

А.А.Грешилов,
В.А.Стакун, А.А.Стакун

СТАТИСТИЧЕСКИЕ
МЕТОДЫ
ПРИНЯТИЯ РЕШЕНИЙ
С ЭЛЕМЕНТАМИ
КОНФЛЮЕНТНОГО
АНАЛИЗА



Москва "Радио и связь"
1998

УДК 621.391

Грешилов А.А., Стакун В.А., Стакун А.А. Статистические методы принятия решений с элементами конфлюэнтного анализа. - М.: Радио и связь, 1998. 112 с. ISBN 5-256-01420-X.

Рассмотрены статистические задачи решения (распознавания образов) с учетом погрешностей результатов наблюдений признаков, а как следствие с учетом интервальных оценок функций распределений, решающих функций, ошибок первого и второго рода. Рассмотрены задачи при фиксированном объеме выборки и методы последовательного принятия решений.

Книга рассчитана на специалистов занимающихся задачами принятия решений, на студентов вузов и на слушателей системы дополнительного профессионального образования, изучающих подобные задачи.

Табл. 7. Ил. 14. Библиогр. 76 назв.

Заказное издание

Научное издание

**Грешилов Анатолий Антонович, Стакун Виталия Анатольевна,
Стакун Андрей Анатольевич**

**СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРИНЯТИЯ РЕШЕНИЙ С ЭЛЕМЕНТАМИ
КОНФЛЮЭНТНОГО АНАЛИЗА**

Редактор **Г р е ш и л о в А. А.**

ИБ № 2843

ЛР № 010164 от 29.01.97

Сдано в набор 16.02.98

Подписано в печать 22.05.98

Формат 60x90/16 Бумага офс. № 1 Гарнитура литературная Печать офсетная
Усл.печ.л. 7,0 Усл.-кр.отг. 7,25 Уч.-изд. Л. 6,35 Тираж 500 экз.

Издательство "Радио и связь". 101000 Москва, Почтамт. а/я 693
Типография издательства "Радио и связь". 101000 Москва, Почтамт. а/я 693

ISBN 5-256-01420-X

© Грешилов А.А., Стакун В.А., Стакун А.А., 1998

ВВЕДЕНИЕ

Теория принятия решений в последние годы приобретает все большее значение в различных областях знания: в технике, экономике, медицине и др. Она перешагнула свои границы в узком смысле и все шире проникает в область управления, исследования операций, радиолокацию и т.д. Число статей и книг, посвященных этой проблеме, публикуемых в самых разных изданиях, резко выросло.

Алгоритмы распознавания образов по совокупности признаков различаются: этапностью принятия решений; степенью и характером учета статистики признаков, помех, сигналов.

По этапности различают алгоритмы одноэтапного и многоэтапного [8, 39, 40, 41] принятия решений. Одноэтапное принятие решений предусматривает обязательную выдачу оценки $k=i$ класса i известного алфавита с приемлемой достоверностью. К многоэтапности принятия решений могут вести следующие соображения: целесообразность отказа от выдачи решения на первом этапе (первых этапах) до дополнительного набора признаков (последовательные алгоритмы типа Вальда); целесообразность принятия загрубленного решения до дополнительного набора признаков; необходимость обобщения предварительных решений, полученных в различные моменты времени или от различных источников.

По степени учета статистических закономерностей различают лингвистические и статистические алгоритмы. По характеру учета статистических закономерностей из статистических алгоритмов выделяют параметрические (байесовские и небайесовские), непараметрические и нейрокомпьютерные алгоритмы.

Лингвистические алгоритмы [11, 38, 52, 55] не учитывают статистики признаков. Вводимые признаки описывают качественно, часто двоичными цифрами 0,1. Описание признаков на языке алгебры логики (языковое, кодовое, синтаксическое) служит при этом основой распознавания.

Байесовские параметрические алгоритмы, в отличие от параметрических небайесовских, учитывают не только статистику параметров помех, флуктуаций сигналов и признаков, но и определенные гипотезы об априорных вероятностях P , различных элементов алфавита классов [18, 51, 56]. Структура алгоритмов и работающих по ним устройств обработки сигналов определяется из математических расчетов. Статистика признаков сигналов, негауссовская в общем случае, устанавливается путем натурального эксперимента, математического или

физического моделирования. Введение этой статистики можно трактовать как обучение распознаванию, адаптацию к конкретным условиям распознавания.

Непараметрические алгоритмы синтезируются эвристически без явного принятия предположений о конкретных статистических распределениях [18, 21, 32, 36, 38, 51, 56]. Их можно рассматривать в ряде случаев как эвристическое упрощение параметрических байесовских.

Нейрокомпьютерные алгоритмы отличаются своей заранее заданной универсальной структурой с большим числом неизвестных параметров, уточняемых в процессе адаптации (обучения) [68, 74]. Возрастание вычислительных затрат как издержку универсализации компенсируют ростом производительности вычислительных средств.

К решению задач распознавания привлекают также ряд математических теорий и методов, развитие которых связано с появлением экспертных систем [23, 50, 57]: теорию нечетких множеств и теорию возможностей [47, 53], теорию игр и другие математические методы [4, 5, 19].

В задачах распознавания образов и принятия решений важную роль играет тот факт, как учтены (насколько строго) неопределенности исходных данных. Использование усредненных величин ведет к смещенным оценкам основных показателей, определяющих решение, а, как следствие, к неверным практическим выводам. Поэтому здесь особое внимание уделено строгому учету погрешностей измерений вектора признака объектов.

Традиционные статистические методы распознавания образов часто не учитывают погрешности в наблюдаемых значениях признаков, что приводит к следующему:

1) оценки функций условных плотностей вероятности признаков образов (если даже вид плотностей известен априори), которые получают по результатам наблюдений, будут смещенными, а их интервальные оценки - неверными;

2) в процессе идентификации образов, имеющих близкие значения координат векторов признаков, из-за влияния погрешностей на результаты наблюдений может быть принято неверное решение;

3) в традиционных методах функции условных плотностей вероятности признаков выступают в процедурах идентификации образов как детерминированные (не учитываются их интервальные оценки), и по этой причине не рассматриваются имеющие место зоны неопределенности (нулевые зоны) в принятии решений.

В отличие от традиционных статистических методов распознавания образов в данной книге разработаны и используются методы, которые позволяют учесть погрешности наблюдаемых значений координат векторов признаков объектов, получить несмещенные точечные и интервальные оценки функций условных плотностей вероятности

признаков оценить "истинные" координаты вектора признаков, по которому ведется идентификация объектов, и включить эти данные в процедуру принятия решений.

Таким образом излагаемый материал представляет интерес для различных областей знаний и практических приложений и является актуальной.

В первой главе книги дается анализ традиционных методов статистических задач решения; рассматриваются как стохастический, так и детерминистский подходы, а также задачи решения с фиксированным объемом выборки и последовательная решающая модель.

Во второй главе описываются способы учета погрешностей вектора признаков в статистических задачах решения с фиксированным объемом выборки, когда вид функций обобщенных условных плотностей вероятности известен априори, но неизвестны оценки их параметров, приводятся алгоритмы получения несмещенных точечной и интервальных оценок разделяющей функции, а также оценок "истинных" значений измеренного вектора признаков, по которым принимается решение. Здесь же приводятся сведения из конъюнктного анализа совокупности алгоритмов, позволяющих учитывать одновременно погрешности как в значениях функции, так и в значениях аргументов. Показаны отличия в решениях, принимаемых согласно традиционным и предлагаемым в диссертации методов.

В третьей главе рассмотрена задача распознавания образов, когда обобщенные условные плотности вероятности не заданы, но предполагается вид разделяющих функций, параметры которых подлежат определению с учетом погрешности наблюдаемых значений признаков.

Работа выполнена при общем научном руководстве профессора Грешилова А.А., процедура принятия решений по выборке фиксированного объема с учетом погрешности признаков разрабатывалась Стакун В.А., процедуры последовательных методов принятия разрабатывалась Стакуном А.А.

Книга рассчитана на специалистов занимающихся задачами принятия решений, на студентов вузов и на слушателей системы дополнительного профессионального образования, изучающих подобные задачи.

Г Л А В А 1. АНАЛИЗ ТРАДИЦИОННЫХ МЕТОДОВ В СТАТИСТИЧЕСКИХ ЗАДАЧАХ ПРИНЯТИЯ РЕШЕНИЙ И МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ УЧЕТА ПОГРЕШНОСТИ ПРИЗНАКОВ

1.1. Основные понятия и определения

Пусть D пространство решений d , R пространство доходов r , которые можно получить в результате решения d и исхода ω эксперимента; Ω - пространство возможных исходов эксперимента ω . Считаем заданным вероятностное распределение P на пространстве исходов Ω . На множестве R задана функция полезности u . Тогда для всякого вероятностного распределения P_d для которого функция u интегрируема, средняя полезность

$$E(u|P_d) = \int_R u(r) dP_d(r) = \int_{\Omega} u(\omega, d) dP(\omega).$$

Следует выбрать решение d , максимизирующее $E(u|P_d)$. Обычно в задачах решения каждому доходу $r \in R$ принято сопоставлять не полезность, а ущерб, имеющий смысл отрицательной полезности: для всех исходов $\omega \in \Omega$ всех решений $d \in D$ ущерб (потери) $L(\omega, d) = -u(\omega, d)$. Вещественная функция потерь $L(\omega, d)$ задается на произведении $\Omega \times D$. При любом $(\omega, d) \in \Omega \times D$ число $L(\omega, d)$ представляет собой ущерб от принятия решения d , когда значение параметра w равно ω (в случае исхода ω).

Пусть P данное вероятностное распределение параметра w . При всяком решении $d \in D$ средний ущерб $\rho(P, d)$, называемый риском, определяется формулой

$$\rho(P, d) = \int_{\Omega} L(\omega, d) dP(\omega).$$

Выбирается решение d , при котором минимизируется риск $\rho(P, d)$.

Пусть Ω параметрическое пространство с параметром w , принимающем значения ω .

Для всякого распределения P параметра w байесовский риск $\rho^*(P)$ определяется как точная нижняя грань рисков $\rho(P, d)$ по всем решениям $d \in D$, т.е.

$$\rho^*(P) = \inf_{d \in D} \rho(P, d).$$

Каждое решение d^* риск которого равен байесовскому риску, называется байесовским решением при распределении P , т.е.

$$\rho^*(P) = \rho(P, d^*).$$

В ряде случаев байесовское решение может не достигаться.

Во многих задачах удобнее иметь дело с неотрицательными функциями потерь. Оказывается, любую функцию потерь можно заменить неотрицательным ее аналогом. Рассмотрим новую функцию потерь $L_0(\omega, d)$, определяемую по начальной функции потерь как

$$L_0(\omega, d) = \alpha L(\omega, d) + \lambda(\omega); \quad \omega \in \Omega; \quad d \in D.$$

Исходной функции $L(\omega, d)$, соответствует риск $\rho(P, d)$, а функции $L_0(\omega, d)$ риск $\rho_0(P, d)$. Тогда для любых значений $d_1 \in D$ и $d_2 \in D$ соотношения $\rho_0(P, d_1) < \rho_0(P, d_2)$ и $\rho(P, d_1) < \rho(P, d_2)$ равносильны. В частности, решение d^* тогда и только тогда является байесовским при распределении P для исходной задачи, когда решение d^* является байесовским решением при P для новой задачи с функцией $L_0(\omega, d)$. Выбирая $\lambda(\omega)$ и ее знак, можно получить $L_0(\omega, d)$ такое, что при всех $\omega \in \Omega; d \in D$ $L_0(\omega, d) \geq 0; \inf L_0(\omega, d) = 0$.

В любой задаче решения байесовский риск $\rho^*(P)$ является вогнутой функцией от распределения P параметра w , т.е. для любых распределений P_1 и P_2 параметра w и для любого числа α , такого, что $0 \leq \alpha \leq 1$

$$\rho^*[\alpha P_1 + (1 - \alpha) P_2] \geq \alpha \rho^*(P_1) + (1 - \alpha) \rho^*(P_2).$$

В общем случае байесовский риск относительно мало чувствителен к ошибке (к приращению) в выборе значения распределения P параметра w ; если функция $\rho^*(P)$ кусочно-линейна, то приращение $\Delta \rho^*(P)$ равно нулю, когда приращение P содержится в интервале линейности функции $\rho^*(P)$.

1.2. Статистические задачи решения с наблюдениями

Перед тем, как выбрать решение из множества D , наблюдается значение случайной величины или случайного вектора X , связанных с параметром ω . Наблюдение X помогает принять рациональное решение.

Предполагается, что для всех $\omega \in \Omega$ задано условное распределение X при $w = \omega$. Эти задачи называются статистическими задачами решения. Основные элементы статистической задачи решения - параметрическое пространство Ω , пространство решений D , функция потерь $L(\omega, d)$ и семейство условных обобщенных вероятностных плотностей (о.в.п.) $f(\cdot | \omega)$, $\omega \in \Omega$, величины X , наблюдаемой до принятия решения.

Пусть S выборочное пространство возможных значений наблюдения X . Для принятия решения требуется знать решающую функцию δ , заданную для любого возможного значения $x \in S$ решения $d(x) \in D$.

Класс всех решающих функций δ обозначим Δ .

Для любой о.в.п. ξ параметра w и любой решающей функции $\delta \in \Delta$ риск $\rho(\xi, \delta)$ определяется равенством

$$\rho(\xi, \delta) = E\{L[\omega, \delta(x)]\} = \int \int_{\Omega S} L[\omega, \delta(x)] f(x|\omega) \xi(\omega) d\mu(x) d\nu(\omega). \quad (1.1)$$

Предполагается, что при всех $\omega \in \Omega$ функция $L[\omega, \delta(\cdot)]$ измерима и интегрируема на множестве S . Символы $d\mu(x)$ и $d\nu(x)$ указывают на то, что каждый из интегралов может быть как обычным интегралом от о.в.п., так и суммой значений дискретной функции вероятностей (ф.в.).

Риск здесь определяет средний ущерб (потери).

Для каждого решения $d \in D$ риск $\rho(\xi, \delta)$ при о.в.п. $\xi(\omega)$:

$$\rho(\xi, d) = \int_{\Omega} L(\omega, d) \xi(\omega) d\nu(\omega).$$

Для каждого значения $\omega \in \Omega$ риск $\rho(\omega, \delta)$, соответствующий решающей функции δ при $w = \omega$

$$\rho(\omega, \delta) = \int_S L[\omega, \delta(x)] f(x|\omega) d\mu(x). \quad (1.2)$$

Для каждой решающей функции $\delta \in \Delta$ функция $\rho(\cdot, \delta)$, определяемая формулой (1.2), называется ее функцией риска. Из (1.1) и (1.2)

$$\rho(\xi, \delta) = \int_{\Omega} \rho(\omega, \delta) \xi(\omega) d\nu(\omega).$$

Пусть $\delta^* \in \Delta$ - такая решающая функция, что

$$\rho(\xi, \delta^*) = \inf_{\delta \in \Delta} \rho(\xi, \delta) = \rho^*(\xi); ;$$

тогда δ^* называется байесовской решающей функцией при ξ , а $\rho^*(\xi)$ называется байесовским риском.

При заданной о.в.п. $\xi(\omega)$ параметра w надо найти решающую функцию δ , минимизирующую риск $\rho(\xi, \delta)$ /см. (1.1)/. Если L неотрицательна или ограничена, в (1.1) можно изменить порядок интегрирования:

$$\rho(\xi, \delta) = \int_S \left\{ \int_{\Omega} L[\omega, \delta(x)] f(x|\omega) \xi(\omega) d\nu(\omega) \right\} d\mu(x).$$

Поэтому решающую функцию δ , минимизирующую риск, при каждом значении $x \in S$ можно определить из условия минимизации внутреннего интеграла, т.е. байесовская решающая функция $\delta^*(x) = d^*$, где d^* - решение из D , минимизирующее интеграл

$$\int_{\Omega} L[\omega, d] f(x|\omega) \xi(\omega) d\nu(\omega). \quad (1.3)$$

Вместо того, чтобы искать минимум d^* для интеграла (1.3), можно найти то же самое значение d^* из условия минимума интеграла

$$\int_{\Omega} L(\omega, d) \left[\frac{f(x|\omega)\xi(\omega)}{f_1(x)} \right] d\nu(\omega),$$

$$\text{где } f_1(x) = \int_{\Omega} f(x|\omega)\xi(\omega) d\nu(\omega).$$

Поскольку дробь в квадратных скобках является о.в.п. случайной величины ω при $X=x$, то значение интеграла равно условному математическому ожиданию $E[L(\omega, d)/x]$; d^* - байесовское решение при условном распределении w , когда $X=x$. Маргинальное распределение w называется априорным распределением, оно задает распределение w до проведения наблюдений над X . Условное распределение w при известном значении x называется апостериорным распределением w ; оно задает распределение w после наблюдения $X=x$.

Пример. Пространство Ω содержит только точки 0 и 1. Пространство решений D состоит из чисел d интервала $0 \leq d \leq 1$. Функция потерь определена для $\omega \in \Omega$ и $d \in D$ формулой $L(\omega, d) = |\omega - d|$.

Вероятностное распределение P параметра w :

$$Pr(\omega=0)=3/4; Pr(\omega=1)=1/4.$$

Тогда для $\forall d \in D$ риск

$$\begin{aligned} \rho(P, d) &= L(\omega=0, d) \cdot Pr(\omega=0) + L(\omega=1, d) \cdot Pr(\omega=1) = \\ &= \frac{3}{4}d + \frac{1}{4}(1-d) = \frac{1}{2}d + \frac{1}{4} \end{aligned}$$

Отсюда

$$\inf_d \rho(P, d) = \rho(P, d=0) = \frac{1}{4},$$

т.е. $\rho^*(P) = \frac{1}{4}$ - единственное байесовское решение.

Если предположить, что пространство решений полуоткрытый интервал $0 < d \leq 1$, то байесовский риск будет по-прежнему равен $1/4$, но ни одно решение из D не будет байесовским.

Если решение принимается без предварительных наблюдений, то оптимально байесовское решение при априорном распределении w . Если же было предварительное наблюдение, то априорное распределение w заменяется на апостериорное. Отсюда видно, что решение $d^*(x_w)$, задаваемое байесовской решающей функцией $\delta^*(x)$ для наблюдаемого значения $X=x_w$, можно найти без вычисления $\delta^*(x)$, и $\delta^*(x)$ при о.в.п. $\xi(\omega)$ можно определить без расчета байесовского риска $\rho^*(\xi)$

Пусть $c(\omega, x)$ обозначает цену наблюдения значения x величины X , если $w=\omega$. Тогда, если ξ есть о.в.п. случайной величины w , то средняя цена наблюдения

$$E[c(\omega, x)] = \int_{\Omega} \int_S c(\omega, x) f(x|\omega)\xi(\omega) d\mu(x) d\nu(\omega).$$

И эта цена может быть такой, что выигрыш от проведенного наблюдения не окупит стоимости измерения.

Общим риском от наблюдения x и принятия решающей функции δ называется сумма риска $\rho(\xi, \delta)$ и средней цены наблюдения $E[c(\omega, x)]$. Выбирается наблюдение x и соответствующая байесовская решающая функция δ , минимизирующая общий риск.

Таблица 1.1 *Пример.* Рассмотрим $\Omega = \{\omega_1, \omega_2\}$;

Ω	D	
	d_1	d_2
ω_1	0	5
ω_2	10	0

$D = \{d_1, d_2\}$, функция потерь задается табл. 1.1. Случайная величина X принимает значения 0 и 1 со следующими условными вероятностями:

$$Pr(x=1|w=\omega_1)=3/4; Pr(x=0|w=\omega_1)=1/4;$$

$$Pr(x=0|w=\omega_2)=2/3; Pr(x=1|w=\omega_2)=1/3;$$

$$(\text{частный случай ф.в. } f(x| \omega_1) = \frac{3x}{4}; f(x| \omega_2) = \frac{2-x}{3}).$$

Априорное распределение параметра w таково:

$$Pr(w=\omega_1)=p; Pr(w=\omega_2)=1-p; 0 \leq p \leq 1,$$

p - задано. Построить байесовскую решающую функцию.

Пусть $\xi(x)$ апостериорная вероятность события $w=\omega_1$, если наблюдаемое значение $X=x$;

$$\xi(x) = Pr(w=\omega_1|X=x).$$

Пусть $x=1$, тогда

$$\xi(1) = \frac{3/4 p}{3/4 p + 1/3(1-p)}$$

Для $x=0$

$$\xi(0) = \frac{1/4 p}{1/4 p + 2/3(1-p)}$$

После наблюдения $x=x_1=1$ риск от принятия решения d_1 равен $0 \cdot p + 10[1-\xi(x)]$, а от принятия решения d_2 равен $5\xi(x)$. Чтобы d_2 было байесовским решением, должно быть $10[1-\xi(x)] > 5\xi(x)$; $\xi(x) < 2/3$; d_1 будет байесовским решением, если $\xi(x) > 2/3$. При $\xi(x) = 2/3$ и d_1 , и d_2 байесовские решения.

Если наблюдается значение $x=1$, то для байесовской решающей функции δ^* имеем: $\delta^*(1)=d_2$, если $\xi(1) < 2/3$ или $p < 8/17$; $\delta^*(1)=d_1$, если $\xi(1) > 2/3$ или $p > 8/17$; при $p=8/17$ оба решения d_1 и d_2 являются байесовскими.

Если наблюдается значение $x=0$, то: $\delta^*(0)=d_2$ при $\xi(0) < 2/3$ или $p < 16/19$; $\delta^*(0)=d_1$ при $p > 16/19$; при $p=16/19$ оба решения d_1 и d_2 байесовские.

Вычислим значения байесовского риска $\rho^*(p)$ для произвольной априорной вероятности p .

Если $0 \leq p \leq 8/17$, то решение d_2 будет байесовским независимо от наблюдаемых значений x ; для таких p байесовский риск равен $\rho^*(p) = 5p$. При $8/17 < p < 16/19$ $\delta^*(0) = d_2$, $\delta^*(1) = d_1$. Поэтому $\rho^*(p) = p \cdot \rho(\omega_1, \delta^*) + (1-p) \cdot \rho(\omega_2, \delta^*) = p(0 \cdot 3/4 + 5 \cdot 1/4) + (1-p)(10 \cdot 1/3 + 0 \cdot 2/3) = 5/4p + 10/3(1-p) = 10/3 - 25/12p$.

Если $16/19 \leq p \leq 1$, то решение d_1 будет байесовским для любого значения x ; $\rho^*(p) = 10(1-p)$.

Таким образом мы провели экстенсивный вид анализа построение байесовской решающей функции (рис. 1.1).

В данном примере не учитывалась цена наблюдения.

Если в рассмотренном примере потребуется принять решение до проведения измерения x , то минимальный риск $\rho_0(p)$ (по априорной информации) будет таким (рис. 1.2)

$$\rho_0(p) = \begin{cases} 5p & \text{при } 0 \leq p \leq 2/3; \\ 10(1-p) & \text{при } 2/3 < p \leq 1 \end{cases}$$

На рис. 1.2 наложена байесовская функция рис. 1.1 $\rho^*(p)$. Если $p \leq 8/17$ или $p \geq 16/19$, то тот же риск может быть достигнут без наблюдения x . Если же $8/17 < p < 16/19$, то $\rho^*(p) < \rho_0(p)$. За возможность наблюдения x может быть уплачена цена $c < \rho_0(p) - \rho^*(p)$. Эта разность максимальна при $p = 2/3$ и равна $25/18$.

Если в задаче наблюдается вектор (несколько случайных величин), то случайные величины могут наблюдаться одновременно или в несколько этапов; тогда апостериорное распределение можно вычислять на каждом этапе, беря в качестве априорного распределения апостериорное распределение, полученное на предыдущем этапе. В конечном итоге получим тоже апостериорное распределение, что и при одновременном учете всех случайных величин.

Например, пусть $f(x, y | \omega)$ - совместная (многомерная) условная о.в.п. случайных величин X и Y при $\omega = \omega$, $\omega \in \Omega$ Апостериорная о.в.п. $\xi(\cdot | x, y)$ параметра ω при $X=x$, $Y=y$ в точке ω :

$$\xi(\omega | x, y) = \frac{f(x, y | \omega)p(\omega)}{\int_{\Omega} f(x, y | \omega)p(\omega)d\nu(\omega)} \quad (1.4)$$

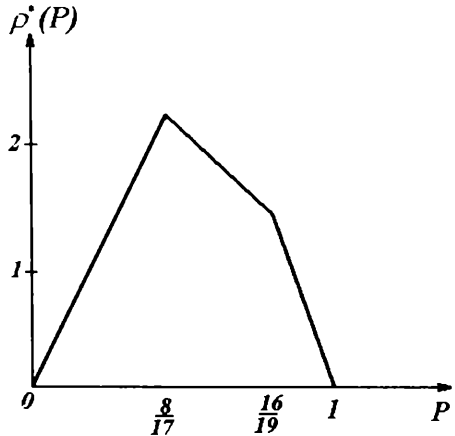


Рис. 1.1 Байесовская решающая функция

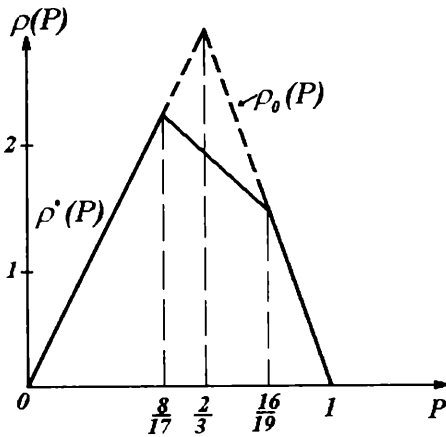


Рис. 1.2. Зависимость минимального риска от p

Апостериорная о.в.п. $\xi(\cdot|x, y)$ параметра w при $Y=y$:

$$\xi(\omega|x, y) = \frac{h(y|\omega, x)\xi(\omega|x)}{\int_{\Omega} h(y|\omega, x)\xi(\omega|x)d\nu(\omega)} \quad (1.7)$$

Выражения (1.5) и (1.6) подставим в (1.7), получим (1.4).

1.3. Статистическая классификация при фиксированном объеме выборки

Пусть результаты измерений $x_i, i=1, \dots, N$ - случайные величины.

Считаем, что для каждого класса образов $\omega_j, j=1, \dots, m$ известны многомерная (N -мерная) функция плотности вероятности (или распределения) вектора признаков $X, P(X|\omega_j)$, и вероятность $P(\omega_j)$ появления $\omega_j, j=1, \dots, m$.

Классификация проводится путем минимизации вероятности ошибочного распознавания с помощью определения решающей функции $\delta(x)$, где $\delta(x) = \delta_j$ означает, что принимается гипотеза $H_j: x \sim \omega_j$.

Пусть принятие решения d_j , когда в действительности входной образ принадлежит ω_i приводит к потере $L(\omega_i, d_j)$. Величина условных потерь (условный риск) для $x \sim \omega_i$ составит

$$r(\omega_i, d) = \int_{\Omega_x} L(\omega_i, d)P(x|\omega_i)dx.$$

Для данного множества априорных вероятностей $P = \{P(\omega_j)\}$ средние потери (средний риск)

$$R(P, d) = \sum_{j=1}^m P(\omega_j)r(\omega_j, d)$$

Пусть наблюдаем случайную величину X , затем Y ; функция $g(\cdot|\omega)$ означает условную о.в.п. X при $w=\omega$. После наблюдения $X=x$ апостериорная о.в.п. $\xi(\cdot|x)$ для w :

$$\xi(\omega|x) = \frac{g(x|\omega)p(\omega)}{\int_{\Omega} g(x|\omega)p(\omega)d\nu(\omega)} \quad (1.5)$$

Условная о.в.п. $h(\cdot|\omega, x)$,

для Y при $w=\omega, X=x$:

$$h(y|\omega, x) = \frac{f(x, y|\omega)}{g(x|\omega)} \quad (1.6)$$

или

$$R(P, d) = \int_{\Omega_x} P(x) r_x(P, d) dx, \quad (1.8)$$

$$\text{где } r_x(P, d) = \frac{\sum_{j=1}^m L(\omega_j, d) P(\omega_j) P(x|\omega_j)}{P(x)};$$

$r_x(P, d)$ - апостериорный условный средний риск решения d при данных замерах признаков X . Надо найти такое решение $d_j, j=1, \dots, m$, которое минимизирует средний риск $R(P, d)$ или минимизирует максимум условного риска $r(\omega_j, d)$ (критерий минимакса).

Оптимальное решающее правило минимизации среднего риска называется байесовским правилом. Из (1.8) следует, что достаточно рассмотреть каждый X в отдельности и минимизировать $r_x(P, d)$. Если d^* является оптимальным решением в смысле минимума среднего риска, то

$$r_x(P, d^*) \leq r_x(P, d),$$

т.е.

$$\sum_{j=1}^m L(\omega_j, d^*) P(\omega_j) P(x|\omega_j) \leq \sum_{j=1}^m L(\omega_j, d) P(\omega_j) P(x|\omega_j).$$

Для функции потерь

$$L(\omega_j, d_i) = 1 - \delta_{ji} = \begin{cases} 0; & i = j, \\ 1; & i \neq j \end{cases}$$

средний риск является также вероятностью ложного распознавания и байесово решающее правило дает $X \sim \omega_i (d^* = d_i)$, если

$$P(\omega_i) P(X|\omega_i) \geq P(\omega_j) P(X|\omega_j) \quad (1.9)$$

для всех $j=1, \dots, m$.

Определим отношение правдоподобия между классами следующим образом:

$$\lambda = \frac{P(X|\omega_i)}{P(X|\omega_j)}.$$

Тогда (1.9) примет вид $d^* = d_i$, если $\lambda \geq \frac{P(\omega_j)}{P(\omega_i)}$ для всех $j=1, \dots, m$.

Если информация об априорных вероятностях $P(\omega_j)$ отсутствует, то классификация строится на основе минимаксного критерия по отношению к наименее благоприятному априорному распределению. Из (1.9) получаем разделяющую функцию

$$D_i(X) = P(\omega_i) P(X|\omega_i), \quad i=1, \dots, m$$

или эквивалентную ей

$$D_i(X) = \log[P(\omega_i) P(X|\omega_i)], \quad i=1, \dots, m.$$

Решающая граница между областями в Ω_x относящимися к ω_i и ω_j , определяется условием

$$P(\omega_i)P(X|\omega_i) - P(\omega_j)P(X|\omega_j) = 0$$

или

$$\log \frac{P(\omega_i)P(X|\omega_i)}{P(\omega_j)P(X|\omega_j)} = 0. \quad (1.10)$$

Пример. Статическая задача решения с Ω и D из двух точек

Ω	D	
	d_1	d_2
ω_1	0	a_1
ω_2	a_2	0

Пусть $\Omega = \{\omega_1, \omega_2\}$; $D = \{d_1, d_2\}$; функция потерь \bar{L} задана табл. 1.2; $\alpha(\delta)$ - условная вероятность для любой решающей функции δ принятия решения d_2 при $w = \omega_1$; $\beta(\delta)$ - условная вероятность

принятия решения d_1 , $a_1, a_2 > 0$, когда $w = \omega_2$.

Другими словами, $\alpha(\delta)$ и $\beta(\delta)$ - вероятности того, что δ предписывает неправильные решения в случаях $w = \omega_1$ и $w = \omega_2$ соответственно. Пусть априорное распределение w задано $Pr(w = \omega_j) = p$, где $0 < p < 1$. С учетом функции потерь L риск $\rho(p, \delta)$ решающей функции δ равен

$$\rho(p, \delta) = a_1 \cdot p \cdot \alpha(\delta) + a_2 (1-p) \cdot \beta(\delta).$$

В каждой конкретной задаче надо минимизировать эту комбинацию. Согласно лемме Неймана-Пирсона, решающая функция, доставляющая минимум линейной комбинации $a_1 p \cdot \alpha(\delta) + a_2 (1-p) \cdot \beta(\delta)$, определяется отношением $f_2(x)/f_1(x)$; $\delta^*(x) = d_1$, если $a f_1(x) > b f_2(x)$; $\delta^*(x) = d_2$, если $a f_1(x) < b f_2(x)$. Здесь $a = a_1 p$; $b = a_2 p$.

Если $P(X|\omega)$ есть многомерное гауссово распределение со средним вектором M_i и ковариационной матрицей K_i , $i = 1, \dots, m$:

$$P(X|\omega_i) = \frac{\exp\left[-\frac{1}{2}(X - M_i)^T K_i^{-1}(X - M_i)\right]}{(2\pi)^{N/2} |K_i|^{1/2}}$$

то решающей границей, согласно (1.10), будет

$$\log \frac{P(\omega_i)}{P(\omega_j)} - \frac{1}{2} \log \frac{|K_i|}{|K_j|} - \frac{1}{2} \left[(X - M_i)^T K_i^{-1} (X - M_i) - (X - M_j)^T K_j^{-1} (X - M_j) \right] = 0.$$

При $K_i = K_j = K$ имеем

$$X^T K^{-1} (M_i - M_j) - \frac{1}{2} (M_i + M_j)^T K^{-1} (M_i - M_j) + \log \frac{P(\omega_i)}{P(\omega_j)} = 0.$$

Получили уравнение гиперплоскости.

При $P(\omega_i) = P(\omega_j)$,

имеем

$$X^T K^{-1} (M_i - M_j) - \frac{1}{2} (M_i + M_j)^T K^{-1} (M_i - M_j) = 0.$$

В действительности должны получить "полосу", так как X , M , K определяются с погрешностями.

Рассмотрим решающую процедуру с фиксированным объемом выборки.

Пример 1. Найти решающую границу для

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2} \frac{(x - M_i)^2}{\sigma_i^2}\right\}; \quad i = 1, 2.$$

В общем случае (при $\sigma_1 \neq \sigma_2$), согласно (1.10), решающая граница определяется из условия

$$\log \frac{P(\omega_1)}{P(\omega_2)} + \log \frac{P(x|\omega_1)}{P(x|\omega_2)} = 0,$$

$$\log \frac{P(\omega_1)}{P(\omega_2)} + \log \frac{\sigma_2}{\sigma_1} - \frac{(x - M_1)^2}{2\sigma_1^2} + \frac{(x - M_2)^2}{2\sigma_2^2} = 0.$$

При $\sigma_1 = \sigma_2 = \sigma$ получим

$$\log \frac{P(\omega_1)}{P(\omega_2)} - \frac{(M_1 - M_2)}{2\sigma^2} (M_1 - M_2 - 2x) = 0,$$

откуда решающая граница

$$x = \frac{\sigma^2 \log \frac{P(\omega_1)}{P(\omega_2)}}{M_2 - M_1} + \frac{M_1 + M_2}{2}$$

Пример 2. Найти решающую границу для двумерных гауссовых распределений $P(X|\omega_i)$, ($i=1, 2$) со средними векторами (m_{ix}, m_{iy}) , ($i=1, 2$) и общей ковариационной матрицей

$$K = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \text{ при } P(\omega_1) = P(\omega_2).$$

Разделяющая граница будет в данном случае прямой линией, уравнение которой имеет вид

$$(x, y) \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix} \begin{pmatrix} m_{1x} - m_{2x} \\ m_{1y} - m_{2y} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} m_{1x} + m_{2x} \\ m_{1y} + m_{2y} \end{pmatrix}^T \times \begin{pmatrix} \sigma_1^{-2} & 0 \\ 0 & \sigma_2^{-2} \end{pmatrix} \begin{pmatrix} m_{1x} - m_{2x} \\ m_{1y} - m_{2y} \end{pmatrix} = 0$$

или

$$x \frac{m_{1x} - m_{2x}}{\sigma_1^2} + y \frac{m_{1y} - m_{2y}}{\sigma_2^2} - \frac{1}{2} \left[\frac{m_{1x}^2 - m_{2x}^2}{\sigma_1^2} + \frac{m_{1y}^2 - m_{2y}^2}{\sigma_2^2} \right] = 0.$$

1.4. Методы детерминистской классификаций

Концепцию классификации образов можно выразить на языке разбиения пространства признаков или отображения пространства признаков в пространство решений. Пусть у каждого входного образа измеряется N признаков, каждое множество из N признаков называют вектором признаков X или точкой в N -мерном пространстве признаков Ω_x .

