

Министерство образования и науки Российской Федерации

Самарский государственный аэрокосмический университет
имени академика С.П. Королева

ЛИНЕЙНЫЕ КЛАССИФИКАТОРЫ

Методические указания к лабораторной работе № 3
по курсу «МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ»

САМАРА

2004

Составители: к.ф.-м.н., доцент Э.И.Коломиец

к.т.н. В.В.Мясников

УДК 681.3

Линейные классификаторы

Методические указания к лабораторной работе № 3

Самарский государственный аэрокосмический университет

имени академика С.П.Королева

Составители: Э.И.Коломиец, В.В.Мясников

Самара, 2004. 21 с.

В лабораторной работе № 3 по курсу «Методы распознавания образов» изучаются методы построения линейных классификаторов, основанные на критериях минимума суммарной вероятности ошибочной классификации, минимума среднеквадратической ошибки решения, максимума критерия Фишера, а также базирующиеся на методе стохастической аппроксимации.

Методические указания предназначены для студентов специальности 01.02.00 «Прикладная математика и информатика», обучающихся по специализации «Математическое обеспечение обработки изображений».

Печатается по решению редакционно-издательского совета
Самарского государственного аэрокосмического университета
имени академика С.П.Королева

Рецензент: д.ф.-м.н., профессор А.И.Жданов

Данные методические указания разработаны 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).

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

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЛАБОРАТОРНОЙ РАБОТЫ

1.1. Постановка задачи построения линейного классификатора

Классификатор называется *линейным*, если его дискриминантная функция является линейной, то есть допускает представление в следующем виде:

$$d(\bar{x}) = \bar{W}^T \bar{x} + w_N \quad (1)$$

где $\bar{x} = (x_0, \dots, x_{N-1})^T$ - вектор признаков, который определяет образ объекта, подлежащего классификации, $\bar{W} = (w_0, \dots, w_{N-1})^T$ - вектор весовых коэффициентов классификатора, w_N - пороговое значение. При этом процесс принятия решения о номере класса текущего объекта производится в соответствии со следующим правилом:

$$d(\bar{x}) = \sum_{i=0}^{N-1} w_i x_i + w_N \begin{matrix} > 0 \\ < 0 \end{matrix} \Rightarrow \bar{X} \in \begin{cases} \Omega_0 \\ \Omega_1 \end{cases} \quad (2)$$

Задача *синтеза линейного классификатора* заключается в поиске вектора коэффициентов \bar{W} и порогового значения w_N оптимальным в каком-либо смысле образом.

Известно [4,5,7], что байесовский классификатор в случае нормальных распределений с равными корреляционными матрицами является линейным. Однако часто равенство корреляционных матриц является неприемлемым предположением, и, более того, возникает необходимость построения линейного классификатора вообще отказываясь от нормальности распределений.

1.2. Линейный классификатор, минимизирующий суммарную вероятность ошибочной классификации

Предположим, что случайная величина $d(\bar{X})$ (как функция случайного вектора \bar{X}) распределена по нормальному или близкому к нему закону

распределения в каждом из классов Ω_0 и Ω_1 . В этом случае для вычисления вероятностей ошибок достаточно знать лишь математические ожидания и дисперсии случайной величины $d(\bar{X})$ в каждом из классов.

Действительно, параметры одномерных нормальных распределений случайной величины $d(\bar{X})$ имеют вид:

$$m_l = M(d(\bar{X})/\Omega_l) = \bar{W}^T \bar{M}_l + w_N, \quad \sigma_l^2 = D(d(\bar{X})/\Omega_l) = \bar{W}^T B_l \bar{W}, \quad l = \overline{0,1}, \quad (3)$$

где

$$\bar{M}_l = M(\bar{X}/\Omega_l), \quad B_l = M((\bar{X} - \bar{M}_l)(\bar{X} - \bar{M}_l)^T / \Omega_l), \quad l = \overline{0,1}$$

- вектор математических ожиданий и корреляционная матрица случайного вектора \bar{X} , соответственно. В силу же (2) вероятности ошибочной классификации первого и второго рода [4,8] записываются следующим образом:

$$p_1 = \int_{-\infty}^0 f_d(u/\Omega_1) du, \quad p_0 = \int_0^{\infty} f_d(u/\Omega_0) du, \quad (4)$$

где $f_d(u/\Omega_l) \sim N(m_l, \sigma_l^2)$, $(l = \overline{0,1})$ - плотности вероятностей нормального закона распределения с соответствующими параметрами.

Зададим критерий оптимальности линейной дискриминантной функции в виде:

$$R = P(\Omega_0)p_0 + P(\Omega_1)p_1 \rightarrow \min_{\bar{W}, w_N}, \quad (5)$$

где $P(\Omega_0), P(\Omega_1)$ - априорные вероятности появления объектов классов Ω_0 и Ω_1 , соответственно. Критерий (5), таким образом, определяет оптимальную линейную дискриминантную функцию, которая дает минимальное значение суммарной вероятности ошибочной классификации.

