

Методическая разработка
для проведения лекции

Занятие 12. Непараметрические методы распознавания

Учебные вопросы занятия:

1. Оценка функций плотности распределения
2. Метод линейной комбинации базисных функций при непараметрическом распознавании образов

Заключительная часть

Введение.

Решение задачи распознавания с использованием статистических методов предполагает наличие определенного объема исходных данных, в качестве которых выступают плотность распределения вероятностей $\omega(x/A_i)$ признака x , априорная вероятность $P(A_i)$ появления событий класса A_i , величина потерь L_{ij} , связанных с отнесением анализируемого значения признака x к классу j , когда в действительности он принадлежит классу i . При этом наличие исходных данных определенного объема определяет выбор критерия принятия решения при распознавании. Вместе с тем, знание плотности распределения вероятностей для каждого класса является обязательным условием реализуемости статистических методов распознавания, поэтому очевидно, что *оценка плотностей – одна из основных проблем распознавания.*

1. Оценка функций плотности распределения

Если в результате проведения анализа наблюдаемой совокупности выборочных значений (или на основе имеющейся априорной информации) оказывается возможным хотя бы с некоторым приближением установить вид закона их распределения, то для получения функции плотности распределения требуется определить априорно-неизвестные параметры. Методы распознавания, применяемые в этих условиях, носят название *параметрических*.

В наиболее общем случае отсутствия априорных сведений не только о параметрах, но и самом виде закона распределения наблюдаемой совокупности выборочных значений, априорная неопределенность носит название непараметрической, а сами методы распознавания, применяемые в этом случае, именуются *непараметрическими*.

При параметрическом распознавании ПРВ информативного признака формируется путем выбора наиболее близкого к анализируемому процессу закона распределения с точки зрения одного из критериев согласия с последующим установлением параметров указанного закона. Эти параметры могут быть определены либо на основе анализа физических закономерностей рассматриваемого процесса, либо путем получения статистических оценок.

Оценка – приближенное значение величины или параметра, найденное по экспериментальным данным.

Для оценивания ПРВ и ее параметров необходимо сформировать обучающую выборку, требования к объему выборки определяются заданной точностью и надежностью оценивания закона распределения и его параметров. При этом для оценивания параметров законов распределения необходимо использовать такие аналитические выражения, которые удовлетворяют основным требованиям, предъявляемым к оценкам.

Вопрос: Какие требования предъявляются к оценкам? (состоятельность, эффективность, несмещенность).

Наибольшее распространение при описании информативных признаков в условиях РР получил нормальный закон распределения

$$\omega(x) = \frac{1}{\sqrt{2\pi\sigma[x]}} e^{-\frac{1}{2} \frac{(x-m[x])^2}{\sigma^2[x]}} \quad (1)$$

Вопрос: Какие параметры полностью описывают указанный закон распределения? (математическое ожидание и дисперсия)

$$m[x] \approx \frac{1}{N} \sum_{j=1}^N x_j, \quad \sigma^2[x] \approx \frac{1}{N-1} \sum_{j=1}^N (x_j - m[x])^2, \quad (2)$$

где N – объем анализируемой выборки.

После того, как значения параметров установлены, они подставляются в аналитические выражения соответствующих ПРВ, которые используются в отношении правдоподобия для получения порога разделения гипотез (уравнения разделяющей поверхности)

$$\frac{\omega(x/A_1)}{\omega(x/A_2)} = \frac{\frac{1}{\sqrt{2\pi\sigma_1[x]}} e^{-\frac{1}{2} \frac{(x-m_1[x])^2}{\sigma_1^2[x]}}}{\frac{1}{\sqrt{2\pi\sigma_2[x]}} e^{-\frac{1}{2} \frac{(x-m_2[x])^2}{\sigma_2^2[x]}}} \geq 1. \quad (3)$$

При непараметрическом распознавании используются не сами плотности вероятности $\omega(\bar{x}/A_i)$, а их оценки $\tilde{\omega}(\bar{x}/A_i)$.

