности ошибок, на ситуацию совпадающих средних признаков в классах. Для последней ситуации показано, что задача сводится к нахождению ортогонального векторного пространства к векторам – строкам матрицы, полученной как взвешенная сумма ковариационных матриц признаков в классах.

Введение

Процедура вычисления параметров линейной дискриминантной функции $d(x)$, минимизирующей величину общего риска, была предложена D. Peterson и R. Mattson в работе [1] и описана также в монографии [2]. Эта процедура разрабатывалась в предположении, что значения дискриминантной функции имеют нормальный закон распределения в каждом из классов. Кроме того, известная процедура нахождения параметров имеет ряд ограничений для достаточно типичных на практике ситуаций, например, когда средние признаки в классах близки или совпадают. Данная работа расширяет область применения известного метода построения дискриминантной функции за счет модификации процедуры Петерсона-Маттсона.

Работа организована следующим образом. Метод построения линейной дискриминантной функции, основанный на известной процедуре Петерсона-Маттсона, представлен в первом разделе. Во втором разделе будет доказано, что известная процедура дает решение не только для задачи минимума общего риска (линейной оптимизации), но также и для задачи Неймана-Пирсона построения линейной дискриминантной функции. Некоторые замечания о диапазоне параметра, используемого в процедуре Петерсона-Маттсона, будут сделаны в третьем разделе. Будет показано, что ограничения, налагаемые на область значений параметра в известной процедуре, не являются корректными. Будут указаны корректные границы для этого параметра, а также достаточные условия, при которых известные границы являются корректными. Наконец, в четвертом и пятом разделах будет предложена процедура получения параметров линейной дискриминантной функции для ситуации, когда математические ожидания вектора признаков, вычисленные для различных классов совпадают.

1. Метод построения линейной дискриминантной функции, основанный на процедуре Петерсона-Маттсона

Пусть линейная дискриминантная функция (ЛДФ) $d(\vec{x}) = \vec{W}^T \vec{x} + w_N$ имеет нормальный закон распределения в каждом из классов. Это может быть обусловлено следующими причинами:

- если вектор признаков \vec{X} имеет нормальный закон распределения в каждом из классов, тогда линейная комбинация элементов вектора также имеет нормальный закон распределения;
- если вектор признаков \vec{X} не имеет нормальный закон распределения, но количество признаков достаточно большое, тогда (при достаточно общих условиях) в силу центральной предельной теоремы линейная комбинация элементов этого вектора также имеет нормальный закон распределения.

При выполнении указанного допущения поведение случайных значений функции $d(\vec{X})$ в каждом из классов полностью характеризуется математическим ожиданием и дисперсией этой функции в каждом из классов, а именно:

$$\begin{aligned} m_i &= M\left(d(\vec{X})/\Omega_i\right) = \vec{W}^T \vec{M}_i + w_N, \\ \sigma_i^2 &= D\left(d(\vec{X})/\Omega_i\right) = \vec{W}^T B_i \vec{W}, \quad i = \{0,1\}. \end{aligned} \quad (1)$$

Здесь \vec{M}_i, B_i – это вектор математических ожиданий и ковариационная матрица вектора признаков в классе Ω_i , соответственно.

Пусть дискриминантная функция строится из условия минимума общего риска классификации (по параметрам дискриминантной функции):

$$R_{\min} = r_0 P(Q/\Omega_0) + r_1 P(\bar{Q}/\Omega_1) \rightarrow \min_{\vec{W}, w_N}, \quad (2)$$

где событие $Q = \{d(\bar{X}) > 0\}$. С учетом того, что дискриминантная функция имеет нормальный закон распределения, данный критерий эквивалентен следующему:

$$r_1 \Phi\left(-\frac{m_1}{\sigma_1}\right) - r_0 \Phi\left(-\frac{m_0}{\sigma_0}\right) \rightarrow \min_{W, w_N}.$$

Здесь $\Phi(\cdot)$ – функция Лапласа. Необходимым условием минимума, как известно, является равенство нулю производной критерия:

$$r_1 \Phi'\left(-\frac{m_1}{\sigma_1}\right) - r_0 \Phi'\left(-\frac{m_0}{\sigma_0}\right) = 0.$$