Задача классификации заключается в распределении всех возможных векторов (или точек) в пространстве признаков по соответствующим классам образов: пространство признаков рассекается на взаимно непересекающиеся области, каждая из которых соответствует некоторому классу образов. Математическая задача классификации может быть сформулирована с помощью разделяющей функции.

Пусть имеется m возможных классов образов ω_j , $j=1, \dots, m$, $X=(x_1, \dots, x_N)^T$ вектор замеров признаков. Тогда разделяющая функция $D_j(x)$, относящаяся к классу образов ω_j , $j=1, \dots, m$, такова, что если входной образ, представленный вектором признаков X , принадлежит классу ω_i , то величина $D_i(x)$ должна быть наибольшей ($x \sim \omega_i$). Для всех ($x \sim \omega_i$)

$$D_i(x) > D_j(x); \quad i, j=1, \dots, m, \quad i \neq j. \quad (1.11)$$

В пространстве признаков Ω_x граница разбиений, называемая решающей границей, между областями, относящимися соответственно к классу ω_i и ω_j , выражается уравнением $D_i(x) - D_j(x) = 0$. Можно предложить много различных форм для $D_i(x)$, удовлетворяющих (1.11).

Линейные разделяющие функции. В качестве $D_i(x)$ берется линейная комбинация измеренных признаков x_1, \dots, x_N , т.е.

$$D_i(x) = \sum_{k=1}^N w_{ki} x_k + w_{i, N+1}; \quad i = 1, \dots, m.$$

Решающая граница между областями ω_i и ω_j в Ω_x имеет вид

$$D_i(x) - D_j(x) = \sum_{k=1}^N w_k x_k + w_{i, N+1} - w_{j, N+1} = 0; \quad (1.12)$$

где $w_k = w_{ik} - w_{jk}$, $w_{i, N+1} = w_{i, N+1} - w_{j, N+1}$.

Уравнение (1.12) - уравнение гиперплоскости.

При $m=2$ получаем уравнение прямой линии.

Полиномиальные разделяющие функции. Полиномиальная разделяющая функция g -й степени имеет вид

$$D_i(x) = w_{i1} f_1(x) + w_{i2} f_2(x) + \dots + w_{il} f_l(x) + w_i;$$

где $f_j(x)$ является формой

$$x_{k_1}^{n_1} \cdot x_{k_2}^{n_2} \cdot \dots \cdot x_{k_r}^{n_r} \text{ при } \begin{cases} k_1, k_2, \dots, k_r = 1, \dots, N, \\ n_1, n_2, \dots, n_r = 0 \text{ или } 1. \end{cases}$$

Если $r = 2$, разделяющая функция называется квадратичной. Здесь

$$f_j(x) = x_{k_1}^{n_1} x_{k_2}^{n_2} \text{ при } \begin{cases} k_1, k_2 = 1, \dots, N, \\ n_1, n_2 = 0 \text{ или } 1. \end{cases}$$

$$D_i(x) = \sum_{k=1}^N w_{kk} x_k^2 + \sum_{j=1}^{N-1} \sum_{k=j+1}^N w_{jk} x_j x_k + \sum_{j=1}^N w_j x_j + w_{l+1}.$$

где $l = \frac{1}{2} N(N + 3)$.

В общем случае границей для квадратичных разделяющих функций является поверхность второго порядка. В частных случаях это будет гиперболоид, гиперэллипсоид и гиперэллипсоидальный цилиндр.

Классификатор по минимальному расстоянию. Пусть задано m опорных векторов $R_i, i=1, \dots, m$, где $R_i \sim \omega_i$. Входной сигнал X предполагается принадлежащим ω_i , если расстояние между X и R_i , равно $\|X - R_i\|$, минимально. Это расстояние можно определить, например, как

$$\|X - R_i\| = \left[(X - R_i)^T (X - R_i) \right]^{1/2}$$

или

$$\|X - R_i\|^2 = X^T X - X^T R_i - X R_i^T + R_i^T R_i,$$

где T - транспонирование.

Поскольку $X^T X$ не зависит от i , то разделяющая функция для классификатора по минимальному расстоянию имеет вид

$$D(X) = X^T R_i + X R_i^T - R_i^T R_i, i = 1, \dots, m.$$

Это линейная функция (линейный классификатор).

Обучение в линейном классификаторе. Введем дополнительный вектор $Y = [x_1, \dots, x_N, 1]^T = [x, 1]^T$ Пусть в результате наблюдений получены два множества дополнительных векторов, принадлежащих различным ω_1 и ω_2 . Для разделимых множеств

$$Y^T w > 0; Y \approx \omega_1$$

$$Y^T w < 0; Y \approx \omega_2; w = [w_1, \dots, w_N, w_{N+1}]^T$$

Если классификатор дает ошибочный или неопределенный результат при выбранных w , то берут новый вектор весов $w' = w \pm \alpha Y, \alpha > 0$ (поправочное дополнение). Вначале w выбирается произвольно. В результате итерационного процесса получаем значение w' которое считается истинным.

1.5. Последовательная решающая модель для классификации образов

Ранее была рассмотрена решающая процедура с фиксированным объемом выборки. Если стоимость измерений признаков высока или требуется сложное оборудование, большое время или сложные и рискованные операции, уместно применить последовательную решающую процедуру. Последовательный процесс измерений признаков заканчивается (принимается решение), когда достигнута достаточная или необходимая точность классификации.

Естественно, признаки должны быть расположены в таком порядке, чтобы измерения давали окончательное решение возможно раньше. Задача упорядочения признаков является специальной задачей в системах последовательного распознавания.

В случае двух классов образов можно применить последовательный критерий отношения вероятностей Вальда (п.к.о.в.). Он построен для решения о выборе между двумя простыми гипотезами.

Пусть случайная величина x обладает функцией плотности $P(x, \theta)$, где θ - испытуемый параметр. Задача заключается в проверке гипотезы H_0 о том, что $\theta = \theta_1$, против гипотезы $H_1: \theta = \theta_2$. Критерий решает дело в пользу θ_1 или θ_2 на основе наблюдений x_1, x_2, \dots, x_n . Допустим, что если гипотеза H_0 истинна, мы хотим получить решение в пользу H_0 с вероятностью, не меньшей $(1 - \alpha)$, а если H_1 истинна, то решение в пользу H_1 должно иметь вероятность, не меньшую $(1 - \beta)$.

Для последовательного критерия с постоянным объемом выборки оптимальное решение этой задачи дается критерием Неймана - Пирсона: при данном числе наблюдений n критерий, обеспечивающий наименьшее β (наиболее мощный критерий), определяется отношением правдоподобия λ_n имеющим вид:

$$\lambda_n = \prod_{i=1}^n \frac{P(x_i | H_0)}{P(x_i | H_1)} = \frac{P_n(x | H_0)}{P_n(x | H_1)}. \quad (1.13)$$

Гипотеза H_0 принимается или отвергается, когда λ_n соответственно меньше или больше некоторой постоянной. Значение этой постоянной может быть выбрано так, чтобы критерий давал нужную величину α (ошибку 1-го рода). В принципе можно выбрать и так, чтобы критерий имел заданную мощность $(1 - \beta)$.

Последовательный критерий отношения вероятностей аналогичен этому и обладает аналогичными оптимальными свойствами: наблюдения производятся до тех пор, пока выполняется условие

$$B < \lambda_n < A.$$

Наблюдения прекращаются и принимается решение в пользу гипотезы H_0 , как только будет выполнено неравенство $\lambda_n \geq A$; решение принимается в пользу гипотезы H_1 , когда $\lambda_n \leq B$.

Постоянные A и B называются соответственно верхними и нижними порогами (останавливающими границами). Они могут быть выбраны так, чтобы приблизительно получить заданные вероятности ошибок α и β .

Пусть на n -м шаге оказалось, что

$$\lambda_n = A. \tag{1.14}$$

Это указывает на окончательное решение о принятии H_0 . Из (1.13) и (1.14)

$$P(x|H_0) = AP(x|H_1),$$

что эквивалентно равенству

$$\int P(x|H_0) dx = A \int P(x|H_1) dx. \tag{1.15}$$

Интегралы берутся по области, которая содержит все замеры, приводящие к решению H_0 . Условие (1.15) дает $1 - \alpha = A\beta$.

Пусть теперь $\lambda_n = B$, тогда $\alpha = B(1 - \beta)$. Отсюда

$$A = \frac{1 - \alpha}{\beta}; \quad B = \frac{\alpha}{1 - \beta}.$$

Эти соотношения строго выполняются для непрерывных замеров; для дискретных замеров могут быть получены вероятности ошибок, отличающиеся от α и β вследствие некоторого превышения порогов. Это отличие несущественно.

Тем не менее в общем случае

$$A \leq \frac{1 - \alpha}{\beta}; \quad B \geq \frac{\alpha}{1 - \beta}.$$

Рассмотрим пример, как изменяются решающие границы с числом замеров признаков n .

Пример. Пусть x_1, x_2, \dots, x_n независимые замеры признаков с одномерной гауссовой функцией плотности $P(x_j|\omega_i); j=1 \dots n; i=1, 2$, со средним значением m_i и дисперсией σ^2 . Для простоты будем вычислять $\log \lambda_n$. После получения x_1

$$\begin{aligned} \log \lambda_1 &= \log \frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} = \log \frac{(\sigma\sqrt{2\pi})^{-1} \exp\left\{-\frac{(x_1 - m_1)^2}{2\sigma^2}\right\}}{(\sigma\sqrt{2\pi})^{-1} \exp\left\{-\frac{(x_1 - m_2)^2}{2\sigma^2}\right\}} = \\ &= \frac{1}{\sigma^2} \left[(m_1 - m_2)x_1 - \frac{1}{2}(m_1^2 - m_2^2) \right]. \end{aligned}$$

Сравниваем $\log \lambda_1$ с $\log A$ и $\log B$

Если

$$x_1 > \frac{\sigma^2}{m_1 - m_2} \log A + \frac{1}{2}(m_1 + m_2),$$

то $X \sim \omega_1$.

Если

$$x_1 \leq \frac{\sigma^2}{m_1 - m_2} \log B + \frac{1}{2}(m_1 + m_2),$$

то $X \sim \omega_2$.

Если

$$\frac{\sigma^2}{m_1 - m_2} \log B + \frac{1}{2}(m_1 + m_2) < x_1 < \frac{\sigma^2}{m_1 - m_2} \log A + \frac{1}{2}(m_1 + m_2),$$

то следует продолжить измерения.

После второго измерения

$$\log \lambda_1 = \log \frac{P(x_1|\omega_1)}{P(x_1|\omega_2)} + \log \frac{P(x_2|\omega_1)}{P(x_2|\omega_2)} = \frac{m_1 - m_2}{\sigma^2} [x_1 + x_2 - (m_1 + m_2)].$$

Если

$$x_1 + x_2 \geq \frac{\sigma^2}{m_1 - m_2} \log A + m_1 + m_2,$$

то $X \sim \omega_1$.

Если

$$x_1 + x_2 \leq \frac{\sigma^2}{m_1 - m_2} \log B + m_1 + m_2,$$

то $X \sim \omega_2$.

Если

$$\frac{\sigma^2}{m_1 - m_2} \log B + m_1 + m_2 < x_1 + x_2 < \frac{\sigma^2}{m_1 - m_2} \log A + m_1 + m_2,$$

то берется x_3 .

На n -м шаге процесса

$$\log \lambda_n = \sum_{i=1}^n \log \frac{P(x_i|\omega_1)}{P(x_i|\omega_2)} = \frac{m_1 - m_2}{\sigma^2} \sum_{i=1}^n \left[x_i - \frac{1}{2}(m_1 + m_2) \right].$$

Процедура классификации будет следующей:

если

$$\sum_{i=1}^n x_i \geq \frac{\sigma^2}{m_1 - m_2} \log A + \frac{n}{2}(m_1 + m_2), \quad (a)$$

то $X \sim \omega_1$;

если

$$\sum_{i=1}^n x_i \leq \frac{\sigma^2}{m_1 - m_2} \log B + \frac{n}{2}(m_1 + m_2), \quad (б)$$

то $X \sim \omega_2$;

и если

$$\frac{\sigma^2}{m_1 - m_2} \log B + \frac{n}{2}(m_1 + m_2) < \sum_{i=1}^n x_i < \frac{\sigma^2}{m_1 - m_2} \log A + \frac{n}{2}(m_1 + m_2),$$

то берется x_{n+1} .

Решающие границы, определяемые формулами (а), (б) при знаках равенства, представляют собой две параллельные гиперплоскости в пространстве признаков. Ширина области неопределенности равна

$$\frac{\sigma^2}{m_1 - m_2} \log \frac{A}{B}, \text{ т.е. пропорциональна } \frac{\sigma^2}{m_1 - m_2}. \text{ При заданных}$$

вероятностях ошибок α и β среднее число измерений, необходимое для завершения процесса, пропорционально σ^2 и обратно пропорционально $(m_1 - m_2)$.

П.к.о.в. дает минимум среднего числа наблюдений

$$E_1(n) = \frac{(1 - \alpha) \log A + \beta \log B}{E_1(z)},$$

когда H_0 истинно,

$$E_2(n) = \frac{\beta \log A + (1 - \beta) \log B}{E_2(z)},$$

когда H_1 истинно, где

$$z = \log \frac{P_n(x|H_0)}{P_n(x|H_1)}.$$

Процедура п.к.о.в. по существу не зависит от априорных вероятностей $P(H_i)$; $i = 0, 1$, хотя вероятность ошибок зависит от априорных данных.

В ряде случаев необходимо прервать процедуру наблюдений и принять решение при $n=N$. Это может быть сделано усечением последовательного процесса при $n=N$, выполняется обычная процедура п.к.о.в. либо до получения решения, либо до N -го шага. Если на N -м шаге решение не получено, принимается гипотеза H_0 при $\lambda_N > 1$ или принимается гипотеза H_1 при $\lambda_N < 1$.

Для большого числа гипотез предложен обобщенный последовательный критерий отношения вероятностей (о.п.к.о.в.). Пусть имеется m гипотез. На n -м шаге обобщенные последовательные отношения вероятностей для каждой гипотезы определяются следующим образом:

$$U_n(x|H_i) = \frac{P_n(x|H_i)}{\left[\prod_{q=1}^m P_n(x|H_q) \right]^{1/m}}; \quad i = 1, \dots, m.$$

Решающее правило о.п.к.о.в. состоит в следующем. Отношение

$U_n(x|H_i)$ сравнивается с останавливающей границей $A(H_i)$ для гипотезы H_i .
 Гипотеза H_i исключается из рассмотрения, если

$$U_n(x|H_i) < A(H_i); \quad i=1, \dots, m.$$

Останавливающая граница определяется соотношением

$$A(H_i) = \frac{1 - e_{ii}}{\left[\prod_{q=1}^m (1 - e_{iq}) \right]^{1/m}}; \quad i=1, \dots, m,$$

где e_{iq} - вероятность принятия H_i когда в действительности истинна H_q .

После исключения гипотезы H_i общее число гипотез становится на единицу меньше, и составляется новое множество обобщенных последовательных отношений вероятностей. Последовательно исключаются гипотезы, последняя принимается за истинную. При $m=2$ о.п.к.о.в. эквивалентен п.к.о.в. и сохраняет его свойство оптимальности: при заданных α и β не существует другой процедуры, которая обладает меньшими значениями вероятностей ошибок или среднего риска и дает выигрыш в среднем числе измерений признаков по сравнению с последовательной процедурой классификации.

Между A и B заключена область неопределенности, когда не может быть принято окончательного решения (нулевая область).

В классификации на основе о.п.к.о.в. при $n=N$ входной образ полагают принадлежащим к тому классу, обобщенное последовательное отношение вероятностей для которого имеет наибольшее значение.

В последовательном анализе при повышении верхнего порога A и понижении нижнего порога B по меньшей мере одна из вероятностей ошибок α и β уменьшается, если новый п.к.о.в. не оказывается эквивалентен старому.

Поэтому можно начинать с относительно больших значений порогов, и, постепенно уменьшая их, мы получим среднее значение испытаний не таким большим, как в случае, когда в течение всего процесса используется малая величина порога.

1.6. Байесова последовательная решающая процедура

При фиксированном объеме выборки и испытании m статистических гипотез оптимальное решение d^* выбирается так, чтобы минимизировать величину средних потерь

$$R(P, d) = \sum_{i=1}^m P(H_i) r(H_i, d),$$

т.е. $R(P, d^*) = \min R(P, d_i);$

$$r(H_i, d) = \int_{\Omega_x} L(H_i, d)P(x|H_i) - \text{условные потери (условный риск)}.$$

При

$$L(H_i, d_j) = 1 - \delta_{ij} = \begin{cases} 0; & i = j, \\ 1; & i \neq j \end{cases}$$

байесово решение $d^* = d_i$, если

$$P(H_i)P(x|H_i) \geq P(H_j)P(x|H_j), \quad \text{для } \forall j = 1, \dots, m.$$

При

$$\lambda = \frac{P(x|H_i)}{P(x|H_j)} \quad d^* = d_i$$

если

$$\lambda \geq \frac{P(H_j)}{P(H_i)}$$

для всех $j = 1, \dots, m$.

На каждом шаге на основе информации, собранной до этого, принимается решение: принять окончательное решение или провести следующее наблюдение. Проблема в стоимости наблюдений. На каждом шаге приходится соизмерять стоимость выполнения будущих наблюдений со средним выигрышем, даваемым им.

Пусть имеем m гипотез с априорными вероятностями $P(H_i)$, $i = 1, \dots, m$. Для любого плана последовательной выборки S с элементами s_j ; $j = 0, \dots, N$,

$$R(P, S, d) = \sum_{i=1}^m P(H_i) \sum_{j=1}^N \int \left\{ c_j(x) + L[H_i, d_j(x)] \right\} P(x|H_i) dx,$$

где $c_j(x)$ - стоимость наблюдений x_1, \dots, x_j ; $d_j(x)$ - решающая функция, основанная на замерах x_1, \dots, x_j .

1.7. Последовательные решения для двухточечной модели

Пусть проводятся поочередно наблюдения x_1, \dots, x_n над (неизвестным объектом) некоторым распределением, зависящим от неизвестного параметра ω . Получают последовательную выборку. После каждого наблюдения x_n можно оценить информацию о ω , полученную на основе наблюдений x_1, \dots, x_n . Каждое наблюдение стоит некоторую сумму. Предположим, что можно проводить наблюдения последовательно, делая после каждого наблюдения вывод о том следует ли принять решение $d \in D$ или провести очередное наблюдение.

Последовательный выбор может дать, например, выигрыш при проверке большой партии изделий путем извлечения m изделий, из которых допускается бракованных k изделий. Последовательный отбор

прекращается, если будет отобрано k бракованных изделий или $m-k$ годных изделий. Как правило, это происходит до отбора всех m изделий. В первом случае вся партия будет забракована, во втором - принята.

Таблица 1.3

Ω	D	
	d_1	d_2
ω_1	0	λ_1
ω_2	λ_2	0

Рассмотрим задачу последовательного решения с параметрическим пространством $\Omega=\{\omega_1, \omega_2\}$ и пространством решений $D\{d_1, d_2\}$.

Функция потерь задана табл.1.3, где $\lambda_i > 0, i=1,2$. Мы можем наблюдать последовательную выборку x_1, x_2, \dots ; каждое наблюдение стоит с ед. Пусть $f_i(\bullet | \omega_i)$ - условная о.в.п. наблюдения x при $w=\omega_i$. Так как параметр w может принимать всего два значения, то его распределение на каждом шаге процесса выбора задается одним числом $P(w=\omega_1)=p, 0 \leq p \leq 1$.

Риск принятия решения до наблюдения равен

$$\rho_0(p) = \min[\lambda_1 p; \lambda_2(1-p)].$$

Обозначим Δ класс всех процедур последовательного решения δ , требующих хотя бы одного наблюдения, и $\rho'(p) = \inf_{\delta \in \Delta} \rho(p, \delta)$.

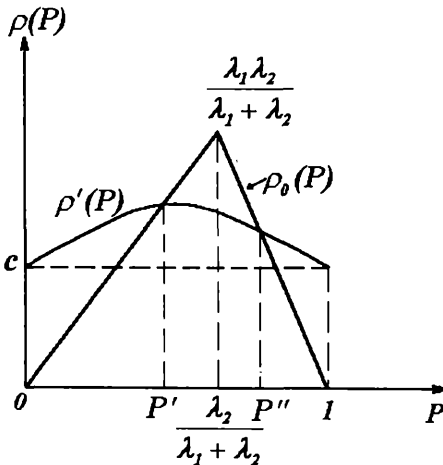


Рис. 1.3. Связь минимального риска и цены наблюдения

Таким образом, байесовский риск $\rho^*(p)$ удовлетворяет соотношению

$$\rho^*(p) = \min\{\rho_0(p), \rho'(p)\}.$$

Уже указывалось, что $\rho'(p)$ - вогнутая непрерывная функция на интервале $0 \leq p \leq 1$. Так как в риск входит по крайней мере цена c , то $\rho'(0)=c$ и $\rho'(p) \geq c$ при всех $p(0 \leq p \leq 1)$ (рис.1.3).

Пусть Σ^* - множество тех значений p , для которых оптимальным является окончание выбора, т.е.

$$\Sigma^* = \{p; \rho(p) \leq \rho'(p)\}.$$

Предположим, что

$$\rho'\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right) < \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2},$$

тогда Σ^* является объединением интервалов $0 \leq p \leq p'$ и $p'' \leq p \leq 1$; p' и p'' удовлетворяет уравнениям $\lambda_1 p' = \rho(p')$; $\lambda_2(1-p'') = \rho(p'')$.

Если это условие не имеет места, то Σ^* совпадает со всем интервалом $0 \leq p \leq 1$. В этом случае проводить наблюдения не имеет смысла.

Множество Σ^* задает байесовскую процедуру последовательного решения. Трудность состоит в получении информации о функции $\rho'(p)$, достаточной для того, чтобы явным образом определить ρ' и p''

Предположим, что априорная вероятность p удовлетворяет условию $p' < p < p''$. В этом случае целесообразно провести первое наблюдение. Если $\xi(x_1, \dots, x_n)$ априорная вероятность того, что $w = \omega_1$ при проведенных $X_i = x_i (i=1 \dots n)$ наблюдениях, то дальнейшие наблюдения надо проводить, если выполнено соотношение

$$p' < \xi(x_1, \dots, x_n) < p''$$

Если нарушается первое неравенство, то принимается решение d_2 , если второе - принимается решение d_1 .

Апостериорную вероятность можно записать так:

$$\xi(x_1, \dots, x_n) = \left[1 + \frac{1-p}{p} \frac{f_2(x_1) \dots f_2(x_n)}{f_1(x_1) \dots f_1(x_n)} \right]^{-1}$$

Введем постоянные A и B :

$$A = \frac{p(1-p'')}{(1-p)p''}; \quad B = \frac{p(1-p')}{(1-p)p'}$$

Так как $p' < p < p''$, то $A < 1$, $B > 1$. Дальнейшие наблюдения следует проводить, если

$$A < \frac{f_2(x_1) \dots f_2(x_n)}{f_1(x_1) \dots f_1(x_n)} < B$$

При нарушении неравенств принимаются соответственно решения d_1 и d_2 .

Процедура последовательного решения такого вида называется последовательным критерием отношения вероятностей.

Вычислим вероятность принятия решений d_1 и d_2 и найдем среднюю стоимость выбора для последовательной процедуры. В большинстве задач точные формулы получить не удастся, но есть простые приближенные формулы. Для последовательности наблюдений x_1, \dots, x_n , приводящих к d_1 , выполняется соотношение

$$\prod_{i=1}^n f_2(x_i) \leq A \prod_{i=1}^n f_1(x_i) \quad \text{или}$$

$$Pr(d_1 | w = \omega_2) \leq A Pr(d_1 | w = \omega_1).$$

Для последовательности наблюдений, приводящей к d_2 , получим

$$\prod_{i=1}^n f_2(x_i) \geq B \prod_{i=1}^n f_1(x_i);$$

$$Pr(d_2 | w = \omega_2) \geq B Pr(d_2 | w = \omega_1).$$

Но $Pr(d_1|w = \omega_2) + Pr(d_2|w = \omega_i) = 1; i=1,2;$

т.е. точка $\{Pr(d_2|w = \omega_i); Pr(d_2|w = \omega_2)\}$ лежит в области Φ .

Можно считать приведенные соотношения приближенными равенствами, т.е.

$$Pr(d_1|w = \omega_1) = 1 - Pr(d_2|w = \omega_1) \approx \frac{B-1}{B-A};$$

$$Pr(d_1|w = \omega_2) = 1 - Pr(d_2|w = \omega_2) \approx A(B-1)/(B-A)$$

Эти приближенные значения являются координатами точки $M[Pr(d_2|w = \omega_1); Pr(d_1|w = \omega_2)]$ (рис. 1.4).

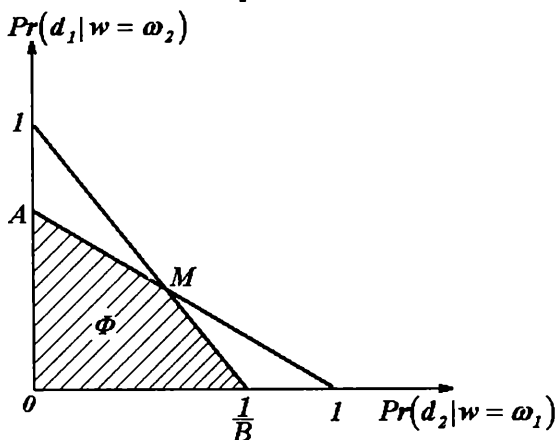


Рис. 1.4. Вероятности принятия решений d_1 и d_2

Введем случайную величину

$$z_i = \log \frac{f_2(x_i)}{f_1(x_i)}; i=1 \dots n,$$

и положим $\log A = a < 0; \log B = b > 0$, тогда, согласно последовательному критерию отношения вероятностей, выбор надо продолжать при

$$a < \sum_{i=1}^n z_i < b$$

При фиксированном значении $w = \omega_i (i=1,2)$ случайные величины x_1, x_2, \dots независимы и одинаково распределены. Следовательно, этими же свойствами обладают и случайные величины z_1, z_2, \dots, z_n . Для любых заданных значений a и b все моменты распределения случайного числа наблюдений N конечны и процесс выбора заканчивается с вероятностью 1, т.е. $Pr(N < \infty) = 1; E(N^k) < \infty$ для $k=1,2, \dots$

Если z_1, z_2, \dots, z_n последовательность независимых одинаково распределенных случайных величин, для которых $E(z_i) = m; i=1 \dots N$, то для

всякой последовательной процедуры с моментом $E(N) < \infty$ имеет место равенство

$$E\left(\sum_{i=1}^N z_i\right) = mE(N)$$

В общем случае границы a и b можно найти, минимизируя риск $\rho(p, \delta)$. Как правило, это сделать сложно. При малой цене с оптимальная процедура обычно предписывает проведение большого числа наблюдений, границы a и b - большие числа. Тогда получают следующие приближения:

$$Pr(d_2 | w = \omega_1) \approx e^{-b}; \quad Pr(d_1 | w = \omega_2) \approx e^{-a}$$

$$E(N | w = \omega_1) \approx \frac{a}{E(z | w = \omega_1)}; \quad E(N | w = \omega_2) \approx \frac{b}{E(z | w = \omega_2)};$$

$$a \approx \log c - \log \frac{I_1 \lambda_2 (1-p)}{p}; \quad b \approx \log \frac{1}{c} + \log \frac{I_2 \lambda_1 p}{1-p},$$

где $I_1 = -E(z | w = \omega_1)$; $I_2 = E(z | w = \omega_2)$;

I_1 и I_2 называют информационными числами.

1.8. Байесовские методы обучения

Обучение с поощрением. Параметры известной функции плотности $P(x | \omega_j)$, могут быть оценены путем итерационного применения теоремы Байеса.

Пусть существует априорная функция плотности с неизвестным параметром θ , $P_0(\theta)$, отражающая первоначальную информацию о θ . Когда наблюдается последовательность независимых и одинаково распределенных векторов признаков x_1, x_2, \dots, x_n , относящихся к одному классу образов, функция стремится к апостериорной плотности $P(\theta | x_1, \dots, x_n)$. После первого наблюдения (символ ω_1 опущен) априорная функция плотности будет

$$P(\theta | x_1) = \frac{P(x_1 | \theta) P_0(\theta)}{P(x_1)}.$$

После наблюдения x_2

$$P(\theta | x_1, x_2) = \frac{P(x_2 | x_1, \theta) P(\theta | x_1)}{P(x_2 | x_1)}.$$

В общем случае

$$P(\theta | x_1, \dots, x_n) = \frac{P(x_n | x_{n-1}, \dots, x_1, \theta) P(\theta | x_1, \dots, x_{n-1})}{P(x_n | x_1, \dots, x_{n-1})}. \quad (1.16)$$

Искомая функция плотности может быть вычислена как $P(x_{n+1}|x_1, \dots, x_n, \omega_i) = \int P(x_{n+1}|x_1, \dots, x_n, \omega_i, \theta) P(\theta|x_1, \dots, x_n, \omega_i) d\theta$.
 $n=1, 2, \dots$,

где член

$P(x_{n+1}|x_1, \dots, x_n, \omega_i, \theta)$ известен, а $P(\theta|x_1, \dots, x_n, \omega_i)$ получается из (1.16).

Известно, что в среднем апостериорная функция плотности улучшает оценку параметра, которая сходится к истинному значению, если только оно не исключено из априорной функции плотности.

Пример. Имеем n обучающих наблюдений x_1, \dots, x_n взятых из двух классов образцов ω_1 и ω_2 , которые имеют одномерные гауссовы распределения с неизвестными параметрами. Оптимальная решающая граница, минимизирующая вероятность ошибочного распознавания, является функцией априорных вероятностей, средних значений наблюдений и дисперсий.

Совместное распределение в этом случае имеет вид

$$P(x) = \sum_{i=1}^2 P(\omega_i) P(x|\omega_i) = \\ = \frac{1}{2} \frac{1}{\sqrt{2\pi} \cdot \sigma} \left\{ \exp \left[-\frac{(x - m_1)^2}{2\sigma^2} \right] + \exp \left[-\frac{(x - m_2)^2}{2\sigma^2} \right] \right\}$$

Тогда оптимальная решающая граница определяется условием

$$m_n = \frac{1}{n} \sum_{i=1}^n x_i$$

Обучение без поощрения. Поскольку невозможно точно отнести каждое наблюдение к правильному классу образов, то задача формируется как задача оценки параметров общего распределения. Обучающие наблюдения считаются принадлежащими к этому совместному распределению, его составляющие могут быть распределениями каждого класса образов, или распределениями, соответствующими различным разбиениям пространства наблюдений (признаков). Пусть множество случайных наблюдений x_1, x_2, \dots, x_n может быть разбито на t ячеек z_1^n, \dots, z_m^n .

Совместное распределение определяется следующим образом:

$$P(x) = \sum_{i=1}^m P(x|z_i^n) P(z_i^n), \quad (1.17)$$

где

$P(x|z_i^n)$ условное распределение для i -го разбиения; $P(z_i^n)$ параметр i -го разбиения.

Параметрически условное совместное распределение $P(x|\theta, p)$ может быть построено на основе семейства параметрически условных распределений $P(x|\theta_i, x_i^n); i = 1, \dots, m$, с двумя множествами параметров $\theta = \{\theta_1, \dots, \theta_m\}$ и $P = \{p(z_1^n), p(z_2^n), \dots, p(z_m^n)\}$.

Основное уравнение (1.17) примет вид

$$P(x|\theta, p) = \sum_{i=1}^m P(x|\theta_i, z_i^n) P(z_i^n).$$

Известно, что класс совместных распределений, для которых существует единственное решение для θ и P , ограничен. Метод оценки параметров θ и P зависит от частной задачи.

Оценка параметров совместных распределений. Имеем два класса образов ω_1 и ω_2 . Вид $P(x|\omega)$ известен, неизвестный параметр θ , полностью известна $P(x|\omega_2)$. По наблюдениям x_1, \dots, x_n надо найти оценку θ . Если последовательность x_1, \dots, x_n разбить на все возможные комбинации по классам ω_1 и ω_2 , то будет 2^n комбинаций. Пусть z_i^n означает i -е разбиение последовательности x_1, \dots, x_n . Тогда апостериорная плотность вероятности будет

$$P(\theta|x_1, \dots, x_n) = \sum_{i=1}^{2^n} P(\theta|x_1, \dots, x_n, z_i^n) P(z_i^n|x_1, \dots, x_n).$$

Задача свелась к обучению с поощрением для каждого из 2^n разбиений. Оценка параметров производится путем взятия взвешенной суммы результатов для каждого разбиения с весами, равными $P(z_i^n|x_1, \dots, x_n)$, $i = 1, \dots, 2^n$. Объем вычислений при этом растет экспоненциально.

Другой метод. По теореме Байеса

$$P(\theta|x_1, \dots, x_n) = \frac{P(x_n|\theta, x_1, \dots, x_{n-1}) P(\theta|x_1, \dots, x_{n-1})}{P(x_n|x_1, \dots, x_{n-1})}. \quad (1.18)$$

Пусть обучающие наблюдения условно независимы, т.е.

$$P(x_{n+1}|\theta, x_1, \dots, x_n) = P(x_{n+1}|\theta).$$

Тогда

$$\begin{aligned} P(x_n|\theta, x_1, \dots, x_{n-1}) &= \\ &= P(\omega_1) \cdot P(x_n|\theta, \omega_1) + P(\omega_2) \cdot P(x_n|\theta, \omega_2). \end{aligned} \quad (1.19)$$

Подставляя (1.18) в (1.19), получим рекуррентное соотношение для оценки θ .

$$\begin{aligned} P(\theta|x_1, \dots, x_n) &= \\ &= P(\theta|x_1, \dots, x_{n-1}) \left[\frac{P(\omega_1) P(x_n|\theta, \omega_1)}{P(x_n|x_1, \dots, x_{n-1})} + \frac{P(\omega_2) P(x_n|\theta, \omega_2)}{P(x_n|x_1, \dots, x_{n-1})} \right]. \end{aligned}$$

Если $P(\theta|x_1, \dots, x_n)$, $P(x_n|\theta, \omega_1)$, $P(x_n|\theta, \omega_2)$, $P(\omega_1)$, $P(\omega_2)$ известны, то можно получить $P(\theta|x_1, \dots, x_n)$. Пусть $P(\omega_1)$ и $P(\omega_2)$ - известны. Тогда для отыскания $P(x_n|\theta, \omega_1)$ и $P(\theta|x_1, \dots, x_n)$ при всех значениях θ следует предположить, что к величине θ можно применить конечное квантование, чтобы объем вычислений был конечным.