Прежде чем излагать методы оценки плотностей распределения, целесообразно остановиться на мотивах выбора плотности распределения определенного вида. Основой для обсуждения будет являться понятие энтропии.

Энтропия (от греч. поворот, превращение) – мера неопределенности какого-либо испытания (опыта), который может иметь разные исходы.

Энтропия совокупности образов с плотностью распределения $\tilde{\omega}(\bar{x}/A_i)$ определяется как

$$H = - \int_{\bar{x}} \tilde{\omega}(\bar{x}/A_i) \cdot \ln \tilde{\omega}(\bar{x}/A_i) d\bar{x}. \quad (4)$$

Принцип максимума энтропии утверждает, что если плотность распределения некоторой случайной величины неизвестна, то из логических соображений следует выбрать такую плотность распределения, которая обеспечивает максимизацию энтропии случайной величины при учете всех известных ограничений.

Применение этого критерия приводит к решению, отличающемуся минимальным смещением, так как плотность распределения любого другого вида будет обладать большим смещением «в сторону» информации, содержащейся в известном наборе данных. Плотность распределения, обеспечивающую максимум энтропии, особенно легко определить в тех случаях, когда все известные ограничения представлены в форме средних оценок, таких как математические ожидания и дисперсии плотности распределения.

Итак, пусть априорная информация о случайной величине \bar{x} задается в виде

$$\int_{\bar{x}} \tilde{\omega}(\bar{x}/A_i) d\bar{x} = 1 \quad (5)$$

$$\int_{\bar{x}} b_k(\bar{x}) \tilde{\omega}(\bar{x}/A_i) d\bar{x} = a_k, \quad \text{где } k = 1, 2, \dots, Q. \quad (6)$$

Наша цель заключается в таком задании плотности распределения $\tilde{\omega}(\bar{x}/A_i)$, чтобы величина энтропии при выполнении ограничений (5),(6) была максимальной.

Вопрос: Каким образом исследовать функцию на максимальное значение? (найти производную, приравнять ее к нулю).

Для решения указанной задачи можно воспользоваться классическим методом отыскания условного экстремума функций нескольких переменных, в частности, методом множителей Лагранжа.

Использование множителей Лагранжа $\lambda_0, \lambda_1, \dots, \lambda_Q$ с учетом имеющихся ограничений позволяет записать функцию Лагранжа в следующем виде:

$$F = - \int_{\bar{x}} \tilde{\omega}(\bar{x}/A_i) \cdot \left[\ln \tilde{\omega}(\bar{x}/A_i) - \sum_{\kappa=0}^Q \lambda_{\kappa} b_{\kappa}(\bar{x}) \right] d\bar{x} - \sum_{\kappa=0}^Q \lambda_{\kappa} a_{\kappa}, \quad (7)$$

где $a_0 = 1$ и $b_0(\vec{x}) = 1$ для всех образов \vec{x} .

Взяв частные производные от функции F по плотности распределения $\tilde{\omega}(\vec{x}/A_i)$, имеем:

$$\frac{\partial F}{\partial \tilde{\omega}(\vec{x}/A_i)} = - \int_{\vec{x}} \left\{ \left[\ln \tilde{\omega}(\vec{x}/A_i) - \sum_{\kappa=0}^Q \lambda_{\kappa} b_{\kappa}(\vec{x}) \right] + 1 \right\} d\vec{x}. \quad (8)$$

Приравняв подинтегральное выражение к нулю и выразив из этого уравнения $\tilde{\omega}(\vec{x}/A_i)$, получим:

$$\tilde{\omega}(\vec{x}/A_i) = \exp \left[\sum_{\kappa=0}^Q \lambda_{\kappa} b_{\kappa}(\vec{x}) - 1 \right]. \quad (9)$$

Здесь $Q+1$ параметров $\lambda_0, \lambda_1, \dots, \lambda_Q$ следует выбирать так, чтобы они соответствовали априорной информации о случайной величине \vec{x} , содержащейся в отношениях (5), (6). Исходя из выражения (9) следует, что когда известно, что случайная величина отлична от нуля только в конечном интервале, следует выбирать равномерное распределение.

