Лекция 7. Распознавание образов в условиях непараметрической неопределенности на основе обучения с учителем

Важный класс задач оценивания в интересах разработки алгоритмов распознавания связан с определением неизвестных плотностей и функций распределения вероятностей, т.е. в ситуации, когда **неизвестно практически все.**

Такие задачи называют задачами **непараметрического оценивания**, поскольку неизвестными являются не только параметры распределения, но сам вид плотности или функции.

Задача оценивания неизвестной плотности распределения $f_{\mathbf{x}}(x) = p(x) = p(x/\omega_*)$ обычно сводится к нахождению приближенной функции $\tilde{f}_{\mathbf{x}}(x) = \tilde{p}(x)$, которая бы обладала определенными свойствами, характеризующими качество аппроксимации. К ним можно отнести следующие свойства.

1. Сходимость по вероятности в точке x: оценка $\tilde{p}(x)$ должна быть состоятельной в этой точке, когда для любого $\varepsilon > 0$ выполняется $\lim\limits_{N\to\infty} P[|p(x) - \tilde{p}(x)| < \varepsilon] = 0$.

2. Сходимость в среднеквадратичном в точке x: оценка $\tilde{p}(x)$ должна быть асимптотически несмещенной в этой точке $\lim\limits_{N\to\infty} M[\tilde{p}(x)] = p(x)$ и $\lim\limits_{N\to\infty} M[|p(x) - \tilde{p}(x)|^2] = 0$.

Разделяют методы локального оценивания, при реализации которых оценка ищется в отдельных точках на основе хранения всей обучающей выборки, и методы аппроксимации плотности распределения вероятностей во всей возможной области значений.

1.Оценка плотности распределения вероятностей на основе метода Парзена

Прежде всего, рассмотрим случай оценивания плотности распределения одномерной СВ. Пусть $X^N = \{x^{(1)},...,x^{(N)}\}$, $x^{(i)} \in R^1$ независимые и одинаково распределенные наблюдения значений случайной величины $\mathbf{x}$.

Изначально рассмотрим возможную оценку функции распределения СВ. Ее естественно представить в следующем виде:

$$\tilde{F}_{\mathbf{x}}(x) = \left\{ \frac{l(X^N)}{N} = \frac{1}{N}\sum_{t=1}^{N} u^{(t)},\ \ u^{(t)} = 1 : x^{(t)} \le x,\ \ u^{(t)} = 0 : x^{(t)} > x \right\},$$

т.е. как относительное количество наблюдений, значения которых меньше x.

Величина $\tilde{F}_{\mathbf{x}}(x)$ является дискретной случайной величиной, которая распределена по биномиальному закону

$$\Pr\left(\tilde{F}_{\mathbf{x}}(x) = \frac{l}{N} \right) = C_N^l [F_{\mathbf{x}}(x)]^l [1 - F_{\mathbf{x}}(x)]^{N-l},$$

где $F_{\mathbf{x}}(x)$ является истинным значением оцениваемой функции распределения.

Выражение для $\tilde{F}_{\mathbf{x}}(x)$ является оценкой максимального правдоподобия функции распределения, при этом для ее математического ожидания и дисперсии выполняется

$$M[\tilde{F}_{\mathbf{x}}(x)] = \sum_{l=1}^{N} \frac{l}{N} C_N^l [F_{\mathbf{x}}(x)]^l [1 - F_{\mathbf{x}}(x)]^{N-l} = F_{\mathbf{x}}(x),$$

$$D[\tilde{F}_{\mathbf{x}}(x)] = \sum_{l=1}^{N} \left[\frac{l}{N} - F_{\mathbf{x}}(x) \right]^2 C_N^l [F_{\mathbf{x}}(x)]^l [1 - F_{\mathbf{x}}(x)]^{N-l} = \frac{F_{\mathbf{x}}(x)(1 - F_{\mathbf{x}}(x))}{N}.$$

Следовательно, оценка функции распределения является **несмещенной, состоятельной и имеет место сходимость в среднеквадратичном**.