Для t классов и ряда неизвестных параметров θ_i , относящихся соответственно к ω_i , полагая условную независимость обучающих наблюдений и значений θ_i параметров, получим рекуррентное соотношение для оценки θ_i :

$$P(\theta_i|x_1, \dots, x_n) = P(\theta_i|x_1, \dots, x_{n-1}) \times \\ \times \frac{P(\omega_i)P(x_n|\theta_i, \omega_i) + \sum_{j=1}^m P(\omega_j)P(x_n|\omega_j, x_1, \dots, x_{n-1})}{\sum_{j=1}^m P(\omega_j)P(x_n|\omega_j, x_1, \dots, x_{n-1})}.$$

В общем случае либо $P(\theta_i)$, либо $P(\omega_i)$, $i=1, \dots, m$, должны быть различны, иначе вычисления для всех θ_i будут оценивать одно и то же значение (так как вычисляется одна величина) и система в целом ничему не будет обучена.

1.9. Обучение с помощью стохастической аппроксимации

Обучение с поощрением (с учителем). Имеем наблюдение x_1, \dots, x_n , n - число случаев, когда наблюдение соответствует ω_i , $n = \sum_{i=1}^m n_i$. Надо

оценить неизвестную вероятность $P(\omega_i)$, если $\sum_{i=1}^m P(\omega_i) = 1$

Имеем начальную оценку $P_0(\omega_i)$; $\sum_{i=1}^m P_0(\omega_i) = 1$

Последовательные оценки

$$P_{n+1}(\omega_i) = P_n(\omega_i) + \gamma_{n+1} \left[\frac{n_i}{n} - P_n(\omega_i) \right].$$

Для оценки неизвестной функции плотности по наблюдениям x_1, \dots, x_n представляют $P(x)$ в виде:

$$P(x) = \sum_{i=1}^m c_i \varphi_i(x); \quad \int \varphi_i(x) \varphi_j(x) dx = \begin{cases} 0; & i \neq j; \\ 1; & i = j; \end{cases}$$

$$c_{i,n+1} = c_{i,n} + \gamma_n [\varphi_i(x_n) - c_{i,n}]; \quad i = 1 \dots m;$$

$$l > \gamma_n > 0; \quad \sum_{n=1}^{\infty} \gamma_n = \infty; \quad \sum_{n=1}^{\infty} \gamma_n^2 < \infty.$$

В частном случае можно найти оценки неизвестных параметров по наблюдениям x_1, \dots, x_n .

Обучение без поощрения (без учителя). Рассматривается задача оценки параметров в совместном распределении с помощью стохастической аппроксимации.

Предполагается, что:

1. Существуют m классов распределений вероятностей, соответствующих m классам образцов, априорные вероятности которых

фиксированы, но неизвестны, $P(\omega_i); \sum_{i=1}^m P(\omega_i) = 1$

2. Функции распределения вероятностей (или плотности) каждого класса ω_i характеризуются некоторым множеством параметров θ_i .

3. Предполагается, что обучающие наблюдения взяты из совместного распределения, построенного из составляющих распределений, т.е.

$$P(x|\theta, p) = \sum_{i=1}^m P(x|\omega_i, \theta_i) P(\omega_i);$$

θ и $P(\omega_i)$ - неизвестны.

4. Существуют несмещенные оценки некоторых статистик $H = \{H(x)\}$, например, моменты. Функциональное отношение $F(H, \theta, p) = 0$ между H и множествами параметров θ и p известно для каждого шага процесса обучения.

5. Имеются дополнительные соотношения $G(\theta, p) = 0$, обеспечивающие получение единственного решения для неизвестных параметров θ и p .

Из равенств $F(H, \theta, p) = 0$ и $G(\theta, p) = 0$ получают оценки θ и p ; $F(\bullet)$ может быть найдено из последовательных оценок $\{H(x)\}$, а $G(\bullet)$ задано априори или с помощью вспомогательных процедур оценки.

Пример 1 Пусть $m=2$; $P(\omega_1) = p_1$, $P(\omega_2) = 1 - p_1$.

Каждая составляющая функция плотности характеризуется средним значением m_i и дисперсией σ_i^2 , т.е.

$$P(x|\omega_i) \sim P(x|m_i; \sigma_i^2; \omega_i); \quad i=1..2.$$

Совместная функция плотности характеризуется параметрами

$\theta = \{m_1; m_2; \sigma_1^2; \sigma_2^2\}$, p_1 задано так:

$$P(x|\theta, p) = p_1 P(x|m_1; \sigma_1^2; \omega_1) + (1 - p_1) P(x|m_2; \sigma_2^2; \omega_2). \quad (1.20)$$

Надо найти оценки параметров θ и p_1 по классифицированным обучающим наблюдениям x_1, x_2, \dots, x_n , полученным из $P(x|\theta, p)$.

Из (1.20) следуют моменты

$$\begin{aligned} E(x) &= p_1 m_1 + (1 - p_1) m_2 = p_1(m_1 - m_2) + m_2; \\ E(x^2) &= p_1(m_1^2 + \sigma_1^2) + (1 - p_1)(m_2^2 + \sigma_2^2); \\ E(x^3) &= p_1(m_1^3 + 3m_1\sigma_1^2) + (1 - p_1)(m_2^3 + 3m_2\sigma_2^2). \end{aligned} \quad (1.21)$$

Из системы (1.21) находим искомые параметры. Как правило, параметры последовательно оцениваются посредством оценок моментов совместных распределений. В случае, когда σ_1^2 и σ_2^2 неизвестны и не равны друг другу, потребуются моменты смешанного распределения более высокого порядка для получения достаточных функциональных связей между неизвестными параметрами.

Обычно неоднозначности решения следует ожидать при совместном решении нелинейных уравнений, полученных в результате применения метода моментов. Единственное решение может быть только в случае, если имеется дополнительная информация о параметрах.

Пример 2. Пусть $m_2=0$, $\sigma_1=\sigma_2=\sigma$. Требуется оценить p_1, σ, m_1 .

Из системы (23)

$$\begin{aligned} E(x) &= p_1 m_1; \\ E(x^2) &= p_1 m_1^2 + \sigma^2 \\ E(x^3) &= p_1(m_1^3 + 3m_1\sigma^2). \end{aligned}$$

Получим $m_1^2 - 3E(x)m_1 + 3E(x^2) - \frac{E(x^3)}{E(x)} = 0$ квадратное

уравнение относительно m_1 . Единственное решение будет при $p=2/3$, что следует из равенства нулю дискриминанта квадратного уравнения

$$9E^2(x) - 12E(x^2) + 4\frac{E(x^3)}{E(x)} = m_1^2(3p_1 - 2)^2$$

В других случаях решение будет не единственно.

1.10. Математическая постановка задачи учета погрешности признаков

Как показано в предыдущих параграфах, в традиционных методах (алгоритмах) решения задач распознавания образов (общем случае в статистических задачах решения) не в полной мере учитываются

статистическая природа наблюдений (Рис. 1.5). В частности, в процедурах классификации функции условных плотностей вероятности векторов признаков и результаты наблюдений, по которым ведется классификация образов, выступают как детерминированные функции и детерминированные величины.

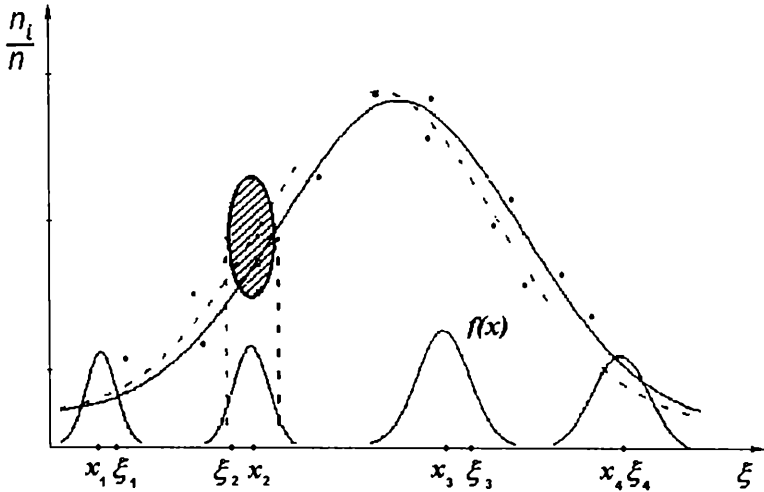


Рис. 1.5. Оценки условной плотности вероятности признака ξ при плотности вероятности наблюдаемых значений $f(x)$

На рис. 1.5

$$\begin{cases} \frac{n_i}{n} = p(\xi_i, \theta | \omega_j) + \varepsilon_i, \\ X = \xi + \delta, \end{cases}$$

$p(\xi, \theta | \omega)$ - условная плотность;

$f(x_j)$ - плотность вероятности наблюдения x_j (в условных единицах),

$$p(\hat{\xi}, \hat{\theta} | \omega_j),$$

$$p(X, \hat{\theta} | \omega_j).$$

Подобные ограничения приводят к смещенным оценкам, например, разделяющих функций, а, как следствие, к увеличению числа неправильных решений.

В процессе полного учета статистических характеристик наблюдаемых векторов признаков в статистических задачах методы приведут к уменьшению числа ложных решений. Пусть задано D пространство решений d ; Ω - пространство образов ω (или пространство параметров в статистических задачах решения); R - пространство потерь, которые могут быть получены в результате решения d и исхода ω . Каждый образ $\omega \in \Omega$ характеризуется своим вектором признаков ξ а также

считаются известными для каждого класса образов ω_j , $j=1, \dots, m$, многомерные условные функции плотности вероятностей векторов признаков $P(\xi|\omega_j)$, которые оцениваются по результатам наблюдений векторов признаков ξ , координаты которых в силу различных случайных помех не наблюдаются, а наблюдаются случайные величины X , связанные с ξ следующим соотношением: $X=\xi+\delta$, где δ - случайная аддитивная помеха с известным законом распределения. В частности, будем считать, что δ подчиняется многомерному нормальному закону распределения с нулевым математическим ожиданием и известной ковариационной матрицей $\delta^2 I$ (I - единичная матрица).

В рассмотренных традиционных методах статистических задач распознавания образов (статистические задачи решения) с фиксированным объемом выборки можно выделить два этапа: на первом этапе по результатам наблюдений случайного вектора признаков x_1, \dots, x_n получают оценку условных распределений в виде $P(X|\omega_j)$ и функции распределения образов $P(\omega_j)$, $j=1, \dots, m$; на втором этапе, считая известными $P(X|\omega_j)$ и $P(\omega_j)$, $j=1, \dots, m$, наблюдаемый вектор X относят к некоторому классу ω_j , $j=1, \dots, m$ (Рис. 1.6). Классификация проводится путем минимизации вероятности ошибочного распознавания, в процессе которой получается решающее правило $\delta(X)$.

В традиционных методах на втором этапе $P(X|\omega_j)$, $j=1, \dots, m$, выступают в роли детерминированные функций, а случайная величина (вектор) X как детерминированная величина (вектор). В силу этого получаемые разделяющие функции также выступают как детерминированные функции.

В действительности же следует оценить "истинные значения" признаков ξ и функции $P(\xi|\omega_j)$, $j=1, \dots, m$, которые в процессе обработки результатов наблюдений x_1, \dots, x_n должны иметь интервальные оценки, а если вид самих функций $P(\xi|\omega_j)$ был выбран априори, то должны быть указаны интервальные оценки их параметров.

Соответственно должны быть представлены интервальные оценки разделяющих функций и наблюдавшейся перед классификацией случайной величины X .

На рис. 1.6 $L(d, \omega_j)$ ущерб от принятия решения d , когда выбран объект ω_j , $P(\omega_j)$ - вероятность появления ω_j -го объекта.

Задача классификации при этом усложняется, но уменьшается число ложных распознаваний. Учет интервальных оценок разделяющих функций и наблюдавшихся перед классификацией величин чрезвычайно важен в практических приложениях, когда координаты вектора признаков достаточно близки для различных объектов, а точность измерения этих признаков соизмерима с разностью значений признаков. Например, при радиолокационных измерениях характеристик различных типов самолетов точность измерения соизмерима с разностью значений этих характеристик.

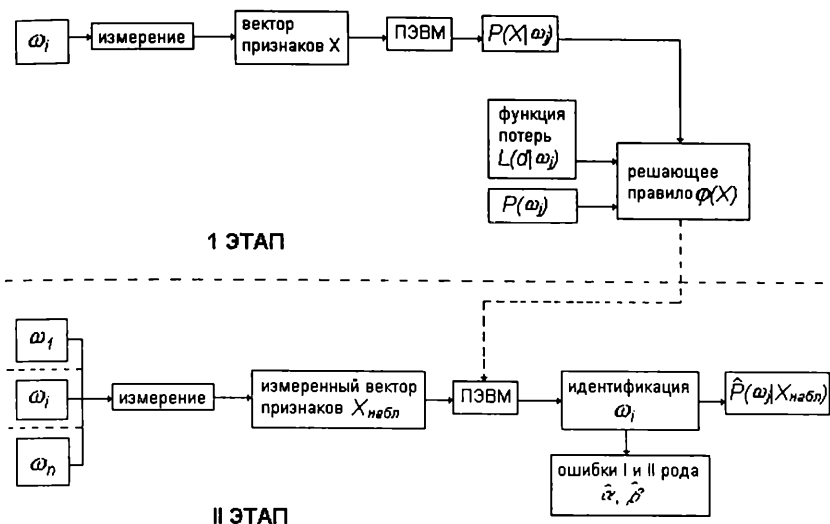


Рис. 1.6 Блок-схема традиционной процедуры обучения системы и распознавания образов

В общем случае байесовский риск (который определяет разделяющую функцию) относительно мало чувствителен к ошибке (приращению) в выборе значения распределения $P(\omega_j)$, $j=1, \dots, m$; если функция байесовского риска кусочно-линейна, то приращение байесовского риска равно нулю, когда приращение $P(\omega_j)$ содержится в интервале линейности функции байесовского риска. Поэтому в дальнейшем $P(\omega_j)$ мы рассматривать не будем.

В действительности в процессе наблюдений измеренные значения вектора признаков имеют две компоненты: истинное значение ξ и помеха δ . Здесь под понятием "помеха" понимается и погрешность измерения, вызванная действием помех, и другие источники неопределенности в значении X . Измеренное значение $X = \xi + \delta$. Неучет помехи δ приводит к смещенным оценкам всех характеристик, в расчете которых участвует значение X . Кроме этого, необходимо учесть и интервальные оценки всех последующих рассчитываемых функций и величин: $P(\xi|\omega_j)$, ошибок I и II рода - чего не делается в традиционных методах. Этот учет приводит к появлению зон неопределенности в принятии решений. Наблюдаемые значения признаков $X_{набл}$ в процедуре идентификации также несут в себе помеху, и идентификацию объектов следует проводить по оценке истинного значения $\xi_{набл}$, а не по наблюдаемому значению $X_{набл}$. Получить оценку $\xi_{набл}$ можно только используя всю информацию, полученную на I этапе. Таким образом задача идентификации практически не разделяется на независимые этапы, а имеет общую схему (Рис. 1.7).

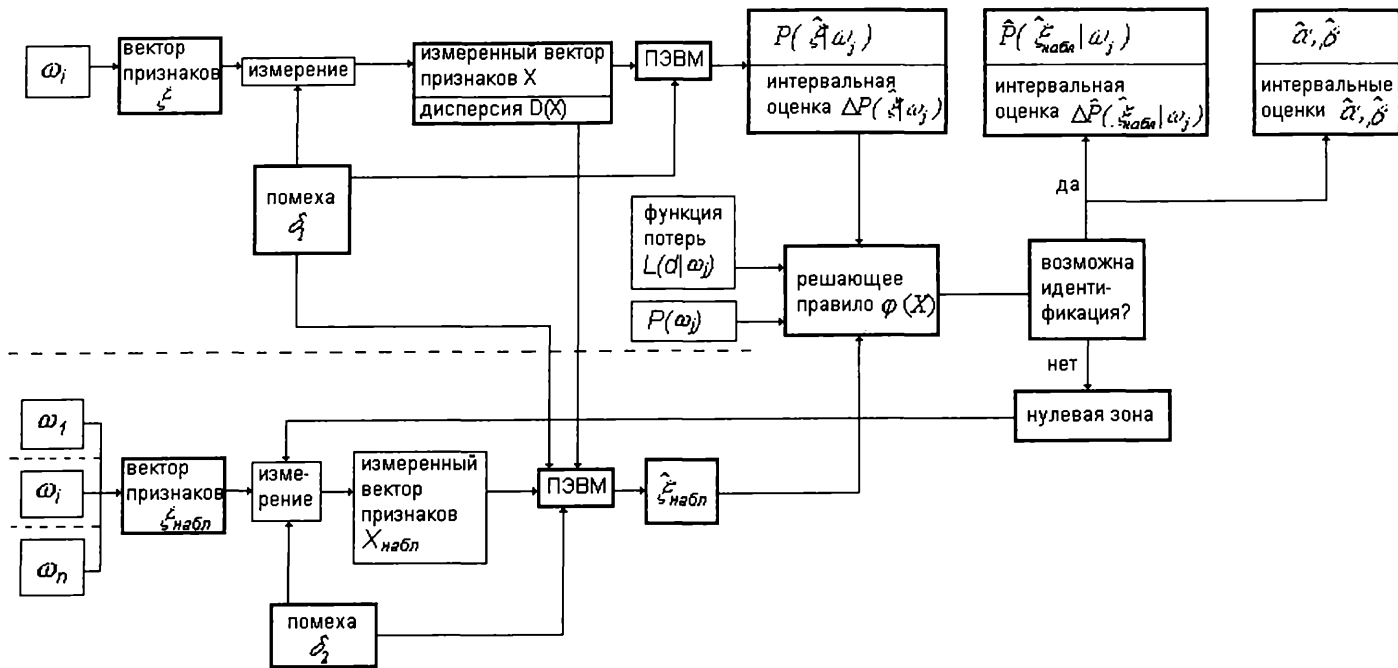


Рис. 1.7 Блок-схема единой процедуры обучения системы и распознавания образов

На рис. 1.6 и 1.7 прямоугольником с надписью ПЭВМ, обозначена процедура обработки результатов наблюдений и ее алгоритм; на рис.1.7 утолщенной обводкой прямоугольников обозначены блоки, которых либо нет в традиционных методах, либо эти блоки претерпевают существенные изменения по сравнению с традиционным подходом.

Из алгоритма распознавания образов, следующего из блок-схемы рис. 1.7, видно, что не только изменяются конкретные значения оцениваемых функций и параметров по сравнению с традиционными алгоритмами (а это приводит и к другим практическим выводам), но появляются и принципиальные моменты: появляется область неопределенности, когда идентификацию объектов в принципе произвести невозможно.

Показано, что в процессах принятия решений даже при фиксированном объеме выборки появляются зоны неопределенности, где необходимо применять процедуры методов последовательного принятия решений. Естественно, что в «чистых» методах принятия решений учет погрешности наблюдаемых значений признаков приведет к изменению верхних и нижних останавливающих границ и среднего числа наблюдений до принятия решений.

Отсюда приходим к математической постановке задачи.

Математическая постановка задачи

Введем некоторые обозначения. Каждый объект в задаче распознавания характеризуется своим вектором признаков ξ . В эксперименте из-за помех вместо величины ξ наблюдается случайная величина X : $X = \xi + \delta$. Для величины ξ может быть задана как плотность вероятности признаков (для непрерывных случайных величин), так и функция вероятности (для дискретных случайных величин).

Сформулируем математическую постановку задачи.

Пусть до принятия решения d , принадлежащего множеству решений $D(d \in D)$ задано параметрическое пространство исходов Ω с параметром $w(w = \omega \in \Omega)$ и для всех $\omega \in \Omega$ задана обобщенная вероятностная плотность $P(\omega)$; Ω - пространство образов ω (или пространство параметров w в статических задачах решения).

Задана вещественная функция полезности $U(w = \omega, d)$ или функция потерь $L(w = \omega, d) = -U(w = \omega, d)$ на произведении $\Omega \times D$.

Пусть S выборочное пространство возможных значений наблюдений признака ξ исхода ω , когда значение признака не наблюдается, а наблюдается случайная величина X , связанная со значением признака ξ соотношением

$$X = \xi + \delta,$$

где δ - аддитивная помеха, закон распределения которой $f(X)$ и его числовые характеристики считаются известными, в частности - это может

быть нормальный закон распределения с математическим ожиданием $M(\delta)=0$, и дисперсионной матрицей $D(\delta)=\sigma^2 \cdot I$ (I - единичная матрица).

Пусть задано с точностью до значений параметров параметрическое семейство условных обобщенных вероятностных плотностей (о.в.п.) $p(\xi, \theta, \omega)$, $\omega \in \Omega$, при неизвестных о.в.п. $p(\xi, \theta, \omega)$ задано параметрическое семейство разделяющих функций $h(\xi, \theta)$.

Выбрана из класса решающих функций Δ решающая функция $\varphi(\xi) \in \Delta$, определяющая для любого возможного значения $\xi \in S$ решение $d(\xi) \in D$, минимизирующее средний риск

$$P[\omega, d(\xi)] = M\{L[\omega, d(\xi)]\} = \int \int_{\Omega S} L[\omega, d(\xi)] p(\xi, \theta, \omega) \cdot P(\omega) \cdot d\mu(\xi) \cdot dv(\omega),$$

где для всех $\omega \in \Omega$ функция $L[\omega, d(\xi)]$ измерима и интегрируема на множестве S .

Требуется по выборке фиксированного объема результатов наблюдений X и в методах последовательного принятия решений найти:

1) точечные и интервальные оценки параметров θ условных о.в.п. $p(\xi, \theta, \omega)$, а также точечные и интервальные оценки условных о.в.п. $p(\xi, \theta, \omega)$; при неизвестных о.в.п. $p(\xi, \theta, \omega)$ точечные и интервальные оценки параметров θ и разделяющих функций $h(\xi, \theta)$;

2) точечные и интервальные оценки решающих функций $\varphi(\xi)$,

3) оценки "истинных" значений $\hat{\xi}$ вектора признаков ξ , по которым впоследствии принимается решение $d \in D$;

4) произвести идентификацию наблюдаемого объекта, т. е. поставить в соответствие наблюдаемому значению вектора признаков X некоторый образ ω_j и определить при этом точечные и интервальные оценки суммарной ошибки и ошибок I и II родов, возникающих при принятии решения $d \in D$.

Здесь обозначения $d\mu(\xi)$ и $dv(\xi)$ показывают, что записанный интеграл $P[\omega, d(\xi)]$ понимается как обычный интеграл для непрерывных случайных величин ξ и ω и как сумма для дискретных величин. При любом наборе $(\omega, d) \in \Omega \times D$ число $L(\omega, d)$ представляет собой ущерб лица, принимающего решение d , когда значение параметра ω равно ω .

В отличие от этой постановки, в традиционной постановке задачи распознавания не учитываются помехи δ , величины ξ и X отождествляются; не определяются интервальные оценки функций $p(\xi, \theta, \omega)$, $h(\xi, \theta)$ и $\varphi(\xi)$, не определяются оценки $\hat{\xi}$ и интервальные оценки ошибок I и II родов (α и β).

Неучет помех δ и оценок $\hat{\xi}$ приводит к смещенным точечным оценкам $p(\xi, \theta, \omega)$, $h(\xi, \theta)$ и $\varphi(\xi)$, к неверным интервальным оценкам и к ложным практическим выводам.

Г Л А В А 2. ПРИНЯТИЕ РЕШЕНИЙ ПО ВЫБОРКЕ ФИКСИРОВАННОГО ОБЪЕМА С УЧЕТОМ ПОГРЕШНОСТИ ПРИЗНАКОВ В ОБОБЩЕННЫХ УСЛОВНЫХ ПЛОТНОСТЯХ ВЕРОЯТНОСТИ

2.1. Анализ методов оценки параметров функции известного вида с учетом ошибок в значениях функций и аргументов

При определении вида функции плотности вероятности (ф.п.в.) распознаваемых классов или разделяющих (решающих) функций в качестве исходных данных используют выборки, полученные в результате конкретных экспериментов. И, как любые экспериментальные значения, исходные данные содержат ошибки, присутствие которых может серьезно повлиять на результаты анализа. Поскольку в настоящей работе рассматриваемые алгоритмы учета погрешностей в значениях признаков будут применяться к различным функциям: условным плотностям вероятности, разделяющим (решающим) функциям, то измеренные значения признаков обозначим переменной X , а значения оцениваемых функций - $y(x)$. Будем считать, что общий вид оцениваемых функций априори известен, но надлежит найти точные и интервальные оценки свободных параметров этих функций, по которым сможем определить точные и интервальные оценки самих функций.

Модели, позволяющие учитывать ошибки в значениях функций и аргументов, рассматриваются в [5, 9, 13, 16, 20, 24, 58, 63 - 76]. Наиболее часто [2, 7, 26, 49, 51, 62] ставится задача определения оценок θ параметров модели $y_i = f(x_i, \theta) + \varepsilon_i$, $i, j = 1 \dots n$; где ε_i случайная ошибка, имеющая нормальное распределение с параметрами $M(\varepsilon_i) = 0$, $D(\varepsilon_i) = \sigma_i^2$, $D(\varepsilon_i, \varepsilon_j) = 0$, $i, j = 1 \dots n$. Данная постановка является классической регрессионной задачей, которая решается методом максимума правдоподобия или методом наименьших квадратов (МНК).

Если $\sigma_i^2 = \sigma^2$, $i = 1 \dots n$ и значение σ^2 не задано, то оценку параметра σ можно найти по формуле $\hat{\sigma}^2 = \frac{s}{(n-p)}$, где $s = \sum_{i=1}^n [y_i - f(x_i, \hat{\theta})]^2$; p число параметров.

Регрессионный анализ предполагает, что переменные x являются "детерминированными" - измерены с погрешностью много меньшей, чем погрешность значений функции. На практике это требование очень часто

не выполняется, поэтому возникает необходимость учета погрешностей аргумента x .

Рассмотрим пассивный эксперимент определения оценок параметров функции $\eta=f(\xi, \theta)$ где θ - неизвестный вектор параметров.

В процессе измерений значения η и ξ не наблюдаются, а получают набор значений x_i и y_i , определяемых как $y_i = \eta_i + \varepsilon_i$, $x_i = \xi_i + \delta_i$, $i = 1 \dots n$,

где ε_i и δ_i - ошибки измеренных значений функции и аргумента.

Предположим, что ошибки измерений ε_i и δ_i - нормально распределенные случайные величины с нулевыми средними значениями, с дисперсиями $\sigma^2(y_i)$ и $\sigma^2(x_i)$ соответственно и коэффициентом корреляции $\rho_i=0$.

Рассмотрим несколько подходов к решению данной задачи.

I. Один из подходов подразумевает использование вместо истинных значений ξ_i значения наблюдаемых x_i при оценивании параметров МНК. То есть ошибка измерения δ_i игнорируется и получаем следующую модель: $y=f(x, \theta)$ +ошибка.

В основном, как показано в [15, 20, 24], этот подход дает несостоятельную оценку с большим асимптотическим смещением.

В [65] рассмотрен случай, когда ξ - случайная величина, выбранная независимо от ε_i и δ_i , с характеристиками $E(\xi) = \mu$, $cov(\xi) = \sum_{\xi}$ значения ξ_i аппроксимируются результатами измерений x_i и при нормальном законе распределения случайной величины ξ_i

$$E(\xi|x) = x \cdot \Lambda + \mu(I - \Lambda),$$

$$\text{где } \Lambda = \left(\sum_{\xi} + \sigma^2(x) \cdot I \right)^{-1} \sum_{\xi}$$

Получаем модель

$$y = f\{x\Lambda + \mu(I - \Lambda), \theta\} + \text{ошибка}$$

Когда μ и Λ неизвестны, то, используя исходную выборку x_i , $i=1 \dots n$, определяем оценки $\hat{\mu}$, $\hat{\Lambda}$ и в итоге получаем модель

$$y = f\left[x \cdot \hat{\Lambda} + \hat{\mu}(I - \hat{\Lambda}), \hat{\theta}\right] + \text{ошибка} = f(\hat{\xi}, \hat{\theta}) + \text{ошибка}. \quad (2.1)$$

Оценки параметров модели (2.1) могут быть определены МНК. Например, для прямой линии $\eta=a\xi+b$, имеем

$$\hat{\xi} = x \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\theta^2} + \mu \left(I - \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\theta^2} \right),$$

где σ_θ^2 - дисперсия ошибки δ_i .

Минимизируемый функционал примет вид

$$F = \frac{\sum_{i=1}^n [y_i - \eta(\xi_i, \theta)]^2}{\sigma^2(y)} = \frac{\sum_{i=1}^n [(y_i - b - a\mu)(\sigma_x^2 + \sigma_0^2) - (x_i - \mu)a\sigma^2 x]^2}{(\sigma_x^2 + \sigma_0^2)^2 \cdot \sigma^2(y)}$$

В случае, когда функция $f(\xi, \theta)$ нелинейна по параметрам, используют разложение $f(\xi, \theta)$ в ряд Тейлора в окрестности точки ξ [20, 65]. Например, для нормального закона распределения получим

$$\begin{aligned} \eta &\approx \frac{1}{\sqrt{2\pi} \theta_2} \exp\left[-\frac{(x - \theta_1)^2}{2\theta_2^2}\right] \Big|_{x=\hat{\xi}} + \\ &+ \frac{1}{\sqrt{2\pi} \theta_2} \exp\left[-\frac{(x - \theta_1)^2}{2\theta_2^2}\right] \cdot \left[-\frac{(x - \theta_1)}{\theta_2^2}\right] \Big|_{x=\hat{\xi}} (x - \hat{\xi}) = \\ &= \frac{1}{\sqrt{2\pi} \theta_2} \exp\left[-\frac{(\hat{\xi} - \theta_1)^2}{2\theta_2^2}\right] \cdot \left[1 + \frac{\theta_1 - \hat{\xi}}{\theta_2^2} \cdot (x - \hat{\xi})\right] \end{aligned}$$

Оценки параметров θ_1 и θ_2 найдем из условия минимума функционала

$$F = \frac{\sum_{i=1}^n [y_i - \eta(\xi_i, \theta)]^2}{\sigma^2(y)}$$

II. Поставленная задача может быть решена итеративным методом наименьших квадратов с уточняемыми весами [20, 48, 49, 71].

Предполагается, что x_i является выборкой из некоторой генеральной совокупности с функцией плотности распределения $f(x_n, x_{oi})$. При этом первые моменты функции $f(x_n, x_{oi})$ известны и конечны.

В этом случае для построения минимизируемого функционала F необходимо знать вид функции плотности распределения $f(x_n, x_{oi})$ или ее моменты.

Для определения моментов в [71] предлагается использовать разложение функции $f(x_n, \theta)$ в ряд Тейлора в окрестности точки x_{oi} . Тогда первый и второй моменты функции плотности распределения $f(y_n/x_{oi})$ находят по формулам [71]:

$$E(y_i | x_{oi}) \approx f(x_i, \theta),$$

$$\sigma^2(y_i | x_{oi}) \approx \sigma^2(y_i) + \left[\frac{\partial f(x_i, \theta)}{\partial x_i} \Big|_{x_i=x_{oi}} \right]^2 \cdot \sigma^2(x_i)$$

Итерационный процесс нахождения оценок параметров строится следующим образом:

1. Составляется функционал

$$F_0 = \sum_{i=1}^n [y_i - E(y_i|x_{0i})]^2 \cdot w_i^0$$

где $w_i^0 = \sigma^{-2}(y_i)$.

Отыскиваются оценки параметров θ^0 при которых достигается минимум F_0 , т.е. решается регрессионная задача.

2. Подсчитываются величины $w_i^1 = \sigma^2(y_i|x_{0i})$.

3. Составляется сумма

$$F_1 = \sum_{i=1}^n [y_i - E(y_i|x_{0i})]^2 \cdot w_i^1, \quad (2.2)$$

и определяются оценки θ^1

Операции 2 и 3 повторяются до тех пор, пока относительные изменения параметров на соседних итерациях не будут меньше некоторой малой величины γ ,

$$\max \left| \frac{\theta_j^s - \theta_j^{s-1}}{\theta_j^s} \right| \leq \gamma \quad j=1 \dots m.$$

Рассмотрим применение данного подхода для линейной функции

$$\eta_i = ax_i + b$$

$$E(y_i|x_{0i}) \approx ax_{0i} + b, \quad \sigma^2(y_i|x_{0i}) \approx \sigma^2(y_i) + a^2 \sigma^2(x_i)$$

Функционал (2.2) примет вид

$$F = \sum_{i=1}^n \frac{(y_i - ax_i - b)^2}{\sigma^2(y_i) + a^2 \sigma^2(x_i)}$$

Для нелинейного случая, например, нормального закона, получим

$$E(y_i|x_{0i}) \approx \frac{1}{\sqrt{2\pi} \theta_2} \exp \left[-\frac{(x_{0i} - \theta_1)^2}{2\theta_2^2} \right],$$

$$\sigma^2(y_i|x_{0i}) \approx \sigma^2(y_i) + \left[\frac{1}{\sqrt{2\pi} \theta_2} \exp \left[-\frac{(x_{0i} - \theta_1)^2}{2\theta_2^2} \right] \frac{(\theta_1 - x_{0i})}{\theta_2^2} \right]^2 \sigma^2(x_i).$$

Подставив эти выражения в функционал (2.2), определим оценки параметров θ_1 и θ_2 .

В статье [71] приведены асимптотические свойства получаемых оценок. В [20, 48, 49, 71] исследуются также статистические свойства оценок и вопрос о сходимости итерационного алгоритма получения оценок.