Если случайная величина может принимать любое действительное значение, а единственными характеристиками считаются математическое ожидание и дисперсия, то следует выбирать нормальное распределение.

Пример 1.

Пусть некоторая случайная величина отлична от нуля только в конечном интервале. Необходимо выбрать вид функции ПРВ на основании принципа максимума энтропии.

В нашем случае априорная информация о случайной величине заключается в том, что

$$\alpha \leq x \leq \beta \quad \text{и} \quad \int_{-\infty}^{\infty} \omega(x) dx = 1,$$

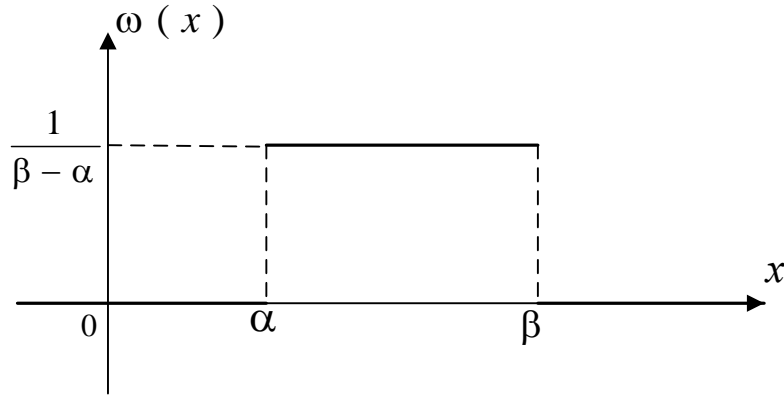
тогда из (9) следует, что

$$\tilde{\omega}(x) = \exp(\lambda_0 - 1).$$

Так как $\int_{\alpha}^{\beta} \exp(\lambda_0 - 1) dx = 1$, то $\exp(\lambda_0 - 1) \int_{\alpha}^{\beta} dx = \exp(\lambda_0 - 1)(\beta - \alpha) = 1$, а

$$\exp(\lambda_0 - 1) = \frac{1}{\beta - \alpha}, \text{ следовательно}$$

$$\tilde{\omega}(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{при } \alpha \leq x \leq \beta; \\ 0 & \text{при } x \notin (\alpha, \beta). \end{cases}$$



Применимость рассмотренного подхода ограничивается необходимостью представления исходных данных о ПРВ в виде математических оценок ее параметров, что не всегда реализуемо.

Таким образом, при непараметрическом оценивании необходимо построить оценку ПРВ, сходящуюся к истинной плотности с точки зрения какого-либо вероятностного критерия.

В настоящее время при непараметрическом оценивании большее распространение получили гистограммный метод, метод Парзена, полигональный метод, метод разложений по базисным функциям.

Гистограммный метод применяется обычно для априорно непрерывных плотностей вероятности, отличных от нуля на всем рассматриваемом множестве значений аргумента \vec{x} , и в случае, когда объем выборки достаточно велик.

Область возможных значений признаков разбивается на непересекающиеся одинаковые или различные по размерам области I_L группировки наблюдений (интервалы при $p=1$ или параллелепипеды при $p>1$), подсчитывается число наблюдений K_L обучающей выборки объема N , попавших в L -ю область и строится оценка плотности в виде функции

$$\tilde{\omega}(\vec{x}) = \Gamma(\vec{x}) = \frac{1}{N} \sum_L (k_L(\vec{x})/V_L), \quad (10)$$

где $k_L(\vec{x}) = K_L$, если $\vec{x} \in I_L$ и $k_L(\vec{x}) = 0$, если $\vec{x} \notin I_L$,

V_L – объем L -й области I_L .

В одномерном случае $\Gamma(\vec{x})$ представляет собой ступенчатую функцию, сумма площадей отдельных столбиков гистограммы равна единице с учетом масштаба оси абсцисс.