С учетом (4) показатель качества критерия (5) может быть переписан в следующем виде:

$$R = P(\Omega_0) \int_{-\frac{m_0}{\sigma_0}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du + P(\Omega_1) \int_{-\infty}^{\frac{m_1}{\sigma_1}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du. \quad (6)$$

Необходимым условием минимума показателя (6), как известно, является

равенство нулю его частных производных по \bar{W} и w_N :

$$\begin{cases} 0 = \frac{\partial R}{\partial \bar{W}} = -P(\Omega_0) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) \cdot \frac{\partial}{\partial \bar{W}}\left(-\frac{m_0}{\sigma_0}\right) + P(\Omega_1) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right) \cdot \frac{\partial}{\partial \bar{W}}\left(-\frac{m_1}{\sigma_1}\right), \\ 0 = \frac{\partial R}{\partial w_N} = -P(\Omega_0) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) \cdot \frac{\partial}{\partial w_N}\left(-\frac{m_0}{\sigma_0}\right) + P(\Omega_1) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right) \cdot \frac{\partial}{\partial w_N}\left(-\frac{m_1}{\sigma_1}\right). \end{cases} \quad (7)$$

Подставляя в (7) выражения для частных производных

$$\frac{\partial}{\partial \bar{W}}\left(-\frac{m_l}{\sigma_l}\right) = -\frac{\bar{M}_l}{\sigma_l} + \frac{\bar{W}^T B_l m_l}{\sigma_l^3}, \quad \frac{\partial}{\partial w_N}\left(-\frac{m_l}{\sigma_l}\right) = -\frac{1}{\sigma_l}, \quad l = \overline{0,1},$$

окончательно получим следующую систему уравнений:

$$\begin{cases} P(\Omega_1) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right) \left(\bar{M}_1 - \frac{m_1}{\sigma_1^2} \bar{W}^T B_1\right) = P(\Omega_0) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right) \left(\bar{M}_0 - \frac{m_0}{\sigma_0^2} \bar{W}^T B_0\right), \\ P(\Omega_1) \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{m_1^2}{2\sigma_1^2}\right) = P(\Omega_0) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{m_0^2}{2\sigma_0^2}\right). \end{cases} \quad (8)$$

При выполнении второго равенства в (8) первое уравнение системы можно переписать в виде:

$$\bar{M}_1 - \bar{M}_0 = \left[\frac{m_1}{\sigma_1^2} B_1 - \frac{m_0}{\sigma_0^2} B_0 \right] \bar{W}. \quad (9)$$

Решение уравнения (9) может быть получено с помощью следующей итеративной процедуры (процедура Петерсона-Матсона):

$$\bar{W} = [sB_0 + (1-s)B_1]^{-1} (\bar{M}_1 - \bar{M}_0), \quad (10)$$

где

$$s = \frac{-\frac{m_0}{\sigma_0^2}}{\frac{m_1}{\sigma_1^2} - \frac{m_0}{\sigma_0^2}}, \quad 0 \leq s \leq 1$$

Из выражения для s следует справедливость равенства $s\sigma_0^2 m_1 + (1-s)\sigma_1^2 m_0 = 0$,

откуда, учитывая соотношения (3), получаем выражение для w_N :

$$w_N = -\frac{s\sigma_0^2 \overline{M}_1^T + (1-s)\sigma_1^2 \overline{M}_0^T}{s\sigma_0^2 + (1-s)\sigma_1^2} \overline{W} \quad (11)$$

С использованием выражений (10) и (11) для \overline{W} и w_n можно построить график зависимости суммарной вероятности ошибочной классификации R от единственного параметра s ($0 \leq s \leq 1$) и в качестве окончательного ответа принять то значение параметра, при котором $R(s)$ имеет минимум. Пример подобного решения приведен на рис.1.

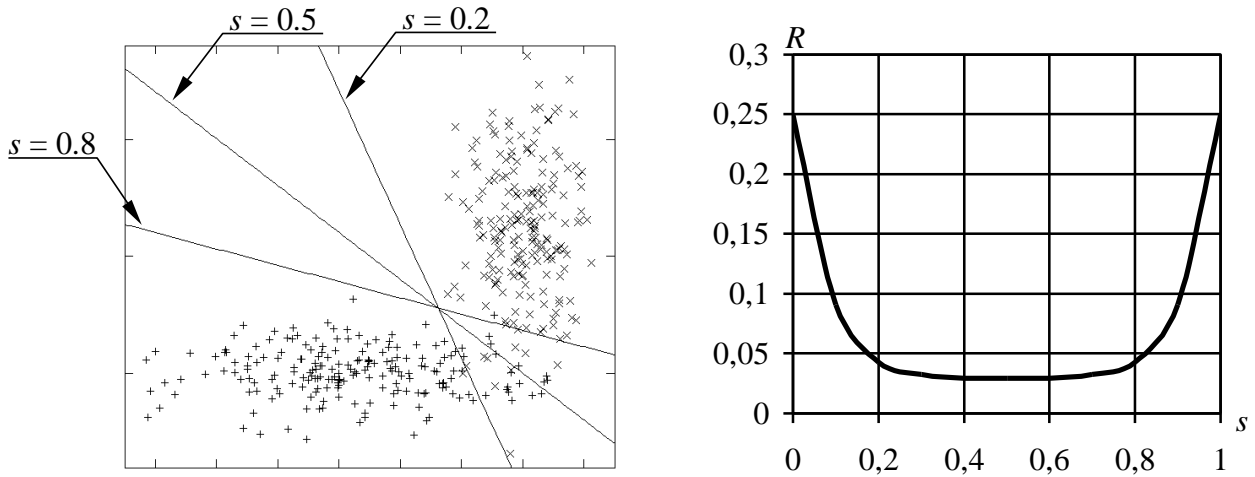


Рис.1. Построение линейного классификатора, минимизирующего суммарную вероятность ошибочной классификации.