Тогда, учитывая, что выполняются равенства [1, 2]

$$\begin{aligned} \frac{\partial}{\partial W} \left(-\frac{m_i}{\sigma_i}\right) &= -\frac{M_i}{\sigma_i} + \frac{W^T B_i m_i}{\sigma_i^3}, \\ \frac{\partial}{\partial w_N} \left(-\frac{m_i}{\sigma_i}\right) &= -\frac{1}{\sigma_i}, \quad i = \overline{0,1}, \end{aligned} \quad (3)$$

получим систему уравнений:

$$\begin{cases} \frac{r_0}{\sqrt{2\pi\sigma_0}} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) \left(M_0 - \frac{m_0}{\sigma_0^2} W^T B_0\right) = \\ = \frac{r_1}{\sqrt{2\pi\sigma_1}} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right) \left(M_1 - \frac{m_1}{\sigma_1^2} W^T B_1\right), \\ \frac{r_0}{\sqrt{2\pi\sigma_0}} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) = \frac{r_1}{\sqrt{2\pi\sigma_1}} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right). \end{cases} \quad (4)$$

Из этой системы уравнений очевидно следующее равенство:

$$\bar{M}_0 - \frac{m_0}{\sigma_0^2} W^T B_0 = \bar{M}_1 - \frac{m_1}{\sigma_1^2} W^T B_1.$$

или

$$\left(\frac{m_1}{\sigma_1^2} B_1 - \frac{m_0}{\sigma_0^2} B_0\right) W = \bar{M}_1 - \bar{M}_0. \quad (5)$$

Явного выражения для решения такого соотношения получить не удастся, поскольку величины m_i, σ_i^2 ($i = \overline{0,1}$), задаваемые соотношениями (1), зависят от вектора коэффициентов \bar{W} линейной дискриминантной функции.

Для численного решения выражения (5) D. Peterson и R. Mattson в 1966 году предложили подход, позволяющий свести решение системы нелинейных уравнений (5) к итерационной поисковой процедуре, которая зависит от одного скалярного параметра. Предложенный ими подход состоит в следующем.

Вектор \bar{W} определяется как функция некоторого числового параметра s :

$$\bar{W} = a[sB_0 + (1-s)B_1]^{-1}(\bar{M}_1 - \bar{M}_0), \quad (6)$$

где

$$s = \frac{-m_0/\sigma_0^2}{m_1/\sigma_1^2 - m_0/\sigma_0^2}, \quad a = m_1/\sigma_1^2 - m_0/\sigma_0^2. \quad (7)$$

Тогда для порогового значения дискриминантной функции можно получить следующее выражение:

$$w_N = -\frac{s\sigma_0^2 \bar{M}_1^T + (1-s)\sigma_1^2 \bar{M}_0^T}{s\sigma_0^2 + (1-s)\sigma_1^2} \bar{W}. \quad (8)$$

Оно может быть использовано для расчета w_N при известном векторе коэффициентов \bar{W} линейной дискриминантной функции. Из этого выражения очевидно, что умножение вектора коэффициентов \bar{W} на произвольное положительное число ведет к пропорциональному увеличению величины w_N . Таким образом, при расчете параметров классификатора произвольным положительным множителем у коэффициентов \bar{W} линейной дискриминантной функции можно пренебречь, поскольку этот множитель меняет только «масштаб» значений дискриминантной функции, но не меняет классификацию.

Анализируя второе соотношение в системе уравнений (4), авторы предположили, что оптимальное решение приведет к тому, что величины m_0 и m_1 будут иметь разные знаки, то есть $m_0 < 0, m_1 > 0$. В этом случае величина a в выражении (6) – это положительный масштабный множитель. В соответствии с ранее сделанными замечаниями этим масштабным множителем можно пренебречь.

Из предположения авторов о разных знаках средних дискриминантной функции в классах $m_0 < 0, m_1 > 0$ также следует, что диапазон величины $s \in [0,1]$. Тогда возможные значения коэффициентов \bar{W} линейной дискриминантной функции задаются соотношением:

$$\bar{W} = [sB_0 + (1-s)B_1]^{-1}(\bar{M}_0 - \bar{M}_1), \quad s \in [0,1]. \quad (9)$$

Окончательно значение параметров линейной дискриминантной функции определяется численно из решения оптимизационной задачи, в которой всего один независимый параметр оптимизации s :

$$r_1 \Phi\left(-\frac{m_1}{\sigma_1}\right) - r_0 \Phi\left(-\frac{m_0}{\sigma_0}\right) \rightarrow \min_s, \quad (10)$$

а параметры линейной дискриминантной функции вычисляются в соответствие с выражениями (1), (8) и (9). Значение общего риска при таком подходе может также оцениваться по выборке.