Гистограммный метод оценивания плотности следует отнести к асимптотическим (приближенным). При конечных объемах обучающих выборок он не позволяет получить оценку точности приближения $\Gamma(\bar{x})$ к $\omega(\bar{x})$ и, следовательно, вычислить вероятности ошибочных решений, характеризующие качество распознавания.

При группировке данных приходится учитывать два взаимно исключающих обстоятельства. С одной стороны, число интервалов группировки должно быть достаточно велико, чтобы детально описать поведение плотности распределения, но, с другой стороны, число точек, попадающих в интервал, также должно быть достаточным, чтобы надежно представлять данный интервал. Если интервалов будет много, то некоторые из них могут оказаться пустыми, а плотность распределения – изрезанной, многомодальной.

В литературе имеется много рекомендаций по выбору оптимального числа интервалов группировки. Рассмотрим две распространенные формулы.

Формула Старджеса:

$$L = 1 + \lfloor \log_2 N \rfloor. \quad (11)$$

Формула Брукса и Каррузера:

$$L = \lfloor \sqrt{N} \rfloor. \quad (12)$$

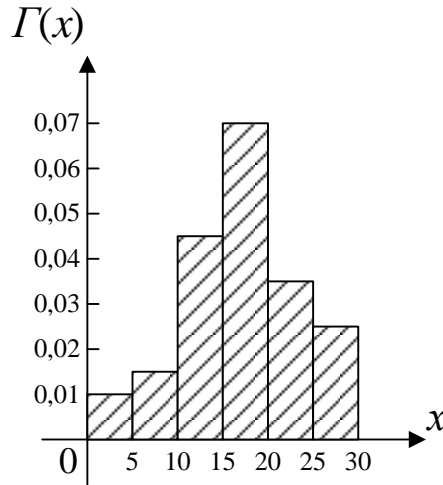
При малых объемах выборки ($N \leq 20$) некоторые интервалы могут оказаться пустыми. В таком случае необходимо уменьшить их количество. В таких условиях считается приемлемым выбирать число интервалов так, чтобы выполнялось условие $K_L \leq 5$.

Пример 2.

Пусть существует некоторая случайная величина x , представленная выборкой значений

x	(0, 5]	(5, 10]	(10, 15]	(15, 20]	(20, 25]	(25, 30]
K_L	10	15	45	70	35	25

Необходимо оценить ПРВ гистограммным методом.



Для одномерного случая часто используют полигональные оценки.

Полигональные оценки получают путем сглаживания гистограммы, соединяя прямыми линиями крайние левые, средние или крайние правые точки верхов столбцов, в результате получаются кусочно-линейные функции (ломанные).

Подобные аппроксимации необязательно имеют вид обычных ломаных, концы отдельных прямых могут не совпадать, а соединяться вертикальными прямыми. Плавную кривую можно получить, используя локальную аппроксимацию кривой полиномами или сглаживание результатов.

Развитием гистограммного метода является *метод Парзена*. Он обеспечивает более гладкую аппроксимацию плотности распределения вероятностей.

Данная функция связана с функцией распределения через оператор дифференцирования

$$\omega(x) = dF(x)/dx. \quad (13)$$

Это уравнение берется за основу при построении оценки $\tilde{\omega}(x)$ плотности распределения вероятности. Заменяем в предыдущем уравнении производную конечной разностью

$$\tilde{\omega}(x) \approx \frac{F(x+h) - F(x-h)}{2h}, \quad (14)$$

а функции распределения $F(x+h)$, $F(x-h)$ – их простейшими кусочно-постоянными оценками:

$$F_N(x+h) = \frac{1}{N} \sum_{i=1}^N 1(x+h-x_i), \quad (15)$$

$$F_N(x-h) = \frac{1}{N} \sum_{i=1}^N 1(x-h-x_i), \quad (16)$$

построенными на основе выборки x_1, x_2, \dots, x_N для одномерной случайной величины X , где h – интервал произвольной длины, $1(z)$ – единичная функция:

$$1(z) = \begin{cases} 1, z \geq 0, \\ 0, z < 0. \end{cases}$$

Оценка плотности тогда приобретает вид

$$\omega_N(x) = \frac{\frac{1}{N} \sum_{i=1}^N 1(x+h-x_i) - \frac{1}{N} \sum_{i=1}^N 1(x-h-x_i)}{2h} = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} \frac{1(x+h-x_i) - 1(x-h-x_i)}{2}. \quad (17)$$