Значительно сложнее обстоит ситуация с оценкой плотности распределения вероятностей. Поскольку теоретическая плотность $f_{\mathbf{x}}(x) = p(x)$ определяется как производная функции распределения, а в данном случае оценка $\tilde{F}_{\mathbf{x}}(x)$ является не дифференцируемой функцией, то напрямую так получить оценку плотности распределения $\tilde{p}(x)$ нельзя.

Однако можно предположить, что эту оценку можно записать как

$$\tilde{p}(x) = [\tilde{F}_{\mathbf{x}}(x+h) - \tilde{F}_{\mathbf{x}}(x-h)]/2h,$$

где h положительная скалярная величина, определяющая размер ячейки $(2h)$, в пределах которой истинная плотность меняется предположительно незначительно.

Остается вопрос о том, как выбирать размер этой ячейки в конкретных случаях. Очевидно, что если зафиксировать размер ячейки и увеличивать объем обучающей выборки X^N, то полученная величина будет равна отношению $\tilde{P} = k_N / 2hN$, где k_N – количество наблюдений, попавших в ячейку. Это отношение сходится к величине, определяющей вероятность попадания в ячейку

$$P = [F_{\mathbf{x}}(x+h) - F_{\mathbf{x}}(x-h)]/2h = \int_{x-h}^{x+h} F_{\mathbf{x}}(u)du,$$

что не совсем то, что мы хотим получить.

С другой стороны, если при фиксированном N уменьшать размер ячейки, приближая его к нулю, чтобы получить более близкое приближение плотности, может возникнуть ситуация, при которой в ячейку не попадает никаких наблюдений и оценка плотности становится равной нулю, что совершенно неприемлемо.

Отсюда следует, что размер ячейки должен быть функцией от $h = h(N)$, для которой требуется выполнение следующих условий:

$$\lim_{N\to\infty} h(N) = 0,\ \lim_{N\to\infty} k_N = \infty,\ \lim_{N\to\infty} k_N / N = 0.$$

Первое условие обеспечивает сходимость $P/2h$ к величине $p(x)$ при однородном сокращении ячеек.

Второе условие обеспечивает сходимость по вероятности (при $p(x) \neq 0$) эмпирического отношения частот $\tilde{P}$ к величине P.

Третье условие определяет то, что интересующая оценка является состоятельной оценкой плотности распределения, поскольку гарантирует, что в каждую ячейку попадает небольшое относительно N количество элементов обучающей выборки.

При обосновании метода Парзена выполняются следующие преобразования:

$$\tilde{p}(x)=\frac{1}{2h}\int_{x-h}^{x+h}d\tilde{F}_{\mathbf{x}}(\xi)=\int_{-\infty}^{+\infty}\frac{1}{h}\varphi\left(\frac{x-\xi}{h}\right)d\tilde{F}_{\mathbf{x}}(\xi)=\frac{1}{Nh}\sum_{i=1}^{N}\varphi\left(\frac{x-x^{(i)}}{h}\right),$$

где

$$\varphi\left(\frac{u}{h}\right)=\begin{cases}1/2, & |u/h|\le 1\\ 0, & |u/h|>1\end{cases}.$$

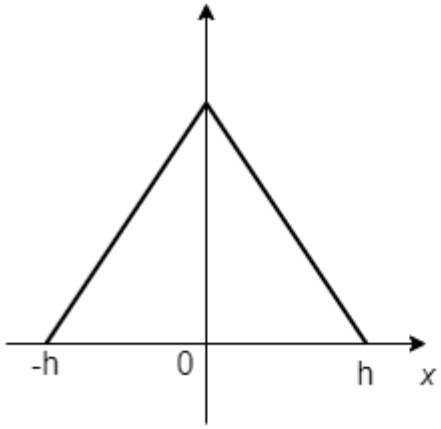
Функция $h^{-1}\varphi(u/h)$ **называется ядром**, или оконной функцией так как имеет прямоугольный вид, при этом ее форма определяет размер ячейки, относительно которой проводится оценка плотности.