Существует обобщение изложенного подхода для произвольного вида функционала качества [2], которое, однако, приводит только к системе уравнений и не дает окончательного решения. Одним из частных решений, полученных таким образом, является классификатор Фишера.

**2. Определение параметров ЛДФ,
оптимальной по критерию Неймана-Пирсона.
Другие критерии оптимальности**

Во многих задачах вместо линейного критерия оптимизации используется критерий Неймана-Пирсона, который состоит в следующем. Фиксируется вероятность ложного обнаружения, и для этого фиксированного значения находятся такие параметры классификатора, которые минимизируют вероятность пропуска объекта. Таким образом, при построении ЛДФ, оптимальной по критерию Неймана-Пирсона решается следующая задача оптимизации:

$$\begin{cases} p_1 \rightarrow \min, \\ p_0 = p_0^*. \end{cases} \quad (11)$$

Заметим, что различия в критериях Неймана-Пирсона и критерия минимума общего риска возникают только тогда, когда Паретто-оптимальное множество значений вероятностей ошибок (p_0, p_1) имеет невыпуклый характер. Если множество значений вероятностей ошибок является выпуклым, то критерии Неймана-Пирсона и минимума риска эквивалентны при соответствующем выборе параметров критериев – величин r_1, r_0 и p_0^* [2]. В альтернативной ситуации линейная задача оптимизации не позволяет получить все решения на границе множества, а задача оптимизации по критерию Неймана-Пирсона позволяет это сделать.

Итак, пусть решается задача (11) для ЛДФ, и все описанные выше допущения на законы распределения значений ДФ справедливы. Тогда задача оптимизации может быть переписана с использованием метода множителей Лагранжа в виде следующей безусловной целевой функции вида:

$$p_1 + \mu(p_0 - p_0^*)^2 \rightarrow \min_{\bar{W}, w_N},$$

здесь μ – множитель Лагранжа ($\mu \gg 0$). Необходимое условие минимума запишется:

$$p_1' = 2\mu(p_0^* - p_0)p_0'.$$

Тогда если

$$p_1 = F_d\left(-\frac{m_1}{d_1}/\Omega_1\right), \quad p_0 = 1 - F_d\left(-\frac{m_0}{d_0}/\Omega_0\right),$$

получим следующее уравнение:

$$\begin{aligned} f_1\left(-\frac{m_1}{d_1}\right)\left(-\frac{m_1}{d_1}\right)' &= \\ &= 2\mu\left(1 - p_0^* - F_d\left(-\frac{m_0}{d_0}\right)\right)f_0\left(-\frac{m_0}{d_0}\right)\left(-\frac{m_0}{d_0}\right)'. \end{aligned}$$

Подставляя в это равенство выражения для производных $\left(-m_i/d_i\right)'$ по вектору коэффициентов ЛДФ \bar{W} и порогу w_N из (3), снова приходим к аналогичной (4) системе уравнений:

$$\begin{cases} \frac{1}{\sigma_0} f_0\left(-\frac{m_0}{\sigma_0}\right)\left(\bar{M}_0 - \frac{m_0}{\sigma_0^2} \bar{W}^T B_0\right) = \\ = 2\mu\left(1 - p_0^* - F_d\left(-\frac{m_0}{d_0}\right)\right)f_1\left(-\frac{m_1}{\sigma_1}\right)\frac{1}{\sigma_1} \cdot \\ \cdot \left(\bar{M}_1 - \frac{m_1}{\sigma_1^2} \bar{W}^T B_1\right), \\ \frac{1}{\sigma_0} f_0\left(-\frac{m_0}{\sigma_0}\right) = \\ = 2\mu\left(1 - p_0^* - F_d\left(-\frac{m_0}{d_0}\right)\right)f_1\left(-\frac{m_1}{\sigma_1}\right)\frac{1}{\sigma_1}. \end{cases}$$

Учитывая второе равенство, из первого соотношения получаем уравнение (5) относительно вектора \bar{W} . Результатом рассуждений снова окажется итерационная процедура Петерсона-Матсона.

