МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени академика С.П.КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

Байесовская классификация

Электронные методические указания
к лабораторной работе № 2

САМАРА
2010

Составители: КОЛОМИЕЦ Эдуард Иванович,

МЯСНИКОВ Владислав Валерьевич

В лабораторной работе № 2 по дисциплине «Математические методы распознавания образов и понимания изображений» изучаются методы построения классификаторов, основанных на оптимальных стратегиях, используемых при наличии различного количества априорной информации. В качестве примеров приводятся типовые задачи распознавания образов и изображений.

Методические указания предназначены для магистров направления 010400.68 “Прикладная математика и информатика”, обучающихся по программе «Математические и компьютерные методы обработки изображений и геоинформатики».

Цель работы - изучение теоретических основ и экспериментальное исследование методов построения оптимальных классификаторов для распознавания образов.

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЛАБОРАТОРНОЙ РАБОТЫ

1.1. Постановка задачи классификации

Пусть задано некоторое множество из I подлежащих распознаванию *объектов*:

$$\Omega = \{\omega_0, \omega_1, \ldots, \omega_{I-1}\},$$

и задано его разбиение на L непересекающихся подмножеств, называемых в дальнейшем *образами* или *классами*:

$$P_\Omega = \{\Omega_0, \Omega_1, \ldots, \Omega_{L-1}\}, \qquad \bigcup_{l=0}^{L-1} \Omega_l = \Omega .$$

Пусть каждый из объектов $\omega \in \Omega$ представляется набором числовых характеристик, называемого *вектором признаков:*

$$\bar{x} = (x_0, x_1, \ldots, x_{N-1})^T .$$

Задача классификации заключается в отыскании решающего правила, которое по заданному вектору признаков $\bar{x}(\omega)$ указывает, какому классу Ω_l принадлежит соответствующий объект ω. Построение такого решающего правила эквивалентно разбиению *метрического пространства признаков* $D = \{\bar{x} : \bar{x} \in D\}$ на множество непересекающихся областей:

$$P_D = \{D_0, D_1, \ldots, D_{L-1}\}, \qquad \bigcup_{l=0}^{L-1} D_l = D . \tag{1}$$

При этом решение о принадлежности некоторого объекта $\omega \in \Omega$ к классу Ω_l принимается в том случае, если соответствующий объекту вектор признаков $\bar{x}(\omega) \in D$ принадлежит области D_l.

- *Решающее правило, предназначенное для указания, какой области D_l признакового пространства D принадлежит предъявленный вектор признаков $\bar{x}$, называется классификатором.*

В идеале классификатор должен быть таким, чтобы области, выделяемые в пространстве признаков, соответствовали классам, то есть в идеале для элементов множеств P_Ω и P_D должно выполняться следующее условие: объект ω принадлежит классу Ω_l тогда и только тогда, когда соответствующий объекту вектор признаков $\overline{x}(\omega)$ принадлежит области D_l:

$$\forall\ \omega \in \Omega: \quad \omega \in \Omega_l \Leftrightarrow \overline{x}(\omega) \in D_l. \tag{2}$$

Как правило, на практике данное условие не выполняется, и существует вероятность неверно проклассифицировать объект или допустить ошибку при распознавании.

Обозначим p_{lj} $\left(l, j = \overline{0, L-1}\right)$ вероятность того, что классификатор принимает решение об отнесении вектора признаков некоторого объекта к области D_j, в то время как сам объект принадлежит классу Ω_l:

$$p_{lj} = P\left(\overline{X} \in D_j / \Omega_l\right). \tag{3}$$

При $l \neq j$ вероятности p_{lj} характеризуют ошибки распознавания и называются *вероятностями неверной* или *ошибочной классификации*, а при $l = j$ вероятности p_{ll} задают вероятности *верной (правильной) классификации* представителей соответствующего класса. Уменьшение вероятностей ошибочной классификации – это основная задача, которая возникает при построении классификатора.

Обычно классификатор задается не в виде областей признакового пространства (1), а в виде набора, так называемых, *дискриминантных* или *решающих функций* $d_l(\overline{x}(\omega))$, $\left(l = \overline{0, L-1}\right)$. При этом процесс принятия решения осуществляется по следующему правилу: объект считается принадлежащим тому классу, дискриминантная функция которого для соответствующего вектора признаков является максимальной:

$$\forall\ j \neq l: \quad d_l(\overline{x}(\omega)) \geq d_j(\overline{x}(\omega)) \Rightarrow \overline{x}(\omega) \in D_l. \tag{4}$$

Замечание 1. Выбор решающих функций не единственен. Так, очевидно, наряду с функциями $d_l(\overline{x})$, $\left(l = \overline{0, L-1}\right)$, решающими функциями также являются:

- $g_1(\overline{x})d_l(\overline{x}) + g_2(\overline{x})$, где $g_1(\overline{x})$ - любая неотрицательная функция, а $g_2(\overline{x})$ - любая функция, не зависящие от номера класса *l*;
- $\varphi(d_l(\overline{x}))$, где $\varphi(\ldots)$ - любая монотонно возрастающая функция, не зависящая от

номера класса.

Часто за счет приведенных преобразований удается существенно упростить вид классификатора.

1.2. Качество классификатора

Качество классификатора характеризуется величиной, называемой в теории статистических решений *условным средним риском*. Она задает среднюю величину потерь, связанных с принятием классификатором решения об отнесении данного вектора признаков $\bar{x}$ к классу с номером *j*:

$$R_j(\bar{x}) = \frac{1}{f(\bar{x})} \sum_{l=0}^{L-1} c_{lj} P(\Omega_l) f(\bar{x}/\Omega_l). \tag{5}$$

В данном выражении:

- $P(\Omega_l)$ - *априорная вероятность* появления объектов из класса Ω_l, причем $\sum_{l=0}^{L-1} P(\Omega_l) = 1$;
- $f(\bar{x}/\Omega_l)$ - условная плотность вероятностей случайного вектора признаков $\overline{X}$ для объектов класса Ω_l (в теории распознавания образов ее называют *функцией правдоподобия* для соответствующего класса);
- $f(\bar{x})$ - безусловная плотность вероятностей случайного вектора $\overline{X}$;
- элементы квадратной матрицы

$$C = \left\| c_{lj} \right\|_{l,j=0}^{L-1} \tag{6}$$

характеризуют величины *штрафов* или *потерь* за ошибки классификатора. Матрица C может быть достаточно произвольной. Единственным ограничением на ее элементы является то, что штраф за ошибочное решение должен быть больше, чем штраф за решение правильное, то есть: $c_{lj} > c_{ll}$.

Интегральной величиной, характеризующей качество классификатора является *математическое ожидание потерь* или *общий риск*, который с учетом (5) и (3) имеет вид:

$$R = \sum_{j=0}^{L-1} \int_{D_j} R_j(\bar{x}) f(\bar{x}) d\bar{x} = \sum_{j=0}^{L-1} \sum_{l=0}^{L-1} c_{lj} P(\Omega_l) p_{lj}, \tag{7}$$

1.3. Оптимальные стратегии классификации

Процесс классификации аналогичен игре двух лиц с нулевой суммой, в которой одним из игроков является классификатор. В такой игре выигрыш (проигрыш) одного из участников равен проигрышу (выигрышу) другого. Выбор оптимальной стратегии в игре зависит от количества исходной информации. Могут использоваться *байесовская*, *минимаксная* стратегии или *стратегия Неймана-Пирсона*. В зависимости от того, какая из стратегий используется для построения классификатора, последний называют, соответственно, *байесовским классификатором*, *минимаксным классификатором* или *классификатором Неймана-Пирсона*.