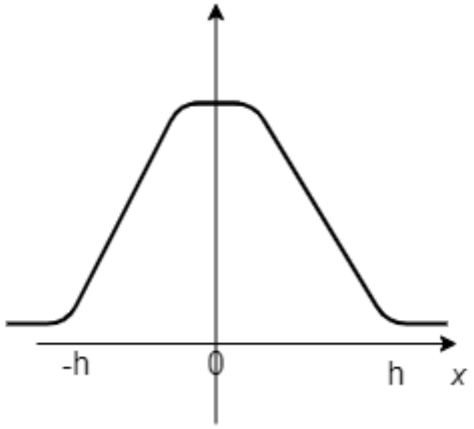
Важно отметить, что в качестве $\varphi(u/h)$ могут использоваться функции другой формы, обладающие схожими свойствами, а именно оценка плотности в методе Парзена обладает свойством асимптотической несмещённости, если для ядра и функции $h(N)$ выполняются следующие условия:

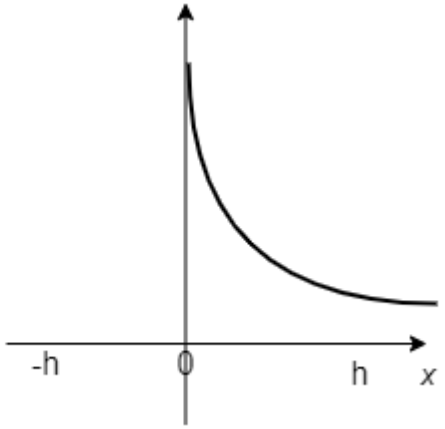
$$\int_{-\infty}^{\infty}h^{-1}\varphi(u/h)du=1,\quad \int_{-\infty}^{\infty}h^{-1}\left|\varphi(u/h)\right|du<\infty,\quad \sup\left|\varphi(u/h)\right|<\infty,\ \lim_{u/h\to\infty}(u/h)\varphi(u/h)=0,$$

$$\lim_{N\to\infty}h(N)=0,\qquad \lim_{N\to\infty}Nh(N)=\infty$$

Помимо оконной функции ядра, существует много ядер, удовлетворяющих данным условиям. К их числу относятся треугольная, гауссовская, показательная







$$\frac{1}{h}\varphi\left(\frac{u}{h}\right)=\begin{cases}\frac{1}{h}\left(1-|u/h|\right), & |u/h|\le 1\\ 0, & |u/h|>1\end{cases},\qquad \frac{1}{h}\varphi\left(\frac{u}{h}\right)=\frac{1}{\sqrt{2\pi}h}\exp\left(-\frac{u^2}{2h^2}\right),\qquad \frac{1}{h}\varphi\left(\frac{u}{h}\right)=\frac{1}{2h}\exp\left(-\left|\frac{u}{h}\right|\right).$$

Рассмотрим теперь обобщение метода Парзена на многомерный случай. Пусть теперь $X^N=\{x^{(1)},...,x^{(N)}\},\ x^{(i)}\in R^n$ независимые и одинаково распределенные наблюдения значений случайного вектора $\mathbf{x}$. Тогда для оценки плотности распределения применяются следующая оценка:

$$\tilde{p}(x)=\frac{1}{Nh^n}\sum_{i=1}^{N}\varphi\left(\frac{x-x^{(i)}}{h}\right)=\int\frac{1}{h^n}\varphi\left(\frac{x-\xi}{h}\right)d\tilde{F}_{\mathbf{x}}(\xi).$$

Условия сходимости оценки и ее другие свойства определяются аналогичным образом. Используемые ядра также являются аналогами ядер для одномерного случая.

Наиболее просто в многомерном случае использовать ядро в виде гауссианы

$$\frac{1}{h^n}\varphi\left(\frac{x-x^{(i)}}{h}\right)=\frac{1}{(2\pi)^{n/2}h^n\left|\Xi\right|^{1/2}}\exp\left(-\frac{(x-x^{(i)})^T\Xi^{-1}(x-x^{(i)})}{2h^2}\right),$$

где Ξ – симметричная неотрицательно определенная матрица, например, единичная.