1.3. Обобщенная формула построения линейных классификаторов для различных критериев. Классификатор Фишера

Изложенный выше подход построения линейного классификатора можно обобщить следующим образом. Рассмотрим некоторый критерий $f(m_0, m_1, \sigma_0^2, \sigma_1^2)$, экстремум которого соответствует оптимальному набору параметров линейной дискриминантной функции. Тогда справедливы следующие соотношения:

$$\begin{cases} 0 = \frac{\partial f}{\partial \overline{W}} = \frac{\partial f}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial \overline{W}} + \frac{\partial f}{\partial \sigma_0^2} \cdot \frac{\partial \sigma_0^2}{\partial \overline{W}} + \frac{\partial f}{\partial m_1} \cdot \frac{\partial m_1}{\partial \overline{W}} + \frac{\partial f}{\partial m_0} \cdot \frac{\partial m_0}{\partial \overline{W}}, \\ 0 = \frac{\partial f}{\partial w_N} = \frac{\partial f}{\partial \sigma_1^2} \cdot \frac{\partial \sigma_1^2}{\partial w_N} + \frac{\partial f}{\partial \sigma_0^2} \cdot \frac{\partial \sigma_0^2}{\partial w_N} + \frac{\partial f}{\partial m_1} \cdot \frac{\partial m_1}{\partial w_N} + \frac{\partial f}{\partial m_0} \cdot \frac{\partial m_0}{\partial w_N}. \end{cases} \quad (12)$$

Из соотношений (3) следует, что

$$\frac{\partial \sigma_l^2}{\partial \bar{W}} = 2B_l \bar{W}, \quad \frac{\partial \sigma_l^2}{\partial w_N} = 0, \quad \frac{\partial m_l}{\partial \bar{W}} = \bar{M}_l, \quad \frac{\partial m_l}{\partial w_N} = 1, \quad l = \overline{0,1}.$$

Подставляя эти выражения для частных производных в (12), окончательно получаем следующую общую систему уравнений:

$$\begin{cases} 2 \left[\frac{\partial f}{\partial \sigma_0^2} \cdot B_0 + \frac{\partial f}{\partial \sigma_1^2} \cdot B_1 \right] \bar{W} = (\bar{M}_0 - \bar{M}_1) \frac{\partial f}{\partial m_1}, \\ \frac{\partial f}{\partial m_0} = - \frac{\partial f}{\partial m_1}. \end{cases} \quad (13)$$

Задавая конкретный вид критерия $f(m_0, m_1, \sigma_0^2, \sigma_1^2)$, из (13) можно определить вектор весовых коэффициентов \bar{W} и пороговое значение w_N линейной дискриминантной функции, оптимальные в смысле этого критерия.

Классификатор Фишера

Выберем в качестве критерия $f(m_0, m_1, \sigma_0^2, \sigma_1^2)$ функцию вида:

$$f = \frac{(m_1 - m_0)^2}{\sigma_1^2 + \sigma_0^2}. \quad (14)$$

Критерий (14) называется *критерием Фишера* и представляет собой меру отличия значений линейной дискриминантной функции в классах Ω_1 и Ω_0 . Для наилучшего разделения классов необходимо определить \bar{W} и w_N , которые доставляли бы этому критерию максимум. Получаемый при этом линейный классификатор называется классификатором Фишера.

Подставляя выражение (14) в общую систему уравнений (13) и игнорируя масштабный множитель линейной функции, получим следующее выражение для вектора весовых коэффициентов дискриминантной функции:

$$\bar{W} = \left(\frac{1}{2} (B_1 + B_2) \right)^{-1} (\bar{M}_1 - \bar{M}_0) \quad (15)$$

Выражение (10) совпадает с (15) при значении параметра $s = 0.5$. Делая подстановку этого значения s в (11), получим выражение для порогового значения дискриминантной функции w_N :

$$w_N = - \frac{(\overline{M}_1 - \overline{M}_0)^T \left(\frac{1}{2} (B_1 + B_0) \right)^{-1} (\sigma_1^2 \overline{M}_0 + \sigma_0^2 \overline{M}_1)}{\sigma_1^2 + \sigma_0^2}. \quad (16)$$

Примечание.

Для ситуации равных корреляционных матриц $B \equiv B_1 = B_0$ выражения (15) и (16) преобразуются к следующим:

$$\overline{W} = B^{-1}(\overline{M}_1 - \overline{M}_0), \quad (15')$$

$$w_N = -\frac{1}{2}(\overline{M}_1 - \overline{M}_0)^T B^{-1}(\overline{M}_0 + \overline{M}_1). \quad (16')$$

Из (15') и (16') классификатор Фишера совпадает с байесовским классификатором для нормального закона распределения с равными корреляционными матрицами и равными априорными вероятностями классов:

$$d(\bar{x}) = (\overline{M}_1 - \overline{M}_0)^T B^{-1} \bar{x} - \frac{1}{2}(\overline{M}_1 - \overline{M}_0)^T B^{-1}(\overline{M}_0 + \overline{M}_1).$$

Пример классификатора Фишера приведен на рис.2.