Тот факт, что, используя рассматриваемую процедуру, для любого значения вероятности ложной тревоги можно найти соответствующее ей наименьшее значение вероятности пропуска объекта означает, что процедура Петерсона-Матсона позволяет получить все Паретто-оптимальное множество решений. Следовательно, любую задачу построения ЛДФ с критерием, требующим совместного понижения значений вероятностей p_0 и p_1 , можно решить, используя процедуру Петерсона-Матсона – то есть простым перебором значений величины s . Этот факт очень удобен при построении ЛДФ на базе других критериев – нет необходимости разрабатывать для них новую процедуру поиска параметров ЛДФ.

**3. Диапазон параметра оптимизации
процедуры Петерсона-Матсона**

В работах [1,2] диапазон параметра оптимизации s , определенного выражением (7), принимается следующим: $s \in [0,1]$. Такое ограничение следовало из выполнения в системе уравнений (4) равенства

$$\frac{r_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) = \frac{r_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right),$$

которое можно интерпретировать, как

$$\begin{aligned} \frac{r_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(u - m_0)^2}{2\sigma_0^2}\right) &= \\ &= \frac{r_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(u - m_1)^2}{2\sigma_1^2}\right) \Big|_{u=0}. \end{aligned}$$

Последнее равенство имеет ясную геометрическую трактовку, представленную на рис. 1.

А именно, процедура Петерсона-Матсона так выбирает параметры ЛДФ, чтобы математические ожидания законов распределений различных классов располагались по разные стороны от начала ко-

ординат, причем их значения в начале координат совпадали. Тогда оптимальным пороговым значением, следующим также и из байесовской теории оптимизации, будет начало координат.

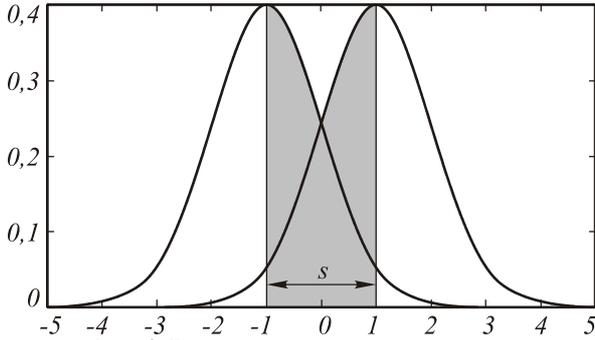


Рис. 1. Геометрическая интерпретация необходимого условия экстремума: ситуация высокой разделимости классов

В такой ситуации можно считать справедливым неравенство $m_0 < 0 < m_1$, и значение s действительно принадлежит вышеуказанному диапазону. На рис. 2 представлено изменение положения ДЛФ в двумерном пространстве признаков по мере изменения значения параметра.

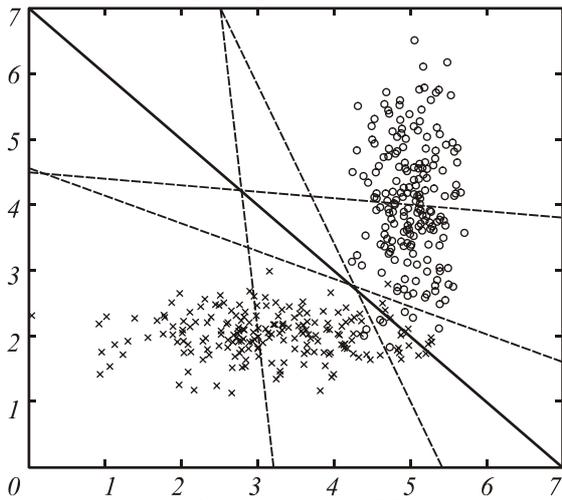


Рис. 2. Расположение ДЛФ, построенной с использованием процедуры Петерсона-Маттсона для значений параметра $s=0, 0.25, 0.5, 0.75$ и 1

К сожалению, на практике часто возникают ситуации, когда классы плохо разделимы. Примеры таких ситуаций для одномерного и двумерного случаев приведены на рис. 3 и 4, соответственно. В частности, для одномерного случая, представленного на рис. 3, изменение параметра s в диапазоне $[0,1]$ приведет к изменению реального порогового значения дискриминантной функции только в выделенном на рисунке отрезке. Очевидно, такое положение порогового значения не является оптимальным. В частности, байесовская теория оптимизации предполагает простановку границ в точках выполнения равенства:

$$r_0 f\left(\frac{x}{\Omega_0}\right) = r_1 f\left(\frac{x}{\Omega_1}\right).$$

Ситуация, приведенная на рис. 3 реально возникает, если, например, одно из весовых значений в целевой функции существенно больше другого, или если один из классов существенно больше «размыт» в признаковом пространстве, а также если средние векторы признаков в классах близки. А поскольку это достаточно реальные ситуации, то необходимо, оставаясь в рамках ЛДФ, уметь находить наилучшие решения. При этом заметим, что в этом случае возможно получения не одного, а нескольких локально-оптимальных решений, среди которых можно отобрать наилучшую ЛДФ.