III. В [64] предложен метод, позволяющий учитывать ошибки переменных x_i для линейных моделей.

Допустим, что задана модель с известными дисперсиями ошибок σ_ε^2 и σ_δ^2 .

$$\begin{cases} y_i = a\xi_i + b + \varepsilon_i \\ x_i = \xi_i + \delta_i, \quad i = 1 \dots n \end{cases}$$

Традиционными методами, например МНК, определяют оценки параметров a и b модели, считая, что x_i - детерминированные величины. То есть теперь мы имеем модель с известными параметрами

$$\begin{cases} y_i - \hat{b} = \hat{a}\xi_i + \varepsilon_i \\ x_i = \xi_i + \delta_i, \quad i = 1 \dots n \end{cases}$$

Или в матричном виде

$$\begin{bmatrix} y_i - \hat{b} \\ x_i \end{bmatrix} = \begin{bmatrix} \hat{a} \\ 1 \end{bmatrix} \xi_i + \begin{bmatrix} \varepsilon_i \\ \delta_i \end{bmatrix}$$

Получили регрессионную модель, где ξ_i - неизвестный параметр, подлежащий оцениванию. Оценки ξ_i определяют из выражения

$$\hat{\xi}_i = \left[(\hat{a}; 1) \cdot \sum^{-1} \cdot (\hat{a}; 1)^T \right]^{-1} (\hat{a}; 1) \cdot \sum^{-1} \cdot (y_i - \hat{b}; x_i)^T$$

где $\sum = \text{diag}(\sigma_\varepsilon^2, \sigma_\delta^2)$.

Последнее выражение можно преобразовать:

$$\hat{\xi}_i = \left[a^2 \sigma_\delta^2 + \sigma_\varepsilon^2 \right]^{-1} \left[\sigma_\delta^2 (y_i - b) a + \sigma_\delta^2 x_i \right].$$

Дисперсии оценок определяют по формулам

$$D(\hat{\xi}_i) = \left[(a; 1) \cdot \sum^{-1} \cdot (a; 1)^T \right]^{-1}$$

или

$$D(\hat{\xi}_i) = \left[a^2 \sigma_\delta^2 + \sigma_\varepsilon^2 \right]^{-1}$$

Далее уточняют оценки параметров a и b при новых значениях аргументов ξ_i . Затем снова определяют оценки ξ_i .

Итерационный процесс продолжается до тех пор, пока относительные изменения параметров на соседних итерациях не будут

меньше некоторой малой величины γ , т.е. $\max \left| \frac{\theta_j^s - \theta_j^{s-1}}{\theta_j^s} \right| \leq \gamma, \quad j = 1, 2$

IV. Рассмотрим методы конъюгентного анализа [13-16, 24]. Запишем выражение совместной плотности вероятности для статистически независимых экспериментальных данных

$$L(x, y|\theta) = \prod_{i=1}^n f_1(x_i|\theta) \cdot f_2(y_i|\theta),$$

где $f_1(x_i|\theta)$ и $f_2(y_i|\theta)$ - соответственно плотности распределения случайных величин x_i и y_i .

В случае, когда экспериментальные значения x_i и y_i случайные величины, каждая из которых имеет ф.п.в., описываемую функцией Гаусса с математическими ожиданиями ξ_i и η_i дисперсиями $\sigma^2(x_i)$ и $\sigma^2(y_i)$ и коэффициентом корреляции $\rho_i=0$, логарифм совместной плотности вероятности примет вид

$$\ln L(x, y) = -\frac{1}{2} \sum_{i=1}^n \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] + const$$

Оценки искомых параметров θ находят из условия минимума функционала

$$F = \frac{1}{2} \cdot \sum_{i=1}^n \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right]. \quad (2.3)$$

Сложность определения оценок параметров для данного случая заключается в том, что неизвестны истинные значения ξ_i , а известны лишь их доверительные интервалы. Поэтому перед тем, как найти оценки параметров θ , необходимо каким-либо образом оценить ξ_i . Истинные значения ξ_i вычисляют по условию:

$$\left. \frac{\partial F}{\partial \xi_i} \right|_{\xi_i = \hat{\xi}_i} = 0, \quad i = 1 \dots n \quad (2.4)$$

Оценки параметров находят из условия минимума функционала (2.3)

$$\left. \frac{\partial F}{\partial \eta} \cdot \frac{\partial \eta}{\partial \theta_j} \right|_{\theta_j = \hat{\theta}_j} = 0, \quad j = 1 \dots m$$

При определении ξ_i необходимо следить за тем, чтобы значения ξ_i принадлежали области неопределенности измеренных величин x_i .

При условии, что случайные величины x_i распределены по нормальному закону, ограничение примет следующий вид: $|x_i - \xi_i| \leq k\sigma(x_i)$, где k - коэффициент, который определяют исходя из условия доверия.

Таким образом, задача минимизации функционала (2.3) при условии (2.4) эквивалентна решению системы уравнений:

$$\sum_{i=1}^n \frac{y_i - \eta_i}{\sigma^2(y_i)} \cdot \frac{\partial \eta}{\partial \theta_j} = 0, \quad j = 1 \dots m \quad (2.5)$$

при

$$\frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta_i}{\sigma^2(y_i)} \cdot \frac{\partial \mathcal{F}}{\partial \xi_i} = 0, \quad i = 1 \dots n \quad (2.6)$$

Для функций, линейных по параметрам θ , система уравнений (2.5) - система линейных алгебраических уравнений. Система (2.6) для линейных по ξ функций представляет п не связанных между собой систем из т линейных уравнений.

Таким образом, практически сначала решают регрессионную задачу нахождения оценок θ при значениях $\xi_i = x_i, i = 1 \dots n$. Получают первое приближение для оценок θ . Затем определяют точные значения ξ_i , при этом проверяют принадлежность новых значений ξ_i области возможных значений переменной X . Эти действия повторяют до тех пор, пока не выполнится одно из условий:

а) на очередном шаге значение функционала (2.3) меньше заданного числа γ ,

б) на соседних итерациях значение функционала F и значения оценок параметров θ отличаются несущественно, т.е.

$$\left| \frac{F_v - F_{v+1}}{F_v} \right| \leq \gamma_1; \quad \max \left| \frac{\theta_j^s - \theta_j^{s-1}}{\theta_j^s} \right| \leq \gamma_2, \quad j = 1 \dots m,$$

где γ_1 и γ_2 - заданные числа;

в) исчерпан лимит итераций.

Дисперсии оценок параметров находят с помощью матрицы вторых производных от функционала (2.3) по θ .

В [13-16] рассмотрено применение конфлюэнтного анализа для основных элементарных функций: прямой линии, линейных функций, полиномов, систем линейных алгебраических уравнений, сигномов, кубических сплайнов.

Например, для линейной функции $\eta = a\xi + b$ получим следующий минимизируемый функционал:

$$F = \sum_{i=1}^n \frac{(y_i - ax_i - b)^2}{\sigma^2(y_i) + a^2 \sigma^2(x_i)}$$

Значения ξ_i определим по формуле

$$\hat{\xi}_i = \frac{\sigma^2(y_i)x_i + a\sigma^2(x_i)(y_i - b)}{\sigma^2(y_i) + a^2\sigma^2(x_i)}. \quad (2.7)$$

В случае, когда функция η соответствует нормальному закону, возникают трудности нахождения оценок параметров, так как функционал (2.3) имеет сложный вид. Поэтому оценки параметров и истинные значения ξ_i определяют итерационными методами [13-16].

Для линейных моделей мы получили формулу (2.7) для расчета значений ξ_i , аналогичную ранее приведенной в п. III. При определении

параметров прямой линии функционал (2.3) соответствует функционалам п. II.

Таким образом, в качестве единого подхода к задаче нахождения оценок параметров моделей с учетом погрешностей в значениях функции и аргументов могут быть использованы методы конфлюэнтного анализа, описанные в [13-16, 24].

2.2. Задача оценивания свободных параметров функций в пассивной схеме наблюдений

Пассивная схема конфлюэнтного анализа наиболее часто применяется при сборе экспериментального материала (информации). В пассивной схеме требуется найти интервальную оценку параметра θ функции $\eta=f(\xi, \theta)$, когда точные значения η и ξ определить нельзя, а вместо них измеряются случайные величины x и y , связанные с η и ξ следующим образом:

$$x_i = \xi_i + \delta_i \quad y_i = \eta_i + \varepsilon_i \quad i = 1 \dots n \quad (2.8.)$$

где δ_i и ε_i соответственно ошибки значений переменных и функции, т.е. случайные величины.

Пусть имеем статистический ряд экспериментальных значений $X = \{x_i\} \in X$ и соответствующий им ряд значений функции $Y = \{y_i\} \in Y$ где $i = 1 \dots n$; $n \geq m$, где m - число оцениваемых параметров θ . Будем считать, что переменные x_i и y_i не являются детерминированными, но являются выборками из генеральных совокупностей X и Y с известными функциями распределений. Переменные $x_i = \xi_i + \delta_i$ и $y_i = \eta_i + \varepsilon_i$ могут быть статистически зависимы или независимы, коррелированы или не коррелированы.

В основном будем иметь дело с выборками из n независимых наблюдений из одного и того же распределения. Пусть $f_1(x_i|\theta)$ и $f_2(y_i|\theta)$ соответственно плотности распределения случайных величин x_i и y_i , если x_i и y_i непрерывны, либо соответственно вероятности значений x_i и y_i , если распределения x_i и y_i дискретны; x_i и y_i и соответственно распределения $f_1(x_i|\theta)$ и $f_2(y_i|\theta)$ могут быть как одномерными, так и многомерными.

Найдем выражение для совместной плотности вероятности экспериментальных данных при условии, что ξ_i и η_i связаны функциональной зависимостью, но их погрешности δ_i и ε_i являются независимыми при переходе от одной точки (x_i, y_i) к другой. Тогда совместная плотность вероятности получить одновременно значения x_i и y_i равна

$$p_i = f_1(x_i|\theta)f_2(y_i|\theta).$$

Совместная плотность вероятности получить n статистически независимых точек (x_i, y_i)

$$L = \prod_{i=1}^n P_i = \prod_{i=1}^n f_1(x_i|\theta)f_2(y_i|\theta)$$

Аналогично можно получить формулы совместной плотности вероятности для зависимых или коррелированных экспериментальных точек. Эти формулы могут быть взяты из книг по теории вероятностей [6, 25, 29, 58, 59, 62].

Для нас важен тот факт, что в выражения для совместной плотности вероятности входят математические ожидания экспериментальных данных, экспериментальные значения и оцениваемые параметры, так как $f_1(x_i|\theta)$ функция математического ожидания ξ_i , экспериментальных значений x_i и параметров θ , $f_2(y_i|\theta)$ функция математического ожидания η_i , экспериментальных значений y_i , x_i и параметров θ . Кроме того, нам известно функциональное соотношение

$$\eta_i = f(\xi_i, \theta),$$

которое порождает структурные соотношения между наблюдаемыми случайными величинами x_i , и y_i :

$$y_i = \psi(x_i, \theta, \delta_i, \varepsilon_i), \text{ или } y_i = f(x_i - \delta_i, \theta) + \varepsilon_i$$

при аддитивных помехах (2.8).

Таким образом, в поставленной задаче следует отметить две проблемы: первая - каким образом ввести в рассмотрение погрешность аргумента; вторая состоит в том, что функционалы, которые требуется минимизировать при отыскании оценок параметров, имеют сложную форму, и соответствующие системы уравнений для определения этих же оценок нелинейны. Известно, что каждая линейная система для своего решения требует особого рассмотрения.

Наиболее часто в практике встречаются распределения Гаусса. Найдем (в качестве примера) для них вид функционалов, из которых затем могут быть получены оценки искомых параметров. Для других функций распределений экспериментальных данных процедура получения минимизируемого функционала будет аналогичной.

Пусть экспериментальные значения x_i и y_i случайные величины, каждая из которых имеет ф.п.в., описываемую функцией Гаусса соответственно с математическими ожиданиями ξ_i и η_i , дисперсиями $\sigma^2(x_i)$ и $\sigma^2(y_i)$ и коэффициентом корреляции $\rho(x_i, y_i) = \rho_i$. Тогда плотность вероятности получить точку с координатами (x_i, y_i)

$$P_i = \frac{1}{2\sigma(x_i)\sigma(y_i)\sqrt{1-\rho_i^2}} \exp \left[-\frac{u_{1i}^2 - 2\rho_i u_{2i} + u_{2i}^2}{2(1-\rho_i^2)} \right],$$

где

$$u_{1i} = \frac{x_i - \xi_i}{\sigma(x_i)}; \quad u_{2i} = \frac{y_i - \eta_i}{\sigma(y_i)}.$$

Совместная плотность вероятности получить n независимых таких точек $L = \prod_{i=1}^n P_i$ и

$$\ln L = -\frac{1}{2} \sum_{i=1}^n \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i - \xi_i)(y_i - \eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] \frac{1}{1 - \rho_i^2} + const.$$

Оценки искоемых параметров находятся из условия минимума функционала

$$F = \frac{1}{2} \sum_{i=1}^n \frac{1}{1 - \rho_i^2} \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} - 2\rho_i \frac{(x_i - \xi_i)(y_i - \eta_i)}{\sigma(x_i)\sigma(y_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] \quad (2.9)$$

Для важного частного случая, когда погрешности δ_i и ε_i некоррелированы, выражение (2.9.) примет вид

$$F = \frac{1}{2} \sum_{i=1}^n \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right] \quad (2.10)$$

Таким образом, получен явный вид одного из функционалов, которые вводятся по определению в методе обобщенной невязки [29, 30].

Аналогично получают вид минимизируемого функционала при других законах распределения исходных данных. Переход в выражениях (2.9) (2.10) к случаю, когда переменные X являются детерминированными, очевиден.

Теперь рассмотрим задачу отыскания минимума функционалов типа (2.9) (2.10) по параметрам θ при условии $\eta = f(\xi, \theta)$. В отличие от регрессионного анализа, нам неизвестны истинные значения абсцисс экспериментальных точек, а известны только их доверительные области. Причем случайная величина X коррелирована с ошибкой функции η (например, для прямой линии $\eta = \theta_1 + \theta_2 \xi$ ошибка функции η равна $(\varepsilon - \theta_2 \delta)$). Следовательно, перед тем, как приступить к определению точки минимума функционалов (2.9), (2.10) по θ требуется каким-то образом определить ξ_i и только затем, подставив выражения для ξ_i и η_i в функционал, приступить к отысканию точки минимума получившейся функции нескольких переменных.

Искомые значения ξ_i и оценки параметров θ в [13-16] определяются из условий:

$$\frac{\partial F}{\partial \xi_i} = 0, \quad i = 1 \dots n;$$

$$\frac{F}{\partial \eta} \frac{\partial \eta}{\partial \theta_j} = 0, i = 1 \dots m.$$

Очевидно, что получаемые оценки ξ_i должны принадлежать области неопределенности D_i измеренных значений x_i , т. е. $\xi_i \in D_i$. В условиях, когда известен закон распределения погрешности измерения x_i , это условие может быть выражено в более конкретной форме: при нормальном законе распределения случайной величины x_i , можно взять $|x_i - \xi_i| \leq k\sigma(x_i)$, где коэффициент k определяется выбранным уровнем доверия. Таким образом, например, задача минимизации функционала (2.10) при условии $\eta_i = f(\xi_i, \theta)$ эквивалентна решению системы уравнений

$$\sum_{i=1}^n \frac{(y_i - \eta_i)}{\sigma^2(y_i)} \frac{\partial f}{\partial \theta_j} = 0, j = 1 \dots m; i = 1 \dots n; n \geq m, \quad (2.11)$$

при

$$\frac{x_i - \xi_i}{\sigma^2(x_i)} + \frac{y_i - \eta_i}{\sigma^2(y_i)} \frac{\partial f_i}{\partial \xi_i} = 0 \quad (2.12)$$

Для функций, линейных по параметрам θ , система уравнений (2.11) - система линейных алгебраических уравнений. Система (2.12) для линейных по ξ функций представляет n не связанных между собой систем из m линейных уравнений. Замена нормального закона распределения другим, например законом Пуассона, Лапласа и др., приводит к системе нелинейных уравнений (2.11).

Учет корреляции погрешностей x_i и y_i не вносит принципиальных трудностей, а лишь усложняет формулы. Так, учет корреляции погрешности значений x_{i_1} и x_{i_2} в разных точках не позволяет разбить систему из m уравнений на n независимых. В этом случае придется решать систему линейных уравнений в n раз большей размерности, чем при независимых переменных x_{i_1} и x_{i_2} . Кроме того, во всех случаях учет корреляции погрешностей ухудшает обусловленность системы уравнений (2.11).

Условие (2.12) для нелинейных функций $\eta = f(\xi, \theta)$ дает соответственно n независимых систем из m нелинейных уравнений, которые можно решать методом линеаризации. Сходимость итерационного процесса при этом обеспечивается малостью допустимых интервалов для ξ .

В нашем случае важно и другое: условие (2.12) позволяет получить оценки "истинных" значений признаков объектов, по которым и должно приниматься решение.

2.3. О единственности оценок параметров. Состоятельность оценок и алгоритм их получения.

Самый простой функционал (2.10) нелинеен по θ даже в случае оценки параметров прямой линии. Поэтому встает вопрос о единственности оценок искомых параметров, который тесно связан с выбором нулевого приближения. Естественно взять в качестве нулевого приближения решение регрессионной задачи (например, по методу наименьших квадратов).

Из условия (2.12) следует, что для одной и той же задачи минимальное значение функционала в методе наименьших квадратов больше минимального значения функционала (2.10), поскольку условие (2.12) - условие ортогональности вектора касательной к функции $\eta=f(\xi, \theta)$, в точке, принимаемой за "истинную", и вектора, проведенного через исходную точку $(x_{i_1}, \dots, x_{i_m}, y_i)$ и точку, принимаемую за "истинную" Таким образом, минимизация функционала (2.10) при условии (2.12) приводит к минимизации суммы квадратов "наикратчайших расстояний" от точки $(x_{i_1}, \dots, x_{i_m}, y_i)$ до кривой, тогда как в методе наименьших квадратов минимизируется сумма квадратов отклонений при фиксированных значениях x_{ij} .

Для конкретных видов функции $\eta=f(\xi, \theta)$ в работах [13, 16, 24] получены условия существования единственного решения. Покажем существование единственного решения для линейных функций

$$\eta = \sum_{j=1}^m \theta_j x_j + \theta_0 \quad \text{в предположении, что 1) погрешности } \varepsilon \text{ и } \delta$$

статистически независимы и подчиняются нормальному закону распределения; 2) за нулевое приближение взято решение регрессионной задачи [13, 16]. Эти ограничения обусловлены тем, что с помощью преобразований координат функция Гаусса, описывающая нормальный закон распределения, сводится к уравнению прямой линии и оценки параметров этой прямой будут однозначно связаны с оценками параметров функции Гаусса.

Теорема 2.1. Если погрешности ε и δ измеренных случайных величин X и Y статистически независимы и распределены по нормальному закону с известными числовыми характеристиками ф.п.в. и если в качестве нулевого приближения принято решение соответствующей регрессионной задачи, то достаточным условием единственности оценок параметров линейной функции

$$\eta = \theta_0 + \sum_{j=1}^m \theta_j \xi_j \quad \text{является выполнение неравенства } \overline{x^2} > (1 - m/n) \hat{\sigma}^2(y)$$

Здесь $\hat{\sigma}^2(y)$ оценка дисперсии значений y_i при равноточных измерениях.

Для простоты формул $\sigma(x_{ij})$ и $\sigma(y_i)$ приняты равными единице. Функционал (2.10) с учетом условия (2.12) для линейных функций имеет вид

$$F = \frac{1}{2} \frac{\sum_{i=1}^n \left(y_i - \theta_0 - \sum_{j=1}^m \theta_j x_{ij} \right)^2}{1 + \sum_{j=1}^m \theta_j^2}. \quad (2.13)$$

При фиксированных θ_j ($j=1...m$) функционал (2.13) относительно переменной θ_0 есть парабола и любое начальное значение θ_{00} приведет к единственной оценке. С другими переменными θ_j ($j=1...m$) дело обстоит сложнее. Функционал (2.13) монотонно растет с увеличением $(\theta_j - \hat{\theta}_j)$ от

точки минимума вправо до горизонтальной асимптоты, равной $\sum_{i=1}^n x_{ij}^2$, а влево - до точки максимума, а затем монотонно "опускается" сверху до той же горизонтальной асимптоты. Для доказательства теоремы достаточно указать, когда соответствующий регрессионный функционал

$\Phi_p < \sum_{i=1}^n x_{ij}^2$ В точке минимума $\Phi_p = (n-m)\hat{\sigma}^2(y)$ Для единственности

оценок θ_j должно выполняться условие $\Phi_p = (n-m)\hat{\sigma}^2(y) < \sum_{i=1}^n x_{ij}^2 = \overline{x^2} n$

или $\overline{x^2} > \left(1 - \frac{m}{n}\right) \hat{\sigma}^2(y)$ В нашем случае дисперсии значений y , известны и могут быть использованы для проверки достаточного условия.

Можно сформулировать и более строгие условия, когда значение Φ_p в точке минимума будет под асимптотой и в то же время точка минимума Φ_p будет принадлежать строго выпуклой области функционала (2.13). Для этого следует проверить любое достаточное условие строгой выпуклости функционала (2.13) в точке минимума регрессионного функционала.

Условия строгой выпуклости функционала F , а тем самым и единственность его точки минимума [25, 28] устанавливаем согласно теоремам [28], которые гласят, что достаточным условием строгой выпуклости функционала F является положительная определенность

матрицы $M = \left\| \frac{\partial^2 F}{\partial \theta_1 \partial \theta_r} \right\|$, а для того чтобы матрица M была положительно

определена, необходимо и достаточно, чтобы каждый из определителей

$\det \left\| \frac{\partial^2 F}{\partial \theta_j^2} \right\|, \dots, \det \left\| \frac{\partial^2 F}{\partial \theta_1 \partial \theta_r} \right\|$ был положителен.

Сформулируем достаточные условия того, что регрессионное решение принадлежит области строгой выпуклости функционала F (2.13) в случае оценки свободных параметров прямой линии.

Теорема 2.2. Достаточным условием того, что регрессионное решение при оценке параметров прямой линии $\eta = \theta_0 + \theta_1 \xi$ принадлежит строго выпуклой области функционала (2.13) является

$$\sum_i x_i^2 - \frac{1}{n} \left(\sum_i x_i \right)^2 > (n-2) \hat{\sigma}^2(y).$$

Действительно, определитель $\det \left\| \frac{\partial^2 F}{\partial \theta_0^2} \right\|$ положителен всегда.

Определитель $\det \left\| \frac{\partial^2 F}{\partial \theta_0 \partial \theta_1} \right\|$ в точке минимума регрессионного функционала, когда $\sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i) = 0$ и $\sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i) x_i = 0$ равен

$$(1 + \theta_1^2) \left[n \sum_i x_i^2 - \left(\sum_i x_i \right)^2 \right] + n(3\theta_1^2 - 1) \sum (y_i - \theta_0 - \theta_1 x_i)^2$$

Это выражение положительно для любого θ_0 и $|\theta_1| > 1/\sqrt{3}$. Для $|\theta_1| \leq 1/\sqrt{3}$ функционал будет выпуклым, если

$$(1 + \theta_1^2) \left[n \sum_i x_i^2 - \left(\sum_i x_i \right)^2 \right] > \left| n(3\theta_1^2 - 1) \sum (y_i - \theta_0 - \theta_1 x_i)^2 \right|.$$

Заменим величину $\sum_i (y_i - \theta_0 - \theta_1 x_i)^2$ на $(n-2) \hat{\sigma}^2(y)$. Тогда

$$n \sum_i x_i^2 - \left(\sum_i x_i \right)^2 > \left[n(n-2)(3\theta_1^2 - 1) \hat{\sigma}^2(y) \right] / (1 + \theta_1^2).$$

Правая часть последнего условия имеет максимум при $\theta_1 = 0$, т.е. для любого θ_1 условие того, что регрессионное решение принадлежит участку выпуклости функционала (2.13), имеет вид

$$\sum_i x_i^2 - \frac{1}{n} \left(\sum_i x_i \right)^2 > (n-2) \hat{\sigma}^2(y)$$

Условия, обеспечивающие попадание регрессионного решения в выпуклую область функционала (2.10), для других функций несложно проверить на ЭВМ, но достаточно громоздко выразить в явном виде. Методом статистических испытаний было показано, что при оценке

параметров нормального закона распределения (от одномерного до пятимерного) без применения каких либо преобразований координат, итерационный процесс сходился, если нулевой приближение отличалось от истинного значения не более чем на 40% при относительных погрешностях наблюдаемых значений признаков $\approx 10\%$. Такую точность получения нулевых приближений обеспечивает регрессионный анализ.

Для функций $\eta=f(\xi, \theta)$, линейных по параметрам θ , единственность оценки и сходимость нулевого приближения к оцениваемым параметрам при нормально распределенных исходных данных определяются только способом нахождения $\hat{\xi}$ так как при фиксированных $\hat{\xi}$ система уравнений (2.11) система линейных алгебраических уравнений с детерминантом, не равным нулю. Для этих функций из условия (2.12), в принципе, может быть найден не единственный набор значений $\hat{\xi}$ каждый из которых принадлежит области возможных значений при заданном наборе значений $\{x_i\}$. В такой ситуации у нас нет оснований отдавать предпочтение какому-либо набору значений $\hat{\xi}$ и, строго говоря, решение будет не единственно.

Рассмотрим вопрос о состоятельности оценок θ в конъюнктном анализе, если оценки $\hat{\xi}$ находятся из условия (2.12).

Теорема 2.3. Если оценки параметров прямой $\eta=\theta_0+\theta_1\xi$ находятся как координаты точки минимума функционала (2.13), то они равны

$$\hat{\theta}_0 = \frac{1}{n} \sum (y_i - \hat{\theta}_1 x_i) \text{ и } \hat{\theta}_1 = \max(\hat{\theta}_{11}, \theta_{12i}),$$

где

$$\hat{\theta}_{11,12} = \left[(w - d) \pm \sqrt{(d - w)^2 + 4c^2} \right] / 2c;$$

$$c = \sum_i x_i y_i - \frac{1}{n} \sum_i x_i \sum_i y_i; \quad w = \frac{1}{n} \left(\sum_i x_i \right)^2 \quad d = \sum_i x_i^2 - \sum_i y_i^2 + \frac{1}{n} \left(\sum_i y_i \right)^2$$

Для доказательства решим систему алгебраических уравнений, обеспечивающую необходимое условие экстремума функционала (2.13):

$$\begin{cases} \partial F / \partial \theta_0 = 0; \\ \partial F / \partial \theta_1 = 0. \end{cases}$$

Из первого уравнения следует $\hat{\theta}_0 = \frac{1}{n} \sum (y_i - \theta_1 x_i)$. Исключив из полученной системы уравнений θ_0 получим для оценки $\hat{\theta}_1$ квадратное уравнение $c\theta_1^2 + (d - w)\hat{\theta}_1 - c = 0$, корни которого и есть $\hat{\theta}_{11}$ и $\hat{\theta}_{12}$. В данном случае нет необходимости применять достаточные признаки,

чтобы определить, какой корень отвечает точке минимума функционала (2.13). Из сечений функционала F видно, что меньший корень отвечает точке максимума, а больший - точке минимума.

Теорема 2.4. Если оценки параметров прямой $\eta = \theta_0 + \theta_1 \xi$ находятся как координаты точки минимума функционала (2.13), то они состоятельны.

Докажем состоятельность оценки $\hat{\theta}_1$. Обозначим S_x^2 , S_y^2 и S_{xy} соответственно выборочные дисперсии $\{x_i\}$, $\{y_i\}$ и выборочную ковариацию величин x_i и y_i . Тогда

$$w - d = \frac{1}{n} \left(\sum_i x_i \right)^2 - \sum_i x_i^2 + \sum_i y_i^2 - \frac{1}{n} \left(\sum_i y_i \right)^2 = S_y^2 - S_x^2;$$

$$c = \sum_i x_i y_i - \frac{1}{n} \sum_i y_i \sum_i x_i = S_{xy}$$

В соответствии с результатами гл. 10 в [24] выборочные дисперсии S_x^2 , S_y^2 и выборочная ковариация S_{xy} сходятся по вероятности к своим математическим ожиданиям

$$S_x^2 \xrightarrow{n \rightarrow \infty} S_\xi^2 + \sigma^2(x) = S_\xi^2 + 1;$$

$$S_y^2 \xrightarrow{n \rightarrow \infty} \theta_1^2 S_\xi^2 + \sigma^2(y) = \theta_1^2 S_\xi^2 + 1;$$

$$S_{xy} \xrightarrow{n \rightarrow \infty} \theta_1 S_\xi^2,$$

где S_ξ^2 - дисперсия ненаблюдаемой величины ξ . Отсюда

$$\theta_{11,12} = \frac{S_\xi^2(\theta_1^2 - 1) \pm \sqrt{S_\xi^4(\theta_1^2 - 1)^2 - 4S_\xi^4\theta_1^2}}{2\theta_1 S_\xi^2} = \frac{(\theta_1^2 - 1) \pm (\theta_1^2 + 1)}{2\theta_1};$$

$$\hat{\theta}_{11} = \theta_1; \quad \hat{\theta}_{12} = -\frac{1}{\theta_1}$$

Видим, что оценка $\hat{\theta}_{11}$, соответствующая точке минимума функционала (2.13), состоятельна. Докажем теперь состоятельность оценки $\hat{\theta}_0$

$$\hat{\theta}_0 = \frac{\sum_i y_i}{n} - \hat{\theta}_1 \frac{\sum_i x_i}{n} = \bar{y} - \hat{\theta}_1 \bar{x} \rightarrow \bar{y} - \theta_1 \bar{x}$$

Согласно [24] $\bar{y} \rightarrow \theta_0 + \theta_1 \xi$; $\bar{x} \rightarrow \xi$,

т.е. $\hat{\theta}_0 = \theta_0 + \theta_1 \xi - \theta_1 \xi = \theta_0$ - оценка состоятельна.

Если в качестве $\hat{\xi}_i$ берутся значения x_i , то оценка $\hat{\theta}_1$ не будет состоятельной. Для простоты возьмем случай, когда $\theta_0 = 0$. Тогда

$$\hat{\theta}_l = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{S_{xy}}{S_x^2} \xrightarrow{n \rightarrow \infty} \frac{\theta_l \cdot S_\xi'^2}{S_\xi^2 + 1} = \theta_l \left(1 - \frac{1}{S_\xi^2 + 1} \right) \quad \text{оценка}$$

несостоятельна.

Следствие 2.1. Если оценки параметров θ в уравнении $\eta = \theta_0 + \sum_{j=1}^m \theta_j \xi_j$ находятся как координаты точки минимума функционала (2.10) и оценки ξ находятся из условия (2.12), то полученные оценки $\hat{\theta}$ будут состоятельными.

Следствие 2.2. В условиях теоремы 2.4 нулевое приближение от истинного значения оценок искомых параметров может отличаться до 100%.

Действительно, согласно теореме 2.4 разность оценок параметра θ_l составляет $\theta_l + \frac{1}{\theta_l} = \frac{\theta_l^2 + 1}{\theta_l} > \theta_l$.

Учитывая свойства функционала F и получающихся оценок, задачу минимизации функционала F при условии (2.12) будем решать по следующей схеме

Шаг 1. Задаются начальные значения переменных ξ (результаты наблюдений X) и определяется минимум по θ функционала F при фиксированных ξ . В результате получают первое приближение для оценок искомых параметров.

Шаг 2. Точные значения переменных пересчитываются с учетом ограничения (2.12) при полученных приближениях для оценок параметров θ . Для уменьшения числа итераций полезно проверять условие принадлежности нового значения ξ_{ij} области возможных значений при заданном x_{ij} .

Шаг 3. Минимизируют функционал F по θ при новых точных значениях переменных ξ .

Шаги 2 и 3 выполняются поочередно до тех пор, пока не будет выполнено хотя бы одно из трех условий:

на очередном шаге значение функционала (2.10) меньше заданного числа ε_1 ,

на соседних итерациях значения функционала F и оценок параметров θ отличаются незначительно, т.е.

$$\left| \frac{F_v - F_{v+1}}{F_v} \right| \leq \delta_1 \quad \text{и} \quad \max \left| \frac{\theta_l^{(v)} - \theta_l^{(v+1)}}{\theta_l^{(v)}} \right| \leq \delta_2, \quad l = 1 \dots k$$

где δ_1 и δ_2 - заданные числа;
исчерпан заданный лимит итераций.

Шаг 4. После нахождения оценок $\hat{\theta}$ определяют дисперсии оценок.

Из приведенного алгоритма следует, что на первой итерации получается регрессионное решение (иными словами, решение любым из применявшихся ранее традиционных методов). На последующих шагах это решение уточняется. Такой переход от традиционных методов к предлагаемому имеет принципиальное значение в практических приложениях.

2.4. Статистические свойства свободных параметров функции Гаусса, определенных через параметры уравнения прямой линии и непосредственно.

В приложениях часто функцию Гаусса $y = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$ путем

преобразования координат сводят к уравнению прямой $u = \frac{1}{\sigma}x - \frac{a}{\sigma}$,

где u_p - p -квантиль случайной величины X с функцией вероятности $F(x)$, т.е значение аргумента u_p функции $F(x)$, для которого вероятность события $X < u_p$ равна заданному значению p .

Точка пересечения X_L прямой с осью абсцисс дает оценку параметра

a , а разность абсцисс $u = \frac{x-a}{\sigma}$ точки $N(x_N, -1)$ построенной прямой с $u = -1$ и точки X_L определяет оценку σ .

Таким образом, рассматривая уравнение прямой в виде $\eta = \theta_1 + \theta_2 \xi$,

оценка θ_1 будет определять оценку $-\frac{a}{\sigma}$, а оценка $\theta_2 - \frac{1}{\sigma}$.

При определении свободных параметров θ_1 и θ_2 уравнение прямой линии часто записывают в виде $\eta = \theta_1 + \theta_2(\xi - \bar{\xi})$. Оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ статистически независимы, что несправедливо для оценок $\hat{\theta}_1$ и $\hat{\theta}_2$.

Оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ параметров можно получить, не решая совместных систем связанных уравнений, как это приходится делать при другой форме записи модели.

Для нормально распределенных статистически независимых погрешностей экспериментальных точек (x_i, y_i) минимизируемый функционал имеет следующий вид

$$F = \sum_{i=1}^n \left[\frac{(x_i - \xi_i)^2}{\sigma^2(x_i)} + \frac{(y_i - \eta_i)^2}{\sigma^2(y_i)} \right].$$

Из условий (2.11) и (2.12) получим

$$\xi_i = \left[\sigma^2(y_i)x_i + \theta_2 \sigma^2(x_i)(y_i - \theta_1) \right] / \left[\sigma^2(y_i)x_i + \theta_2^2 \sigma^2(x_i) \right].$$

Тогда

$$F = \sum_{i=1}^n \frac{(y_i - \theta_1 - \theta_2 x_i)^2}{\sigma^2(y_i) + \theta_2^2 \sigma^2(x_i)} = \sum_{i=1}^n w(y_i - \theta_1 - \theta_2 x_i)^2$$

где $w^{-1} = \sigma^2(y_i) + \theta_2^2 \sigma^2(x_i)$

Ковариационную матрицу оценок $D(\hat{\theta})$ найдем как матрицу M' , обратную матрице M , элементы которой есть

$$-\frac{\partial^2 F}{\partial \theta_1^2}; \quad -\frac{\partial^2 F}{\partial \theta_1 \partial \theta_2} = -\frac{\partial^2 F}{\partial \theta_2 \partial \theta_1}; \quad -\frac{\partial^2 F}{\partial \theta_2^2},$$

вычисленные при найденных значениях оценок $\hat{\theta}'_1$ и $\hat{\theta}'_2$

После определения оценок $\hat{\theta}'_1$ и $\hat{\theta}'_2$ и их дисперсий $D(\hat{\theta}'_1)$ и $D(\hat{\theta}'_2)$

приступим к нахождению доверительных интервалов. Если число экспериментальных точек достаточно велико и можно считать, что оценки параметров распределены относительно их математических ожиданий по нормальному закону, то для получения доверительных интервалов применяют безразмерную t -статистику Стьюдента, которая подчиняется t -распределению с $\nu = n - 1$ степенями свободы.