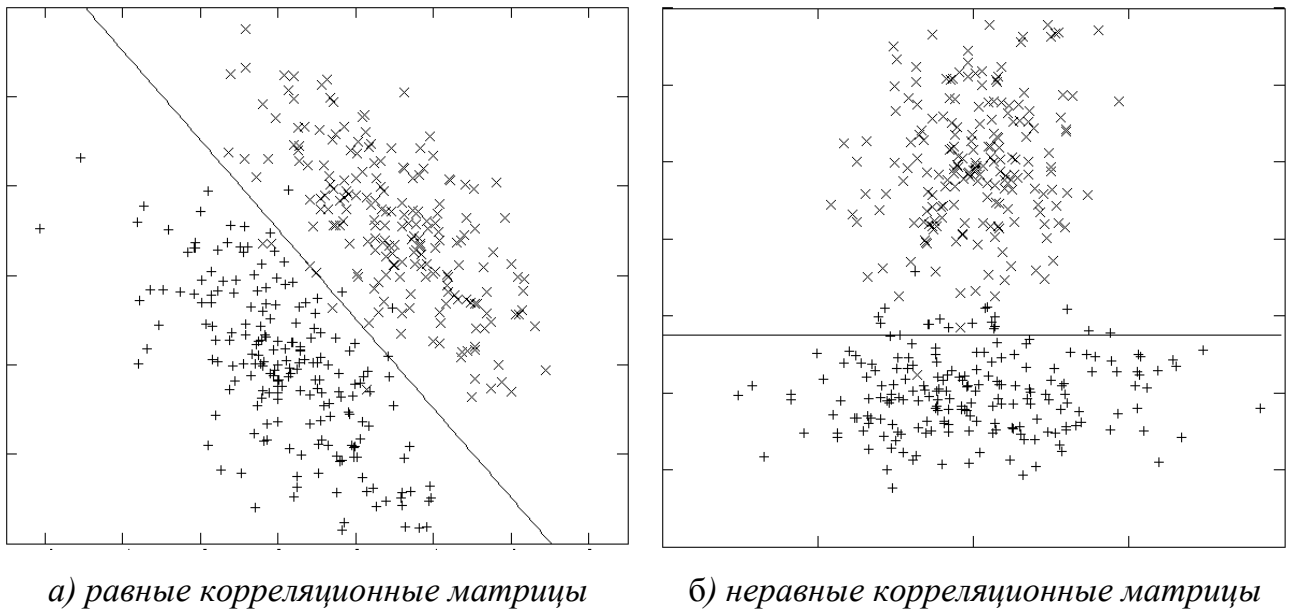


Рис.2. Классификатор Фишера

1.4. Линейный классификатор, минимизирующий СКО решения

При нахождении линейной разделяющей функции, минимизирующей суммарную вероятность ошибочной классификации предполагалось, что в классах Ω_1 и Ω_0 случайная величина $d(\bar{X})$ имеет нормальный закон распределения. Метод, основанный на минимизации *среднеквадратического отклонения* (СКО) решения позволяет получить аналогичные результаты без этого предположения.

Пусть при построении классификатора нам для наблюдения доступен набор из K значений векторов–реализаций $\bar{x}^1, \dots, \bar{x}^K$ случайного вектора признаков \bar{X} , относительно каждого из которых известно, какому из классов Ω_1 и Ω_0 он принадлежит. В этом случае говорят, что задана *обучающая выборка*.

Примем далее следующие обозначения.

- Введем в рассмотрение новый вектор $\bar{z} = (z_0, \dots, z_{N-1}, 1)^T$, формируемый следующим образом. Для объектов класса Ω_1 вектор $\bar{z} = (x_0, \dots, x_{N-1}, 1)^T$, а для объектов класса Ω_0 $\bar{z} = (-x_0, \dots, -x_{N-1}, -1)^T$.
- Дополним вектор весовых коэффициентов значением пороговой величины, то есть перейдем к пополненному вектору весовых коэффициентов:

$$\bar{W} = (w_0, \dots, w_{N-1}, w_N)^T,$$

Проведение таких преобразований позволяет записать линейную разделяющую функцию в виде:

$$d(\bar{z}) = \bar{W}^T \bar{z}.$$

Построение классификатора при этом сводится к определению пополненного вектора коэффициентов \bar{W} так, чтобы для любых известных векторов \bar{z}^k ($k = \overline{1, K}$) было справедливо неравенство:

$$d(\bar{z}) = \bar{W}^T \bar{z} \geq 0. \quad (17)$$

Обозначим $\gamma(\bar{z})$ *наилучшую разделяющую функцию*, то есть такую функцию, значения которой можно рассматривать в качестве требуемого выхода идеальной дискриминантной функции. Как правило $\gamma(\bar{z})$ не известна, но ее можно предположительно определить на основе обучающей выборки из условия (17). Например, можно взять $\gamma(\bar{z}) \equiv 1$, что соответствует случаю, когда на основе обучающей выборки определяется *переменная правильной классификации*:

$$r(\bar{x}) = \begin{cases} -1, & \bar{x} \in \Omega_0 \\ 1, & \bar{x} \in \Omega_1 \end{cases} \quad (18)$$

СКО между требуемым и действительным значением разделяющей функции определяется выражением:

$$\varepsilon^2 = M \left(\gamma(\bar{Z}) - \bar{W}^T \bar{Z} \right)^2 \quad (19)$$

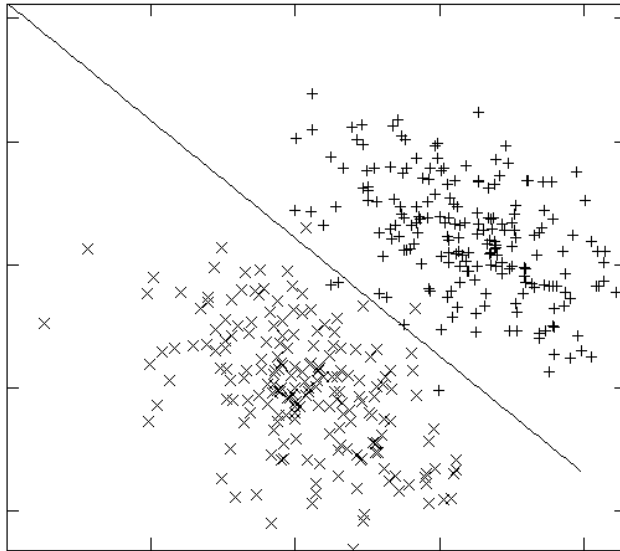
(здесь математическое ожидание соответствует распределению случайного вектора \bar{Z}). Если вместо математического ожидания используется среднее по обучающей выборке, то имеем:

$$\varepsilon^2 = \frac{1}{K} \sum_{k=1}^K \left(\bar{W}^T \bar{z}^k - \gamma(\bar{z}^k) \right)^2 = \frac{1}{N} \left(\bar{W}^T U - \bar{\Gamma} \right) \cdot \left(U^T \bar{W} - \bar{\Gamma} \right), \quad (20)$$