Под знак суммы этой оценки входит *селектирующее прямоугольное ядро*

$$\frac{1(x+h-x_i) - 1(x-h-x_i)}{2} = I\left(\frac{x-x_i}{h}\right), \quad (18)$$

где

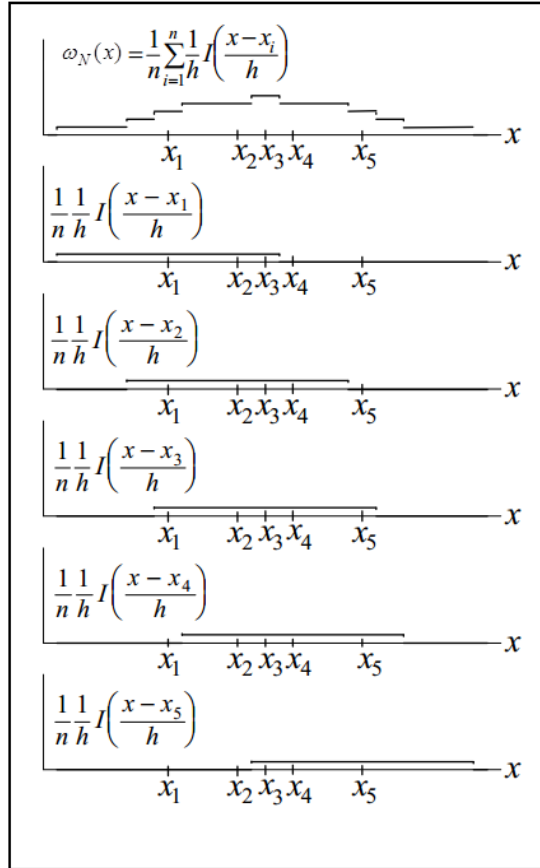