1.3.1. Байесовский классификатор

Байесовская стратегия используется при наличии полной априорной информации о классах, то есть когда известны:

- функции правдоподобия для каждого из классов;
- матрица штрафов;
- априорные вероятности для каждого из классов.

Стратегия решения выбирается таким образом, чтобы обеспечить минимум общего риска (7). Минимальный общий риск при этом называется *байесовским риском*. В соответствии с выражениями (5) и (7), минимум общего риска R будет обеспечен, если разбиение пространства признаков D будет осуществляться по следующему правилу: вектор $\bar{x} \in D$ относится к области D_l только тогда, когда соответствующий условный средний риск $R_l(\bar{x})$ минимален:

$$\forall\ j \neq l \quad R_l(\bar{x}) < R_j(\bar{x}) \Rightarrow \bar{x} \in D_l, \tag{8}$$

Графическая иллюстрация байесовской стратегии приведена на Рис.1а.

Если матрица потерь (6) является простейшей[1], то после подстановки в (8) выражения для условного среднего риска (5) и с учетом Замечания 1 имеем следующий явный вид байесовского классификатора (см. Рис.1б):

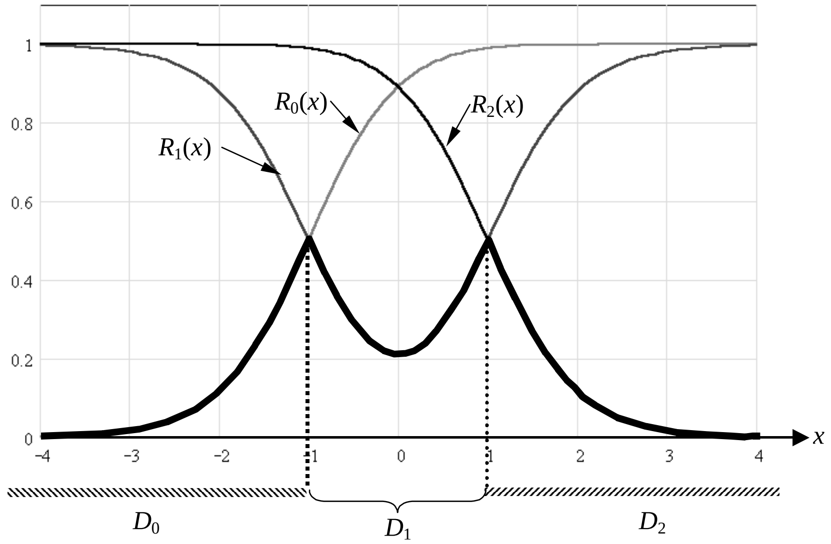
---

[1] Матрица потерь C называется *простейшей*, если ее элементы удовлетворяют равенству $c_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$.

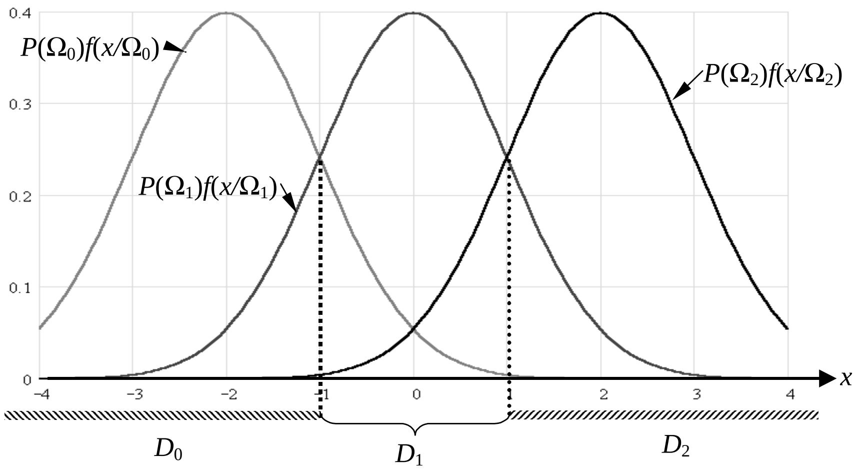
$$\forall \ j \neq l \quad P(\Omega_l) f(\bar{x}/\Omega_l) \geq P(\Omega_j) f(\bar{x}/\Omega_j) \Rightarrow \bar{x} \in D_l . \tag{9}$$

Из (9), в частности, видно, что решающими функциями байесовского классификатора являются функции:

$$d_l(\bar{x}) = P(\Omega_l)\, f(\bar{x}/\Omega_l), \quad l = \overline{0, L-1} . \tag{10}$$



а) байесовская стратегия минимизации общего риска;



б) байесовская классификатор.

Рис.1 *Построение байесовского классификатора для простейшей матрицы штрафов*

Часто используют также следующую форму записи байесовского классификатора:

$$\forall \ j \neq l \quad \frac{f(\bar{x}/\Omega_l)}{f(\bar{x}/\Omega_j)} \geq \frac{P(\Omega_j)}{P(\Omega_l)} \Rightarrow \bar{x} \in D_l . \tag{11}$$

При этом функция $\Lambda_{lj}(\bar{x}) = \dfrac{f(\bar{x}/\Omega_l)}{f(\bar{x}/\Omega_j)}$ называется *отношением правдоподобия*, а величина $\lambda_{jl} = \dfrac{P(\Omega_j)}{P(\Omega_l)}$ - *пороговым значением*. Таким образом, байесовский классификатор основан на сравнении отношения правдоподобия с пороговым значением:

$$\forall\ j \neq l \quad \Lambda_{lj}(\bar{x}) \geq \lambda_{jl} \Rightarrow \bar{x} \in D_j\,,$$

и называется поэтому *классификатором отношения правдоподобия*.

Легко показать, что при произвольном виде матрицы штрафов в случае двух классов байесовский классификатор имеет вид:

$$\frac{f(\bar{x}/\Omega_1)}{f(\bar{x}/\Omega_0)} \begin{matrix} > \\ < \end{matrix} \frac{P(\Omega_0)(c_{01} - c_{00})}{P(\Omega_1)(c_{10} - c_{11})} \Rightarrow \bar{x} \in \begin{cases} D_1 \\ D_0 \end{cases}$$

с дискриминантными функциями:

$$d_j(\bar{X}) = P(\Omega_j)\left(c_{j(1-j)} - c_{jj}\right) f(\bar{x}/\Omega_j), \quad j = \overline{0,1}\,.$$

1.3.2. Минимаксный классификатор

Классификатор, основанный на минимаксной стратегии, используется для случая двух классов и если известны:

- функции правдоподобия для каждого из классов;
- матрица штрафов.