Пусть уровень значимости выбран α . Тогда $t = (\hat{\theta}'_1 - \theta_1) / \sqrt{D(\hat{\theta}'_1)}$

и $100(1 - \alpha)\%$ - ный доверительный интервал для $\hat{\theta}'_1$

$$\hat{\theta}'_1 - t_{1-\alpha/2} \sqrt{D(\hat{\theta}'_1)} \leq \theta_1 < \hat{\theta}'_1 + t_{1-\alpha/2} \sqrt{D(\hat{\theta}'_1)}$$

Аналогично доверительный интервал для θ_2 .

$$\hat{\theta}'_2 - t_{1-\alpha/2} \sqrt{D(\hat{\theta}'_2)} \leq \theta_2 < \hat{\theta}'_2 + t_{1-\alpha/2} \sqrt{D(\hat{\theta}'_2)}, \text{ т. к. оценки } \hat{\theta}'_1 \text{ и } \hat{\theta}'_2$$

не коррелированы.

Интервальные границы в случае прямой линии для $M(\eta | \xi_0)$ будут

$$\hat{\eta} - t_{1-\alpha/2} \sigma / (\eta) \leq M(\eta | \xi_0) < \hat{\eta} + t_{1-\alpha/2} \sigma / (\eta),$$

где $\sigma^2(\hat{\eta}) = (1, \xi_0) D(\hat{\theta}) (1, \xi_0)^T$

Огибающая семейства всех возможных прямых будет кривой второго порядка, уравнение которой получим из последней формулы

$$\sum_{i=1}^n (y_i - \theta_1 - \theta_2 x_i)^2 \leq C_v (1 + \theta_2^2).$$

Принимая за начало отсчета наблюдаемые средние значения x и y , перепишем это неравенство в виде [24]

$$\theta_2^2 (S_x^2 - C) - 2\theta_2 S_{xy} + \theta_1^2 \leq C_v - S_y^2$$

где S_x^2 и S_y^2 и S_{xy} - соответственно выборочные дисперсии x , y и выборочная ковариация величин x и y .

Это условие можно считать ограничением, которому удовлетворяет истинная прямая. Найдем огибающую семейства всех возможных прямых. Уравнение огибающей [24]

$$(y - \theta_2 x)^2 / (C_v - b_1) - (\hat{\theta}_2 y + x)^2 / (b_2 - C_v) = 1 + \hat{\theta}_2^2$$

где

$$b_1 = S_x^2 - S_{xy} / \hat{\theta}_2; \quad b_2 = S_x^2 + \hat{\theta}_2 S_{xy}; \quad \hat{\theta}_2 > 0; \quad b_2 > b_1$$

Программы для получения оценок в конъюнктном анализе свободных параметров прямой линии функций Гаусса и других функций приведены в [13].

Знание статистики получаемых оценок необходимо для дальнейшего анализа экспериментальных данных и правильной трактовки результатов обработки, а также для выявления последствий сделанных допущений. Приведем результаты числовых экспериментов по выявлению закона распределения оценок параметров θ_1 и θ_2 для функции Гаусса.

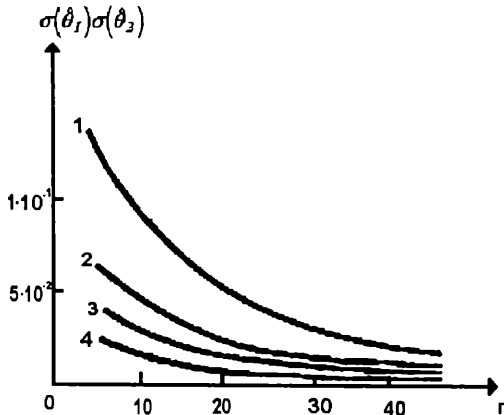


Рис.2.1 Зависимость среднеквадратических отклонений оценок параметров функции Гаусса от числа точек n :

- 1 - $\sigma(\hat{\theta}_1)$ при $\delta x=10\%$, $\delta y=10\%$; 2 - $\sigma(\hat{\theta}_1)$ при $\delta x=1\%$, $\delta y=5\%$; 3 - $\sigma(\hat{\theta}_2)$ при $\delta x=10\%$, $\delta y=10\%$; 4 - $\sigma(\hat{\theta}_2)$ при $\delta x=1\%$, $\delta y=5\%$

Для различных погрешностей по X и Y исследовалась зависимость среднеквадратических отклонений оценок параметров функции Гаусса, соответствующей прямой линии $\eta=2+\xi$, от числа заданных точек на

интервале $[0,2; 9,8]$ и закона их распределения. Как и следовало ожидать из общих соображений, группировка точек к концам прямой приводит к уменьшению дисперсии оценок. Зависимость среднеквадратических отклонений от числа точек приведена на рис. 2.1.

Затем для различного числа экспериментальных точек (от 6 до 50) и различных комбинаций погрешностей по обеим осям координат ($\delta(x_i) = 1 \div 10\%$; $\delta(y_j) = 1 \div 10\%$) с помощью датчика нормально распределенных случайных чисел исследовался вид функции распределения оценок. На рис. 2.2. показаны гистограммы распределения значений оценки θ_1 и θ_2 для разного числа экспериментальных точек, имевших погрешность 10% по обеим осям координат.

В каждом варианте вычислялись также дисперсии оценок и определялись доверительные интервалы дисперсий. Приведем результаты, полученные для 17 экспериментальных точек. Как показал анализ по критерию χ^2 распределения оценок параметров с вероятностью порядка 0,80, можно считать нормальными со среднеквадратическими отклонениями $\sigma(\hat{\theta}_1) = 0,060$; $\sigma(\hat{\theta}_2) = 0,017$.

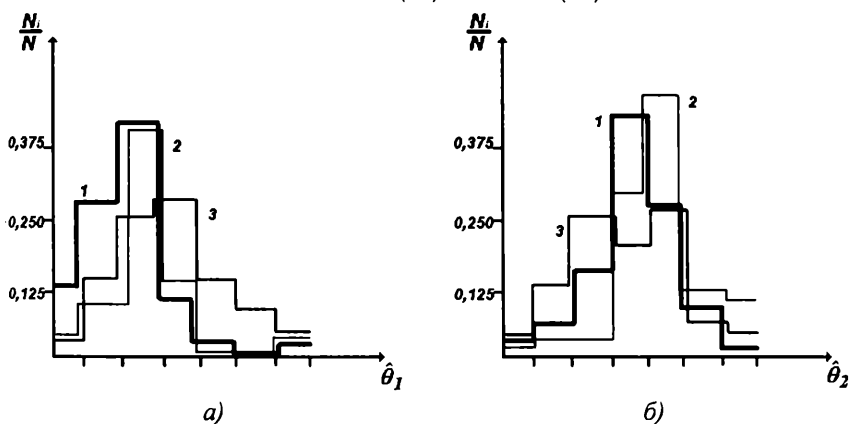


Рис.2.2. Гистограммы распределения оценок параметров функции Гаусса для различного числа экспериментальных точек при погрешности 10% по обеим осям координат: а - распределение параметра θ_1 , б - распределение параметра θ_2 . Число точек 1- 49, 2 - 17, 3 - 6.

Процедура линеаризации имеет и свои недостатки. При обработке нелинейных функций таким путем следует иметь в виду, что, для того чтобы оценки $\hat{\theta}$ соответствующих параметров θ , полученные из преобразованного уравнения, обладали оптимальными свойствами (несмещенностью, минимальной дисперсией и т.д.), необходимо, чтобы предположение об аддитивной ненаблюдаемой случайной ошибке было справедливым для преобразованной, а не для первоначальной модели.

Влияние аддитивной ошибки в преобразованной и исходных моделях можно установить, только исследуя каждую конкретную модель.

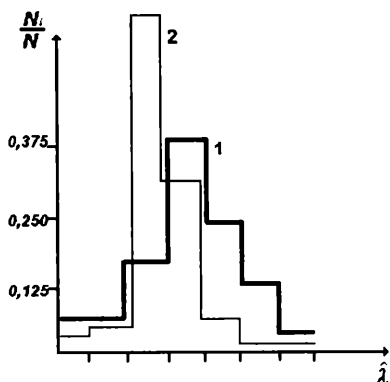


Рис.2.3. Относительная частота оценок параметра σ при применении операции линеаризации:
 1-0,980; 0,005 (начало отсчета и шаг соответственно); 2-0,920; 0,027; $\delta x = \delta y = 10\%$

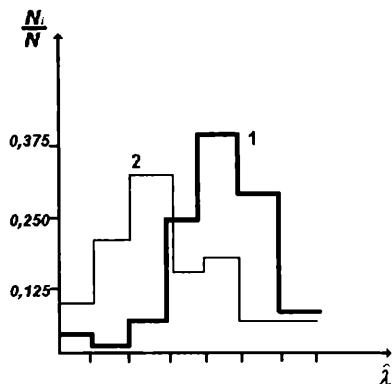


Рис.2.4. Относительная частота появления оценок параметра σ непосредственной оценке
 1-0,709; 0,065 (начало отсчета и шаг); 2-0,653; 0,109, % $\delta x = \delta y = 10\%$

На рис. 2.3 и 2.4 приведены гистограммы относительных частот появления значений параметра σ функции Гаусса при обработке экспериментальных данных непосредственно и с помощью операции линеаризации.

Исходные данные после линеаризации соответствовали прямой линии $\eta = 2 + \xi$. Относительная среднеквадратическая погрешность в исходных данных составляла 10%. Значения ξ принадлежали интервалу $[0,2; 9,8]$. Гистограммы получены для 6, 17 и 49 экспериментальных точек.

Параллельно в каждом варианте определялись дисперсии оценок в предположении, что справедлива квадратичная аппроксимация минимизируемой функции, и относительная частота появления оценок дисперсий, а также находилось среднее значение $\sigma(\hat{\sigma})$. Интервальные оценки при доверительной вероятности 0,68 и средние значения $\sigma(\hat{\sigma})$ при различном числе экспериментальных точек равны:

1) для операции линеаризации

$$0,972 \leq \hat{\sigma}_6 < 1,062; 0,030 \leq \sigma(\hat{\sigma}_6) < 0,100; \overline{\sigma(\hat{\sigma}_6)} = 0,040$$

$$0,987 \leq \hat{\sigma}_{17} < 1,010; 0,008 \leq \sigma(\hat{\sigma}_{17}) < 0,031; \overline{\sigma(\hat{\sigma}_{17})} = 0,018$$

$$0,997 \leq \hat{\sigma}_{49} < 1,008; 0,003 \leq \sigma(\hat{\sigma}_{49}) < 0,019; \overline{\sigma(\hat{\sigma}_{49})} = 0,006$$

2) непосредственная оценка

$$0,762 \leq \hat{\sigma}_6 < 1,092; 1,174 \leq \sigma(\hat{\sigma}_6) < 0,984; \overline{\sigma(\hat{\sigma}_6)} = 0,992$$

$$0,939 \leq \hat{\sigma}_{17} < 1,070; 0,124 \leq \sigma(\hat{\sigma}_{17}) < 0,312; \overline{\sigma(\hat{\sigma}_{17})} = 0,154$$

Индексы у параметра σ определяют число точек.

Подобные числовые эксперименты проводились для различных исходных данных и при различных комбинациях погрешностей от 1 до 10%.

В табл. 2.1 приведены относительные изменения дисперсии оценки математического ожидания в нормальном законе распределения при различных комбинациях относительной погрешности признака δx_i и частоты δn_{0i} .

Т а б л и ц а 2.1.

Относительное изменение дисперсии оценки математического ожидания в нормальном законе распределения при различных комбинациях относительной погрешностей признака δx_i и частоты δn_{0i}

δn_{0i} %	δx_i , %				
	0,5	1,5	2,5	3,0	5,0
0,5	1	1,078	1,083	1,084	1,089
1,5	5,467	9,064	9,557	9,655	9,753
3,0	9,310	29,16	35,02	36,25	38,27
5,0	10,29	42,96	57,14	60,59	66,04

Важен и другой результат. Разные алгоритмы приводят к различным интервальным оценкам. Наиболее оптимальные оценки должны получаться в тех алгоритмах, в которых операции ведутся непосредственно с исходными экспериментальными данными, когда не нарушается принцип аддитивности и не изменяется искусственно закон ее распределения.

Большое практическое значение имеет положение экспериментальных точек на исследуемом интервале возможных значений. Оптимальный закон их распределения (в смысле минимума дисперсии оценок) зависит от конкретных значений экспериментальных величин и их погрешностей, и без упрощающих предположений о последних получить его в явном виде не удастся. Расчеты показали, что при одинаковых относительных ошибках в каждой точке для всех рассмотренных алгоритмов предпочтительнее выбрать равномерный шаг по оси ординат, чем абсцисс. В этом случае дисперсии оценок будут минимальными.

2.5 Оценка параметров функции плотности вероятности с учетом погрешности измерения вектора признаков

Приложим полученные результаты к статистической задаче распознавания образов.

Рассмотрим первый этап задачи классификации. Пусть нам известен априори вид функции $P(\xi, \theta | \omega_j)$, $j = 1 \dots m$, но неизвестны значения вектора параметров θ . По результатам наблюдений случайных векторов

x_1, \dots, x_n требуется найти оценки свободных параметров θ и дисперсии оценок $D(\theta)$. Проблема здесь состоит в том, что наблюдаемые значения x_1, \dots, x_n имеют соответствующие погрешности и аппроксимирующая функция $P(\xi, \theta | \omega_j)$ не строго пройдет по наблюдаемым точкам, т.е. в каждой наблюдаемой точке образуется область неопределенности, форма которой определяется законами распределения координат наблюдаемой точки. Для получения оценок параметров и их дисперсий в данной ситуации применим изложенные методы конfluентного анализа [16]. В названных ранее традиционных методах классификации [3-8, 10-12] определялись только точечные оценки параметров в законах распределения, причем погрешности наблюдаемых случайных величин (векторов) x_1, \dots, x_n при этом не учитывались.

Будем считать, что координаты $\{x_{ik}\}$ наблюдений случайных векторов x_1, \dots, x_n для j -го класса образов непрерывные случайные величины с известными функциями плотностей вероятностей $f(x_{ik})$; $i=1 \dots n$; $k=1 \dots l$. Также будем считать, что в процессе аппроксимации результатов наблюдений x_1, \dots, x_n с помощью функции известного вида $P(\xi, \theta | \omega_j)$ предварительно вычислены (способом группировки, например) эмпирические значения вероятностей $P(x_{ik} | \omega_j)$; $i=1 \dots n$; $k=1 \dots l$, с которыми сравниваются соответствующие значения $P(\xi_{ik} | \theta | \omega_j)$; $i=1 \dots n$; $k=1 \dots l$. Определен также вид функции плотности вероятности $\varphi(p_{ik} | \theta)$ для $P(x_{ik} | \omega_j)$. Тогда, согласно методу максимума правдоподобия, функция правдоподобия будет иметь вид [15, 16]

$$L(X, p|\theta) = \prod_{i=1}^n \prod_{k=1}^l \varphi(p_{ik} | \theta) f(x_{ik}).$$

Положим, что здесь все случайные величины подчиняются многомерному гауссовому распределению соответственно с математическими ожиданиями $P(\xi_{ik}, \theta | \omega_j)$ и ξ_{ik} ковариационной матрицей K и дисперсиями $\sigma^2(x_{ik})$. В этом случае

$$\ln L(X, p|\theta) = -\frac{1}{2} \left\{ \left[P - P(\xi, \theta | \omega_j) \right]^T K^{-1} \left[P - P(\xi, \theta | \omega_j) \right] + \sum_{i=1}^n \sum_{k=1}^l \frac{(x_{ik} - \xi_{ik})^2}{\sigma^2(x_{ik})} \right\} + const, \quad (2.14)$$

где P матрица с элементами $P(x_{ik} | \omega_j)$; $i=1 \dots n$; $k=1 \dots l$; $P(\xi, \theta | \omega_j)$ - матрица с элементами $P(\xi_{ik}, \theta | \omega_j)$; $i=1 \dots n$; $k=1 \dots l$.

Оценки параметров θ функции $P(\xi, \theta | \omega_j)$ и их дисперсии $D(\theta)$ найдем из условия максимума (2.14) или как точку минимума функционала

$$F(X, p|\theta) = \left[P - P(\xi, \theta | \omega_j) \right]^T K^{-1} \left[P - P(\xi, \theta | \omega_j) \right] + \sum_{i=1}^n \sum_{k=1}^l \frac{(x_{ik} - \xi_{ik})^2}{\sigma^2(x_{ik})}. \quad (2.15)$$

Функционал (2.15) можно было бы сразу записать как функционал ортогональной регрессии [15, 16], позволяющий учесть как погрешность в определении P , так и погрешности измерений x_1, \dots, x_n . В ортогональной регрессии минимизируется сумма квадратов расстояний от наблюдававшихся точек до линии (поверхности) регрессии.

Точечные оценки вектора параметров θ в и неизвестных ξ_{ik} можно найти из условий:

$$\frac{\partial F}{\partial \theta_v} \Big|_{\theta = \hat{\theta}} = 0; \quad v = 1 \dots N, \quad (2.16)$$

$$\left[\frac{p_{ik} - p(\hat{\xi}_{ik}, \theta | \omega_j)}{\sigma^2(p_{ik})} \right] p'_{\xi_{ik}}(\hat{\xi}_{ik}, \theta | \omega_j) + \frac{x_{ik} - \hat{\xi}_{ik}}{\sigma^2(x_{ik})} = 0, \quad i = 1 \dots n; \quad k = 1 \dots l, \quad (2.17)$$

где $\sigma^2(p_{ik})$ - дисперсия p_{ik} , $p'_{\xi_{ik}}$ - производная от функции $p(\xi_{ik}, \theta | \omega_j)$ по ξ_{ik} . Решая системы уравнений (2.16) и (2.17) совместно, находим точечные оценки θ и ξ . Условия (2.17) система алгебраических уравнений относительно ξ_{ik} из которой можно найти явные выражения для ξ_{ik} подставить ξ_{ik} в (2.15) и получить более простую форму функционала (2.15).

Для прямой $p = a + bx$ функционал (2.15) при $\sigma^2(p_{ik}) \equiv \sigma^2(p)$ и $\sigma^2(x_{ik}) \equiv \sigma^2(x)$, $i = 1 \dots n$; $k = 1 \dots l$ примет вид:

$$F(\bar{x}, p | \theta) = \frac{1}{\sigma^2(p) + b^2 \sigma^2(x)} \sum_{i=1}^n (p_i - a - bx_i)^2 \quad (2.18)$$

для гиперплоскости - $p = a + \sum_{k=1}^l b_k x_k$ при тех же предположениях -

$$F(x, p | \theta) = \frac{1}{\sigma^2(p) + \sum_{k=1}^l b_k^2 \sigma^2(x_k)} \sum_{i=1}^n \left(p_i - a - \sum_{k=1}^l b_k x_k^2 \right) \quad (2.19)$$

Функционалы (2.18) и (2.19) позволяют непосредственно (согласно (2.16)) найти оценки свободных параметров θ и их дисперсий $D(\theta)$. Дисперсионную матрицу оценок свободных параметров θ находим из следующего условия [15, 16]:

$$D(\hat{\theta}) = M^{-1} \quad M = \left(- \frac{\partial^2 F}{\partial \theta_{j_1} \partial \theta_{j_2}} \right) \Big|_{\theta = \hat{\theta}}; \quad j_1, j_2 = 1 \dots N$$

Пользуясь полученными формулами, покажем теперь, что линейная разделяющая функция, предлагаемая в традиционных методах для двух образов, для которых $P(X | w_i)$, $i = 1 \dots 2$, - функция двумерного гауссова распределения с математическими ожиданиями (x_i, y_i) , $i = 1 \dots 2$, и одной

ковариационной матрицей $K = \begin{pmatrix} \sigma^2(x) & 0 \\ 0 & \sigma^2(y) \end{pmatrix}$, есть смещенная оценка

реальной разделяющей функции.

Традиционная оценка разделяющей функции в данном случае прямая, ортогональная к середине отрезка, проходящего через точки (x_i, y_i) , т.е. - прямая с угловым коэффициентом $\frac{(x_2 - x_1)\sigma^2(y)}{\sigma^2(x)(y_2 - y_1)}$

Найдем уравнение прямой $y = a + bx$, проходящей через точки (x_i, y_i) , $i = 1 \dots n$, учитывая дисперсии оценок $\sigma^2(x)$ и $\sigma^2(y)$. Оценки параметров a и b искомой прямой определим как точку минимума функционала (2.18) из системы уравнений (2.16):

$$\begin{cases} \frac{1}{\sigma^2(y) + b^2 \sigma^2(x)} \sum_{i=1}^n (y_i - a - bx_i) = 0, \\ \frac{\sum_{i=1}^n (y_i - a - bx_i) x_i}{\sigma^2(y) + b^2 \sigma^2(x)} + \frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{[\sigma^2(y) + b^2 \sigma^2(x)]^2} b \sigma^2(x) = 0. \end{cases}$$

Из первого уравнения получим

$$\hat{a} = \frac{1}{n} \left(\sum_i y_i - b \sum_i x_i \right), \quad i = 1 \dots n.$$

Подставим \hat{a} во второе уравнение, которое станет квадратным относительно \hat{b} . Оценку углового коэффициента b определим из квадратного уравнения

$$\hat{b}^2 + \hat{b} \frac{\sigma^2(y) \left[\left(\sum_i x_i \right)^2 - n \sum_i x_i^2 \right] + \sigma^2(x) \left[n \sum_i y_i^2 - \left(\sum_i y_i \right)^2 \right]}{\sigma^2(x) \left(\sum_i y_i \sum_i x_i - n \sum_i x_i y_i \right)} - \frac{\sigma^2(y)}{\sigma^2(x)} = 0$$

Для двух точек это уравнение примет вид

$$\hat{b}^2 + \hat{b} \frac{\sigma^2(x)(y_1 - y_2)^2 - \sigma^2(y)(x_1 - x_2)^2}{\sigma^2(x)(y_1 - y_2)(x_1 - x_2)} - \frac{\sigma^2(y)}{\sigma^2(x)} = 0$$

Угловым коэффициентом b однозначно связан с угловым коэффициентом разделяющей прямой: их произведение равно -1 . Однако в общем случае $\frac{(x_2 - x_1)\sigma^2(y)}{\sigma^2(x)(y_2 - y_1)}$ не является корнем данного квадратного

уравнения за исключением редких случаев, когда не различимы различные линии регрессии. Этот факт подтверждает смещенность оценки.

Далее оценим смещение разделяющих функций при учете погрешностей признаков на примере одномерных гауссовых функций $P(X|\omega_1)$ и $P(X|\omega_2)$ с математическими ожиданиями m_1 и m_2 соответственно, а также с одинаковыми дисперсиями σ^2

В традиционных методах не различаются X и ξ поэтому берется $P(X|\omega_j)$. В процессе оценки свободных параметров m_1 , m_2 и σ^2 этих распределений должны быть получены соответствующие погрешности Δm_1 , Δm_2 и $\Delta \sigma$.

Разделяющая граница в традиционных методах может быть определена согласно выражению

$$\ln \frac{P(X|\omega_1)}{P(X|\omega_2)} = 0, \quad (2.20)$$

$$\text{где } P(X|\omega_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - m_i)^2\right\}; \quad i = 1, 2,$$

и в данном случае при $\Delta m_1 = \Delta m_2 = 0$ равна $m_1 + m_2/2$.

В действительности из-за наличия погрешностей $\Delta m_1, \Delta m_2$ и $\Delta \sigma$ вместо (2.20) следует записать

$$\ln \frac{P(X|\omega_1) \pm \Delta P(X|\omega_1)}{P(X|\omega_2) \pm \Delta P(X|\omega_2)} = 0 \quad (2.21)$$

$\Delta P(X|\omega_i)$ возьмем в виде

$$\Delta P(X|\omega_i) = P'_{m_i}(X|\omega_i)\Delta m_i + P'_{\sigma}(X|\omega_i)\Delta \sigma; \quad i = 1, 2 \quad (2.22)$$

Здесь $P'_{m_i}(X|\omega_i)$ и $P'_{\sigma}(X|\omega_i)$ - производные от $P(X|\omega_i)$ по m_i и σ соответственно.

Найдем максимальное смещение разделяющей границы (в числителе выражения (2.21) прибавляется погрешность $\Delta P(X|\omega_1)$, а в знаменателе - вычитается $\Delta P(X|\omega_2)$ только за счет Δm_1 и Δm_2 . С учетом (2.22) выражение (2.21) примет вид

$$\ln \frac{P(X|\omega_1)}{P(X|\omega_2)} + \ln \frac{1 + \frac{(x - m_1)}{\sigma^2} \Delta m_1}{1 - \frac{(x - m_2)}{\sigma^2} \Delta m_2} = 0$$

Разложим в ряд $\ln\left(1 + \frac{(x - m_1)}{\sigma^2} \Delta m_1\right)$ и $\ln\left(1 - \frac{(x - m_2)}{\sigma^2} \Delta m_2\right)$. Так как

здесь не обязательно выполняется условие разложения функции $\ln(1+x)$ в ряд Маклорена $|x| < 1$, то воспользуемся разложением

$\ln \frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \dots\right)$; $|x| < 1$ и сделаем замену переменной

$\frac{1+x}{1-x} = \frac{N+k}{k}$ [28]. Тогда $x = k/(2N+k)$ - правильная положительная дробь

и

$$\ln \frac{N+k}{N} = 2\left[\frac{k}{2N+k} + \frac{1}{3}\left(\frac{k}{2N+k}\right)^3 + \dots\right]. \quad (2.23)$$

Выбираем удобное значение N , по которому найдем k . В нашем случае

$$N = 1; \quad k_i = \frac{x - m_i}{\sigma^2} \Delta m_i; \quad i = 1, 2$$

Согласно (2.23), выражение (2.21) при $\Delta m_1 = \Delta m_2 = \Delta m$ примет вид

$$x = \frac{\frac{1}{2}(m_1^2 - m_2^2) + 2\Delta m(m_1 + m_2)}{m_1 - m_2 + 4\Delta m} \approx \frac{m_1 + m_2}{2} + \frac{2\Delta m(m_1 + m_2)}{m_1 - m_2 + 4\Delta m}. \quad (2.24)$$

Второе слагаемое здесь определяет возможное смещение разделяющей границы. При другой комбинации знаков смещение будет $2\Delta m$.

Найдем теперь максимальное смещение границы только за счет $\Delta\sigma$. Выражение (2.21) с учетом (2.23) примет вид

$$\begin{aligned} & 8\Delta\sigma x^2 + x[2\sigma(m_1 - m_2) - 8\Delta\sigma(m_1 + m_2)] - \\ & - \sigma(m_1^2 - m_2^2) - 4\sigma^2\Delta\sigma + 4\Delta\sigma(m_1^2 - m_2^2) = 0. \end{aligned} \quad (2.25)$$

Смещение разделяющей границы в данном случае следует из того, что $x = (m_1 + m_2)/2$ не является корнем (32а).

Обратимся к алгоритму классификации, способному учесть погрешности априорной информации $P(\xi|\omega_i)$, $i = 1 \dots m$.

Показали, что разделяющая функция, которую получают в традиционных методах из условия

$$\ln \frac{P(\omega_1)P(X|\omega_1)}{P(\omega_2)P(X|\omega_2)} = 0,$$

ведет к смещенной оценке разделяющей функции. Несмещенная оценка разделяющей функции может быть получена путем

стохастической аппроксимации по граничным точкам, найденным из условия

$$\ln \frac{P(\omega_1) \hat{P}(\xi|\omega_1)}{P(\omega_2) \hat{P}(\xi|\omega_2)} = 0, \quad (2.26)$$

где $P(\xi|\omega_i)$ наблюдаемые значения вероятности появления в выбранном интервале X измеренных значений x_1, \dots, x_n для i -го образа. Значения $\hat{P}(\xi|\omega_i) \equiv \hat{p}$ известны, они использовались для определения вида функциональной зависимости $p(x = \xi|\omega_i)$ на первом этапе.

Из априорных сведений определяют общий вид (класс) разделяющих функций $\Psi(\xi, \theta)$ (линейная, гиперквадратная и т.п.), свободные параметры θ которых оценивают по точкам из условия (2.26), решая совместно системы уравнений (2.16) и (2.17), заменяя в последних $p(x = \xi|\omega_i)$ на

$$\Psi(\xi, \theta) \text{ и } \hat{p}_i \text{ на числовые значения } \ln \frac{P(\omega_1) \hat{P}(\xi|\omega_1)}{P(\omega_2) \hat{P}(\xi|\omega_2)}$$

Другая возможность найти разделяющую функцию $\Psi(\xi, \theta)$

$$\text{использовать сразу условие } \ln \frac{P(\omega_1) \hat{P}(\xi|\omega_1)}{P(\omega_2) \hat{P}(\xi|\omega_2)} = 0 \text{ и получить}$$

аналитический вид разделяющей функции $\Psi(\xi, \theta)$.

Наиболее часто в задачах классификации образов применяют линейные разделяющие функции (гиперплоскости). Оценки свободных параметров гиперплоскости находят как точку минимума функционала (2.19). В данном случае в системе алгебраических уравнений (2.16), служащей для оценки параметров, уравнение для оценки параметра a будет линейным, а остальные уравнения образуют систему квадратных уравнений относительно параметров $b_k, k=1 \dots l$. Методы решения подобных систем, как и ряда задач аппроксимации результатов наблюдений элементарными функциями при учете погрешности во всех координатах, рассмотрены в работе [15].

Интервальную оценку разделяющей функции $\Psi(\xi, \theta)$ после нахождения оценок $\hat{\theta}$ и их дисперсий $D(\hat{\theta})$ можно получить для любого $x = \xi$ из выражения

$$p \left[\psi(x, \hat{\theta}) - t_\gamma \sqrt{D[\psi(x, \hat{\theta})]} \leq \psi(x, \theta) \leq \psi(x, \hat{\theta}) + t_\gamma \sqrt{D[\psi(x, \hat{\theta})]} \right] = \gamma,$$

где γ доверительная вероятность; t_γ квантиль распределения Стьюдента для доверительной вероятности γ ; $D[\psi(x, \hat{\theta})]$ дисперсия

значения оценки функции $\psi(\xi, \hat{\theta})$ при $\xi=x$, которая может быть найдена по следующей формуле [16]:

$$D[\psi(x, \hat{\theta})] = \sum_{j=1}^l \left(\frac{\partial \psi(x, \theta)}{\partial \theta_j} \right)_{\theta=\hat{\theta}}^2 D(\hat{\theta}_j) + 2 \sum_{i>j} \frac{\partial \psi(x, \theta)}{\partial \theta_j} \frac{\partial \psi(x, \theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} D(\hat{\theta}_j, \hat{\theta}_i).$$

Для линейных функций $\psi(\xi, \theta)$ дисперсия

$$D[\psi(x, \hat{\theta})] = (1, x_1, \dots, x_n) \cdot D(\hat{\theta})(1, x_1, \dots, x_n)^T$$

Аналогично определяется интервальная оценка для условной плотности вероятности $p(x = \xi | \omega_j)$ как огибающая семейства всевозможных значений $p(\hat{\xi}, \hat{\theta} | \omega_j)$, полученных при различных $\hat{\theta}$, принадлежащих интервалу $\hat{\theta} \pm t_\gamma \sqrt{D(\hat{\theta})}$

$$\begin{aligned} p \left\{ p(\hat{\xi}, \hat{\theta} | \omega_j) - t_\gamma \sqrt{D[p(\hat{\xi}, \hat{\theta} | \omega_j)]} \leq p(\xi, \theta | \omega_j) \leq \right. \\ \left. \leq p(\hat{\xi}, \hat{\theta} | \omega_j) + t_\gamma \sqrt{D[p(\hat{\xi}, \hat{\theta} | \omega_j)]} \right\} = \gamma \end{aligned}$$

где

$$\begin{aligned} D[p(\hat{\xi}, \hat{\theta} | \omega_j)] = \sum_{v=1}^l \left[\frac{\partial p(\xi, \theta | \omega_j)}{\partial \theta_v} \right]_{\theta=\hat{\theta}}^2 D(\hat{\theta}_v) + \\ + 2 \sum_{v_1 > v_2} \frac{\partial p(\xi, \theta | \omega_j)}{\partial \theta_{v_1}} \frac{\partial p(\xi, \theta | \omega_j)}{\partial \theta_{v_2}} \cdot D(\hat{\theta}_{v_1}, \hat{\theta}_{v_2}) \end{aligned}$$

$D(\hat{\theta}_{v_1}, \hat{\theta}_{v_2})$ - корреляционный момент.

Изложенные результаты получены в предположении, что функция потерь $L(\omega, d)$ имеет вид

$$L(\omega_i, d_i) = \begin{cases} 0; & i = j, \\ 1; & i \neq j \end{cases}$$

В общем случае, когда функция потерь равна нулю для правильного решения; равна l_1 , если ошибочно выбирается образ ω_2 вместо ω_1 , равна l_2 , если ошибочно выбирается ω_1 вместо ω_2 ; решение с минимальным риском находят по выражению

$$l_i P(\omega_i) P(\xi | \omega_i); \quad i=1,2.$$

Объект ω_i выбирается из условия

$$l_1 P(\omega_1) P(\xi|\omega_1) > l_2 P(\omega_2) P(\xi|\omega_2).$$

Уравнение разделяющей функции получают из выражения

$$\ln \frac{l_1 P(\omega_1) P(\xi|\omega_1)}{l_2 P(\omega_2) P(\xi|\omega_2)} = 0$$

2.6 Классификация образов по измеренному с ошибкой вектору признаков

Перейдем к процедуре классификации образов второму этапу задачи классификации.

Мы имеем все характеристики наблюдавшихся случайных величин x_1, \dots, x_n , по которым на первом этапе нашли вид функции $P(\xi, \theta | \omega_j)$, $j=1 \dots m$ эмпирические значения вероятностей $p_i(\xi)$, $i=1 \dots n$, точечную и интервальную оценки разделяющей функции $\psi(x=\xi, \theta)$. Требуется провести измерение случайной величины (вектора признаков) X , по которой должен быть классифицирован объект $\omega_j \in \Omega$. Считаем, что интервальная оценка для случайной величины X известна.