Для выбора $h(N)$ при малой размерности вектора признаков рекомендуется формула:

$$h(N)=N^{-r/n},\ \ 0<r<1.$$

При большой размерности вектора признаков используются более сложные методики, среди которых наибольшее распространение получила методика максимизации функционала правдоподобия методом скользящего контроля (кросс-валидации).

Проще всего выбрать $\Xi=I$, т.е. как единичную матрицу. Тогда ядро будет иметь вид колокольной функции с осевой (вращательной) симметрией относительной оси ординат.

В качестве матрицы также Ξ можно использовать выборочную матрицу ковариации полученной выборки $\tilde{C}$. Если потом подставить ее в выражение для оценки плотности распределения и выполнить интегрирование будет получен следующий результат:

$$\Xi_r=\sum_{i=1}^{N}(x^{(i)}x^{(i),T}+h^2\tilde{C})=(1+h^2)\tilde{C}\,.$$

На рис.1 приведены графики истинной плотности и ее оценки для показательного распределения вероятностей при различных объемах обучающей выборки $N=100;1000;10000$ и ядра в виде гауссианы. На рис.2 приведены графики графики для ядра в виде экспоненты, а на рис.3 – для ядра в виде окна прямоугольной формы.

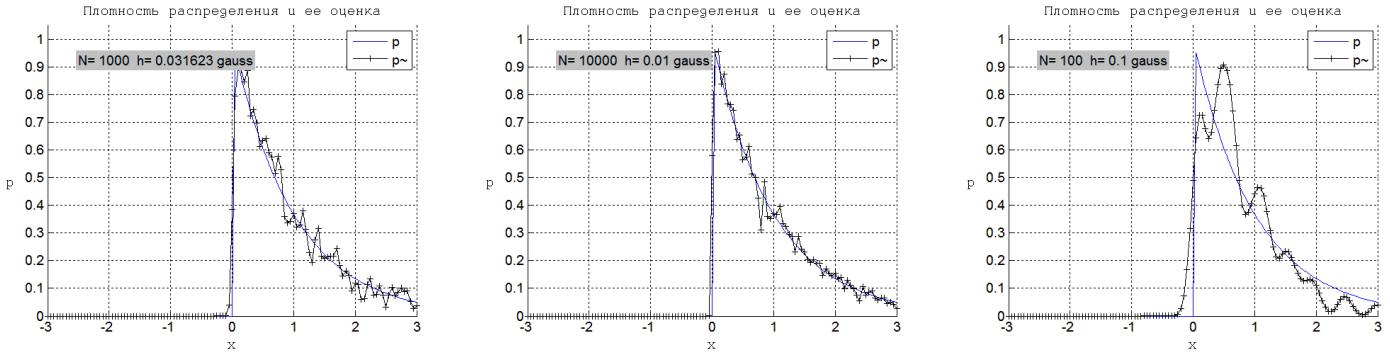


Рис.1. Примеры оценки плотности распределения с использованием ядра в виде гауссианы

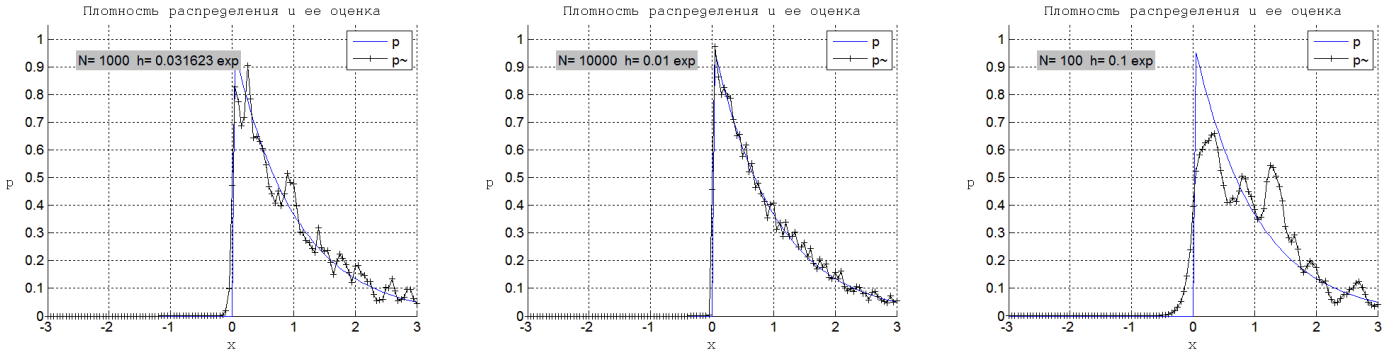


Рис.2. Примеры оценки плотности распределения с использованием ядра в виде экспоненциальной функции