Минимизировать величину общего риска при отсутствии информации об априорных вероятностях классов, очевидно, невозможно. В то же время, предполагая возможность произвольного изменения значений априорных вероятностей классов, можно минимизировать максимально возможное значение риска. Действительно, общий риск (7) в случае двух классов может быть представлен в следующем виде:

$$R = \left(c_{11} + p_{10}(c_{10} - c_{11})\right) + P(\Omega_0) \cdot \left[\left(c_{00} + p_{01}(c_{01} - c_{00})\right) - \left(c_{11} + p_{10}(c_{10} - c_{11})\right)\right]. \qquad (12)$$

При фиксированном классификаторе изменение априорной вероятности приводит к изменению величины общего риска, причем характер зависимости в (12) линейный (см. Рис.2). Поэтому поиск классификатора, минимизирующего максимально возможную величину общего риска, эквивалентен поиску такого байесовского классификатора, для которого величина (12) является постоянной, не зависящей от значения априорной вероятности $P(\Omega_0)$ величиной. Таким классификатором, очевидно, является байесовский классификатор, удовлетворяющий следующему дополнительному условию:

$$(c_{00} + p_{01}(c_{01} - c_{00})) - (c_{11} + p_{10}(c_{10} - c_{11})) = 0\,. \tag{13}$$

Из рисунка Рис.2 видно, что значение величины общего риска для минимаксного классификатора равно максимальному значению байесовского (минимального) риска. Пара априорных вероятностей $(P^*(\Omega_0), 1 - P^*(\Omega_0))$, при которых байесовский риск принимает максимальное значение, называется *наименее благоприятным распределением априорных* вероятностей. Таким образом

- <u>*минимаксный классификатор – это байесовский классификатор, полученный для пары наименее благоприятных априорных вероятностей.*</u>

В более простой ситуации, когда элементы матрицы штрафов таковы, что

$$c_{00} = c_{11} = 0, \quad c_{10} = c_1, \quad c_{01} = c_0,$$

условие (13) преобразуется в следующее:

$$p_{01} c_0 = p_{10} c_1\,. \tag{14}$$

Последнее выражение представляет собой условие выбора областей D_0, D_1 в байесовском классификаторе.

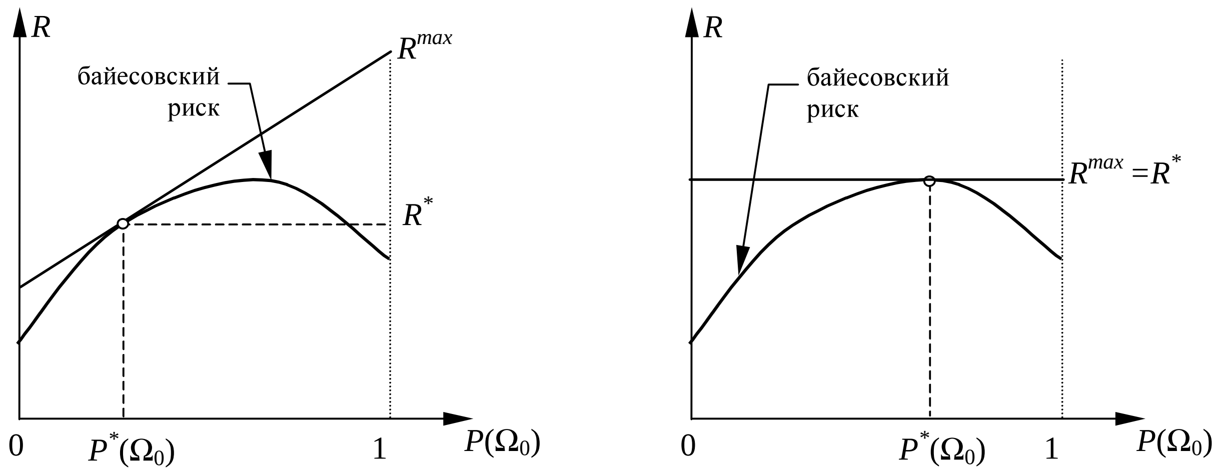


Рис.2. *Иллюстрация минимаксной стратегии построения классификатора*

1.3.3. Классификатор Неймана-Пирсона

Классификатор, основанный на стратегии Неймана-Пирсона, используется для случая двух классов, и если известны только функции правдоподобия для каждого из классов. Суть стратегии Неймана-Пирсона состоит в следующем: задается допустимое значение вероятности ошибки первого рода[2] p_0, а затем классификатор строится таким образом, чтобы обеспечить минимум вероятности ошибки второго рода p_1:

$$\begin{cases} p_1 \to \min\limits_{D_0, D_1} \\ p_0 = p_0^*. \end{cases} \tag{15}$$

Решением задачи Неймана-Пирсона является классификатор вида:

$$\Lambda(\bar{x}) = \frac{f\left(\bar{x}/\Omega_1\right)}{f\left(\bar{x}/\Omega_0\right)} \begin{array}{c} > \\ < \end{array} \lambda \quad \Rightarrow \quad \begin{cases} \bar{x} \in D_1 \\ \bar{x} \in D_0 \end{cases}, \tag{16}$$

где значение пороговой величины λ определяется, исходя из условия: $p_0 = p_0^*$ (см. Рис.3). Из выражения (16) следует, что

- <u>*классификатор Неймана-Пирсона – это классификатор отношения правдоподобия*</u>.

1.4. Байесовский классификатор в типовых задачах распознавания образов

На практике часто возникает задача распознавания детерминированных объектов или сигналов в условиях помех. Она стала традиционной в таких дисциплинах, как теория сигналов, обработка изображений, распознавание образов. Ниже приведены два достаточно типичных примера постановки подобной задачи и ее решения с использованием байесовской стратегии.

---

[2] Критерий Неймана-Пирсона в теории статистических решений традиционно используется для проверки гипотез. Поскольку в классической постановке задачи используется только две возможные гипотезы, то различают два типа ошибок:

- *ошибку первого рода* p_0 - в контексте настоящего изложения $p_0 = p_{01}$,
- *ошибку второго рода* p_1 - в контексте настоящего изложения $p_1 = p_{10}$.

Заметим, что в общем случае $p_1 + p_0 \neq 1$. В дальнейшем изложении данная терминология и

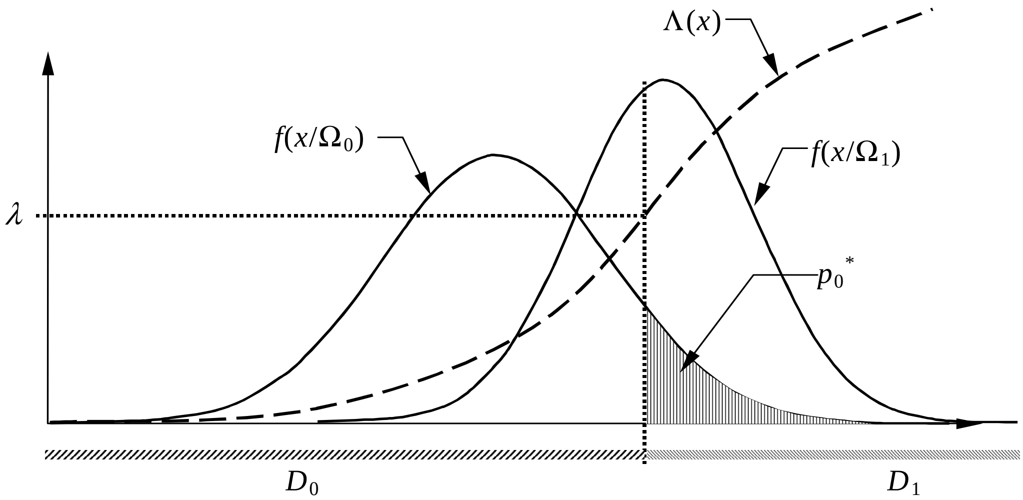


Рис.3. *Иллюстрация стратегии Неймана-Пирсона построения классификатора*

1.4.1. Байесовский классификатор для нормально распределенных векторов признаков

Пусть входной сигнал, задаваемый вектором $\bar{x}=\left(x_0,\ldots,x_{n-1}\right)^T$ и подлежащий распознаванию, представляет собой аддитивную смесь детерминированной и шумовой составляющих. Будем считать, что наблюдаемые вектора имеют нормальный закон распределения в каждом из L классов, то есть имеют плотность вероятностей вида:

$$f\left(\bar{x}/\Omega_l\right)=\frac{1}{(2\pi)^{n/2}\sqrt{|B_l|}}\exp\left(-\frac{1}{2}\left(\bar{x}-\overline{M_l}\right)^T B_l^{-1}\left(\bar{x}-\overline{M_l}\right)\right),\quad l=\overline{0,L-1}. \tag{17}$$

Здесь

$$B_l=M\left(\left(\overline{X}-\overline{M_l}\right)\left(\overline{X}-\overline{M_l}\right)^T/\Omega_l\right),\qquad \overline{M_l}=M\left(\overline{X}/\Omega_l\right)$$

корреляционная матрица и *математическое ожидание* вектора признаков из класса Ω_l, соответственно. Математические ожидания или *средние* характеризуют детерминированные составляющие распознаваемых сигналов, а корреляционные матрицы – характер шумовой составляющей. Считаются также известными априорные вероятности $P(\Omega_l)$ появления векторов из каждого класса. Требуется по реализации $\bar{x}$ случайного вектора $\overline{X}$ определить класс, к которому данный вектор принадлежит.