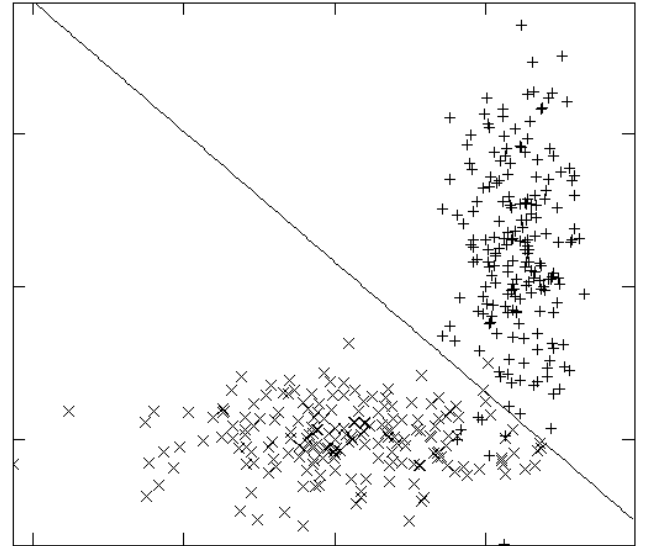
где $U = (\bar{z}^1, \bar{z}^2, \dots, \bar{z}^K)$ - матрица выборочных данных, а $\bar{\Gamma} = (\gamma(\bar{z}_1), \dots, \gamma(\bar{z}_N))^T$ - вектор требуемых значений выхода. Дифференцируя выражение (20) по пополненному вектору коэффициентов и приравнявая частные производные нулю, получим следующее выражение для \bar{W} :

$$\bar{W} = (UU^T)^{-1} U\bar{\Gamma}.$$

Пример классификатора, минимизирующего СКО решения, приведен на рис.3.



а) равные корреляционные матрицы



б) неравные корреляционные матрицы

Рис.3. Линейный классификатор, минимизирующий СКО решения

1.5. Последовательная корректировка линейного классификатора

Рассмотрим алгоритм, использование которого для расчета вектора параметров линейного классификатора не требует запоминания одновременно всех векторов признаков. Вместо этого в памяти ЭВМ следует хранить только текущие оценки параметров, которые обновляются всякий раз при поступлении очередного вектора наблюдений. Система распознавания такого типа носит название *перцептрон*, а процесс ее итерационной настройки – *обучение*. Основное преимущество подобных алгоритмов в том, что они позволяют использовать бесконечное число наблюдений, располагая конечным объемом памяти.

Итак, пусть решается задача распознавания объектов двух классов, заданных своими пополненными векторами признаками:

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

Предположим, что в качестве дискриминантной функции выбрана *линейная*, и классификатор работает в соответствии со следующим правилом:

$$d(\bar{x}) = \bar{W}^T \bar{x} = \sum_{i=0}^N w_i x_i \begin{matrix} < 0 \\ > 0 \end{matrix} \Rightarrow \bar{x} \in \begin{cases} \Omega_0 \\ \Omega_1 \end{cases},$$

где $\bar{W} = (w_0, w_1, \dots, w_{N-1}, w_N)^T$ – пополненный вектор весовых коэффициентов.

1.5.1. Алгоритм перцептрона

Пусть $\{\bar{x}^k\}_{k=1}^K$ – обучающая выборка объемом K . Обозначим $\bar{x}(k)$ – элемент выборки, используемый на k -ом шаге алгоритма настройки, $\bar{W}(k)$ – оценка искомого вектора \bar{W} на k -ом шаге алгоритма.

Классический алгоритм обучения перцептрона, предложенный Розенблатом в [5], выглядит следующим образом:

$$\bar{W}(k+1) = \begin{cases} \bar{W}(k), & \bar{x}(k) \in \Omega_1 \text{ и } \bar{W}^T(k) \bar{x}(k) > 0 \text{ или } \bar{x}(k) \in \Omega_0 \text{ и } \bar{W}^T(k) \bar{x}(k) < 0, \\ \bar{W}(k) + c\bar{x}(k), & \bar{x}(k) \in \Omega_1 \text{ и } \bar{W}^T(k) \bar{x}(k) < 0, \\ \bar{W}(k) - c\bar{x}(k), & \bar{x}(k) \in \Omega_0 \text{ и } \bar{W}^T(k) \bar{x}(k) > 0. \end{cases}$$

Выбор параметра c в алгоритме обучения перцептрона производится в соответствии с одним из нижеследующих правил.

Правило1. Правило фиксированного приращения

Выбирается произвольное постоянное значение $c > 0$.

Правило2. Правило полной коррекции

Значение параметра c выбирается таким, чтобы текущий вектор признаков был проклассифицирован верно. А именно:

$$c > \frac{|\bar{W}^T(k)\bar{x}(k)|}{\bar{x}^T(k)\bar{x}(k)}.$$

В этом случае параметр c является переменным: $c = c(k)$.