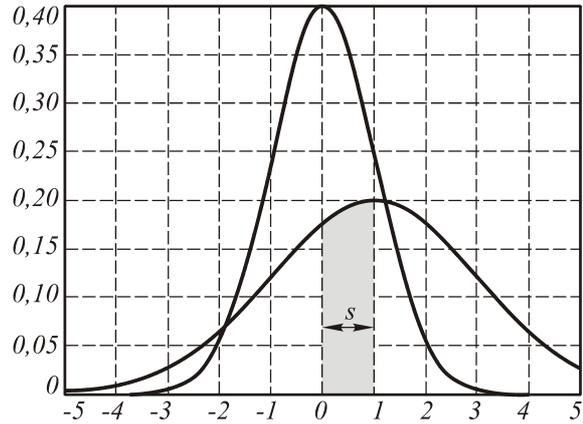


Рис. 3. Геометрическая интерпретация необходимого условия экстремума: ситуация низкой разделимости классов: $M_0=0, D_0=1, M_1=1, D_1=4, r_0=r_1$

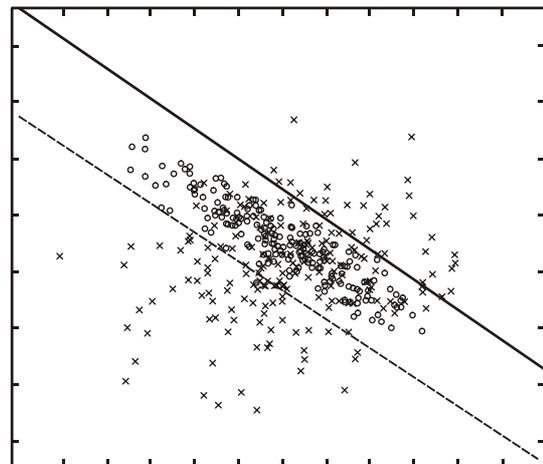


Рис. 4. Множество локальных решений ЛДФ, соответствующих параметру оптимизации s со значениями $-0,024$ и $-0,094$

Для обобщения известной процедуры необходимо расширить область параметра оптимизации s на все вещественную ось: $s \in R$. На рис. 4 приведен пример для одной из таких ситуаций для двумерного вектора признаков и при близких значениях средних признаков в классах. Как видно, оптимальное значение параметра оптимизации выходит за границы $[0,1]$, используемые в известном методе.

Из графических иллюстраций, приведенных на рис. 1 и 3 легко видеть, что традиционные границы

параметра оптимизации $[0,1]$ справедливы лишь в той ситуации, когда максимальное значение плотности вероятностей, умноженной на соответствующее значение штрафа, одного класса больше значения плотности вероятностей, умноженной на соответствующее значение штрафа, другого класса при тех же аргументах. Таким образом, достаточное условие того, что известные границы параметра оптимизации $s \in [0,1]$ являются корректными представимо формально в виде системы неравенств:

$$\begin{cases} \frac{r_0}{\sigma_0} > \frac{r_1}{\sigma_1} \exp\left(-\frac{(m_0 - m_1)^2}{2\sigma_1^2}\right), \\ \frac{r_1}{\sigma_1} > \frac{r_0}{\sigma_0} \exp\left(-\frac{(m_0 - m_1)^2}{2\sigma_0^2}\right). \end{cases} \Rightarrow s \in [0,1].$$

Следует отметить, что расширение границ значений параметра оптимизации решает проблему нахождения оптимальные значения ЛДФ для двух обозначенных выше ситуаций: когда одно из весовых значений в целевой функции существенно больше другого, и когда один из классов существенно больше «размыт» в признаковом пространстве. Для ситуации совпадающих средних данный подход не позволяет получить решения в силу однородности системы линейных уравнений (5). В этой ситуации, очевидно, следует использовать другой подход. Он изложен в следующем разделе.