приведенные обозначения также используются.

Решением данной задачи является байесовский классификатор с дискриминантными функциями следующего вида:

$$d_l(\bar{x}) = \ln P(\Omega_l) - \ln\sqrt{|B_l|} - \frac{1}{2}\left(\bar{x} - \overline{M}_l\right)^T B_l^{-1}\left(\bar{x} - \overline{M}_l\right), \quad l = \overline{0, L-1}. \tag{18}$$

Выражение (18) может быть существенно упрощено в некоторых частных случаях.

Случай 1

Предположим, что компоненты наблюдаемого вектора $\overline{X}$ являются независимыми и имеют одинаковую дисперсию D_X, то есть $B_l = D_X I$, где I – единичная $N \times N$ матрица. Тогда законы распределения (17) отличаются только средними значениями, а решающие функции байесовского классификатора преобразуются к следующему виду:

$$d_l(\bar{x}) = 2D_X \ln P(\Omega_l) - \left\|\bar{x} - \overline{M}_l\right\|^2 \quad l = \overline{0, L-1}, \tag{19}$$

здесь $\|\ldots\|$ - евклидова норма. При равных априорных вероятностях данное решающее правило приобретает очевидную трактовку:

- *вектор признаков* $\bar{x}$ *относится к тому классу, расстояние до центра которого минимально.*

Классификатор в этом случае называют *классификатором по минимуму евклидова расстояния*. Пример разбиения пространства признаков при использовании подобного классификатора для случая трех классов приведен на Рис.4а.

Нетрудно видеть, что решающие функции (19) можно преобразовать к линейной форме:

$$d_l(\bar{x}) = \overline{M}_l^T \bar{x} - \frac{1}{2}\overline{M}_l^T \overline{M}_l + D_X \ln P(\Omega_l), \quad l = \overline{0, L-1}$$

В этом случае *разделяющие границы* между различными областями D_l, задаваемые соотношениями вида:

$$d_{lj}(\bar{x}) \equiv d_l(\bar{x}) - d_j(\bar{x}) = 0, \quad 0 \le l < j \le L-1,$$

также являются линейными:

$$d_{lj}(\bar{x}) = \left(\overline{M}_l - \overline{M}_j\right)^T \bar{x} - \frac{1}{2}\left(\overline{M}_l + \overline{M}_j\right)^T\left(\overline{M}_l - \overline{M}_j\right) + D_X \ln\frac{P(\Omega_l)}{P(\Omega_j)}, \quad 0 \le l < j \le L-1,$$

и говорят о линейном классификаторе.

Случай 2

Предположим, что все корреляционные матрицы одинаковы: $B_l = B$. Тогда решающие функции байесовского классификатора представляются в виде:

$$d_l(\overline{x}) = 2\ln P(\Omega_l) - \left(\overline{x} - \overline{M}_l\right)^T B^{-1}\left(\overline{x} - \overline{M}_l\right), \quad l = \overline{0, L-1}.$$

Величина

$$\rho(\overline{x}, \overline{M}_l) = \left(\overline{x} - \overline{M}_l\right)^T B^{-1}\left(\overline{x} - \overline{M}_l\right) \tag{20}$$

называется *расстоянием Махаланобиса* между векторами $\overline{x}$ и $\overline{M}_l$ и является мерой близости вектора $\overline{x}$ к центру класса Ω_l, учитывающей как дисперсии компонент вектора $\overline{X}$, так и их взаимную корреляцию. Очевидно, что в данной ситуации классификатор снова оказывается классификатором по минимуму расстояния Махаланобиса (см. Рис.4б). Кроме того, и решающие функции, и разделяющие границы снова являются линейными:

$$d_l(\overline{x}) = \overline{M}_l^T B^{-1}\overline{x} - \frac{1}{2}\overline{M}_l^T B^{-1}\overline{M}_l + \ln P(\Omega_l), \quad l = \overline{0, L-1}$$

$$d_{lj}(\overline{x}) = \left(\overline{M}_l - \overline{M}_j\right)^T B^{-1}\overline{x} - \frac{1}{2}\left(\overline{M}_l + \overline{M}_j\right)^T B^{-1}\left(\overline{M}_l - \overline{M}_j\right) + \ln\frac{P(\Omega_l)}{P(\Omega_j)}, \quad 0 \le l < j \le L-1,$$

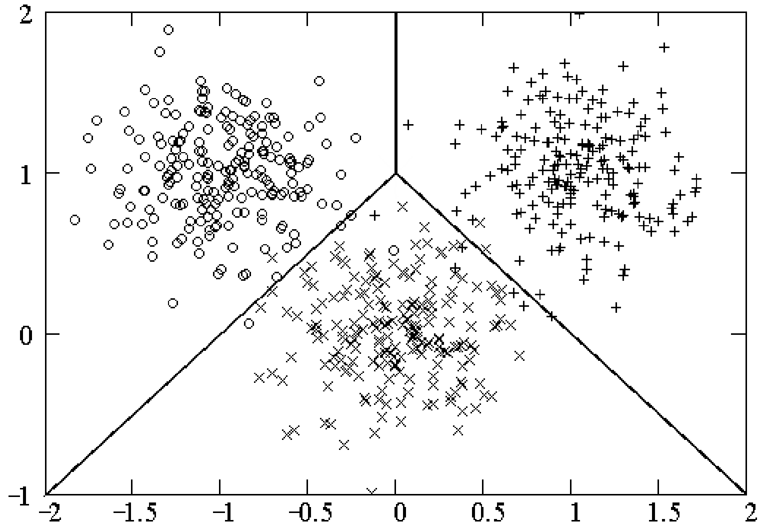
а, следовательно, линейным является и классификатор.

Случай 3

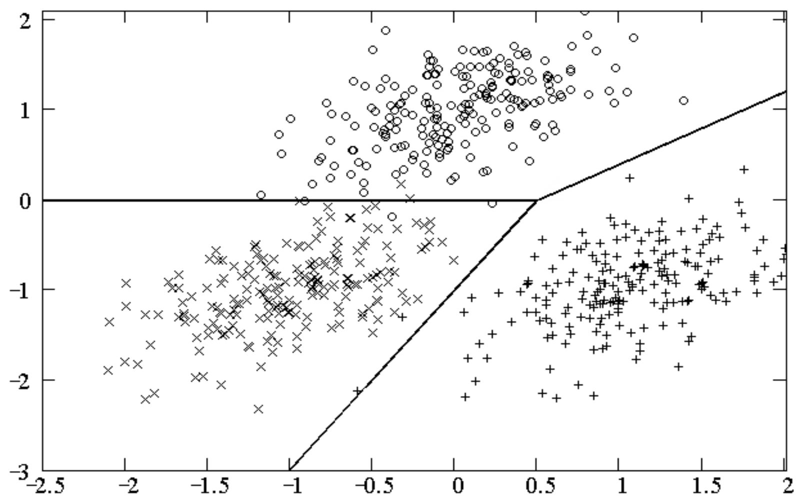
В ситуации, когда все корреляционные матрицы различны, необходимо пользоваться выражением (18) для дискриминантных функций. Разделяющие границы в этом случае представляются в следующем виде:

$$\begin{aligned} d_{lj}(\overline{x}) = {} & \overline{x}^T\left(B_j^{-1} - B_l^{-1}\right)\overline{x} \quad + \quad 2\left(\overline{M}_l^T B_l^{-1} - \overline{M}_j^T B_j^{-1}\right)\overline{x} \quad + \\ & + \left[\ln\frac{|B_l|}{|B_j|} + 2\ln\frac{P(\Omega_l)}{P(\Omega_j)} - \overline{M}_l^T B_l^{-1}\overline{M}_l + \overline{M}_j^T B_j^{-1}\overline{M}_j\right], \quad 0 \le l < j \le L-1 \end{aligned}$$