Правило 3. Градиентное правило коррекции

Данное правило используется, если качество линейной дискриминантной функции определяется некоторым функционалом $J(\bar{W})$, минимизацию или максимизацию которого необходимо произвести в процессе настройки классификатора. Экстремум функционала может быть найден, если двигаться вдоль градиента функции качества, то есть использовать градиентное правило коррекции коэффициентов:

$$\bar{W}(k+1) = \bar{W}(k) - \rho \left. \frac{\partial J(\bar{W})}{\partial \bar{W}} \right|_{\bar{W}=\bar{W}(k)},$$

где $\rho > 0$ - параметр градиентного алгоритма. Например, при минимизации СКО, когда

$$J(\bar{W}) = \frac{1}{2} (\gamma(\bar{x}) - \bar{W}^T \bar{x})^2,$$

алгоритм обучения выглядит следующим образом:

$$\bar{W}(k+1) = \bar{W}(k) + \rho \bar{x}(k) (\gamma(\bar{x}(k)) - \bar{W}^T(k) \bar{x}(k)).$$

В этом случае параметр c , очевидно, равен:

$$c = -\rho (\gamma(\bar{x}(k)) - \bar{W}^T(k) \bar{x}(k))$$

и также, как и в предшествующем случае, зависит от k : $c = c(k)$.

Примечание

Алгоритм обучения перцептрона *сходится* за конечное число итераций (то есть строится классификатор, безошибочно классифицирующий все элементы выборочных данных), если классы являются *линейно разделимыми* (см. рис.4). В противном случае алгоритм обучения *защивается*.

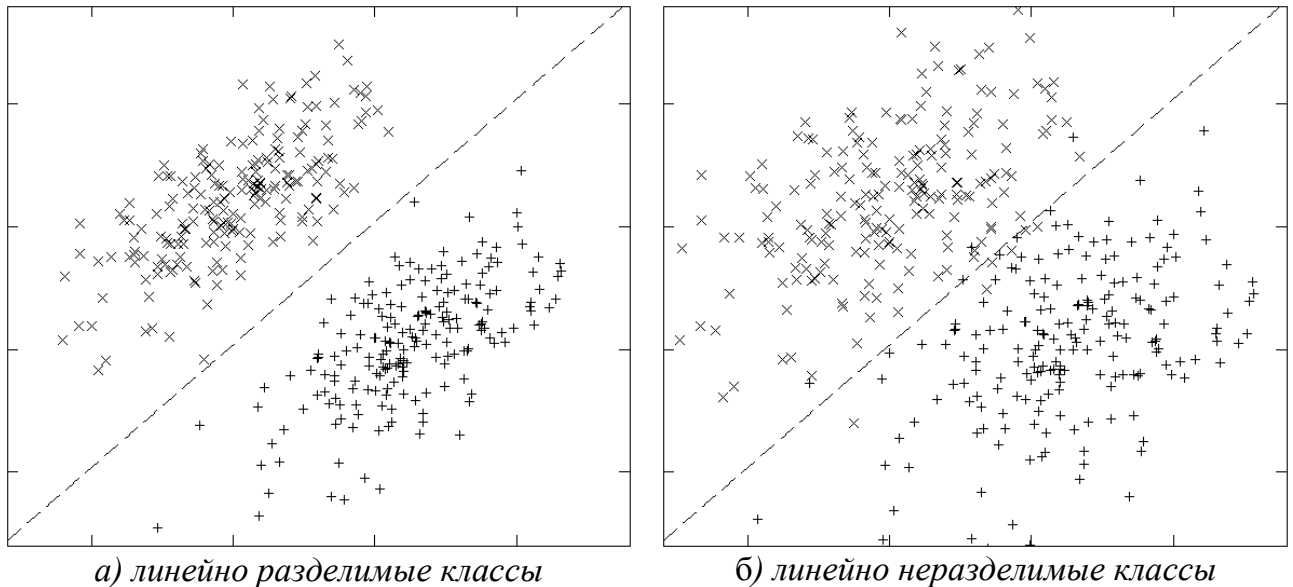


Рис.4. Иллюстрация линейной разделимости классов

1.5.2. Стохастическая аппроксимация. Процедура Роббинса-Монро

Алгоритм обучения перцептрона не сходится, если классы, заданные обучающими выборками, не являются линейно разделимыми. Этот факт выдвигает задачу построения алгоритма оценивания вектора коэффициентов линейной разделяющей функции \bar{W} , сходимость которого была бы гарантирована всегда. Такой алгоритм может быть построен с помощью метода *стохастической аппроксимации*.

Идея метода состоит в том, что критерий $J(\bar{W})$ рассматривают как функцию регрессии вида: $J(\bar{W}) = MF(\bar{W}, \bar{X})$, где $F(\bar{W}, \bar{X})$ - значение функционала качества, наблюдаемого в случайной точке \bar{X} . Метод стохастической аппроксимации позволяет определять по результатам наблюдений корень следующего уравнения, называемого *уравнением регрессии*:

$$M \frac{\partial F(\bar{W}, \bar{X})}{\partial \bar{W}} \equiv Mh(\bar{W}, \bar{X}) = 0.$$

Процедура Роббинса-Монро

Процедура Роббинса-Монро – это итеративная процедура поиска корня уравнения регрессии. Обозначим $\bar{W}(1)$ произвольную начальную оценку корня \tilde{W} и $\bar{W}(k)$ – оценку этого корня, полученного на k -ом шаге итерации. Тогда решение уравнения регрессии может быть получено в результате следующего итерационного процесса:

$$\bar{W}(k+1) = \bar{W}(k) - \alpha_k h(\bar{W}(k), \bar{X}(k)), \quad (21)$$

где α_k – элемент последовательности положительных чисел, удовлетворяющий следующим условиям:

$$\lim_{k \rightarrow \infty} \alpha_k = 0, \quad \sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty. \quad (22)$$