4. Построение ЛДФ

при совпадающих средних признаков в классах

Пусть средние векторов признаков в разных классах совпадают:

$$M_0 = M_1 = M.$$

Подставим это равенство в выражение (5), тогда последнее может быть переписано в виде:

$$m\bar{W}^T B_0 / \sigma_0^2 = m\bar{W}^T B_1 / \sigma_1^2,$$

здесь $m_0 = m_1 \equiv m = \bar{W}^T M + w_N$. Перепишем его следующим образом:

$$(B_0 / \sigma_0^2 - B_1 / \sigma_1^2) \bar{W} = 0. \quad (12)$$

Решить такое уравнения, также как и уравнение (5), аналитически невозможно. Поэтому введем, как в процедуре Петерсона-Матсона, переменную s :

$$s = -(1/\sigma_0^2 - 1/\sigma_1^2)^{-1} / \sigma_1^2 = \sigma_0^2 / (\sigma_0^2 - \sigma_1^2).$$

Тогда, с точностью до ненулевого масштабного множителя, условие (12) переписется в виде однородной системы линейных уравнений, зависящей от параметра:

$$(sB_1 + (1-s)B_0) \bar{W} = 0 \quad (13)$$

Заметим, что диапазон изменения переменной s , определяется всей числовой прямой $s \in R$ за исключением отрезка $[0,1]$. Кроме того, для конкретного значения s вектор \bar{W} может не существовать.

Нетривиальное (ненулевое) решение однородной системы линейных уравнений существует тогда и только тогда, когда определитель матрицы

$$B_\Sigma^s = sB_1 + (1-s)B_0. \quad (14)$$

оказывается нулевым $\det(B_\Sigma^s) = 0$. Тогда уравнение (13) относительно вектора \bar{W} есть определение ортогонального векторного пространства к набору векторов-строк матрицы B_Σ^s . В свою очередь, матрица B_Σ^s есть результат взвешенной суммы ковариационных матриц векторов признаков в классах. Учитывая, что матрицы B_1 и B_0 невырожденные, условие (13) можно преобразовать к виду:

$$\det(B_1 B_0^{-1} + ((1-s)/s) \cdot I) = 0 \quad (15)$$

здесь I – единичная матрица. Значения величины $(1-s)/s$ лежат на всей числовой оси, за исключением точки $s = 0$. Но при этом значении определитель матрицы (14) отличен от нуля, так как $B_\Sigma^s|_{s=0} = B_0$.

Поэтому можно переписать (15) в виде:

$$\det(\check{B}_\Sigma^s) = 0, \text{ где } \check{B}_\Sigma^s = B_1 B_0^{-1} + sI, s \in R. \quad (16)$$

Далее, в задаче используется N признаков, а размер матрицы $B_1 B_0^{-1}$ составляет $N \times N$. Тогда можно переписать условие (16) в виде (степенного) уравнения N -ого порядка относительно величины s . Максимальное количество линейно независимых векторов-решений \bar{W} уравнения (16) не превышает N , причем при одном значении переменной s (корне) число решений равно кратности корня. Сам же вектор \bar{W} может легко быть найден с помощью процедуры Грамма-Шмидта ортогонализации набора векторов: строк матрицы (16) и некоторого произвольного вектора \bar{W} . Окончательно вектор весовых коэффициентов ЛДФ \bar{W} выбирается из множества локальных решений:

$$\bar{W} \in \left\{ \bar{W}: \check{B}_\Sigma^s \bar{W} = 0, \text{ где } s = \arg \det(\check{B}_\Sigma^s) = 0 \right\}_{s \in R}.$$

Пусть далее вектор \bar{W} найден. Для получения окончательного ответа необходимо определить значение порогового значения w_N . В общем случае, когда распределение значений дискриминантной функции отлично от нормального (см. раздел 2), пороговое значение ЛДФ может легко быть найдено численными методами, например, по выборке. Для случая нормального закона распределения значений дискриминантной функции можно получить аналитическое выражение следующим образом. Это выражение следует из второго уравнения системы (5) при условии равных средних $m_0 = m_1 \equiv m$:

$$\frac{r_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{m^2}{2\sigma_0^2}\right) = \frac{r_1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{m^2}{2\sigma_1^2}\right).$$

Преобразуя это выражение, имеем:

$$m = \pm \sigma_0 \sigma_1 \sqrt{2(\sigma_1^2 - \sigma_0^2)^{-1} \ln(r_0 \sigma_1 / r_1 \sigma_0)}.$$