Прежде всего отметим, что с появлением интервальной оценки разделяющей функции появляется нулевая зона - зона неопределенности. Если наблюдавшееся значение X попадает в эту зону, то требуется, строго говоря, следующее дополнительное наблюдение вектора признаков. Получив наблюдение X , надо посмотреть, не пересекается ли допустимый интервал значений случайной величины X с интервальной оценкой разделяющей функции. Если пересечения нет применяется традиционный метод. В противном случае надо предварительно найти "истинное значение" наблюдавшегося признака ξ . Однако по единственному измерению X нельзя найти соответствующие оценки ξ . Поэтому необходимо результаты наблюдений X присовокупить к имеющимся результатам наблюдений, полученным в процессе обучения системы распознавания.

Оценку ξ получим из уравнения (2.17), допуская, что случайная величина X могла принадлежать к обоим соседним распределениям, т.е., получим $\hat{\xi}_1$ и $\hat{\xi}_2$. В большинстве случаев, например, для гауссовых ф.п.в., условие (2.17) приводит к трансцендентному уравнению, которое может быть решено только численными методами. Для численных методов полезно использовать дополнительное ограничение вида

$$\hat{\xi}_i \in [x - 3\sigma(x); x + 3\sigma(x)], \quad i = 1, 2,$$

где $\sigma(x)$ среднеквадратичное отклонение для наблюдавшейся величины x .

Вблизи решающих функций (границы раздела) большинство распределений хорошо аппроксимируются гиперплоскостями (прямыми

линиями). Так, для одномерных гауссовых плотностей условие (2.17) примет вид

$$\frac{p_i - \hat{p}_i(\xi_i|\omega_i)}{\sigma^2(\hat{p})} p'_{i\xi}(\xi_i|\omega_i) + \frac{x - \xi_i}{\sigma^2(x)} = 0, \quad i = 1, 2,$$

где

$$p_i(\hat{\xi}_i|\omega_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left\{-\frac{1}{2} \frac{(m_i - \hat{\xi}_i)^2}{\sigma^2}\right\};$$

$$p_{i\xi}(\hat{\xi}_i|\omega_i) = p_i(\hat{\xi}_i|\omega_i) \frac{m_i - \hat{\xi}_i}{\sigma^2}; \quad i = 1, 2,$$

p_i определяют по результатам всех наблюдений x_1, \dots, x_n (например, путем группировки).

Разложим "теоретическую" функцию плотности $p_i(\xi_i|\omega_i)$ в ряд в окрестности наблюдаемой точки X . Получим прямую

$$p_i(\hat{\xi}_i|\omega_i) \approx p_i(x) \left(1 + \frac{x - m_i}{\sigma^2} x\right) + p_i(x) \frac{x - m_i}{\sigma^2} \hat{\xi}_i.$$

Обозначим

$$p_i(x) \left(1 + \frac{x - m_i}{\sigma^2} x\right) \equiv a_i;$$

$$p_i(x) \frac{x - m_i}{\sigma^2} \equiv b_i$$

и получим в окрестности точки x уравнение $p_i(\hat{\xi}_i|\omega_i) = a + b\hat{\xi}_i$

Тогда $\hat{\xi}_i$ найдем из условия

$$\hat{\xi}_i = \frac{(\hat{p}_i - a_i)b_i + x}{\sigma^2(\hat{p}) + b_i^2 \sigma^2(x)}; \quad i = 1, 2$$

По $\hat{\xi}_1$ и $\hat{\xi}_2$, соотношенными соответственно с первым и вторым распределениями, проводим классификацию (если ξ_1 и ξ_2 не попадают в зону неопределенности, когда требуется еще одно дополнительное наблюдение). В крайнем случае и здесь принимают решение по наиболее вероятному значению $\xi_i, i=1...2$.

Рассмотренный в п.п. 2.5 и 2.6 алгоритм статистической задачи распознавания образов (статистической задачи решения) позволяет полностью учесть статистическую природу результатов наблюдений как на этапе нахождения условных плотностей вероятностей, разделяющих функций, так и на этапе классификации по результатам наблюдений вектора признаков.

Зона неопределенности в принятии решений в задачах распознавания образов, когда применяется решающая процедура с фиксированным объемом выборки, имеет место, когда используется критерий отношения правдоподобия. Рассмотрим, как изменится этот критерий при учете неопределенности исходной информации.

Согласно традиционному подходу определим отношение правдоподобия между классами ω_i и ω_j следующим образом:

$$\lambda = \frac{P(X|\omega_i)}{P(X|\omega_j)} \quad \text{или} \quad \log \lambda = \log \frac{P(X|\omega_i)}{P(X|\omega_j)}$$

Тогда байесово решающее правило дает $d^*=d_i$, т.е. $X \sim \omega_i$, если

$$\lambda \geq \frac{P(\omega_j)}{P(\omega_i)} \quad \text{или} \quad \log \lambda \geq \log \frac{P(\omega_j)}{P(\omega_i)} \quad \text{для всех } j=1...m.$$

Этот же результат следует из критерия Неймана-Пирсона. Согласно ему, при данном числе наблюдений n оптимальное решение задачи в испытании гипотезы H_1 о том, что $X \sim \omega_1$, против гипотезы H_2 , согласно которой $X \sim \omega_2$ с вероятностью не меньшей $(1-\alpha)$ (если верна гипотеза H_1 и с вероятностью, не меньшей $(1-\beta)$ (если верна гипотеза H_2), определяется отношением правдоподобия λ_n имеющим вид

$$\lambda_n = \prod_{i=1}^n \frac{P(x_i|\omega_1)}{P(x_i|\omega_2)} = \frac{P_n(X|\omega_1)}{P_n(X|\omega_2)},$$

где α - ошибка I рода; β - ошибка II рода.

Этот критерий обеспечивает наименьшее β (наиболее мощный критерий).

Рассмотрим, как изменятся результаты решения данной задачи при известных интервальных оценках $P(X|\omega_i)$ и $P(X|\omega_j)$:

$$\lambda = \frac{P(X|\omega_i) \pm \Delta P_i}{P(X|\omega_j) \pm \Delta P_j}$$

или

$$\log \lambda = \log \frac{P(X|\omega_i) \pm \Delta P_i}{P(X|\omega_j) \pm \Delta P_j} \quad (2.26)$$

В последнем случае (см.(2.20)) решение $X \sim \omega_i$ имеет место, если

$$\log \lambda \geq \log \frac{P(\omega_j)}{P(\omega_i)} \quad (2.27)$$

для всех $j=1...m$.

Перепишем (2.26):

$$\log \lambda = \log \frac{P(X|\omega_i) \pm \Delta P_i}{P(X|\omega_j) \pm \Delta P_j} \cong \log \frac{P(X|\omega_i)}{P(X|\omega_j)} \pm \frac{\Delta P_i}{P(X|\omega_i)} \mp \frac{\Delta P_j}{P(X|\omega_j)}.$$

Тогда вместо условия (2.27) имеем

$$\log \frac{P(X|\omega_i)}{P(X|\omega_j)} \pm \frac{\Delta P_i}{P(X|\omega_i)} \mp \frac{\Delta P_j}{P(X|\omega_j)} \geq \log \frac{P(\omega_j)}{P(\omega_i)} \quad (2.28)$$

Учитывая, что погрешности ΔP_i и ΔP_j могут быть разных знаков, для надежной идентификации образов вместо условия (2.28) следует взять

$$\log \frac{P(X|\omega_i)}{P(X|\omega_j)} + \frac{|\Delta P_i|}{P(X|\omega_i)} + \frac{|\Delta P_j|}{P(X|\omega_j)} \geq \log \frac{P(\omega_j)}{P(\omega_i)}$$

Два последних слагаемых в левой части определяют нулевую зону (зону неопределенности в принятии решений).

2.7 Классификация летательных аппаратов с учетом погрешностей в измерениях признаков

Рассмотрим задачу классификации объектов по ряду признаков. Объектами распознавания являются пять классов летательных аппаратов (л.а.).

В результате активной радиолокации были получены следующие радиолокационные характеристики: 1) эффективная площадь рассеяния (ЭПР); 2) спектральные и временные характеристики отраженных от объектов сигналов. Характеристики были измерены радиолокационной станцией (РЛС) с длиной волны 10 м.

Используя эти данные были выделены признаки, определяющие л.а. Такими признаками являлись: скорость движения объекта v , высота полета H , частота модуляции отраженного сигнала f_m и значения ЭПР, измеренные при различных ракурсах.

Способы вычисления скорости движения и высоты полета л.а. по спектральным и временным характеристикам отраженных от объектов сигналов описаны в [12, 40]. Модуляция отраженного сигнала вызывается вращением турбин или винтов л.а. Частота модуляции f_m зависит от типа двигателя (винтовой или реактивный) и от количества двигателей.

ЭПР является оценкой интенсивности вторичного излучения л.а. и зависит от диэлектрической и магнитной проницаемости материала, из которого изготовлен л.а., от ракурса под которым наблюдается л.а., рабочей частоты РЛС, от формы и размеров л.а., от поляризации приемной и передающей антен РЛС [35].

Априорно известно, что признаки статистически независимы, плотности распределения каждого признака каждого класса соответствуют нормальному закону $N(\mu, \sigma^2)$ и априорные вероятности классов равны $P(\omega_j)=1/5; j=1...5$.

В табл. 2.2 приведены значения числовых характеристик плотностей вероятности каждого признака для каждого объекта.

Т а б л и ц а 2.2

Значения числовых характеристик условных плотностей вероятности признаков, принятых в традиционных методах

Номер объекта	скорость движения $v(m/c)$		высота полета $H(m)$		частота модуляции $f_m(kГц)$	
	μ	σ	μ	σ	μ	σ
1	800	200	2000	1583	9,5	1,3
2	1500	366	10000	1816	15	1,6
3	400	100	500	491	1,5	0,25
4	200	100	500	250	15	0,2
5	550	100	10000	166		

Необходимо было, используя данную информацию, осуществить распознавание конечного числа классов по результатам измерений признаков. Критерием распознаванием является вероятность ошибки (в.о.), которая не должна превышать $P(e)_{зад}=0,1$.

Радиолокационные характеристики, как и любые измерения, содержат случайные ошибки. Погрешности радиолокационных измерений подразделяются на ошибки вносимые л.а. и ошибки вносимые аппаратурой. Основные ошибки, вносимые л.а. происходят вследствие флуктуации амплитуды сигнала, а также зависят от состояния окружающей среды. Из числа ошибок, вносимых аппаратурой, наибольшее значение имеют ошибки вследствие собственных шумов приемного устройства. Если заданы параметры РЛС, а также характеристики л.а. и окружающей среды, уровень ошибки для определенной точки пространства обзора РЛС может быть рассчитан как среднеквадратическая сумма всех отдельных составляющих, вычисленных для этой точки. Однако нас интересует определение ошибок во всей зоне обзора РЛС. Для упрощения вычислений в [12, 40] предложена методика расчета оптимальной точности, при которой суммарная ошибка не зависит от погрешностей вносимых л.а., а оценивается по проектным данным или данным испытаний РЛС.

В данной задаче ошибки измерения по каждому признаку определялись ошибками, вносимыми РЛС, и для таких признаков как, скорость движения, высота полета и частота модуляции средние квадратические погрешности соответствуют следующим значениям: $\sigma(v)=100m/c$; $\sigma(H)=250m$; $\sigma(f)=0,2$ кГц. Предполагалось, что средняя

квадратическая погрешность измерения ЭПР составляет 10δ и одинакова при всех ракурсах; точность определения ракурса -10^0

С учетом погрешностей признаков числовые характеристики условных плотностей вероятности изменяются (табл. 2.3).

В табл. 2.3 приводятся также средние квадратические отклонения $\Delta\mu$ и $\Delta\sigma$ числовых характеристик условных плотностей вероятности, с помощью которых легко вычислить интервальные оценки функций условных плотностей вероятности в интересующих нас областях.

Т а б л и ц а 2.3

Одна из реализаций значений числовых характеристик и их средних квадратических отклонений для функций условных плотностей вероятности признаков с учетом погрешности наблюдений

Номер объекта	скорость движения $v(m/c)$				высота полета $H(m)$				частота модуляции $f_m(kГц)$			
	μ	$\Delta\mu$	σ	$\Delta\sigma$	μ	$\Delta\mu$	σ	$\Delta\sigma$	μ	$\Delta\mu$	σ	$\Delta\sigma$
1	720	70	230	40	2150	205	1630	250	10,2	0,9	1,25	0,21
2	1610	120	340	65	9100	850	1750	280	13,5	1,2	1,7	0,3
3	420	30	105	20	465	37	410	95	1,6	0,1	0,24	0,04
4	190	20	97	21	515	46	270	52	15,5	1,4	0,18	0,02
5	510	55	112	22	1080	980	156	29				

Анализ показал, что наилучшим образом классы л.а. по ЭПР разделяются при ракурсах: 30^0 , 60^0 , 120^0 Возьмем в качестве признаков эти значения ракурсов (табл. 2.4).

Т а б л и ц а 2.4

Значения числовых характеристик условной ф.п.в. ЭПР для каждого класса л.а.

Номер объекта	$\varphi=30^0$		$\varphi=60^0$		$\varphi=120^0$	
	μ	σ	μ	σ	μ	σ
1	73	10	74	10	64	10
2	71	10	55	10	71	10
3	62	10	63	10	43	10
4	47	10	61	10	63	10
5	53	10	60	10	61	10

Для вычисления вероятности ошибки (в.о.) воспользуемся способом, предложенном в [7, 10-12].

При большой мерности пространства признаков, непосредственное интегрирование значительно усложняется. Увеличение мерности пространства признаков усложняет выбор пределов интегрирования. Так,

например, для получения в.о. и интервальных оценок в.о. при использовании n - признаков нужно вычислять n - мерные интегралы

$$P(e) = 1 - \sum_{j=1}^S \int \dots \int P(x|\omega_j)P(\omega_j)dx_1 \dots dx_n.$$

Таким образом непосредственное вычисление в.о., описанным способом, вызывает трудности.

В [7] рассмотрен способ вычисления в.о. распознавания классов, использующий каждый признак в отдельности. Формула вычисления в.о. для данного способа имеет вид:

$$P(e) = \sum_{i=1}^T \left[P(x_i) - \max P(\omega_j)P(x_i|\omega_j) \right], \quad (2.29)$$

$$\text{где } P(x_i) = \sum_{j=1}^S P(\omega_j)P(x_i|\omega_j),$$

S - общее число классов; T - количество реализаций; $T=R^N$, где R - число значений признаков; N - общее число признаков.

В случае, когда ф.п.в. соответствуют многомерному нормальному распределению, использование формулы (2.29) дает возможность перейти к серии одномерных нормальных распределений и для одномерных нормальных распределений использовать линейризирующее преобразование.

Границы интервальной оценки в.о. в данном случае будут определяться по формулам:

$$\begin{aligned} P_n(e) &= \sum_{i=1}^T \left[P_n(x_i) - \max_{\omega} P(\omega_i)P_n(x_i|\omega_j) \right] = \\ &= \sum_{i=1}^T \left[\sum_{j=1}^S P(\omega_j)P_n(x_i|\omega_j) - \max_{\omega} P(\omega_j)P_n(x_i|\omega_j) \right] \\ P_b(e) &= \sum_{i=1}^T \left[\sum_{j=1}^S P(\omega_j)P_b(x_i|\omega_j) - \max_{\omega} P(\omega_j)P_b(x_i|\omega_j) \right], \end{aligned}$$

Использование в нашем случае одного любого признака дает в.о. больше заданной $P(e)_{зад} = 0,1$. Например, в.о. при использовании одного признака - скорости - равна $P_1(e) = 0,252$, $P_1(e) > P(e)_{зад}$.

Как указывалось ранее, в процедуре идентификации объектов при известных точечных и интервальных оценках функции плотностей вероятности по наблюдаемому вектору признаков $X_{набл}$ следует определить оценки $\hat{\xi}_{набл}$ и величину эллиптической (гиперэллипсоидной) области, которая покрывает истинное значение $\xi_{набл}$ с вероятностью γ . Эта область резко увеличивается с возрастанием числа координат вектора признаков $\xi_{набл}$. Отклонение от наблюдаемого значения $X_{набл}$ по каждой

координате равно $k, \sigma(X)$: для $\gamma=0,683$ и одной координате $k, \gamma=1$; для двух координат $k, \gamma=2$, для трех координат $k, \gamma=3,4$, для пяти координат $k, \gamma=6$ или другими словами, колебание в одно среднее квадратическое отклонение для одного параметра имеет доверительную вероятность $\gamma=0,683$, для двух параметров $\gamma=0,38$; для трех параметров $\gamma=0,20$ и т.д. Проведем идентификацию л.а. при одновременно наблюдаемых признаках $v \approx 1000$ м/сек, $H \approx 5500$ м, $f_m = 12$ кГц, ЭПР=69 дБ. Традиционные методы идентифицируют объект номер один, тогда как с учетом погрешности измерений признаков это объект номер два при вероятности ошибки $P_{ошиб} = 0,23$ со средним квадратическим отношением вероятности ошибки $\sigma(P_{ошиб}) = 0,03$. Только привлечение дополнительно значений ЭПР позволило уменьшить в.о. до 0,1.

Другой пример распознавания образов при наблюдении двух признаков показан на рис. 2.5. В традиционном подходе решение принимается по значениям (X_1, X_2) , что приводит и в этом случае к неверному решению. Решение следует принимать по оценкам $\hat{\xi}$. Строго говоря, в данном случае решение принято быть не может, поскольку и наблюдаемые значения (X_1, X_2) и оценки $\hat{\xi}$ попали в область неопределенности принятия решений (попали в нулевую зону).

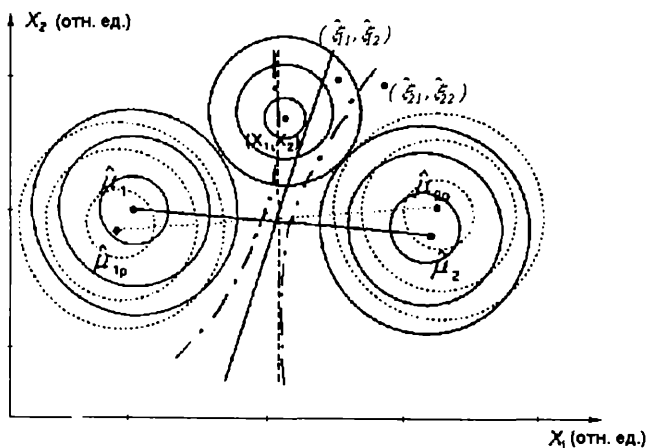


Рис. 2.5 Распознавание образов с учетом погрешности наблюдений признаков при известном с точностью до параметров виде функций условных плотностей вероятности.

На рис. 2.5 принято:

— без учета погрешностей признаков

— с учетом погрешностей признаков

-.-.- интервальная оценка решающей функции

Г Л А В А 3. СТАТИСТИЧЕСКИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ ПРИ НЕИЗВЕСТНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ИЗМЕРЕННЫХ ЗНАЧЕНИЙ ПРИЗНАКОВ

3.1. Оценка значений параметров классификаторов по выборке фиксированного объема

В предыдущих главах было показано, что в задачах распознавания образов необходимо учитывать:

1) погрешности измеряемых признаков при определении условных плотностей вероятности признаков и при идентификации объектов. В противном случае мы получаем смещенные оценки параметров функции условной плотности вероятности и принимаем недостоверные решения;

2) определять не только наиболее вероятные оценки функций условной плотности вероятности и решающих функций, но и их интервальные оценки. Тем самым показывается, что в любой задаче распознавания образов имеется нулевая зона (зона неопределенности), задаваемая, в частности, интервальной оценкой решающей функции.

Были описаны разработанные методы учета статистических характеристик векторов признаков в задачах распознавания образов с фиксированным объемом выборки в предположении, что вид функции условной плотности вероятности $P(\xi, \theta | \omega)$ вектора признаков ξ для класса ω нам известен, но неизвестны значения параметров θ и оценки $\hat{\xi}$ (параметрические методы).

Кроме того, в задачах распознавания образов часто имеет место случаи, когда [3, 7, 17, 18, 52, 54 - 56]:

а) вид распределений вероятностей признаков неизвестен, но статистика наблюдений достаточна для установления вида распределений (непараметрические методы);

б) вид рассматриваемых распределений вероятностей признаков неизвестен, но полагают известным вид (или виды) разделяющих (решающих) функций, а выборки используются для оценок значений параметров классификаторов. Последняя ситуация будет рассмотрена в данной главе.

Классификатор по минимуму расстояния. Пусть имелись выборки для c классов и для каждого класса определены векторы средних значений $\mu_i (i=1...c)$. Чтобы отнести наблюдаемый вектор признаков x к некоторому классу, следует измерить расстояние $\|x - \mu_i\|$ от x до каждого

из s векторов средних значений и отнести наблюдаемый вектор к классу, соответствующему ближайшему среднему значению. Если каждый из векторов средних значений считать идеальным прототипом или эталоном для образов своего класса, то это будет процедура сравнения с эталоном.

Расстояние $\|x - \mu_i\|$ можно определить по разному (разными формулами), но мы будем пользоваться понятием Евклидова расстояния.

Если априорные вероятности не равны, то квадрат расстояния $\|x - \mu_i\|^2$ должен быть нормирован по дисперсии σ^2 и смещен на величину $\log P(\omega_i)$. (Это следует из нормального распределения при одинаковых σ^2).

Когда x одинаково близок к двум различным векторам средних значений, при принятии решения следует предпочесть класс, априорно более вероятный.

Если ковариационные матрицы K для всех классов одинаковы и признаки подчинены нормальному распределению, то для классификации вектора признаков x следует определить квадратичное махаланобисово расстояние

$$(x - \mu_i)^T K^{-1} (x - \mu_i)$$

от x до каждого из s векторов средних значений и отнести x к классу, соответствующему ближайшему среднему значению μ_i . В случае неравных априорных вероятностей при принятии решения несколько большее предпочтение ($\log P(\omega_i)$), отдается классу априорно более вероятному.

Это один из наиболее простых критериев, который, однако, не учитывает возможные виды разделяющих функций, не говоря уже о том, что не учитываются погрешности $\mu_i (i=1 \dots c)$ и x и тем самым не определена область неопределенности принимаемых решений.

Линейная разделяющая функция для двух и многих классов (образов).

Линейные разделяющие функции наиболее просты и удобны с точки зрения аналитического исследования. Часто ими пользуются ради выигрыша в простоте, вычислительный процесс при этом значительно упрощается.

Задача определения линейной разделяющей функции формулируется как задача минимизации некоторой функции критерия, в качестве которого используется *выборочный риск* средние потери при классификации множества конструктивных выборов. Хотя могут использоваться и другие критерии.

Линейная разделяющая функция $g(x)$ для двух классов может быть записана в виде

$$g(x) = w \cdot x + w_0$$

где w - весовой вектор, w_0 - величина порога.

Для 2-х классов ω_1 и ω_2 применяется следующее решающее правило: принять решение ω_1 , если $g(x) > 0$, и ω_2 , если $g(x) < 0$.

В линейном случае уравнение $g(x)$ представляет гиперплоскость H , w - нормаль этой гиперплоскости, направленная в сторону области решений R_1 для ω_1 .

Разделяющая функция $g(x)$ должна бы представлять собой алгебраическое расстояние от x до гиперплоскости H . Поэтому предпочтительнее сразу пользоваться нормальным (нормированным) уравнением гиперплоскости. Для чего общее уравнение гиперплоскости следует поделить на длину вектора w , равную $\|w\|$. Тогда расстояние r от точки x до гиперплоскости H будет равно

$$r = \frac{g(x)}{\|w\|}.$$

При $w_0=0$ - гиперплоскость проходит через начало координат. Если имеется не два, а c классов, то задачу можно свести к $(c-1)$ задачам для двух классов, где решением i -той задачи служит линейная разделяющая функция, определяющая границу между точками, соответствующими решению ω_i , и точками, не соответствующими решению ω_i .

Как другой вариант можно использовать $\frac{c}{2}(c-1)$ линейных разделяющую функций, по одной для каждой пары классов.

В обоих подходах могут образовываться области, в которых классификация не определена. Чтобы избежать этой неприятности применяется классификатор, называемый *линейной машиной*. Здесь определяются c линейных разделяющих функций

$$g_i(x) = w_i^T x + w_{i0}, \quad i = 1 \dots c;$$

x приписывается к ω_i , если $g_i(x) > g_j(x)$ для всех $j \neq i$.

Если области решений R_i и R_j соприкасаются, то границей между ними будет часть гиперплоскости H_{ij} , определяемой соотношением

$$g_i(x) = g_j(x)$$

или

$$(w_i - w_j)^T x + w_{i0} - w_{j0} = 0.$$

Вектор нормали здесь $(w_i - w_j)$, а расстояние от x до плоскости H_{ij} равно

$$\frac{g_i - g_j}{\|w_i - w_j\|}.$$

Параметрические и непараметрические методы лучше "действуют" в пространствах меньшей размерности. Поэтому при распознавании образов, которые имеют признаки размерности d , применяется прием, позволяющий уменьшить размерность c измерений до одного: проецируются d -мерные данные на прямую (для двух классов).

Пусть имеется множество $n \times d$ мерных выборок x_1, x_2, \dots, x_n , из которых n_1 принадлежат X_1 ; помеченном ω_1 , и n_2 лежат в подмножестве

X_2 , намеченном ω_2 . Рассмотрим линейную комбинацию компонент вектора x :

$$y = w^T x$$

или $y = w_1 x_1 + \dots + w_n x_n$.

Если длина $\|w\| = 1$, то каждая компонента y_i проекция соответствующего x_i на прямую в направлении w . Направление имеет существенное значение. Доказано, что наилучшее направление w совпадает с направлением прямой, проходящей через m_1 и m_2 , где

$$m_i = \frac{1}{n_i} \sum_{x_j \in X_i} x_{ij}; \quad i = 1, 2.$$

На плоскости

$$x = \{x_1, x_2\}^T; \quad m_i = \{m_{i1}, m_{i2}\}^T; \quad m_{ik} = \frac{1}{n_i} \sum_{x \in X_i} x_{ik}, \quad k, i = 1, 2$$

На прямой w получим множество n выборок y_1, y_2, \dots, y_n , разделенное на подмножества Y_1, Y_2 .

Вместо того, чтобы проводить идентификацию образов по выборке X идентификацию проводят по проекции и вводится понятие линейного дискриминанта Фишера.

Линейный дискриминант Фишера определяется как такая линейная разделяющая функция $w^T x$, для которой функция критерия

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

максимальна,

где

$$s_i^2 = \sum_{y \in Y_i} (y_i - m_i)^2$$

$$m_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{y \in X_i} w^T x = w^T m_i; \quad i = 1, 2$$

В итоге получают

$$w = S_w^{-1} (m_1 - m_2),$$

где S_w матрица разброса внутри класса, она пропорциональна ковариационной выборочной матрице для совокупности d -мерных данных;

$$S_w = s_1 + s_2; \quad S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T \quad i = 1, 2$$

Например, известно, что для нормально распределенных случайных величин с равными ковариационными матрицами K и математическими ожиданиями μ_1 и μ_2 линейная разделяющая функция будет

$$w = K^{-1}(\mu_1, \mu_2).$$

Для задачи с c классами обобщение линейного дискриминанта Фишера включает $c-1$ разделяющих функций; проекция из d -мерного пространства в $(c-1)$ -мерное пространство (полагаем $d \geq c$).

Линейные разделяющие функции имеют направления m_1, \dots, m_{c-1} , где

$$m_i = \frac{1}{n_i} \sum_{x \in X_i} x;$$

$$m = \frac{1}{n} \sum_x x = \frac{1}{n} \sum_{i=1}^c n_i m_i$$

При редких наблюдениях каждое наблюдение, по которому ведется идентификация образов, должно быть проверено совместно с прежними наблюдениями на изменение решающей функции (или другого критерия).

3.2. Обобщенные линейные разделяющие функции

Линейная разделяющая функция имеет вид

$$g(x) = w^T x + w_0 \equiv w_0 + \sum_{i=1}^d w_i x_i;$$

где w_i - компоненты весового вектора w ; d - размерность признаков x . Можно ввести квадратичную разделяющую функцию

$$g(x) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=1}^d w_{ij} x_i x_j.$$

Причем можно считать, что $w_{ij} = w_{ji}$, поскольку $x_i x_j = x_j x_i$. Для квадратичной разделяющей функции разделяющая поверхность, определяемая уравнением $g(x) = 0$, является поверхностью второго порядка, или гиперквадрикой. Продолжая вводить дополнительные члены вида $w_{ijk} x_i x_j x_k$, можно получить класс полиномиальных разделяющих функций, которыми можно аппроксимировать разделяющую функцию $g(x)$ любого вида.

Однако хотелось бы остаться при этом в классе линейных функций. Вводят понятие *обобщенных линейных разделяющих функций*, которые имеют следующий вид

$$g(x) = \sum_{i=1}^d a_i y_i(x)$$

или $g(x) = a^T y$,

где a - d -мерный весовой вектор; $y_i(x)$ - произвольные функции от x .

Пример: пусть $g(x) = a_1 + a_2 x + a_3 x^2$, обобщенная линейная разделяющая функция имеет вид

$$g(x) = a_1 y_1 + a_2 y_2 + a_3 y_3$$

где $y_1 = 1$; $y_2 = x$; $y_3 = x^2$ или $y = \{1, x, x^2\}^T$

Однородная разделяющая функция $a^T y = 0$ разделяет точки в данном отображенном пространстве посредством гиперплоскости, проходящей через начало координат (что также положительный факт).

Но увеличение размерности d при переходе к обобщенной функции затрудняет практическое использование этого приема. Хотя переход к обобщенной функции удобен для линейных разделяющих функций. В этом случае

$$y = [1, x_1, \dots, x_d]^T = [1, x]^T$$

$$a = [w_0, w_1, \dots, w_d]^T = [w_0, w]^T$$

Переход от d -мерного пространства X к $(d+1)$ -мерному пространству Y не нарушает соотношений в расстояниях между выборками, но гиперплоскость $a^T y = 0$ проходит через начало координат Y -пространства; расстояние от y до гиперплоскости равно $|a^T y| / \|a\|$.

Задача сводится к определению весового вектора a . Как же его определить?

Пусть имеется множество n выборок y_1, \dots, y_n , часть которых относится к образу ω_1 , а другая к ω_2 . Эти выборки мы хотели бы использовать для определения весового вектора a , который правильно классифицировал бы все выборки. Если такой вектор существует, то выборки называются *линейно разделяемыми*. Выборка y_i классифицируется правильно, если $a^T y_i > 0$ и y_i помечен ω_1 , или если $a^T y_i < 0$, но y_i помечен ω_2 . Во втором случае y_i будет также классифицироваться правильно, если $a^T(-y_i) > 0$. Последнее условие дает возможность ввести понятие нормирования для случая двух классов: проводится замена всех выборок, обозначенных ω_2 , их отрицаниями. При ведении нормирования можно искать весовой вектор a , для которого выполняется условие для *всех выборок*. Данный вектор a называется разделяющим вектором или вектором решения. Каждая выборка y_i накладывает ограничение на возможное расположение вектора решения. Уравнение $a^T y_i = 0$ определяет гиперплоскость, проходящую через начало координат в весовом пространстве, и для которой y_i является нормальным вектором, в который должен находиться с положительной стороны гиперплоскости (рис. 3.1).

Система неравенств, образуемая выборками y_1, y_2, \dots, y_n , или неравенствами $a^T y_i > 0$, $i = 1 \dots n$, определяет *область решений*. Но сам вектор решения определяется неоднозначно. Например, вектор решения может быть выбран, как минимизирующий расстояния от выборок до разделяющей плоскости.

Другими словами, надо определить критерий (целевую функцию $J(a)$), которому должен удовлетворять вектор решения.

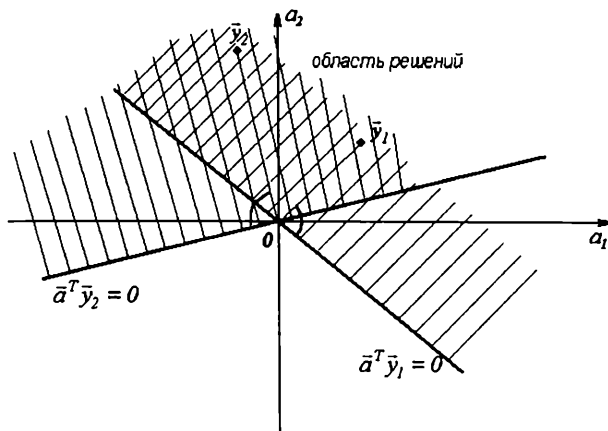


Рис. 3.1. Область решений в задаче распознавания

В качестве целевых функций используются:

- персептронная функция критерия $J_p(a) = \sum_{y \in Y} (-a^T y)$

где множество Y — множество выборок, классифицируемых с ошибкой посредством a . Она не может быть отрицательной и достигает нулевого значения, когда a является вектором решения;

- функции критерия

$$J_q(a) = \sum_{y \in Y} (a^T y)^2; J_r(a) = \frac{1}{2} \sum_{y \in Y} \frac{(a^T y - b)^2}{\|y\|^2}; J_s(a) = \sum_{i=1}^n (a^T y_i - b_i)^2 \text{ и др.}$$

где вектор b — некоторый допуск.

Процесс определения вектора решения a часто является итерационным, построенным на градиентном спуске:

$$\begin{cases} a_1 - \text{произвольно,} \\ a_{k+1} = a_k - p_k \nabla J(a_k); k \geq 1, \end{cases}$$

где $a_k^T y^k \leq 0$ для всех k ,

или $a_k^T y^k \leq b$ для всех k , при заданном допуске b .

Здесь $\nabla J(a_k)$ — градиент целевой функции (критерия) в точке a_k , p_k — положительный скалярный коэффициент, определяющий величину шага.

Для линейно разделяемых выборок итерационный процесс сходится при $k \rightarrow \infty$. В случае неразделяемых множеств сходимости может и не быть.

При малом числе наблюдений можно получить достаточно смещенную оценку вектора a (из-за нарушения условия $k \rightarrow \infty$).

В процедуре нахождения решения по методу наименьших квадратов (последняя целевая функция) весовой вектор находится как в случае

линейно разделяемых, так и в случае линейно неразделяемых выборок.

Найдем вектор a , при котором удовлетворяются равенства

$$a^T y_i = b_i, \quad i = 1 \dots n,$$

где b_i являются произвольно заданными положительными константами (в общем, заранее неизвестными).

Запишем систему $a^T y_i = b_i, \quad i = 1 \dots n$, в матричном виде

$$Ya = b,$$

где Y - матрица размера $n \times \hat{d}$, i -я строка которой является y_i^T , b - вектор столбец: $b = \{b_1, b_2, \dots, b_n\}^T$

Вектор a находится из условия минимума по a

$$J_s(a) = \|Ya - b\|^2 = \sum_{i=1}^n (a^T y_i - b_i)^2$$

Откуда

$$a = (Y^T Y)^{-1} Y^T b = Y^+ b,$$

если матрица $Y^T Y$ размера $\hat{d} \times \hat{d}$ невырождена.