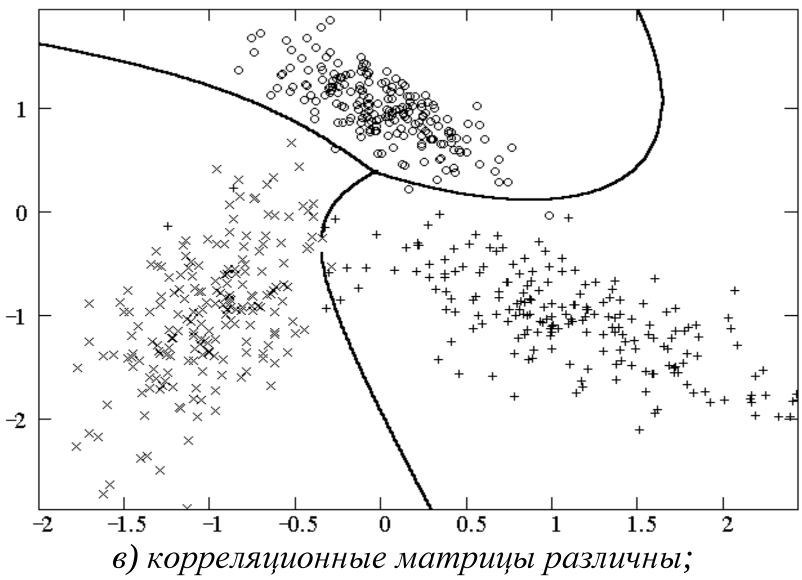
и являются, очевидно, квадратичными функциями. Такие границы называются *гиперквадриками* (гиперсферы, гиперпараболы и т.д., пример их приведен на Рис.4с), а сам классификатор называется *квадратичным*.



а) признаки статистически независимых и одинаково распределены;



б) корреляционные матрицы одинаковы;



в) корреляционные матрицы различны;

Рис.4. *Байесовский классификатор в случае нормально распределенных векторов признаков*

1.4.2. Байесовский классификатор для распознавания бинарных векторов признаков

На практике достаточно часто возникает задача распознавания векторов признаков, компоненты которых являются бинарными. Эта задача, в частности, решается при автоматическом распознавании печатного текста в известных системах CuneiForm и FineReader. Ниже приведено ее решение с использованием байесовской стратегии.

Пусть закон распределения бинарного случайного вектора $\overline{X}$ для каждого из классов Ω_l $\left(l = \overline{0, L-1}\right)$ задан распределением вероятностей $P\left(\overline{X} = \overline{x}/\Omega_l\right)$; пусть также известны априорные вероятности появления представителей каждого класса $P(\Omega_l)$ и матрица штрафов. При наличии этой информации выражение для условного среднего риска (5) переписываются с учетом дискретного характера вектора признаков в следующем виде:

$$R_j(\overline{x}) = \frac{1}{P\left(\overline{X} = \overline{x}\right)} \sum_{l=0}^{L-1} c_{lj} P(\Omega_l) P\left(\overline{X} = \overline{x}/\Omega_l\right).$$

Предположим, что матрица штрафов является простейшей. Тогда байесовский классификатор может быть записан в одной из двух форм: либо в терминах дискриминантных функций (10), либо в терминах отношения правдоподобия (11). С учетом дискретного характера вектора признаков эти выражения имеют следующий вид:

$$\forall\ j \neq l \quad P(\Omega_l) P\left(\overline{X} = \overline{x}/\Omega_l\right) = d_l(\overline{x}) \quad \geq \quad d_j(\overline{x}) = P\left(\Omega_j\right) P\left(\overline{X} = \overline{x}/\Omega_j\right) \Rightarrow \overline{x} \in D_l,$$

$$\forall\ j \neq l \quad \frac{P(\overline{X} = \overline{x}/\Omega_l)}{P(\overline{X} = \overline{x}/\Omega_j)} = \Lambda_{lj} \quad \geq \quad \lambda_{jl} = \frac{P(\Omega_j)}{P(\Omega_l)} \Rightarrow \overline{x} \in D_l. \qquad (21)$$

В общем случае аналитически получить окончательные выражения для байесовского классификатора не представляется возможным. Однако это может быть сделано в предположении независимости компонент вектора признаков.

Итак, пусть компоненты вектора $\overline{X}$ являются независимыми. Тогда:

$$P\left(\overline{X} = \overline{x}/\Omega_l\right) = \prod_{i=0}^{N-1} P\left(X_i = x_i/\Omega_l\right).$$

Учитывая, что возможные значения компонент вектора “0” или “1”, получаем следующее выражение для дискриминантной функции:

$$d_l(\bar{x}) = P(\Omega_l)\prod_{i=0}^{N-1}\left(\left(1-P\left(X_i = 1/\Omega_l\right)\right)(1-x_i) + P\left(X_i = 1/\Omega_l\right)x_i\right), \quad l = \overline{0, L-1}.$$

Окончательным решением задачи является классификатор с дискриминантной функцией вида:

$$d_l(\bar{x}) = \left[\ln(P(\Omega_l)) + \sum_{i=0}^{N-1}\ln\left(1-P\left(X_i = 1/\Omega_l\right)\right)\right] \quad + \quad \sum_{i=0}^{N-1}x_i \cdot \ln\left(\frac{P\left(X_i = 1/\Omega_l\right)}{1-P\left(X_i = 1/\Omega_l\right)}\right),$$

который, очевидно, является линейным.

Аналогичным образом можно получить выражение для байесовского классификатора в терминах отношения правдоподобия (21):

$$\forall\ j \neq l \quad \tilde{\Lambda}_{lj}(\bar{x}) \geq \tilde{\lambda}_{jl} \Rightarrow \bar{x} \in D_l,$$

где

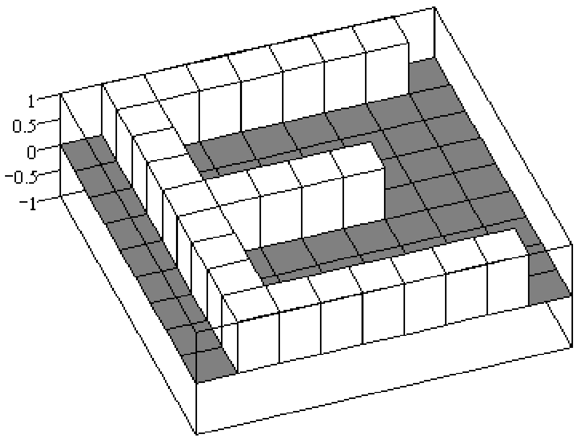
$$\begin{aligned} &\tilde{\Lambda}_{lj}(\bar{x}) = \sum_{i=0}^{N-1} w_{lj}^i \cdot x_i, \qquad w_{lj}^i = \ln\left(\frac{P\left(X_i = 1/\Omega_l\right)}{1-P\left(X_i = 1/\Omega_l\right)} \cdot \frac{1-P\left(X_i = 1/\Omega_j\right)}{P\left(X_i = 1/\Omega_j\right)}\right) \\ &\tilde{\lambda}_{jl} = \ln\left(\frac{P(\Omega_j)}{P(\Omega_l)}\right) + \sum_{i=0}^{N-1}\ln\left(\frac{1-P\left(X_i = 1/\Omega_j\right)}{1-P\left(X_i = 1/\Omega_l\right)}\right). \end{aligned} \tag{22}$$

Очевидно, отношение правдоподобия также является линейной функцией компонент вектора признаков. Пример байесовского классификатора в терминах отношения правдоподобия приведен на Рис.5.