Учитывая, что $m = W^T M + w_N$, получим окончательное явное выражение для искомой пороговой величины ЛДФ:

$$w_N = \pm \sigma_0 \sigma_1 \sqrt{\frac{2}{\sigma_1^2 - \sigma_0^2} \ln\left(\frac{r_0 \sigma_1}{r_1 \sigma_0}\right)} - W^T M. \quad (17)$$

Следует отметить, что задача поиска оптимального значения порогового значения w_N неразрешима, если подкоренное выражение оказывается отрицательным. Это может произойти, если выполняется одно из следующих соотношений:

$$\sigma_1 > \sigma_0 \text{ и } r_0 \sigma_1 < r_1 \sigma_0$$

или

$$\sigma_1 < \sigma_0 \text{ и } r_0 \sigma_1 > r_1 \sigma_0.$$

На рис. 5 представлена графическая интерпретация подобной ситуации, которая возникает, когда штрафы за решения выбраны таким образом, что минимум риска достигается только в ситуации, если все пространство классифицируется в какой-то один класс.

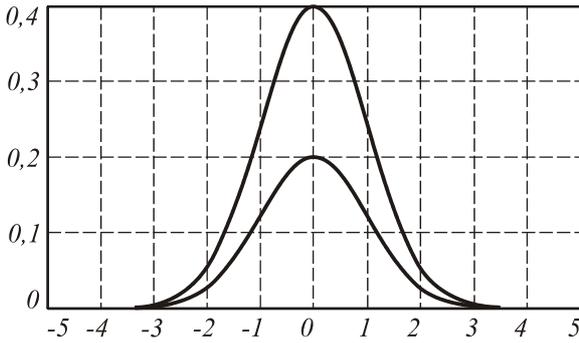


Рис. 5. Пример с отсутствием решения:
 $M_0 = M_1 = 0, D_0 = D_1 = 1, r_0 = 1, r_1 = 0,5$

Подобная вырожденная ситуация невозможна, если риски за ошибки одинаковы и/или решение принимается на основе анализа только плотностей законов распределений. Единственная аномалия возникает, когда плотности совпадают, но тогда абсолютно бессмысленно выбирать какое-либо конкретное решение, так как все они дают одинаковую величину общего риска.

5. Построение ЛДФ при совпадающих средних признаков в классах для 2-D случая

Рассмотрим наиболее простую ситуацию, когда пространство признаков двумерно. Тогда можно получить окончательное аналитическое решение для классификатора. Итак, пусть матрицы B_1 и B_0 имеют вид:

$$B_0 = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \quad B_1 = \begin{bmatrix} d & f \\ f & e \end{bmatrix},$$

причем выполняются равенства $ab > c^2$ и $de > f^2$.

Тогда матрица B_Σ :

$$B_\Sigma = \begin{bmatrix} sa + (1-s)d & sc + (1-s)f \\ sc + (1-s)f & sb + (1-s)e \end{bmatrix}$$

Условие (13) существования ортогонального пространства преобразуется к виду:

$$(sa + (1-s)d)(sb + (1-s)e) - (sc + (1-s)f)^2 = 0.$$

Используем обозначения

$$A = (a-d)(b-e) - (c-f)^2, \quad C = de - f^2,$$

$$B = ae + bd + 2f^2 - 2cf - 2de$$

Тогда можно записать решение уравнения (16):

$$s_{1,2} = \left(-B \pm \sqrt{B^2 - 4AC} \right) / 2A.$$

Параметр w_N может быть определен из равенства (17).

Наиболее простое решение получается в ситуации, когда дисперсии по координатам одинаковы, то есть $a = b$ и $d = e$. Тогда:

$$B|_{s_1} = \gamma_1 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B|_{s_2} = \gamma_2 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

А соответствующие векторы линейной дискриминантной функции, определяющие ортогональное подпространство, таковы:

$$\bar{W}|_{s_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \bar{W}|_{s_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

На рис.6 и 7 приведены примеры для двумерного случая.