Решение с наименьшей квадратичной ошибкой зависит от вектора допуска b и различные способы выбора b приводят к различным свойствам полученного решения. Если b задан произвольно, то нет гарантии того, что указанный вектор a будет разделяющим вектором в случае разделяемых множеств.

Автоматизируем процедуру выбора вектора b , используя метод градиентного спуска. Градиент функции $J_s(a)$ относительно a записывается в виде

$$\nabla_a J_s = 2Y^T(Ya - b),$$

а градиент, взятый относительно b , задается в виде

$$\nabla_b J_s = 2(Ya - b)$$

При любом b всегда можно считать $a = Y^+ b$, получая $\nabla_a J_s = 0$ и минимизируя $J_s(a)$ по a за один шаг.

Для b нужно учитывать условие $b > 0$ (не допуская $b = 0$). Можно положить вначале $b > 0$ и не уменьшать никаких координат вектора. Этого можно достичь при сохранении отрицательного градиента, если вначале все положительные координаты вектора $\nabla_b J_s$ взять равными 0. Приходим к процедуре спуска (алгоритм X_0 -Кашьяна)[18]

$$\begin{cases} b_l > 0, \text{ в остальном произвольный} \\ b_{k+1} = b_k + 2pe_k^T, \end{cases}$$

где e_k - вектор ошибки; $e_k = Ya_k - b_k$,

e_k^T - положительная составляющая вектора ошибки:

$$e_k^T = \frac{1}{2}(e_k + |e_k|),$$

$|e_k|$ - вектор, координаты которого суть абсолютные величины соответствующих координат вектора e_k ;

$$a_k = Y^T b_k ; k = 1, 2, \dots$$

Рассматриваются значения $0 < p < 1$.

Алгоритм образует последовательность векторов допуска b и определяет разделяющий вектор для случая разделяемых множеств или явно обнаруживает неразделимость для случая неразделимых множеств. Если выборки линейно неразделимы, то из этого больше не следует, что при e_k^T , равном 0, e_k тоже должно быть равным нулю. Для задачи с неразделимым множеством можно получить нулевой вектор ошибки с положительными координатами. Если это происходит, то выборки неразделимы.

В литературе рассмотрены другие варианты этого алгоритма, где оптимизируется выбор шага и исключаются процедуры обращения (псевдообращения) Y и вычисления Y^T

Решение по методу наименьшей квадратичной ошибки в пределе при $n \rightarrow \infty$ приближается в смысле минимума среднеквадратичной ошибки к разделяющей функции Байеса

$$g_0(x) = P(\omega_1|x) - P(\omega_2|x)$$

Это следует из того что, если предположить, что выборки взяты независимо в соответствии с вероятностным законом

$$P(x) = P(x|\omega_1) \cdot P(\omega_1) + P(x|\omega_2) \cdot P(\omega_2),$$

то решение по методу наименьшей квадратичной ошибки с использованием расширенного вектора y дает разложение в ряд функции

$$g(x) = a^T y,$$

где $y = y(x)$.

Если определить среднеквадратичную ошибку аппроксимации выражением $\varepsilon^2 = \int [a^T y - g_0(x)]^2 P(x) dx$, то можно доказать, что величина ε^2 минимизируется посредством решения $a = Y^T b$.

Аппроксимируя $g_0(x)$, разделяющая функция $a^T y$ дает непосредственную информацию относительно апостериорных вероятностей

$$P(\omega_1|x) = \frac{1}{2}(1 + g_0(x)) \text{ и } P(\omega_2|x) = \frac{1}{2}(1 - g_0(x)).$$

Качество аппроксимации зависит от функций $y_i(x)$ и числа членов в разложении $a^T y$. К сожалению, критерий среднеквадратической ошибки в основном распространяется не на точки, близкие к поверхности решения $g_0(x)=0$, а на точки, для которых значение $P(x)$ велико. Таким образом, разделяющая функция, которая наилучшим образом аппроксимирует разделяющую функции Байеса, не обязательно минимизирует вероятность ошибки.

Для построения разделяющих функций используется понятие *потенциальной функции* $K(x, x_i)$ с текущей точкой x и выборкой x_i . Вид потенциальной функции определяется видом выборки и существенно влияет на результаты. В общем виде можно указать, что $K(x, x_i)$ обратно пропорциональна $\|x - x_i\|^2$, например

$$K(x, x_i) = \frac{\sigma^2}{\sigma^2 + \|x - x_i\|^2}$$

Разделяющая функция $g(x)$ получается итерационным методом, когда выборки x_1, x_2, \dots рассматриваются последовательно:

$$g_{k+1}(x) = g_k(x) + r_k(x, x_k) \cdot K(x, x_k),$$

где r_k - некоторая функция ошибки.

Использование метода потенциальных функций наиболее оправдано, когда либо невелико число выборок, либо размерность x достаточно мала, чтобы функцию $g(x)$ можно было представить в виде таблицы дискретных значений x .

Например, можно разделяющую функцию $g(x)$ определить по следующему алгоритму:

$$g_{k+1}(x) = \begin{cases} g_k(x) + K(x, x_k), & \text{если } x_k = \omega_1 \text{ и } g(x_k) \leq 0; \\ g_k(x) - K(x, x_k), & \text{если } x_k = \omega_2 \text{ и } g(x_k) \geq 0; \\ g_k(x). & \end{cases}$$

3.3. Оценка разделяющего вектора с помощью процедур линейного и квадратичного программирования

Для оценки вектора a в обобщенной линейной разделяющей функции $a^T y = 0$ применяются методы математического программирования. Пусть имеется множество из n выборок y_1, y_2, \dots, y_n и требуется найти вектор a удовлетворяющий неравенству $a^T y_i \geq b > 0$ при всех i .

Введем искусственную переменную t ; $t \geq 0$, чтобы выполнялось $a^T y_i + t \geq b_i$.

Получим следующую задачу линейного программирования: минимизировать переменную t при условии

$$a^T y_i + t \geq b_i, \quad i = 1 \dots n, \quad t \geq 0.$$

В традиционной задаче линейного программирования полагают, что неизвестный вектор $a \geq 0$. В нашем случае это условие неприемлемо, т.к. вектор решения может иметь как положительные, так и отрицательные координаты.

Представим компоненты вектора a как комбинацию двух неотрицательных компонент:

$$a_j = a_j^l - a_0; j = 1 \dots d,$$

$$a_j^l \geq 0; a_0 \geq 0.$$

Получим следующую классическую задачу линейного программирования для переменных $t, a_1^l, a_2^l, \dots, a_d^l, a_0$ результатом решения которой будет решающий вектор a : минимизировать t при условии

$$\sum_{j=1}^d y_{ij} (a_j^l - a_0) + t \geq b_i, i = 1, n,$$

$$t \geq 0, a_0 \geq 0, a_j^l \geq 0; j = 1 \dots \hat{d}.$$

Методы решения подобных задач хорошо изучены [18, 61].

Можно сформулировать и другую задачу линейного программирования: сведем задачу минимизации перцептронной функции критерия к задаче линейного программирования. Напомним, что минимизация этой функции дает разделяющий вектор в случае разделяемых множеств и приводит к разумному решению в случае, когда выборки неразделимы.

Перцептронная функция критерия задается в виде

$$J_p(a) = \sum_{y \in Y} (-a^T y),$$

где Y множество выборок, классифицируемых с ошибкой посредством вектора a .

Чтобы исключить тривиальное решение $a=0$ введем положительный вектор допуска b и запишем функцию критерия в виде

$$J'_p(a) = \sum_{y_i \in Y'} (b_i - a^T y_i),$$

где $y_i \in Y'$, если $a^T y_i \leq b_i$.

Функция $J'_p(a)$ является кусочно-линейной функцией от a , поэтому следует ввести n искусственных переменных и соответствующих им ограничений и построить некоторую эквивалентную линейную целевую функцию. Рассмотрим задачу отыскания векторов a и t , минимизирующих линейную функцию

$$Z = \sum_{i=1}^n t_i$$

при ограничениях: $t_i \geq 0, t_i \geq b_i - a^T y_i; i = 1 \dots n$. С учетом неотрицательности переменных получим следующую задачу линейного программирования для нахождения вектора a :

минимизировать $Z = \sum_{i=1}^n t_i$ при условии

$$\sum_{j=1}^{\hat{d}} y_{ij} (a_j^I - a_0) + t \geq b_i; \quad t_i \geq 0, a_0 \geq 0, a_j^I \geq 0; \quad j = 1 \dots \hat{d}.$$

Выбор $a=0$; $t_i=b_i$, $i=1\dots n$, обеспечивает допустимое базисное решение, с которого начинается традиционный алгоритм симплекс-метода. Сходимость процессов линейного программирования гарантирована как в случае линейной разделяемости, так и в случае, когда исходные выборки неразделяемы.

Использование процедуры минимизации квадратичной ошибки позволяло находить разделяющий вектор для случая разделяемых множеств и обнаруживать неразделяемость для случая неразделяемых множеств. Поэтому естественно для отыскания решающего вектора \bar{a} применить следующую задачу квадратичного программирования:

$$\begin{aligned} &\text{минимизировать } J_s = \sum_{i=1}^n (a^T y_i - b_i)^2 \\ &\text{при условии } \sum_{j=1}^{\hat{d}} y_{ij} (a_j^I - a_0) \geq b_i; \quad i = 1 \dots n \\ & \quad a_0 \geq 0, a_j^I \geq 0; \quad j = 1 \dots \hat{d}. \end{aligned}$$

Анализ различных алгоритмов для отыскания линейных разделяющих функций показывает, что выбор подходящего алгоритма определяется такими факторами, как простота программирования, количество и размерность выборок. Если линейная разделяющая функция обеспечивает незначительный процент ошибок, любая из этих процедур при корректном ее применении позволит получить хорошее качество решения.

3.4. Разделяющие функции для случая многих классов (образов)

Процедуры получения разделяющих функций для случая двух классов стремятся распространить на случай многих классов. Универсального метода такого распространения не существует, но различными приемами в различных задачах это удастся сделать. Так в случае линейной машины для классов вместо множества разделяющих функций $g_i(x)=a_i^T y$; $i=1\dots c$, вводится разделяющая функция вида $\alpha^T \eta_{ij}$, где

$$\begin{aligned} \alpha &= \{a_1, a_2, \dots, a_c\}^T \\ \eta_{12} &= \{y, -y, 0, \dots, 0\}^T \\ \eta_{13} &= \{y, 0, -y, \dots, 0\}^T \end{aligned}$$

$$\eta_{ic} = \{y, 0, 0, \dots, -y\}^T \text{ и т. д.}$$

В этой процедуре размерность исходных данных увеличивается в c раз, число выборок в $(c-1)$ раз, что делает трудоемким применение данного алгоритма, но такое сведение полезно для доказательства сходимости.

При распространении метода наименьших квадратов на случай многих классов наиболее часто используют прием, при котором исходная задача сводится к рассмотрению множества c задач для двух классов. Если сразу решать задачу минимизации квадрата ошибки в случае многих классов, аналогично случаю двух классов, то вводят матрицу Y размера $n \times \hat{d}$, которую разбивают на блоки следующим образом:

$$Y = \{Y_1, Y_2, \dots, Y_c\}^T,$$

где выборки, помеченные символом ω_i включают в себя строки Y_i . Затем вводят матрицу весовых векторов A размером $\hat{d} \times c$: $A = \{a_1, a_2, \dots, a_c\}^T$ и матрицу B размера $n \times c$: $B = \{B_1, B_2, \dots, B_c\}^T$, где все элементы B_i являются нулевыми, за исключением элементов i -го столбца, равных 1. Тогда след квадратичной матрицы ошибок $(YA-B)^T(YA-B)$ будет минимален при решении $A = Y^T B$, где Y^T - псевдообращение матрицы Y .

Полученный результат может быть обобщен. Пусть λ_{ij} - потери для случая когда принимается решение ω_i вместо истинного ω_j , и предположим, что j -я подматрица B задается выражением

$$B_j = - \left[\begin{array}{c} \lambda_{1j} \lambda_{2j} \dots \lambda_{cj} \\ \lambda_{1j} \lambda_{2j} \dots \lambda_{cj} \\ \text{-----} \\ \lambda_{1j} \lambda_{2j} \dots \lambda_{cj} \end{array} \right] n_j, \quad j = 1 \dots c.$$

Тогда, если число выборок стремится к бесконечности, то решение $A = Y^T B$ дает разделяющие функции $a_i^T y$, которые обеспечивают оптимальное (в смысле минимума среднеквадратичной ошибки) приближение байесовской разделяющей функции

$$g_0(x) = - \sum_{j=1}^c \lambda_{ij} P(\omega_j | x).$$

3.5. Учет погрешностей наблюдений при оценке значений параметров классификаторов по выборке

В п.п. 3.1 3.4 настоящей главы рассмотрены основные традиционные методы оценки значений параметров классификаторов по выборке. В основном они сводятся к получению разделяющих функций между классами (образами), в частности, к линейным разделяющим

функциям или к обобщенным линейным разделяющим функциям. Но все рассмотренные алгоритмы не учитывают погрешности измерений наблюдаемых признаков, на основе которых принимаются решения. При очень богатой статистике этот неучет может быть как-то скомпенсирован. Однако при небольшом числе наблюдений мы получим смещенные оценки разделяющих (решающих) функций без указания интервальных оценок для них. При малом числе наблюдений интервальные оценки разделяющих (решающих) функций будут значительными и зоны неопределенности в принятии решения будут большими. Таким образом, надлежит разработать алгоритмы, которые учитывали бы погрешности измерений наблюдаемых признаков и позволяли бы правильно определять интервальные оценки разделяющих функций, а тем самым и зоны неопределенности принимаемых решений.

Зона неопределенности классификатора по минимуму расстояния. Рассмотрим квадрат Евклидовского расстояния между измеренными вектором признаков x и эталонным вектором признаков μ одного из объектов (образов)

$$r^2 = (x - \mu)^T K^{-1} (x - \mu),$$

где K - ковариационная (дисперсионная) матрица наблюдений x и μ . Например, на плоскости в предположении, что

$$K = \begin{pmatrix} D(x) & \dots & 0 \\ 0 & \dots & D(\mu) \end{pmatrix}, \quad (3.1)$$

$$r^2 = \frac{(x_1 - \mu_1)^2}{D(x)} + \frac{(x_2 - \mu_2)^2}{D(\mu)}$$

Все величины, входящие в r^2 получены в результате эксперимента и имеют свои погрешности.

Введем общий вектор θ размерности n , координатами которого являются все входящие в формулу для r^2 величины; дисперсию ошибки i -го параметра обозначим $D(\theta_i)$. Тогда за счет погрешностей наблюдаемых величин дисперсия ошибки в определении r^2 может быть вычислена по формуле

$$D(r^2) = \sum_{i=1}^n \left(\frac{\partial r^2}{\partial \theta_i} \right)^2 D(\theta_i) + 2 \sum_{j>i} \frac{\partial r^2}{\partial \theta_i} \cdot \frac{\partial r^2}{\partial \theta_j} D(\theta_i, \theta_j),$$

где $D(\theta_i, \theta_j)$ корреляционный момент оценок координат вектора θ . В предположении некоррелированности измерений формула упрощается:

$$D(r^2) = \sum_{i=1}^n \left(\frac{\partial r^2}{\partial \theta_i} \right)^2 D(\theta_i). \quad (3.3)$$

Для приведенного примера (формула (3.1))

$$D(r^2) = \sum_{i=1}^2 \left(\frac{\partial^2}{\partial x_i} \right)^2 D(x_i) + \sum_{i=1}^2 \left(\frac{\partial^2}{\partial \mu_i} \right)^2 D(\mu_i) + \left(\frac{\partial^2}{\partial D(x)} \right)^2 D(D(x)) + \left(\frac{\partial^2}{\partial D(\mu)} \right)^2 D(D(\mu)).$$

Дисперсия $D(r^2)$ определяет величину неопределенности $\sqrt{D(r^2)}$

квадрата расстояния между измеренным вектором признаков и эталонным, учет которой обязателен в процессе принятия решения.

Если вводятся другие определения расстояния между вектором измеренных признаков и эталонов, то величина неопределенности вычисляется аналогичным образом по формулам (3.2) и (3.3).

Зона неопределенности линейных разделяющих функций. Будем считать, что одним из указанных способов получено уравнение линейной разделяющей функции, т.е. определен решающий вектор a или, другими словами, оценки свободных параметров этой линейной функции: $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$.

В процессе получения оценок $\hat{a}_l, l = 1 \dots m$, должны быть получены, как минимум, дисперсии этих оценок. Они могут быть получены:

а) аналитическим путем (этот случай будет рассмотрен ниже);

б) методом статистических испытаний. В этом случае, отталкиваясь от погрешности наблюдаемых значений признаков эталонов, с помощью датчиков случайных чисел можно провести необходимую серию статистических испытаний. Каждое наблюдение породит серию точек, для которых получают оценки параметров $\hat{a}_l, l = 1 \dots m$. Для l -го параметра будет получен вариационный ряд, откуда можно определить наиболее вероятное значение \hat{a}_l и дисперсию этого значения $D(\hat{a}_l)$. При необходимости может быть проведен и более строгий анализ вариационного ряда.

Пусть m -мерный вектор a вектор, координатами которого будут оценки $\hat{a}_l, l = 1 \dots m$; $D(\hat{a})$ - ковариационная матрица оценок $\hat{a}_l, l = 1 \dots m$. Тогда оценка дисперсии $D(y)$ разделяющей линейной функции $y(a)$ для любого значения y находится по формуле

$$D(\hat{y}) = y^T D(\hat{a}) y,$$

где y - вектор с координатами y_1, y_2, \dots, y_m .

Эта формула является частным случаем формулы (3.2). Если матрица $D(\hat{a})$ - диагональная, то получим аналог формулы (3.3).

Пример. По результатам наблюдений найти интервальную оценку обобщенной линейной разделяющей функции $y(a) = a_1 y_1 + a_2 y_2 + a_3 y_3$.

Решение. Положим, что в процессе наблюдений определены оценки \hat{a}_l и их дисперсии $D(\hat{a}_l), l = 1...3$. Считаем, что корреляцией оценок $\hat{a}_l, l = 1...3$ можно пренебречь, т.е. матрица $D(\hat{a})$ имеет вид

$$D(\hat{a}) = \begin{pmatrix} D(\hat{a}_1) & 0 & 0 \\ 0 & D(\hat{a}_2) & 0 \\ 0 & 0 & D(\hat{a}_3) \end{pmatrix}$$

Дисперсия значений $D(\hat{y})$ разделяющей функции $y(a)$ равна

$$\begin{aligned} D(\hat{y}) &= (y_1, y_2, y_3) \begin{pmatrix} D(\hat{a}_1) & 0 & 0 \\ 0 & D(\hat{a}_2) & 0 \\ 0 & 0 & D(\hat{a}_3) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \\ &= y_1^2 D(\hat{a}_1) + y_2^2 D(\hat{a}_2) + y_3^2 D(\hat{a}_3). \end{aligned}$$

Зная значения $y(a)$ и $D(\hat{y})$, для каждого y можно построить доверительные интервалы для значений $\hat{y} = y(\hat{a})$ разделяющей функции. Огибающая линия, проведенная по этим точкам, даст интервальную оценку (зону неопределенности в принятии решений) для данной разделяющей функции. Пусть доверительная вероятность равна γ ($\gamma=0,95; 0,997...$). Интервальная оценка для каждой точки y разделяющей функции определяется условием:

$$P(\hat{y} - t_\gamma \sqrt{D(\hat{y})} < y(a) < \hat{y} + t_\gamma \sqrt{D(\hat{y})}) = \gamma,$$

где t_γ - квантиль распределения Стьюдента; определяется выбранным значением γ и числом степеней свободы $v=n-m$ (n - число наблюдаемых точек, m - число оцениваемых параметров); $\hat{y} = y(\hat{a})$ - значения $y(a)$ при найденных оценках \hat{a} . Задавшись значением γ , выбираем значение t_γ и определяем зону неопределенности принятия решений с помощью предлагаемой разделяющей функции. При малом числе эталонных наблюдений эта зона будет достаточно широкой.

В ряде случаев область неопределенности может быть получена аналитическим путем. Разделяющий вектор a может быть определен по методу наименьших квадратов, когда в качестве критерия выступает функция

$$J_s(a, b) = \|Ya - b\|^2$$

где Y - матрица измеренных значений признаков;
 b - вектор допуска.

Метод наименьших квадратов позволяет получить не только оценки \hat{a} , но и дисперсию этих оценок $D(\hat{a})$. При ковариационной матрице измерений $D(y)$ функцию можно записать в виде

$$J_s(a, b) = (Ya - b)^T D^{-1}(y)(Ya - b)$$

Тогда дисперсия оценок будет равна

$$D(\hat{a}) = [Y^T D^{-1}(y) Y]^{-1}$$

Дальнейшие действия при найденных оценках \hat{a} и $D(\hat{a})$ аналогичны ранее описанным: находятся оценки разделяющей функции $y(\hat{a})$ и дисперсии значений $D(\hat{y})$ этой функции, затем определяются интервальные оценки разделяющей функции, а тем самым и зона неопределенности принимаемых решений.

3.6. Оценка зоны неопределенности в процедурах линейного программирования

В п. 3.3 было показано, что задача оценки решающего вектора \hat{a} в обобщенной линейной разделяющей функции может быть сформулирована как задача линейного программирования. Нам важно установить, как в данном случае учесть неопределенности исходной информации.

Рассмотрим один из основных алгоритмов линейного программирования симплекс-метод, когда исходные данные задачи имеют неопределенности. Прежде всего задача линейного программирования должна быть записана в каноническом виде [15, 61]:

целевая функция $F_0 = CX \rightarrow \max$

условия-ограничения $AX = B; x_j \geq 0, j = 1 \dots n.$

Здесь вектор C имеет размерность m ; вектор X - n ; матрица A - $(l \times n)$; размерность вектора B равна l .

В задаче линейного программирования влияние неопределенности исходных данных определяется согласно зависимостям:

$$\frac{\partial F_0}{\partial a_{ij}} = \lambda_i^* x_j^*;$$

$$\left. \frac{\partial x_i}{\partial a_{ij}} \right|_{x_j^*} = - \frac{\lambda_i^* x_j^*}{c_j};$$

$$\left. \frac{\partial \lambda_i}{\partial a_{ij}} \right|_{x_j^*} = - \frac{\lambda_i^* x_j^*}{b_j}; \lambda_i = \frac{\partial F_0}{\partial b_i} \quad (3.4)$$

Здесь λ решение двойственной задачи линейного программирования; звездочкой отмечены оптимальные решения.

В нашем случае элементы векторов C , B и элементы матрицы A известны приближенно или могут изменяться в заданных пределах. Поставим в соответствие возможным изменениям элементов C , B и A функции плотностей вероятностей (ф.п.в.) этих изменений. Будем

считать, что числовые характеристики ф.п.в. известны. Тогда с помощью метода максимума правдоподобия можно будет получить функционал, точка минимума которого даст "истинные" значения заданных величин. Для исходных данных, подчиняющихся нормальному закону распределения, этот функционал имеет вид

$$\phi = \sum_{i=1}^l (b_i - \hat{b}_i)^2 + \sum_{k=1}^m (c_k - \hat{c}_k)^2 + \sum_{i=1}^l \sum_{j=1}^n (a_{ij} - \hat{a}_{ij})^2 \quad (3.5)$$

Здесь принято для простоты, что среднеквадратические отклонения исходных величин равны единице.

Из условий

$$\begin{cases} \frac{\partial \phi}{\partial c_k} = 0; k = 1 \dots m \\ \frac{\partial \phi}{\partial a_{ij}} = 0; i = 1 \dots l; j = 1 \dots n \end{cases} \quad (3.6)$$

определяем "истинные значения" элементов вектора C и матрицы A . Когда вектор C не связан с A и B , то "истинные" значения C - исходные значения, а "истинные" значения a_{ij} находят из решения n систем линейных алгебраических уравнений, состоящих из уравнений, элементы которых включают исходные данные b_i и решения $x_j, j=1 \dots n$, т.е. получить "истинные" значения a_{ij} можно, только решив задачу линейного программирования (ЛП). Отсюда процесс получения решения задачи ЛП с неточными исходными данными - итерационный процесс.

Неточность значений B учитывается через функционал (3.5) и соответствующее условие (3.6). Неточность в исходных данных повлияет не только на решение, но может принципиально изменить условия допустимого решения и выбора оптимального решения. Определив "истинные" значения исходных величин a_{ij} и b_j на последней итерации можем найти те изменения в решении задачи ЛП, которые вызваны неопределенностью в значениях C .

Выбрав базисные переменные, необходимо вычислить коэффициенты свободных переменных и ошибки этих коэффициентов.

В симплекс-методе процесс начинается с проверки условия оптимальности решения: есть ли положительные элементы C в канонической записи для каждого допустимого базисного решения. Если таковых нет, то решение оптимально. Надо анализировать величины $C \pm \Delta C$: может оказаться, что при $C > 0$ будет $C - \Delta C < 0$ и при $C < 0$ - $C + \Delta C > 0$. В обоих случаях следует продолжать решение задачи ЛП, но отметить возможное оптимальное решение.

Затем составляем отношения свободных членов B к положительным коэффициентам $a_{ij} \pm \Delta a_{ij}$ при свободной переменной в том столбце, для которого $C + \Delta C > 0$, и находим минимальное отношение. Таким образом выбрали базисную переменную, которая станет свободной.

Далее пересчитываем коэффициенты при новом наборе свободных переменных и ошибки этих коэффициентов. Процесс начинается опять с анализа новых $C \pm \Delta C'$. Получив оптимальное решение, пересчитываем коэффициенты a_{ij} согласно условию (3.6) и вновь определяем оптимальное решение при a_{ij}^* . Процесс прекращаем, когда

$$\left| \left(a_{ij}^* \right)_v - \left(a_{ij}^* \right)_{v+1} \right| < \varepsilon.$$

Так получаем оптимальный план, в котором учтена неопределенность исходных данных.

Далее можем найти коэффициенты влияния неопределенности исходных данных, согласно зависимостям (3.4). Однако пользоваться соотношениями (3.4) можно до тех пор, пока не выходим за область допустимых значений или пока не появляется возможное другое оптимальное решение, отмеченное ранее (пока не изменяется базис).

3.7. Распознавание образов по измеренному с погрешностью вектору признаков

Пусть нам известны погрешности измеренных векторов признаков X как в процессе обучения системы распознавания, так и в процессе идентификации образов. Закон распределения погрешностей наблюдений неизвестен, число точек наблюдения ограничено. Будем искать линейную разделяющую функцию.

Рассмотрим обобщенную линейную разделяющую функцию

$$g(x) = a^T y,$$

где a - весовой вектор, y - вектор, координаты которого включают в себя константу, равную 1, и функции от X , определяющие вид разделяющих функций (линейные, полиномиальные и т.д.). Задача сводится к определению весового вектора (вектора решения) a . Для однозначного выбора вектора решения следует ввести некую целевую функцию (или критерий). Чтобы включить в рассмотрение погрешности наблюдений X (а они трансформируются по формулам переноса ошибок в погрешности для y) выберем метод наименьшей квадратичной ошибки. С учетом погрешностей наблюдений признаков, как подробно показано в гл. 2, целевая функция, минимум которой определяет вектор решений a , имеет вид:

$$J(a) = \sum_{i=1}^n \left[\left(a^T \xi_i - b_i \right)^2 + \left(\xi_i - y_i \right)^2 \right], \quad (3.7)$$

где $y_i = \xi_i + \delta_i$; b_i - вектор допуска, который должен быть определен. Функция критерия $J(a)$ включает все выборки.

Функционал (3.7) получен как функционал ортогональной регрессии в предположении, что средние квадратические отношения признаков

равны единице. Необходимо его дополнить условием для нахождения неизвестного вектора ξ :

$$\frac{\partial J(a)}{\partial \xi} \Big|_{\xi=\hat{\xi}} = 0 \quad (3.8)$$

Преимуществом выбранного критерия является тот факт, что решение, следующего из этого критерия, приближается в пределе к разделяющей функции Байеса $g_0(X) = p(\omega_1|X) - p(\omega_2|X)$, хотя может не обеспечивать минимальную ошибку решения.

Рассмотрим два подхода:

- 1) вектор допуска b задается;
- 2) вектор допуска b оценивается в алгоритме построения разделяющей функции.

В первом случае $b=-1$, если наблюдения относятся к $y_1 \sim \omega_1$ и $b=+1$, если наблюдения относятся к $y_2 \sim \omega_2$

Тогда

6.

Объединив условия (3.7) и (3.8), для обобщенной линейной разделяющей функции получим функционал (3.7) в более приемлемом для практического применения виде

$$J(a) = \sum_{j=1}^{n_1} \frac{\left(\sum_{i=1}^k a_i y_{ij} - 1 \right)^2}{\sum_{i=1}^k a_i^2 \sigma^2(y_i)} + \sum_{j=n_1+1}^{n_2} \frac{\left(\sum_{i=1}^k a_i y_{ij} + 1 \right)^2}{\sum_{i=1}^k a_i^2 \sigma^2(y_i)}, \quad (3.9)$$

где

$\sigma^2(y_i) \equiv \sigma^2(x_i)$ средняя квадратическая ошибка i -го вектора наблюдений, одинаковая во всех наблюдениях; $j=1 \dots n_1$ - наблюдения для ω_1 , $j=n_1+1 \dots n_2$ - наблюдения для ω_2 .

Точка минимума функционала (3.9) определяет точечную оценку вектора решения a , а следовательно и точечную оценку разделяющей функции. Дисперсионная матрица оценки вектора решения a определяется из условия

$$D(\hat{a}) = M^{-1} \quad M = \left\| -\frac{\partial^2 J(a)}{\partial a_l \partial a_m} \right\|_{a=\hat{a}} \quad l, m = 1 \dots k,$$

а дисперсия значений разделяющей функции в точке наблюдения y равна

$$D(g(\hat{y})) = y^T D(\hat{a}) y,$$

где

$$Y = (y_1, y_2, \dots, y_k)^T.$$

При неизвестных векторах допуска b , как следует из результатов главы 2 для линейной функции, функционал (3.9) приобретает вид

$$J(a) = \sum_{j=1}^n \frac{\left(\sum_{i=1}^k a_i y_{ij} - b_j \right)^2}{\sum_{i=1}^k a_i^2 \sigma^2(y_i)},$$

где принято, что координаты вектора b детерминированные величины. Для каждой оценки вектора a с помощью градиентного спуска образуется последовательность векторов допуска b :

$$\begin{cases} b_1 > 0, \\ b_{v+1} = b_v + 2pe_v^+ \end{cases}$$

где

$$e_v^+ = \frac{1}{2}(e_v + |e_v|); \quad e_v = a_v^T y - b_v; \quad 0 < p < 2$$

В процедуре идентификации образов по вектору наблюдений $X_{j\text{набл}}$ строится вектор $Y_{\text{набл}}$, который присоединяется к ранее проведенным наблюдениям, и с помощью функционала (3.9) или условия (3.8) для каждой j -ой гипотезы определяется $\hat{\xi}_{j\text{набл}}$, по которым принимается решение.

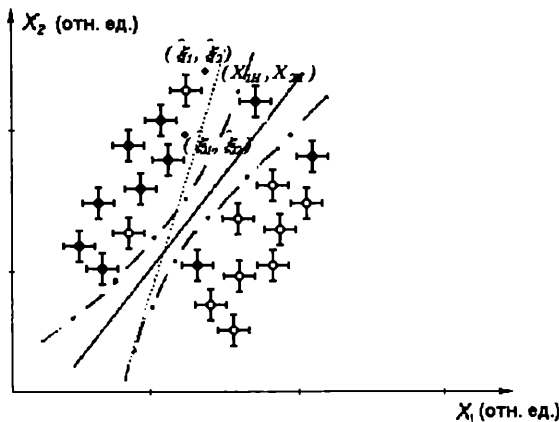


Рис.3.2. Распознавание образов с учетом погрешности наблюдений признаков с помощью линейной разделяющей функции.

На рис. 3.2 принято:

..... - без учета погрешностей признаков

— - с учетом погрешностей признаков

интервальная оценка решающей функции с учетом погрешностей признаков.

Результат определения линейной разделяющей функции и идентификации летательных аппаратов по наблюдаемому вектору (X_{1H}, X_{2H}) приведен на рис. 3.2. Различие решающих функций и отличие оценок "истинных значений" наблюдаемого вектора $(\hat{\xi}_{j1}, \hat{\xi}_{j2})$, $j = 1, 2$, от (X_{1H}, X_{2H}) очевидно.

3.8. Учет интервальных оценок функций плотности вероятности в последовательных методах распознавания образов

Пусть нам известны функции плотности вероятности (ф.п.в.) $P(X|\omega_i)$ некоторого признака X для различных объектов $\{\omega_i\}$, $i=1...n$. Кроме того, нам известны интервальные оценки этих функций. Тем самым нам известны интервальные оценки ошибок I и II родов α и β , которые имеют место при принятии гипотез $H_1: X \sim \omega_1$ и $H_2: X \sim \omega_2$.

Рассмотрим, как изменится задача последовательного распознавания образов при учете погрешностей в оценке условных ф.п.в. В частности, как изменятся верхние и нижние пороги (останавливающие границы) и среднее число наблюдений до принятия решений в данном случае.

Воспользуемся для решения данной задачи последовательным критерием отношения вероятностей Вальда (п.к.о.в.). Критерий предназначен для решения о выборе между двумя простыми гипотезами. Рассматривается отношение правдоподобия λ_n между классами ω_1 и ω_2 , имеющее вид

$$\lambda_n = \prod_{i=1}^n \left[P(x_i|\omega_1) / P(x_i|\omega_2) \right].$$

Наблюдения проводятся до тех пор, пока

$$B < \lambda_n < A \text{ или } \log B < \log \lambda_n < \log A,$$

$$\text{где } A = \frac{1-\alpha}{\beta}; \quad B = \frac{\alpha}{1-\beta}$$

Наблюдения прекращаются и принимается гипотеза $H_1: X \sim \omega_1$, как только выполнены неравенства

$$\lambda_n \geq A \text{ или } \log \lambda_n \geq \log A,$$

Гипотеза $H_2: X \sim \omega_2$ принимается при выполнении неравенств

$$\log \lambda_n \leq \log B \text{ или } \lambda_n \leq B.$$

В нашем случае будем учитывать погрешности ΔP_1 и ΔP_2 в определении условных ф.п.в., т.е. для первого измерения x_1 имеем

$$\log \lambda_1 = \log \frac{P(x_1|\omega_1) + \Delta P_1}{P(x_1|\omega_2) + \Delta P_2}.$$

Если ф.п.в. подчиняются нормальному закону распределения с математическими ожиданиями m_1 и m_2 , одинаковыми дисперсиями σ^2 и погрешностями оценок параметров Δm_1 , Δm_2 , $\Delta \sigma$, то

$$P(x|\omega_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-m_i)^2}{2\sigma^2}}$$

$$P(x|\omega_i) + \Delta P_i = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-m_i)^2}{2\sigma^2}} \left[1 + \frac{\Delta m_i}{\sigma^2} (x - m_i) + \frac{(x - m_i)^2}{\sigma^3} \Delta \sigma - \frac{\Delta \sigma}{\sigma} \right].$$

Здесь мы приращение $\Delta P_i = \Delta P(x|\omega_i)$ заменили дифференциалом $dP(x|\omega_i)$. Тогда

$$\begin{aligned} \log \lambda_1 &= \frac{1}{\sigma^2} \left[(m_1 - m_2)x_1 - \frac{1}{2}(m_1^2 - m_2^2) \right] + \\ &+ \log \frac{\sigma^3 + \sigma \Delta m_1 (x_1 - m_1) + (x_1 - m_1)^2 \Delta \sigma - \sigma^2 \Delta \sigma}{\sigma^3 + \sigma \Delta m_2 (x_1 - m_2) + (x_1 - m_2)^2 \Delta \sigma - \sigma^2 \Delta \sigma} \approx \\ &\approx \frac{1}{\sigma^2} \left\{ x_1 \left[\left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1 - m_2) + \Delta m_1 - \Delta m_2 \right] - \right. \\ &\left. - \frac{1}{2} \left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1^2 - m_2^2) - m_1 \Delta m_1 + m_2 \Delta m_2 \right\}. \end{aligned}$$

Если $\lambda_1 \geq A$ или $\log \lambda_1 \geq \log A$, то истинной является гипотеза $X \sim \omega_1$.