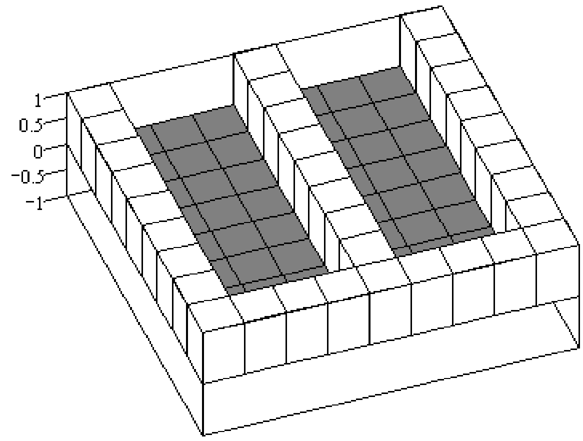
1.5. Вычисление вероятностей ошибочной классификации

Эффективность любого классификатора характеризуется вероятностями ошибок. Однако их нахождение в общем случае оказывается достаточно сложной задачей, поскольку требует вычисления многомерных интегралов:

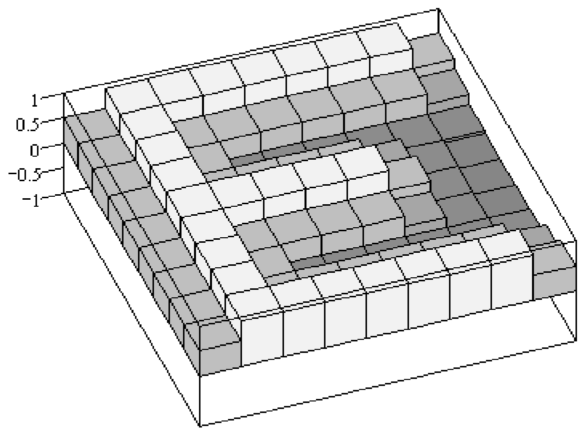
$$p_{lj} = \int_{D_j} f\left(\bar{x}/\Omega_l\right)d\bar{x}, \quad l \neq j, \quad l, j = \overline{0, L-1}. \tag{23}$$



а) представитель класса Ω_0;

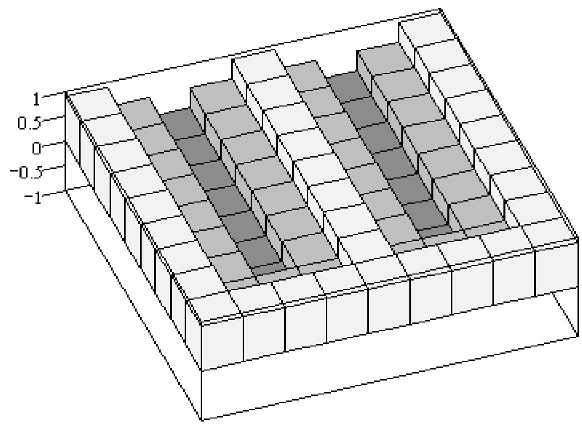


б) представитель класса Ω_1;



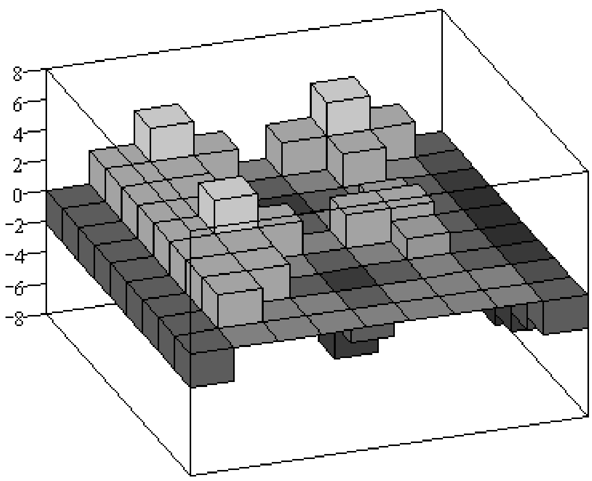
в) распределение вероятностей

$P\left(X_{ij}=1/\Omega_0\right)$;

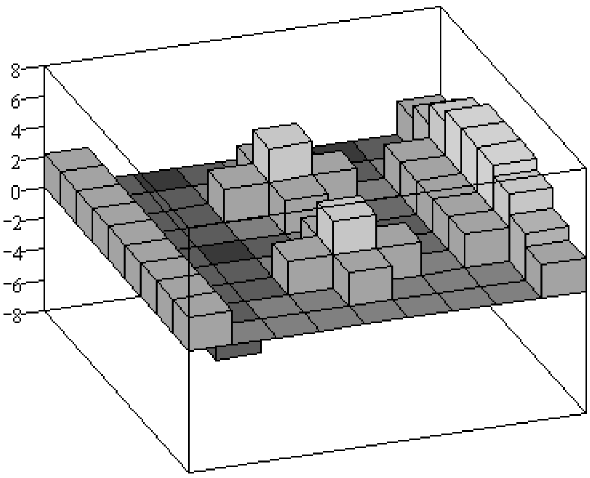


г) распределение вероятностей

$P\left(X_{ij}=1/\Omega_1\right)$;



д) компоненты вектора $\overline{w}_{01}$;



д) компоненты вектора $\left(-\overline{w}_{01}\right)$;

Рис.5. *Пример байесовского классификатора для распознавания бинарных векторов признаков*

При использовании байесовского классификатора, который является классификатором отношения правдоподобия, многомерный интеграл (23) может быть заменен одномерным от плотности вероятностей отношения правдоподобия Λ_{ij} в каждом из классов. В частности, в случае двух классов для вероятностей ошибок имеем следующие выражения:

$$p_0 = \int_{\lambda}^{+\infty} f_\Lambda(u/\Omega_0)du, \quad p_1 = \int_{-\infty}^{\lambda} f_\Lambda(u/\Omega_1)du\,, \tag{24}$$

где $\Lambda = \Lambda(\overline{X}) = \dfrac{f(\overline{X}/\Omega_1)}{f(\overline{X}/\Omega_0)}$, а $\lambda = \dfrac{P(\Omega_0)}{P(\Omega_1)}$ - пороговое значение. Плотность вероятностей отношения правдоподобия удается найти далеко не всегда. Однако, когда случайный вектор $\overline{X}$ имеет нормальный закон распределения, это может быть сделано.