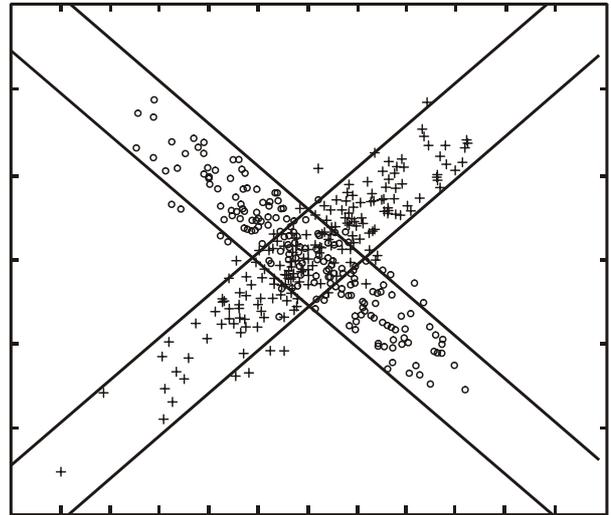


Рис. 6. Иллюстрация для $a = b = d = e$

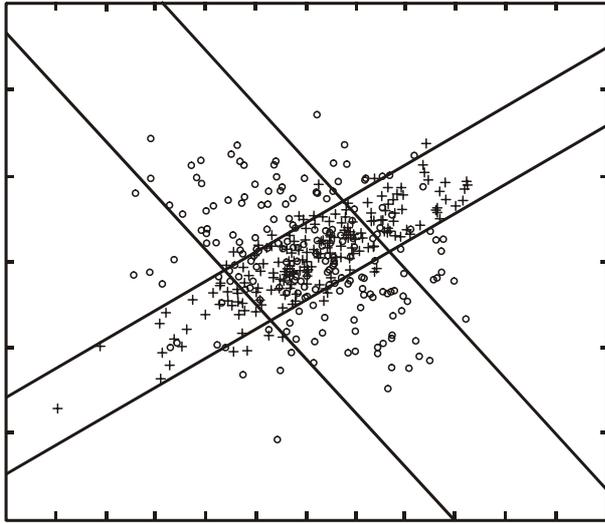


Рис. 7. Иллюстрация для $a = b \neq d = e$

Благодарности

Работа выполнена при поддержке Министерства образования РФ, Администрации Самарской области и Американского фонда гражданских исследований и развития (CRDF Project SA-014-02) в рамках российско-американской программы «Фундаментальные исследования и высшее образование» (BRHE).

Литература

1. Peterson D.YV., Mattson R.T. A method of finding linear discriminant functions for a class of performances criteria // IEEE Trans. Information Theory IT-12, 1966. P. 380-387.
2. Fukunaga K., Introduction to Statistical Pattern Recognition // New York & London.: Academic Press, 1972.
3. Амосов А.А., Дубинский Ю.А., Копленова Н.В. Вычислительные методы для инженеров // М.: Высш. шк., 1994. 544 с.
4. Мудров А.Е. Численные методы для ПЭВМ на языках Бейсик, Фортран, Паскаль // Томск: МП «РАСКО», 1991. 272 с.

On modifications of the method of constructing a linear discriminant function based on the Peterson-Mattson procedure

V.V. Myasnikov^{1,2}

¹ Samara State Aerospace University

² Image Processing Systems Institute of RAS

Abstract

The paper considers a number of new modifications of the well-known method for constructing a linear discriminant function that minimizes the probability of misclassification and is based on the Peterson-Mattson procedure. The modifications are related to the expansion of its scope to include a number of typical situations that are not solved by the known method. In particular, this applies to the situation of developing a classifier according to the Neumann-Pearson criterion or other criteria that simultaneously minimize the probability of errors, and to the situation of coinciding average attributes in classes. For the latter situation, it is shown that the problem is reduced to finding an orthogonal vector space to the vectors being the rows of the matrix obtained as a weighted sum of covariance matrices of attributes in classes.

Keywords: discriminant function, Peterson-Mattson procedure, Neumann-Pearson criterion, orthogonal vector, matrix.

Citation: Myasnikov VV. On modifications of the method of constructing a linear discriminant function based on the Peterson-Mattson procedure. Computer Optics 2004; 26: 74-80.

References

- [1] Peterson D., Mattson R. A method of finding linear discriminant functions for a class of performances criteria. IEEE Trans Inf Theory 1966; 12(3): 380-387.
- [2] Fukunaga K. Introduction to statistical pattern recognition. New York, London: Academic Press; 1972.
- [3] Amosov AA, Dubinsky YA, Koplenova NV. Computational methods for engineers [In Russian]. Moscow: "Vysshaya Shkola" Publisher; 1994.
- [4] Mudrov AE. Numerical methods for PC in Basic, Fortran, and Pascal [In Russian]. Tomsk: "MP Rasko" Publisher; 1991.