Утверждение [8]: Если последовательность $\{\alpha_k\}_{k=1}^{\infty}$ удовлетворяет условиям (22) и выполнены некоторые дополнительные условия, то оценка (21) сходится к корню \tilde{W} в среднеквадратическом и с вероятностью равной единице, то есть

$$\lim_{k \rightarrow \infty} M(W(k) - \tilde{W})^2 = 0, \quad \lim_{k \rightarrow \infty} P(W(k) = \tilde{W}) = 1$$

Примером последовательности, удовлетворяющей условиям (22), является последовательность $\{\alpha_k\}_{k=1}^{\infty}$ с

$$\alpha_k = \frac{1}{k^\beta}, \quad \frac{1}{2} < \beta \leq 1.$$

1.5.3. Общая схема построения линейных классификаторов, основанная на методе стохастической аппроксимации

Зададимся критерием $J(\bar{W})$ вида:

$$J(\bar{W}) = MF(\bar{W}^T \bar{X} - \gamma(\bar{X}))$$

где $F(\dots)$ – некоторая выпуклая функция (например, модуль), $\gamma(\bar{X})$ – требуемый выход разделяющей функции. После дифференцирования по \bar{W} получим:

$$\frac{\partial J(\bar{W})}{\partial \bar{W}} = M \frac{\partial F(\bar{W}^T \bar{X} - \gamma(\bar{X}))}{\partial \bar{W}} = 0.$$

Полученное уравнение является уравнением регрессии. Воспользовавшись теперь процедурой Роббинса-Монро, можно получить последовательность оценок вектора коэффициентов линейного классификатора, положив:

$$h(\bar{W}(k), \bar{X}(k)) = \left. \frac{\partial F(\bar{W}^T \bar{X} - \gamma(\bar{X}))}{\partial \bar{W}} \right|_{\bar{W}=\bar{W}(k), \bar{X}=\bar{X}(k)}$$

и записав алгоритм в виде:

$$\bar{W}(k+1) = \bar{W}(k) - \alpha_k \left. \frac{\partial F(\bar{W}^T \bar{X} - \gamma(\bar{X}))}{\partial \bar{W}} \right|_{\bar{W}=\bar{W}(k), \bar{X}=\bar{X}(k)}$$

где начальный вектор $\bar{W}(1)$ выбирается произвольно, а последовательность $\{\alpha_k\}_{k=1}^{\infty}$ удовлетворяет условиям (22). Рассмотрим два из возможных алгоритмов, основанных на методе стохастической аппроксимации.

Алгоритм корректирующих приращений (АКП-алгоритм)

В АКП-алгоритме критерий качества классификатора задается в виде:

$$J(\bar{W}) = M |r(\bar{X}) - \bar{W}^T \bar{X}|,$$

где $r(\bar{X}) = \begin{cases} 1, \bar{X} \in \Omega_1 \\ -1, \bar{X} \in \Omega_0 \end{cases}$ - случайная переменная правильной классификации.

Поскольку производная критерия:

$$\frac{\partial J(\bar{W})}{\partial \bar{W}} = -M \bar{X} \operatorname{sgn}(r(\bar{X}) - \bar{W}^T \bar{X}),$$

то в соответствии с общей схемой получаем следующий алгоритм настройки коэффициентов линейной дискриминантной функции:

$$\bar{W}(k+1) = \bar{W}(k) + \alpha_k \bar{x}(k) \operatorname{sgn}(r(\bar{x}(k)) - \bar{W}^T(k) \bar{x}(k)).$$

Алгоритм можно переписать также в более наглядном виде:

$$\bar{W}(k+1) = \begin{cases} \bar{W}(k) + \alpha_k \bar{x}(k), & r(\bar{x}(k)) \geq \bar{W}^T(k) \bar{x}(k), \\ \bar{W}(k) - \alpha_k \bar{x}(k), & r(\bar{x}(k)) < \bar{W}^T(k) \bar{x}(k). \end{cases}$$

Очевидно, полученный алгоритм корректирует текущую оценку вектора коэффициентов на каждом шаге.

В случае линейной разделимости классов АКП-алгоритм сходится к точному решению, то есть получающийся в результате классификатор все векторы признаков обучающей выборки классифицирует верно. В случае, когда классы линейно разделимыми не являются, в пределе получается решение, оптимальное в смысле минимизации абсолютной величины его расхождения с переменной правильной классификации.

Алгоритм наименьшей СКО (НСКО-алгоритм)

В НСКО-алгоритме критерий качества классификатора имеет вид:

$$J(\bar{W}) = \frac{1}{2} M(r(\bar{X}) - \bar{W}^T \bar{X})^2$$

Поскольку производная критерия:

$$\frac{\partial J(\bar{W})}{\partial \bar{W}} = -M \bar{X} (r(\bar{X}) - \bar{W}^T \bar{X}),$$

то в этом случае имеется следующий алгоритм настройки коэффициентов линейного классификатора:

$$\bar{W}(k+1) = \bar{W}(k) + \alpha_k \bar{x}(k) (r(\bar{x}(k)) - \bar{W}^T(k) \bar{x}(k)).$$

Аналогично АКП-алгоритму, в НСКО-алгоритме коррекции также производятся на каждом шаге. Отличие заключается только в величине этих коррекций. Для линейно неразделимых классов алгоритм сходится в смысле минимизации величины СКО между решением и переменной правильной классификации.