1.5.1. Вычисление вероятностей ошибочной классификации для нормально распределенных векторов-признаков

Пусть вектор признаков в каждом из двух классов характеризуется нормальным законом распределения, причем все корреляционные матрицы являются равными $B_l = B$ $(l = 0{,}1)$. Тогда случайная величина $\tilde{\Lambda} = \ln(\Lambda(\overline{X}))$ имеет нормальный закон распределения с параметрами:

$$M(\tilde{\Lambda}/\Omega_0) = M(\ln(\Lambda(\overline{X}))/\Omega_0) = -\frac{1}{2}\rho(\overline{M}_0, \overline{M}_1), \quad M(\tilde{\Lambda}/\Omega_1) = M(\ln(\Lambda(\overline{X}))/\Omega_1) = \frac{1}{2}\rho(\overline{M}_0, \overline{M}_1),$$
$$D(\tilde{\Lambda}/\Omega_l) = D(\ln(\Lambda(\overline{X}))/\Omega_l) = \rho(\overline{M}_0, \overline{M}_1), \quad l = \overline{0{,}1}$$

где $\rho(\overline{M}_0, \overline{M}_1)$ - расстояние Махаланобиса между векторами средних $\overline{M}_0$ и $\overline{M}_1$.

Выражения для вероятностей ошибок (24) преобразуются к следующему виду:

$$p_0 = 1 - \Phi\left(\frac{\tilde{\lambda} + \frac{1}{2}\rho(\overline{M}_0, \overline{M}_1)}{\sqrt{\rho(\overline{M}_0, \overline{M}_1)}}\right), \quad p_1 = \Phi\left(\frac{\tilde{\lambda} - \frac{1}{2}\rho(\overline{M}_0, \overline{M}_1)}{\sqrt{\rho(\overline{M}_0, \overline{M}_1)}}\right),$$

где $\Phi(\ldots)$ - функция Лапласа, а

$$\tilde{\lambda} = \ln\lambda = \ln\left(\frac{P(\Omega_0)(c_{01} - c_{00})}{P(\Omega_1)(c_{10} - c_{11})}\right)$$

новая пороговая величина (см. Рис.6). В частном случае, когда матрица штрафов является простейшей и априорные вероятности классов совпадают, имеем:

$$\lambda = 1, \quad \tilde{\lambda} = 0, \quad p_0 = 1 - \Phi\left(\frac{1}{2}\sqrt{\rho\left(\overline{M}_0, \overline{M}_1\right)}\right) \quad p_0 = \Phi\left(-\frac{1}{2}\sqrt{\rho\left(\overline{M}_0, \overline{M}_1\right)}\right).$$

Общий риск при этом определяется формулой:

$$R = 1 - \Phi\left(\frac{1}{2}\sqrt{\rho\left(\overline{M}_0, \overline{M}_1\right)}\right)$$

и монотонно убывает с ростом расстояния Махаланобиса между векторами средних.

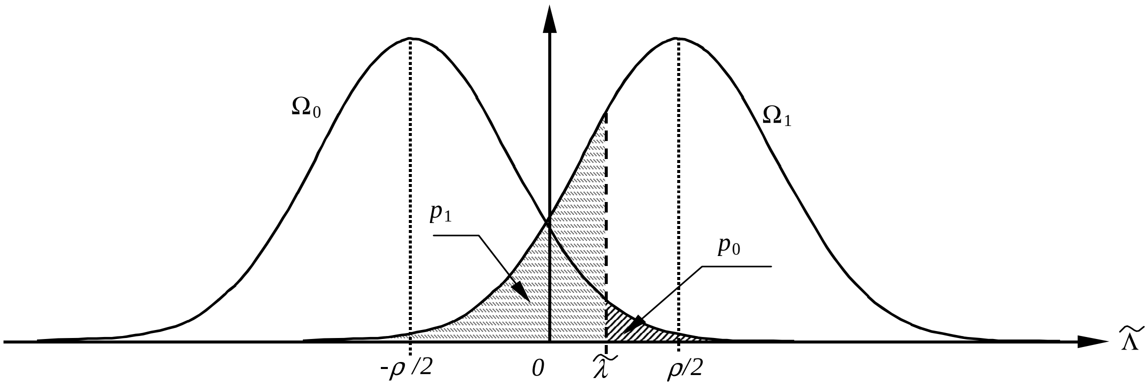


Рис.6. *Плотности вероятностей логарифма отношения правдоподобия для нормально распределенных векторов признаков с равными корреляционными матрицами*

Минимаксный классификатор

Предположим, что матрица штрафов имеет следующие элементы: $c_{01} = c_{10} = c, \quad c_{00} = c_{11} = 0$. Тогда соотношение (14) для выбора разделяющей границы байесовского классификатора, соответствующего минимаксной стратегии, превращается в равенство вероятностей ошибочной классификации: $p_0 = p_1$. С учетом выражений (24) получаем, что пороговое значение для минимаксного классификатора: $\lambda = 1, \quad \tilde{\lambda} = 0$.

Классификатор Неймана-Пирсона

Используя условие (15) и равенства (24), получаем, что пороговое значение классификатора Неймана-Пирсона определяется по формуле:

$$\lambda = e^{\tilde{\lambda}}, \quad \tilde{\lambda} = -\frac{1}{2}\rho\left(\overline{M}_0, \overline{M}_1\right) + \sqrt{\rho\left(\overline{M}_0, \overline{M}_1\right)}\Phi^{-1}\left(1 - p_0^*\right),$$

где p_0^* - заданная величина вероятности ошибки первого рода.

1.5.2. Вычисление вероятностей ошибочной классификации для бинарных векторов признаков

Получить аналитическое выражение для вероятностей ошибочной классификации бинарных векторов признаков в общем случае невозможно. Однако, при небольшой размерности вектора признаков ($N \le 26$) значения этих вероятностей можно вычислить на ПЭВМ, используя дискретный аналог формулы (23):

$$p_{lj} = \sum_{\bar{x} \in D_j} P\left(\overline{X} = \bar{x} / \Omega_l\right), \quad l \ne j, \quad l, j = \overline{0, L-1},$$

поскольку в признаковом пространстве находится всего 2^N элементов. Когда число компонент вектора признаков велико, можно получить приближенные выражения для вероятностей ошибочной классификации, если предположить независимость компонент случайного вектора $\overline{X}$. В подобной ситуации в соответствии с центральной предельной теоремой можно считать закон распределения случайной величины $\tilde{\Lambda}_{lj} \equiv \tilde{\Lambda}_{lj}\left(\overline{X}\right)$ нормальным. Для простоты рассмотрим ситуацию с разделением двух классов. В этом случае числовые характеристики закона распределения случайной величины $\tilde{\Lambda}_{10}$ имеют вид:

$$m_l = M\left(\tilde{\Lambda}_{10}\left(\overline{X}\right) / \Omega_l\right) = \sum_{i=0}^{N-1} \ln\left(\frac{P\left(X_i = 1/\Omega_1\right)}{1 - P\left(X_i = 1/\Omega_1\right)} \frac{1 - P\left(X_i = 1/\Omega_0\right)}{P\left(X_i = 1/\Omega_0\right)}\right) P\left(X_i = 1/\Omega_l\right), \qquad l = 0{,}1,$$

$$\sigma_l^2 = D\left(\tilde{\Lambda}_{10}\left(\overline{X}\right) / \Omega_l\right) = \sum_{i=0}^{N-1} \left[\ln\left(\frac{P\left(X_i = 1/\Omega_1\right)}{1 - P\left(X_i = 1/\Omega_1\right)} \frac{1 - P\left(X_i = 1/\Omega_0\right)}{P\left(X_i = 1/\Omega_0\right)}\right)\right]^2 P\left(X_i = 1/\Omega_l\right)\left(1 - P\left(X_i = 1/\Omega_l\right)\right),$$

а для вероятностей ошибочной классификации получаем следующие приближенные выражения:

$$p_0 = 1 - \Phi\left(\frac{\tilde{\lambda} - m_0}{\sigma_0}\right), \quad p_1 = \Phi\left(\frac{\tilde{\lambda} - m_1}{\sigma_1}\right), \qquad \text{где} \quad \tilde{\lambda} = \ln\frac{P\left(\Omega_0\right)}{P\left(\Omega_1\right)}$$

В ситуации когда условия центральной предельной теоремы не выполняются, а компоненты вектора независимы, можно воспользоваться следующими выражениями для верхних границ вероятностей ошибок, вытекающих из неравенства Чебышева:

$$p_0 \le \frac{\sigma_0^2}{\left(m_0 - \tilde{\lambda}\right)^2}, \qquad p_1 \le \frac{\sigma_1^2}{\left(m_1 - \tilde{\lambda}\right)^2}$$

1.5.3. Экспериментальная оценка вероятностей ошибочной классификации

На практике воспользоваться аналитическими выражениями для вычисления вероятностей ошибок классификации чаще всего не представляется возможным. Поэтому единственным способом определения искомых вероятностей является их статистическое оценивание.