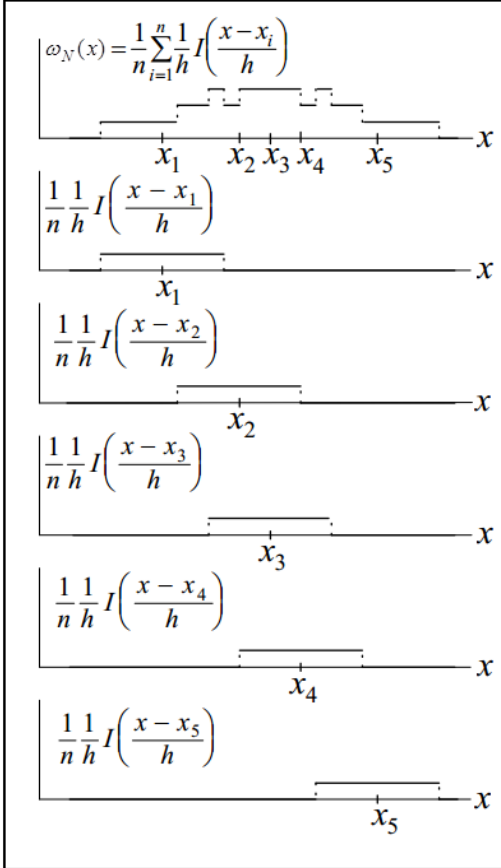
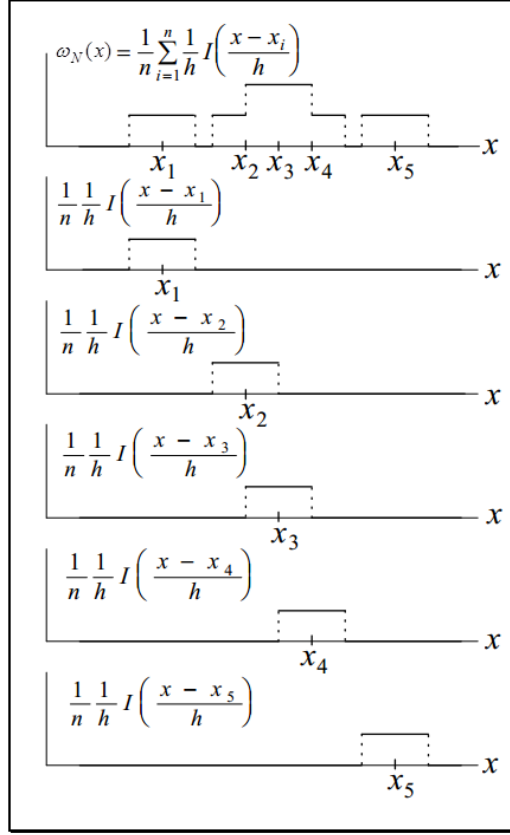
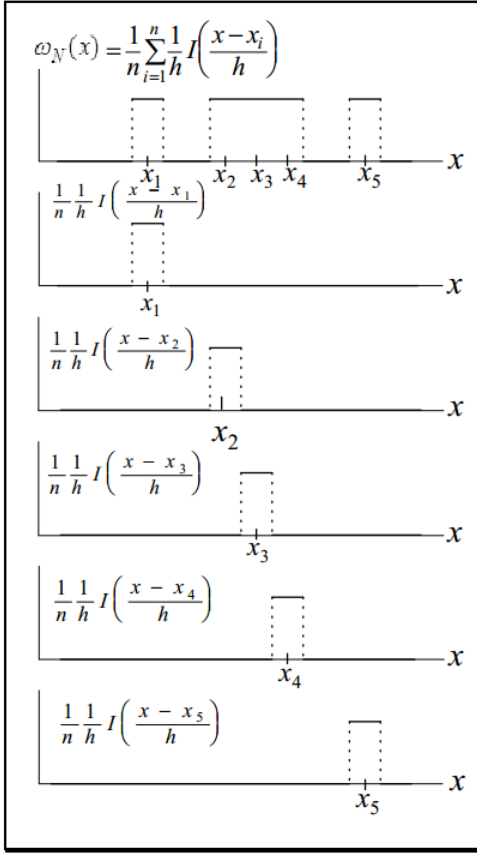
$$I(z) = \begin{cases} 0,5, |z| \leq 1, \\ 0, |z| > 1. \end{cases}$$