Проблемы алгоритмов и способы их разрешения

Одной из основных проблем алгоритмов построения линейных классификаторов на основе метода стохастической аппроксимации является очень медленная скорость сходимости. Основные приемы, используемые для ускорения сходимости следующие:

- выбор медленно убывающей последовательности $\{\alpha_k\}_{k=1}^{\infty}$,
- изменение значения α_k , если только величина $(r(\bar{x}(k)) - \bar{W}^T(k)\bar{x}(k))$ на соседних итерациях изменила знак,
- использование «перцептронного приема», при котором коррекция вектора коэффициентов производится только в случае неверной классификации поступившего вектора признаков.

2. ЛИТЕРАТУРА

1. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений. - М.: Высшая школа, 1983. - 295 с.
2. Горелик А.Л., Скрипкин В.А. Методы распознавания. - М.: Высшая школа, 1984. - 208с.
3. Дуда Р., Харт П. Распознавание образов и анализ сцен: Пер. с англ. - М.: Мир, 1976. - 512 с.
4. Коломиец Э.И., Мясников В.В. Байесовская классификация: Методические указания к лабораторной работе № 2 по курсу «Методы распознавания образов».
5. Розенблатт Ф. Принципы нейродинамики. – М.: Мир., 1965.
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. 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; параметры байесовского классификатора для ситуации равных корреляционных матриц из лабораторной работы №2;
- два 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; параметры байесовского классификатора для ситуации неравных корреляционных матриц из лабораторной работы №2;
- исполняемые в системе MatLab файлы, необходимые при выполнении лабораторной работы: Lab31.m, Lab32.m (предоставляются преподавателем).

3.2. Общий план выполнения работы

1. Построить линейный классификатор, максимизирующий критерий Фишера, для классов Ω_0 и Ω_1 двумерных нормально распределенных векторов признаков для случаев равных и неравных корреляционных матриц. Сравнить качество полученного классификатора с байесовским классификатором.
2. Построить линейный классификатор, минимизирующий среднеквадратичную ошибку, для классов Ω_0 и Ω_1 двумерных нормально распределенных векторов признаков для случаев равных и неравных корреляционных матриц. Сравнить качество полученного классификатора с байесовским классификатором и классификатором Фишера.
3. Построить линейный классификатор, основанный на процедуре Роббинса-Монро, для классов Ω_0 , и Ω_1 двумерных нормально распределенных векторов признаков для случаев равных и неравных корреляционных матриц. Исследовать зависимость скорости сходимости итерационного процесса и качества классификации от выбора начальных условий и выбора последовательности корректирующих коэффициентов. Сравнить качество полученного классификатора с байесовским классификатором.

3.3. Содержание отчета

Отчет по работе должен содержать:

1. Аналитические выражения для классификаторов, полученных в результате выполнения пп.1-2, и графическое изображение соответствующих им решающих границ вместе с элементами обучающих выборок.
2. Параметры классификатора, полученного в результате выполнения п.3 и его графическое изображение.
3. Вероятности ошибочной классификации построенных в пп.1-3 классификаторов, найденные экспериментально. Результаты сравнения построенных классификаторов с байесовским классификатором.
4. Графическая иллюстрация работы итерационного процесса построения классификатора с помощью процедуры Роббинса-Монро; результаты исследования скорости сходимости этого процесса и качества классификации в зависимости от начальных условий и последовательности корректирующих коэффициентов.

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Процедура Петерсона-Матсона построения линейной дискриминантной функции, минимизирующей суммарную вероятность ошибочной классификации.
2. Классификатор Фишера.
3. Линейная разделяющая функция, минимизирующая СКО решения.
4. Алгоритм перцептрона.
5. Стохастическая аппроксимация. Идея метода.
6. Процедура Роббинса-Монро.
7. Схема построения линейных классификаторов, основанная на методе стохастической аппроксимации.
8. АКП-алгоритм.
9. НСКО-алгоритм.
10. Способы ускорения сходимости алгоритмов построения линейных классификаторов, основанных на методе стохастической аппроксимации.

СОДЕРЖАНИЕ

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЛАБОРАТОРНОЙ РАБОТЫ	3
1.1. Постановка задачи построения линейного классификатора	3
1.2. Линейный классификатор, минимизирующий суммарную вероятность ошибочной классификации.....	3
1.3. Обобщенная формула построения линейных классификаторов для различных критериев. Классификатор Фишера.....	6
1.4. Линейный классификатор, минимизирующий СКО решения	9
1.5. Последовательная корректировка линейного классификатора.....	9
1.5.1. Алгоритм перцептрона.....	11
1.5.2. Стохастическая аппроксимация. Процедура Роббинса-Монро.	13
1.5.3. Общая схема построения линейных классификаторов, основанная на методе стохастической аппроксимации.	14
2. ЛИТЕРАТУРА.....	17
3. ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ.....	18
3.1. Исходные данные.....	18
3.2. Общий план выполнения работы	18
3.3. Содержание отчета	19
4. КОНТРОЛЬНЫЕ ВОПРОСЫ	19

Учебное издание

ЛИНЕЙНЫЕ КЛАССИФИКАТОРЫ

Методические указания к лабораторной работе № 3
по курсу «Методы распознавания образов»

Составители: Коломиец Эдуард Иванович
Мясников Владислав Валерьевич

Самарский государственный аэрокосмический университет
имени академика С.П.Королева
443086, Самара, Московское шоссе, 34

Отпечатано на кафедре геоинформатики СГАУ

Тираж 20 экз.