Если

$$\begin{aligned} x_1 &\geq \frac{\sigma^2 \log A}{\left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1 - m_2) + \Delta m_1 - \Delta m_2} + \\ &+ \frac{1}{2} \frac{\left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1^2 - m_2^2) - m_1 \Delta m_1 + m_2 \Delta m_2}{\left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1 - m_2) + \Delta m_1 - \Delta m_2}, \end{aligned}$$

то $X \sim \omega_1$.

При $\Delta m_1 = \Delta m_2 = \Delta m$ данное условие примет вид

$$x_1 \geq \frac{\sigma^2 \log A}{\left(1 - 2 \frac{\Delta \sigma}{\sigma} \right) (m_1 - m_2)} + \frac{1}{2} (m_1 + m_2) - \frac{\Delta m}{2 \left(1 - 2 \frac{\Delta \sigma}{\sigma} \right)}.$$

Если же $\Delta \sigma = 0$, то получим

$$x_1 \geq \frac{\sigma^2 \log A}{m_1 - m_2} + \frac{1}{2}(m_1 + m_2) - \frac{\Delta m}{2}.$$

Мы здесь считали, что погрешности имеют один знак (направлены в одну сторону). Даже в этом случае получили условие, отличающееся от традиционных. Правильнее было определить максимальный разброс, сложив модули погрешностей. Чтобы не усложнять формулы, мы этого здесь делать не будем.

После второго измерения при $\Delta m_1 = \Delta m_2 = \Delta m$

$$\begin{aligned} \log \lambda_2 &= \log \frac{P(x_1|\omega_1) + \Delta P_{11}}{P(x_1|\omega_2) + \Delta P_{21}} + \log \frac{P(x_2|\omega_1) + \Delta P_{12}}{P(x_2|\omega_2) + \Delta P_{22}} = \\ &= \frac{m_1 - m_2}{\sigma^2} \left[(x_1 + x_2) \left(1 - \frac{2\Delta\sigma}{\sigma}\right) - \left(1 - \frac{2\Delta\sigma}{\sigma}\right) (m_1 + m_2) - \Delta m \right]. \end{aligned}$$

При n измерениях x_1, x_2, \dots, x_n , когда истинные значения соответственно равны $\xi_1, \xi_2, \dots, \xi_n$

$$\log \lambda_n = \frac{m_1 - m_2}{\sigma^2} \left[\sum_{i=1}^n x_i \left(1 - \frac{2\Delta\sigma}{\sigma}\right) - \frac{n}{2} \left(1 - \frac{2\Delta\sigma}{\sigma}\right) (m_1 + m_2) - \frac{n}{2} \Delta m \right]$$

и процедура классификации будет следующей:

если

$$\sum_{i=1}^n x_i \geq \frac{\sigma^2 \log A}{(m_1 - m_2) \left(1 - \frac{2\Delta\sigma}{\sigma}\right)} + \frac{n}{2} (m_1 + m_2) + \frac{n\Delta m}{2 \left(1 - \frac{2\Delta\sigma}{\sigma}\right)},$$

то $X \sim \omega_1$;

если

$$\sum_{i=1}^n x_i \leq \frac{\sigma^2 \log B}{(m_1 - m_2) \left(1 - \frac{2\Delta\sigma}{\sigma}\right)} + \frac{n}{2} (m_1 + m_2) + \frac{n\Delta m}{2 \left(1 - \frac{2\Delta\sigma}{\sigma}\right)},$$

то $X \sim \omega_2$;

если

$$\begin{aligned} &\frac{\sigma^2 \log B}{(m_1 - m_2) \left(1 - \frac{2\Delta\sigma}{\sigma}\right)} + \frac{n}{2} (m_1 + m_2) + \frac{n\Delta m}{2 \left(1 - \frac{2\Delta\sigma}{\sigma}\right)} < \\ &< \sum_{i=1}^n x_i < \frac{\sigma^2 \log A}{(m_1 - m_2) \left(1 - \frac{2\Delta\sigma}{\sigma}\right)} + \frac{n}{2} (m_1 + m_2) + \frac{n\Delta m}{2 \left(1 - \frac{2\Delta\sigma}{\sigma}\right)}. \end{aligned}$$

То следует провести новое измерение.

Из условия (2.4) следует, что

$$\lim_{n \rightarrow \infty} \sum_{i=1}^l x_i = \sum_{i=1}^n \xi_i.$$

Поэтому погрешность измерения наблюдений X будет сказываться заметно только при первых наблюдениях. И в процедурах классификации тогда вместо x_i следует брать $x_i \pm 3\sigma(x_i)$, где $\sigma(x_i)$ - среднее квадратичное отклонение погрешности значения x_i .

Останавливающие границы получены при условии, что

$$A = \frac{1 - \alpha}{\beta}; \quad B = \frac{\alpha}{1 - \beta},$$

где α - ошибка 1-го рода, β - ошибка 2-го рода. Пока интервальные оценки α и β не учтены. Чтобы их учесть, надо взять $(\log B)_{\min}$ и $(\log A)_{\max}$.

$$(\log A)_{\max} = \log(A + \Delta A) = \log A + \log\left(1 + \frac{\Delta A}{A}\right) \cong \log A + \frac{|\Delta A|}{A}$$

$$(\log B)_{\min} = \log(B - \Delta B) \cong \log B - \frac{|\Delta B|}{B}$$

Оценим, как изменится среднее число измерений при учете погрешности наблюдений.

В традиционных методах определено, что среднее число наблюдений, когда истинна гипотеза H_1 , равно \bar{n}_1 :

$$\bar{n}_1 = \frac{1}{z} \left[(1 - \alpha) \log A + \alpha \log B \right].$$

Если истинна гипотеза H_2 , то

$$\bar{n}_2 = \frac{1}{z} \left[\beta \log A + (1 - \beta) \log B \right].$$

В нашем случае, когда известны интервальные оценки α и β , надо взять верхние границы $\log A$ и $\log B$ и соответственно верхние или нижние границы α и β . Тогда среднее число измерений

$$\begin{aligned} \bar{n}_{1\text{нов}} &= \frac{1}{z} \left[(1 - \alpha + \Delta\alpha) \left(\log A + \frac{|\Delta A|}{A} \right) + (\alpha + \Delta\alpha) \left(\log B - \frac{|\Delta B|}{B} \right) \right] = \\ &= \frac{1}{z} \left[(1 - \alpha + \Delta\alpha) \frac{|\Delta A|}{A} + \Delta\alpha (\log A + \log B) - \frac{|\Delta B|}{B} (\alpha + \Delta\alpha) \right] + \bar{n}_1. \end{aligned} \quad (3.10)$$

Здесь первое слагаемое - поправка к среднему числу наблюдений.

Рассмотрим

$$\bar{z}_{\text{нов}} = \log \frac{P(X|\omega_1) + \Delta P_1}{P(X|\omega_2) + \Delta P_2} \cong \bar{z} + \frac{\Delta P_1}{P(X|\omega_1)} - \frac{\Delta P_2}{P(X|\omega_2)}, \quad (3.11)$$

$$\text{где } \bar{z} = \log \frac{P(X|\omega_1)}{P(X|\omega_2)}$$

Поскольку мы оцениваем среднее значение (результат многократных наблюдений), то можно считать, что при $\frac{\Delta P_1}{P(X|\omega_1)} = \frac{\Delta P_2}{P(X|\omega_2)}$

одинаковой относительной погрешности для условных п.ф.в., знаменатель дроби (3.10) при определении $\bar{n}_{1нов}$ не изменяется. Поэтому будем считать, что не меняется и $\bar{z}_{нов}$. (Это не означает, что не будут меняться другие параметры, определяемые из условия (3.11)).

Среднее число наблюдений (если верна гипотеза H_2 , с учетом интервальных оценок α и β ; A и B):

$$\begin{aligned} \bar{n}_{2нов} &= \frac{1}{z} \left[(\beta + \Delta\beta) \left(\log A + \frac{|\Delta A|}{A} \right) + (1 - \beta + \Delta\beta) \left(\log B - \frac{|\Delta B|}{B} \right) \right] = \\ &= \bar{n}_2 + \frac{1}{z} \left[(\beta - \Delta\beta - 1) \frac{|\Delta B|}{B} + \Delta\beta \log(AB) + \frac{|\Delta A|}{A} (\beta + \Delta\beta) \right], \end{aligned}$$

где $\frac{\Delta A}{A}$ и $\frac{\Delta B}{B}$ находят из предположения, что погрешности ΔA , ΔB ,

$\Delta\alpha$, $\Delta\beta$ независимы и подчиняются нормальным законам распределения с дисперсиями, соответственно равными $(\Delta A)^2$, $(\Delta B)^2$, $(\Delta\alpha)^2$, $(\Delta\beta)^2$

По формулам переноса ошибок получим

$$\left(\frac{\Delta A}{A} \right)^2 = \left(\frac{\Delta\alpha}{1 - \alpha} \right)^2 + \left(\frac{\Delta\beta}{\beta} \right)^2$$

$$\left(\frac{\Delta B}{B} \right)^2 = \left(\frac{\Delta\alpha}{\alpha} \right)^2 + \left(\frac{\Delta\beta}{1 - \beta} \right)^2$$

3.9. Сравнение зон неопределенности в задачах принятия решений. Общий алгоритм принятия решений.

Как было показано, поскольку в реальных условиях не возможно произвести измерение без погрешностей, процедура методов последовательного принятия решений присутствует и в процессе принятия решений по выборке фиксированного объема. Представляет интерес сравнить зоны неопределенности в процессах принятия решений по выборке фиксированного объема и в последовательных методах. Другими словами, не является ли зона неопределенности (нулевая зона) при анализе выборки фиксированного объема столь незначительной, что ее учет представляет только академический интерес.

Одномерный случай. Пусть некоторый признак объекта x распределен по нормальному закону с математическим ожиданием m_1 и

дисперсией σ_1^2 для гипотезы H_0 ; и - с математическим ожиданием m_2 и дисперсией σ_2^2 для гипотезы H_1 . При равной вероятности гипотез H_0 и H_1 , решающая граница x_0 имеет вид

$$x_0 = \frac{1}{2}(\hat{m}_1 + \hat{m}_2).$$

В этом случае дисперсия $D(x_0) = \frac{1}{4}[\sigma^2(\hat{m}_1) + \sigma^2(\hat{m}_2)]$,

где $\sigma^2(\hat{m}_i)$ дисперсия оценки математического ожидания \hat{m}_i , $i=1,2$. При одинаковых дисперсиях $\sigma^2(\hat{m}_1) = \sigma^2(\hat{m}_2) = \sigma^2(\hat{m})$

$$\sigma(x_0) = \frac{\sigma(\hat{m})}{\sqrt{2}}$$

Во многих случаях, когда инструментальная погрешность измерения не является определяющей, оценкой математического ожидания нормального закона распределения является среднее арифметическое результатов наблюдений \bar{x}

$$\hat{m} = \bar{x} = \sum_{i=1}^n x_i$$

Отсюда $\sigma(\hat{m}) = \sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{n}}$,

где $\sigma^2(x)$ - дисперсия наблюдаемых значений x .

Ширина зоны неопределенности в этом случае будет

$$\frac{2t\sigma(\hat{m})}{\sqrt{2}} = \frac{\sqrt{2}t\sigma(x)}{\sqrt{n}},$$

где t квантиль нормального распределения, определяемая доверительной вероятностью: $t=1,2,3,\dots$

В методе последовательного принятия решений, когда решение принимается по среднему значению \bar{x} , полоса неопределенности в случае нормального распределения признака с математическими ожиданиями m_1 и m_2 для разных гипотез и с одинаковыми дисперсиями σ^2 для обеих гипотез, при числе наблюдений, равном n , определяется выражением

$$\frac{\sigma^2}{|m_1 - m_2|n} \ln \frac{A}{B},$$

где A и B соответственно останавливающие верхние и нижние границы:

$$A = \frac{1 - \alpha}{\beta}; \quad B = \frac{\alpha}{1 - \beta}.$$

Здесь α и β - ошибки первого и второго рода.

Пусть $|m_1 - m_2| = k\sigma$ Тогда полоса неопределенности будет

$$\frac{\sigma}{kn} \ln \frac{(1-\alpha)(1-\beta)}{\beta\alpha}$$

Чем больше разность математических ожиданий m_1 и m_2 , т.е. чем больше k , тем меньше α и β и тем уже полоса неопределенности. Найдем конкретные значения ширины полосы неопределенности для равновероятных гипотез. Пусть $|m_1 - m_2| = 4\sigma$, тогда

$$\alpha = \beta \cong \frac{1-0,95}{2} = 0,025 \quad \text{При} \quad |m_1 - m_2| = 2\sigma \quad \text{ошибки}$$

$$\alpha = \beta = \frac{1-0,68}{2} = 0,16 \quad \text{В первом случае полоса неопределенности будет}$$

$$\approx \frac{1,5\sigma}{n}, \quad \text{во втором} - \approx \frac{2\sigma}{n}.$$

Это показывает, что зоны неопределенности в задачах принятия решений при фиксированном объеме выборки и в последовательных методах соизмеримы, поскольку $\sigma(x)$ наиболее часто просто совпадает с σ . Причем отношение ширины полос в этих методах принятия решений равно

$$\frac{\sqrt{2} t\sigma(x)kn}{\sqrt{n} \sigma \ln \frac{(1-\alpha)(1-\beta)}{\alpha\beta}} \cong \frac{\sqrt{2} tk\sqrt{n}}{\ln \frac{(1-\alpha)(1-\beta)}{\alpha\beta}} \approx \sqrt{n}$$

Это говорит о том, что в процессе принятия решения при фиксированном объеме выборки ширина зоны неопределенности может быть даже больше, чем в методе последовательного принятия решений.

Двумерный случай. Рассмотрим задачу принятия решений по двум признакам x и y (на плоскости). Для простоты будем считать, что решающая граница - прямая линия.

Уравнение линейной решающей функции приводилось:

$$X^T K^{-1} (M_1 - M_2) - \frac{1}{2} (M_1 + M_2)^T K^{-1} (M_1 - M_2) + \log \frac{P(\omega_1)}{P(\omega_2)} = 0,$$

где X - текущий вектор; M_i - вектор математического ожидания для i -ой гипотезы, $i=1,2$; K - ковариационная матрица, одинаковая для обеих гипотез, $P(\omega_i)$ - вероятность i -ой гипотезы, $i=1,2$.

В нашем случае при равновероятных гипотезах и единичной ковариационной матрице уравнение решающей прямой линии имеет вид

$$x(m_{1x} - m_{2x}) + y(m_{1y} - m_{2y}) - \frac{1}{2}(m_{1x}^2 - m_{2x}^2 + m_{1y}^2 - m_{2y}^2) = 0$$

Здесь m_{ij} - математическое ожидание j -ой координаты для i -гипотезы, $i=1,2$; $j=x,y$.

В процессе наблюдений необходимо найти оценки \hat{m}_{ij} и их дисперсии $D(\hat{m}_{ij})$, а также оценки и дисперсии оценок элементов ковариационной матрицы K (если они не исключены из рассмотрения).

Представим данную прямую линию в традиционном виде

$$Y = \theta_0 + \theta_1 X,$$

где

$$\theta_0 = \frac{m_{1x}^2 - m_{2x}^2 + m_{1y}^2 - m_{2y}^2}{2(m_{1y} - m_{2y})};$$

$$\theta_1 = -\frac{m_{1x} - m_{2x}}{m_{1y} - m_{2y}}.$$

Интервальная оценка нашей прямой (интервальная оценка значений y) найдется через дисперсию \hat{y}

$$D(\hat{y}) = (1 \ x) D(\hat{\theta}) \begin{pmatrix} 1 \\ x \end{pmatrix}.$$

$D(\hat{\theta})$ выражается через \hat{m}_{ij} и $D(\hat{m}_{ij})$

Последняя формула определяет дисперсию значений \hat{y} для любого значения x , тем самым и интервальные оценки нашей прямой.

В методе последовательного принятия решений на плоскости, когда признаки независимы и подчиняются нормальному закону распределения с математическими ожиданиями $M_1 = \{m_{1x}, m_{1y}\}$ и $M_2 = \{m_{2x}, m_{2y}\}$ и

общей ковариационной матрицей $K = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$, после проведения

первого измерения (x_1, y_1) $\ln \lambda_1$ будет иметь следующий вид:

$$\begin{aligned} \ln \lambda_1 &= \ln \frac{P(x_1, y_1 | \omega_1)}{P(x_1, y_1 | \omega_2)} = \\ &= \frac{(m_{1x} - m_{2x}) \left(x_1 - \frac{m_{1x} + m_{2x}}{2} \right)}{\sigma_x^2} + \frac{(m_{1y} - m_{2y}) \left(y_1 - \frac{m_{1y} + m_{2y}}{2} \right)}{\sigma_y^2} \end{aligned}$$

После n наблюдений $x_1, y_1, \dots, x_n, y_n$ получим $\ln \lambda_n = \sum_{i=1}^n \ln \frac{P(x_i, y_i | \omega_1)}{P(x_i, y_i | \omega_2)}$

или

$$\ln \lambda_n = \frac{(m_{1x} - m_{2x}) \left[\sum_{i=1}^n x_i - \frac{n}{2}(m_{1x} + m_{2x}) \right]}{\sigma_x^2} + \frac{(m_{1y} - m_{2y}) \left[\sum_{i=1}^n y_i - \frac{n}{2}(m_{1y} + m_{2y}) \right]}{\sigma_y^2}$$

Рассматриваем неравенства $\ln \lambda_n \geq \ln A$ и $\ln \lambda_n \leq \ln B$ и принимаем решение о прекращении или продолжении наблюдений. Хотя выражения получились более громоздкие, чем в одномерном случае, их структура не изменилась: останавливающие границы будут представлять собой две параллельные прямые.

Границы зоны неопределенности в процессе принятия решений по выборке фиксированного объема, когда решающая функция есть прямая линия, описываются гиперболами. Наименьшая ширина зоны неопределенности в данном случае будет в окрестности средних значений признаков \bar{x} и \bar{y} . По мере удаления от точки с координатами (\bar{x}, \bar{y}) зона неопределенности будет резко расти.

Для других видов решающих функций (признака y) $y=f(x_1, \dots, x_m)$ от случайных значений признаков x_1, x_2, \dots, x_m интервальная оценка также находится с помощью вычисления дисперсии значений \hat{y} . Дисперсия $D(\hat{y})$ может быть определена по формуле

$$D(\hat{y}) = \sum_{i=1}^m \left(\frac{\partial y}{\partial x_i} \right)^2 D(x_i) + \sum_{i>j} \frac{\partial^2 y}{\partial x_i \partial x_j} D(x_i, x_j)$$

где $D(x_i, x_j)$ - корреляционный момент для случайных величин x_i и x_j .

Итак, показано, что в любом методе принятия решений из-за погрешностей исходных данных образуется область неопределенности со своими останавливающими границами: верхними $D_g(x)$ и нижними $D_n(x)$. Это могут быть или границы интервальных оценок решающих функций, или обычные останавливающие границы: в методе последовательного принятия решений. Пусть решающая граница между i и j -областями $D(x) = D_i(x) - D_j(x) = 0$.

Тогда для любой задачи принятия решений (для последовательных методов или при фиксированном объеме выборки) справедлив следующий алгоритм:

$$\begin{cases} \text{если } D(x) \geq D_g(x), \text{ то } x \approx \omega_i; \\ \text{если } D(x) \leq D_n(x), \text{ то } x \approx \omega_j; \\ \text{если } D_n(x) < D(x) < D_g(x), \text{ то следует продолжить наблюдения.} \end{cases}$$

Список литературы.

1. Альберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения: Пер. с англ. / Под ред. С.Б.Стечкина. - М.: Мир, 1972. - 316 с.
2. Бард Й. Нелинейное оценивание параметров. М.: Статистика, 1979. - 349 с.
3. Бартон Д., Вард Г. Справочник по радиолокационным измерениям. - М.: Сов. радио, 1976. - 392 с.
4. Борисов А.И., Алексеев А.В. и др. Обработка нечеткой информации в системах принятия решений. - М., 1989.
5. Борисов А.Н., Крумберг О.А., Федоров И.П. Принятие решений на основе нечетких моделей. Рига: Зинатие, 1990. - 184 с.
6. Вентцедь Е.С. Теория вероятностей. Изд. 4-е. - М.: Наука, 1969. - 576 с.
7. Вопросы статистической теории распознавания. /Под ред. Б.В.Варского. - М.: Сов. радио. 1967. - 400 с.
8. Гельман Л.М. В докл. Междунар. конф. "Современная радиолокация", вып. 1, Киев, 1994.
9. Гончарский А.В., Леонов А.С., Ягола А.Г. Некоторое обобщение принципа невязки для случая оператора, заданного с ошибкой. - Докл. АН СССР, 1972, т. 203, N 6, с. 1238-1239.
10. Горелик А.Л., Скрипкин В.А. Методы распознавания. М.: Высшая школа, 1977. - 222 с.
11. Горелик А.Л., Скрипкин В.А. Методы распознавания. М.: Высш. шк., 1989.
12. Горелик А.Л., Барабаш Ю.Л., Кривошеев О.В., Эпштейн С.С. Селекция и распознавание на основе локационной информации / Под ред. А.Л.Горелика - М.: Радио и связь, 1990. - 240 с.
13. Грешилов А.А. Анализ и синтез стохастических систем. Параметрические модели и конфлюентный анализ. М.: Радио и связь, 1990. - 320 с.
14. Грешилов А.А. Метод наименьших квадратов и элементы конфлюентного анализа - М.: Издательство МГТУ, 1980. - 67 с.
15. Грешилов А.А. Некорректные задачи цифровой обработки информации и сигналов. - М.: Радио и связь, 1984. - 161 с.
16. Грешилов А.А., Стакун В.А., Стакун А.А. Математические методы построения прогнозов. - М.: Радио и связь, 1997. - 112 с.
17. Де Гроот М. Оптимальные статистические решения / Пер. с англ. М.: Мир, 1974. - 491 с..
18. Дуда Р., Харт П. Распознавание образов и анализ сцен / Пер. с англ. М.: Мир, 1976. - 511 с.

19. Дюбуа Д., Прад А. Теория возможностей. - М., 1990.
20. **Жилинская Е.И., Товмаченко Н.Н., Федоров В.В.** Методы регрессионного анализа при наличии ошибки в предикторных переменных. М.: АН СССР, 1978. 34с.
21. **Журавлев Ю.И., Никифоров В.В.** - Кибернетика, 1971, N 3.
22. **Зельнер А.** Байесовские методы в эконометрии. М.: Статистика, 1980. - 438с.
23. **Искусственный интеллект.** Справочник, кн. 2. / Под ред. Д.А.Поспелова. - М.: Радио и связь, 1990.
24. **Кендалл М., Стьюарт А.** Статистические выводы и связи: Пер. с англ. / Под ред. А.Н.Колмогорова. - М.: Наука, 1973. - 590 с.
25. **Корн Г., Корн Т.** Справочник по математике для научных работников и инженеров: Пер. с англ. - М.: Наука, 1970. - 720 с.
26. **Круг Г.К., Кабанов В.А., Фомин Г.А., Фомина Г.С.** Планирование эксперимента в задачах нелинейного оценивания и распознавания образов. М.: Наука, 1981. - 172 с.
27. **Круг Г.К., Сосулин Ю.А., Фатуев В.А.** Планирование эксперимента в задачах идентификации и экстраполяции. М.: Наука, 1977. - 208 с.
28. **Кудрявцев Л.Д.** Курс математического анализа. Т. 2. / Изд-е 2-е. М.: Высш. шк., 1988. - 575 с.
29. **Леонов А.С.** Некоторые аспекты реализации регуляризующего алгоритма обобщенной невязки. В кн.: Обработка и интерпретация физических экспериментов. - М.: МГУ, 1976, вып. 4, с. 69-81.
30. **Линник Ю.В.** Метод наименьших квадратов и основы математико-статистической теории обработки информации. М.: Физматгиз, 1958. - 333 с.
31. **Митрофанов Д.Г., Ермоленко В.П.** Распознавание воздушных целей за счет измерения их пространственной протяженности. Зарубежная радиоэлектроника, 1996, N 1, с. 53-56.
32. **Модели. Алгоритмы. Принятие решений:** Сб. статей. - М.: Наука, 1989.
33. **Небабин В.Г., Белоус О.И.** Методы и техника противодействия радиолокационному распознаванию. Зарубежная радиоэлектроника, 1987, N 2.
34. **Небабин В.Г., Гришин В.К.** Методы и техника радиолокационного распознавания: Современное состояние, тенденции развития, перспективы. - Зарубежная радиоэлектроника, 1992, N 10.
35. **Небабин В.Г., Сергеев В.В.** Методы и техника радиолокационного распознавания. - М.: Радио и связь, 1984. - 152 с.
36. **Патрик Э.** Основы теории распознавания образов: Пер. с англ. - М.: Сов. радио, 1980.
37. **Персептрон** система распознавания образов / Под ред. А.Г.Ивахненко. - Киев: Наук. думка, 1975.

38. **Распознавание образов.** Состояние и перспективы: Пер. с англ. / Под ред. И.Б.Гуревича. - М.: Радио и связь, 1985.
39. **Репин В.Г., Тартаковский Г.П.** Статистический синтез при априорной неопределенности и адаптация информационных систем. - М.: Сов. радио, 1977.
40. **Селекция и распознавание на основе локационной информации** / Под ред. А.Л.Горелика. - М.: Радио и связь, 1990.
41. **Сосулин Ю.Г., Фишман М.М.** Теория последовательных решений и ее применения. - М.: Радио и связь, 1985.
42. **Стайнберг Б.Д.** Формирование радиолокационного изображения самолета в диапазоне СВЧ. - ТИИЭР, 1988, т. 76, N 12, с. 26-46.
43. **Статистические методы в экспериментальной физике:** Пер. с англ. / Под ред. А.П.Тяпкина. - М.: Атомиздат, 1976. - 335 с.
44. **Стронгин Р.Г.** Численные методы в многоэкстремальных задачах. - М.: Наука, 1978. - 239 с.
45. **Трубицын Е.Г., Ермоленко В.П.** Алгоритм распознавания сложных воздушных целей. - Зарубежная радиоэлектроника, 1992, N 10, с. 82-84.
46. **Уилкс С.** Математическая статистика / Пер. с англ. М.: Наука, 1967. - 632 с.
47. **Уотермен Д.** Руководство по экспертным системам. - М., Мир, 1989.
48. **Успенский А.Б., Федоров В.В.** Вычислительные аспекты МНК при анализе и планировании регрессионных экспериментов. - М.: Изд-во Моск. ун-та. 1975. - 168 с.
49. **Федоров В.В.** Теория оптимального эксперимента. М.: Наука, 1971. - 312 с.
50. **Финкельштейн М.И., Мендельсон В.А., Кутев В.Л.** Радиолокация слоистых земных покровов. - М.: Сов. радио, 1977.
51. **Фомин Я.А., Тарловский Г.Р.** Статистическая теория распознавания образов. - М.: Радио и связь, 1986.
52. **Фор А.** Восприятие и распознавание образов: Пер. с франц. - М.: Машиностроение, 1989.
53. **Франклин Д., Кармоди У** и др. - ТИИЭР, 1988, т.78, N 10.
54. **Фу К.** Последовательные методы в распознавании образов и обучении машин: Пер. с англ. - М.: Наука, 1971. - 255 с.
55. **Фу К.** Структурные методы в распознавании образов. Пер. с англ. - М.: Мир, 1977.
56. **Фукунага К.** Введение в статистическую теорию распознавания: Пер. с англ. - М.: Наука, 1979.
57. **Хармут Х.Ф.** Несинусоидальные волны и радиолокации и радиосвязи. - М.: Радио и связь, 1985.
58. **Химмельблау Д.** Анализ процессов статистическими методами: Пер. с англ. - М.: Мир, 1973. 957 с.

59. **Худсон Д.** Статистика для физиков: Пер. с англ. / Под ред. Е.М.Лейкина. - М.: Мир, 1970. - 296 с.
60. **Шишов С.А.** Классификаторы на основе нейронных структур. Зарубежная радиоэлектроника, 1992, N 8.
61. **Юдин Д.Б., Гольштейн Е.Г.** Линейное программирование. Теория и конечные методы. - М.: Физматгиз, 1963. - 775 с.
62. **Яноши Л.** Теория и практика обработки результатов измерений: Пер. с англ. / Под ред. Н.П.Клепикова. - М.: Мир, 1968. - 462 с.
63. **Frisch R.** Statistical confluence anakysis by means of complete regression systems. - Oslo, 1934. - 192 s.
64. **Fuller W.A.** Measurement error models. New york ect.: Wiley, 1987. - 440 p.
65. **Gleser L.J.** Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. // Statist. Anal. Meas. Error Models and Appl.: Proc. AMS-IMS-SIAM. It Summer Res. Conf., Arcata, Calif., June 10-16, 1989. Providence, 1990. P. 99-114.
66. **Jefferys W.H.** Robust estimation when more then one variable per equation of condition has error // Biometrika. 1990. V 77. P. 597-607.
67. **Lyons Louis.** On estimating systematic errors from repeated measurements // J. Phys. A. 1992. V. 25. P. 1967-1979.
68. **Parallel Distributed Processing / Ed. by Rummelhart D., Mc. Clelland J., v. I.** - Cambridge MA: MJT Press, 1988.
69. **Spiegelman Cl.H.** Plotting techniques for error-in-variables problems // Statist. Anal. Meas. Error Models and Appl.: Proc. AMS-IMS-SIAM. It Summer Res. Conf., Arcata, Calif, June 10-16, 1989. Providence, 1990. P. 167-168.
70. **Schneeweis H., Mittag H-J.** Lineare Modelle mit fehler behafteten Daten.-Physica-Verlag. Heidelberg Wien, 1986, P. 504.
71. **Schater D.W.** Measurement error model estimation using iteratively weighted teast squares // Statist. Anal. Meas. Error Models and Appl.: Proc. AMS-IMS-SIAM. It Summer Res. Conf., Arcata, Calif., June 10-16, 1989. Providence, 1990. P. 129-138.
72. **Yohai V.J., Zamar R.H.** Bounded influence estimation in the error-in-variables model // Statist. Anal. Meas. Error Models and Appl.: Proc. AMS-IMS-SIAM. It Summer Res. Conf., Arcata, Calif., June 10-16. 1989. Providence, 1990. P. 243-248.
73. **Werbos P.** -Proc. IEEE, 1990, v. 78, N 10.
74. **Widrow B. and Lehr M.A.** - Proc. IEEE, 1990, v. 78, N 9.
75. **Williamson J.H.** Least-squares fitting of straight line. Canad. J. Phys., 1968, 46, 16, 1845.
76. **Whittemore A.S.** Error-variables regression problems in epidemiology // Statist. Anal. Meas. Error Models and Appl.: Proc. AMS-IMS-SIAM. It Summer Res. Conf., Arcata, Calif., June 10-16, 1989. Providence, 1990. P. 17-31.

ОГЛАВЛЕНИЕ

Введение	3
ГЛАВА 1. АНАЛИЗ ТРАДИЦИОННЫХ МЕТОДОВ В СТАТИСТИЧЕСКИХ ЗАДАЧАХ ПРИНЯТИЯ РЕШЕНИЙ И МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ	6
1.1. Основные понятия и определения	6
1.2. Статистические задачи решения с наблюдениями	7
1.3. Статистическая классификация при фиксированном объеме выборки	12
1.4. Методы детерминистской классификаций	16
1.5. Последовательная решающая модель для классификации образов	18
1.6. Байесова последовательная решающая процедура	22
1.7. Последовательные решения для двухточечной модели	23
1.8. Байесовские методы обучения	27
1.9. Обучение с помощью стохастической аппроксимации	30
1.10. Математическая постановка задачи исследований	32
ГЛАВА 2. ПРИНЯТИЕ РЕШЕНИЙ ПО ВЫБОРКЕ ФИКСИРОВАННОГО ОБЪЕМА С УЧЕТОМ ПОГРЕШНОСТИ ПРИЗНАКОВ В ОБОБЩЕННЫХ УСЛОВНЫХ ПЛОТНОСТЯХ ВЕРОЯТНОСТИ	39
2.1. Анализ методов оценки параметров функции известного вида с учетом ошибок в значениях функций и аргументов	39
2.2. Задача оценивания свободных параметров функций в пассивной схеме наблюдений	46
2.3. О единственности оценок параметров. Состоятельность оценок и алгоритм их получения.	50
2.4. Статистические свойства свободных параметров функции Гаусса, определенных через параметры уравнения прямой линии и непосредственно.	56
2.5. Оценка параметров функции плотности вероятности с учетом погрешности измерения вектора признаков	61
2.6. Классификация образов по измеренному с ошибкой вектору признаков	69
2.7. Классификация летательных аппаратов с учетом погрешностей в измерениях признаков	72

ГЛАВА 3. СТАТИСТИЧЕСКИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ ПРИ НЕИЗВЕСТНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ИЗМЕРЕННЫХ ЗНАЧЕНИЙ ПРИЗНАКОВ	77
3.1. Оценка значений параметров классификаторов по выборке фиксированного объема	77
3.2. Обобщенные линейные разделяющие функции	81
3.3. Оценка разделяющего вектора с помощью процедур линейного и квадратичного программирования	86
3.4. Разделяющие функции для случая многих классов (образов)	88
3.5. Учет погрешностей наблюдений при оценке значений параметров классификаторов по выборке	89
3.6. Оценка зоны неопределенности в процедурах линейного программирования	93
3.7. Распознавание образов по измеренному с погрешностью вектору признаков	95
3.8. Учет интервальных оценок функций плотности вероятности в последовательных методах распознавания образов	98
3.9. Сравнение зон неопределенности в задачах принятия решений. Общий алгоритм принятия решений.	102
Список литературы	107