Ядро $I(z)$ симметричное, и площадь под кривой $I(z)$ равна единице. Эта функция ничем не отличается от плотности распределения вероятности случайной величины z , равномерно распределенной в интервале $[-1; 1]$. В итоге оценка для плотности распределения вероятности имеет вид

$$\omega_N(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} I\left(\frac{x-x_i}{h}\right). \quad (19)$$

Впервые оценка предложена в 1956 году Розенблаттом. Дальнейшие обобщения данного вида оценок осуществлены Парзеном. В настоящее время специалистами принято именовать данный класс оценок оценками Розенблатта – Парзена.

На рисунках представлен пример формирования оценки Розенблатта – Парзена при различной ширине h прямоугольных ядер.



На каждом рисунке даны итоговая оценка и все ее составляющие. С ростом h сглаживающие свойства оценки нарастают. По h для каждого конечного объема выборки существует некоторый оптимум, так как при малых h оценка представляет собой набор непересекающихся (или слабо пересекающихся) функций и оценка теряет свой смысл, а при большом h оценка становится сильно заглаженной и не отражает индивидуальных особенностей плотности распределения вероятности.

Степень гладкости оценки плотности зависит от степени гладкости ядра. Заменим в оценке $\omega_N(x)$ прямоугольное ядро $I(z)$ на произвольное $K(z)$ и получим

$$\omega_N(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} K\left(\frac{x - x_i}{h}\right). \quad (20)$$

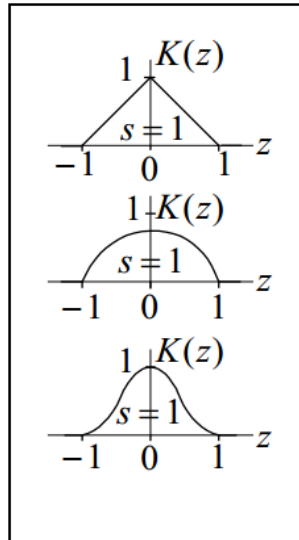
Здесь h – коэффициент размытости ядра. Примеры треугольного, параболического и кубического ядер представлены в следующем виде:

$$K(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & |z| > 1; \end{cases} \quad (21)$$

$$K(z) = \begin{cases} 0,75(1 - z^2), & |z| \leq 1, \\ 0, & |z| > 1; \end{cases} \quad (22)$$

$$K(z) = \begin{cases} (1 + 2|z|)(1 - |z|)^2, & |z| \leq 1, \\ 0, & |z| > 1. \end{cases} \quad (23)$$

Вид этих ядер изображен на рисунке.



Все они нормированы относительно единицы. За счет этого свойства оценка Розенблатта – Парзена удовлетворяет условию нормировки для плотности вероятности:

$$\int_{-\infty}^{\infty} \omega_N(x) dx = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-x_i}{h}\right) dx = \frac{1}{N} \sum_{i=1}^N \int_{-1}^1 K(z) dz = \frac{1}{N} \sum_{i=1}^N 1 = 1. \quad (24)$$

Выводы:

1. При решении задач мониторинга выбор класса методов распознавания определяется уровнем априорной неопределенности относительно закона распределения и его параметров для выбранной системы информативных признаков.

2. Отсутствие информации относительно закона распределения выбранной системы информативных признаков определяет выбор непараметрических методов распознавания, причем оценки ПРВ могут быть получены на основе анализа элементов обучающих выборок достаточного объема.

2. Метод линейной комбинации базисных функций при непараметрическом распознавании образов

Одним из наиболее развитых методов при непараметрическом распознавании является оценивание одномерной функции путем представления ее *линейной комбинацией известных функций*. Воспользуемся разложением оценки в ряд:

$$\tilde{f}(x) = \sum_{l=1}^L c_l \varphi_l(x). \quad (25)$$

При этом c_l , $l = \overline{1, L}$ – коэффициенты, подлежащие оцениванию по обучающей выборке, а система функций $\{\varphi_l(x)\}_{l=1}^L$ является ортогональным базисом

$$\int_{-\infty}^{\infty} \varphi_l(x) \varphi_k(x) dx = \begin{cases} 1, & \text{если } l = k; \\ 0, & \text{если } l \neq k. \end{cases} \quad (26)$$

Последнее условие обеспечивает при любом заданном L минимум интегрального квадратичного отклонения δ

$$\delta = \int_{-\infty}^{\infty} \left| f(x) - \sum_{l=1}^L c_l \varphi_l(x) \right|^2 f(x) dx, \quad (27)$$

если оценка вычисляется по формуле

$$c_l = \frac{1}{N} \sum_{i=1}^N \varphi_l(x_i), \quad (28)$$

где x_i – наблюдение обучающей выборки.

Однако, подобно предыдущим, метод разложения функции по базисным функциям обеспечивает лишь сходимость оценки $\tilde{f}(x)$ к $f(x)$ при неограниченном росте объема обучающей выборки и не позволяет дать хорошую оценку точности аппроксимации для конечных фиксированных N . Кроме того, многомерные обобщения данного метода досконально не изучены.

Выводы:

1. При использовании метода линейной комбинации базисных функций для оценивания закона распределения выбор системы базисных функций осуществляется в результате исследования характера поведения и области существования случайных величин, используемых в качестве информативных признаков.

2. Число используемых базисных функций определяется требуемой точностью оценивания закона распределения.

Заключительная часть.

Подвожу итоги занятия.

Рекомендованная литература:

1. Хемминг Р.В. Численные методы. – М.: Наука, 1972.
2. Левин Б.Р. Теоретические основы статистической радиотехники. – М.: Сов. радио и связь, 1974.
3. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978.