Пусть выборочные данные представлены в виде набора из N объектов $\{\omega_i\}_{i=0}^{N-1}$ класса Ω_0 и N рассчитанных по ним векторов признаков $\{\overline{x}(\omega_i)\}_{i=0}^{N-1}$ (при этом говорят, что задана обучающая выборка объема N из класса Ω_0), а также задан некоторый классификатор, производящий классификацию объектов в соответствии со следующим правилом:

$$d\left(\overline{x}(\omega)\right) \begin{matrix} > \\ < \end{matrix} 0 \quad \Rightarrow \quad \begin{cases} \omega \in \Omega_0 \\ \omega \notin \Omega_0 \end{cases}.$$

Обозначим p_0 - истинное значение вероятности ошибочной классификации объектов класса Ω_0:

$$p_0 = P\left(d\left(\overline{X}/\Omega_0\right) < 0\right).$$

Наилучшей точечной оценкой вероятности p_0, как известно, является относительная частота события $\left(d\left(\overline{X}/\Omega_0\right) < 0\right)$:

$$\widehat{p}_0 = \frac{1}{N}\sum_{i=0}^{N-1} I\left(d\left(\overline{x}(\omega_i)\right) < 0\right). \tag{26}$$

Качество оценки (26) можно охарактеризовать величиной ее относительной погрешности, которая имеет вид:

$$\varepsilon \equiv \frac{\sqrt{D[\widehat{p}]}}{\widehat{p}} = \sqrt{\frac{1-\widehat{p}}{N\widehat{p}}}.$$

Последнее выражение можно использовать также с целью определения необходимого объема N обучающей выборки для получения оценки вероятности с заранее заданной относительной погрешностью ε.

Замечание 2. Аналогично (26) выглядит оценка вероятности p_l ошибочной классификации объектов класса Ω_l и вероятностей p_{lj} ошибочной классификации объектов из класса Ω_l в класс Ω_j по обучающим выборкам из соответствующих классов.

2. ЛИТЕРАТУРА

1. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений. - М.: Высшая школа, 1983. - 295 с.
2. Верхаген К., Дейн Р., Грун Ф., Йостен Й., Вербек П. Распознавание образов: состояние и перспективы: Пер. с англ. – М.: Радио и связь, 1985. 103с.
3. Горелик А.Л., Скрипкин В.А. Методы распознавания. - М.: Высшая школа, 1984. - 208с.
4. Горелик А.Л., Гуревич И.Б., Скрипкин В.А. Современное состояние проблемы распознавания. - М.: Высшая школа, 1985. - 160с.
5. Дуда Р., Харт П. Распознавание образов и анализ сцен: Пер. с англ. - М.: Мир, 1976. - 512 с.
6. Ту Дж., Гонсалес Р. Принципы распознавания образов: Пер. с англ. - М.: Мир, 1978. - 412с.
7. Фомин Я.А., Тарловский Г.Р. Статистическая теория распознавания образов. – М.: Радио и связь, 1986. – 264с.
8. Фор А. Восприятие и распознавание образов: Пер. с англ. - М.: Машиностроение, 1989. - 272с.
9. Фу К. Последовательные методы в распознавании образов и обучении машин: Пер. с англ. -М.: Наука, 1971, 256 стр.
10. Фукунага К. Введение в статистическую теорию распознавания образов: Пер. с англ. - М.: Наука, 1979. - 368с.

3. ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ

3.1. Исходные данные

- два файла данных, полученных в процессе выполнения лабораторной работы №1 и содержащих наборы двумерных нормально распределенных векторов признаков для ситуации равных корреляционных матриц; параметры этих законов распределения;
- три файла данных, полученных в процессе выполнения лабораторной работы №1 и содержащих наборы двумерных нормально распределенных векторов признаков для ситуации неравных корреляционных матриц; параметры этих законов распределения;
- два файла данных, полученных в процессе выполнения лабораторной работы №1 и содержащие наборы бинарных векторов признаков, распределения вероятностей бинарных векторов;
- исполняемые в системе MathCad файлы, необходимые при выполнении лабораторной работы: Lab21.mcd, Lab22.mcd, Lab23.mcd (предоставляются преподавателем).

3.2. Общий план выполнения работы

1. Построить байесовскую решающую границу между классами Ω_0 и Ω_1 двумерных нормально распределенных векторов признаков для случая равных корреляционных матриц и равных априорных вероятностей и изобразить ее графически. Вычислить вероятности ошибочной классификации и суммарную вероятность ошибочной классификации в этом случае.
2. Построить минимаксный классификатор и классификатор Неймана-Пирсона для вероятности ошибки первого рода $p_0^* = 0.05$ для двух классов Ω_0 и Ω_1 двумерных нормально распределенных векторов признаков в случае равных корреляционных матриц. Изобразить решающие границы полученных классификаторов графически.
3. Построить байесовскую решающую границу между классами Ω_0, Ω_1 и Ω_2 двумерных нормально распределенных векторов признаков для неравных корреляционных матриц и равных априорных вероятностей. Изобразить полученные решающие границы графически. Для любых двух классов оценить

экспериментально вероятности ошибочной классификации в этом случае и определить относительную погрешность полученных оценок для заданного объема обучающей выборки N. Определить объем обучающей выборки, обеспечивающий получение оценок вероятностей ошибочной классификации с погрешностью не более 5%.

4. Построить байесовскую разделяющую границу между классами Ω_0 и Ω_1 двумерных бинарных векторов признаков Вычислить вероятности ошибочной классификации аналитически и оценить их экспериментально.

3.3. Содержание отчета

Отчет по работе должен содержать:

1. Аналитические выражения для классификаторов, полученных в результате выполнения пп.1-3, и графическое изображение соответствующих им решающих границ вместе с элементами обучающих выборок.
2. Параметры классификатора, полученного в результате выполнения п.4 и его графическое изображение.
3. Вероятности ошибочной классификации построенных в пп.1-4 классификаторов, найденные аналитически или экспериментально. Для первого случая привести расчетные формулы, для второго - относительную погрешность оценки и объем выборки, гарантирующий величину погрешности не более 5%.

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Постановка задачи классификации.
2. Определение классификатора.
3. Способы задания классификатора.
4. Качество классификатора.
5. Байесовский классификатор.
6. Минимаксный классификатор.
7. Классификатор Неймана-Пирсона.
8. Отношение правдоподобия.
9. Байесовский классификатор для нормально распределенных векторов признаков.
10. Байесовский классификатор для распознавания бинарных векторов признаков.
11. Вычисление вероятностей ошибочной классификации.

СОДЕРЖАНИЕ