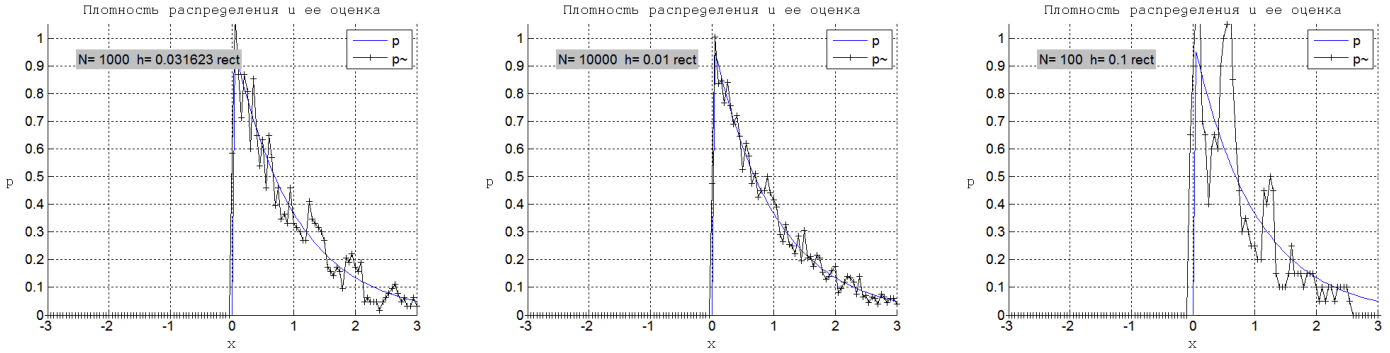
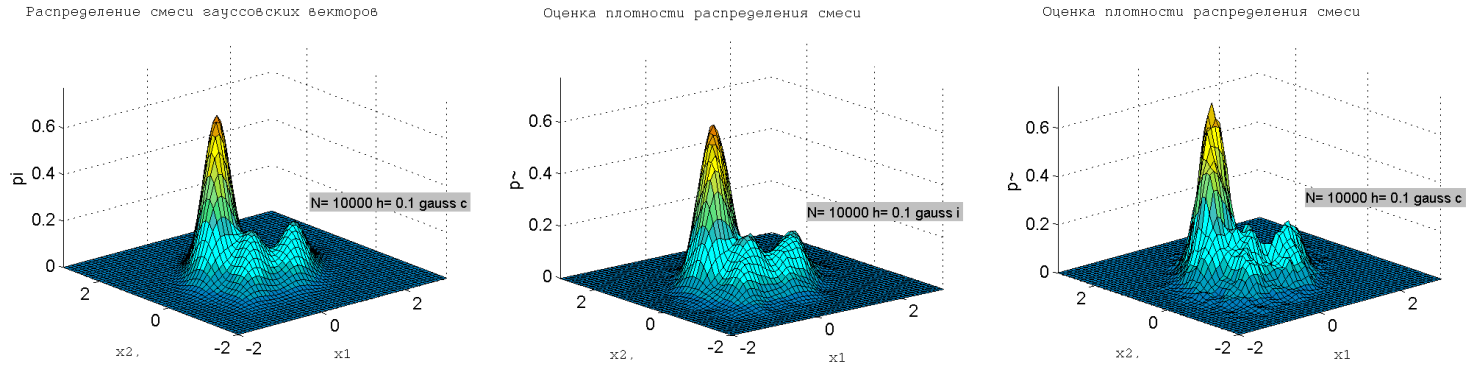


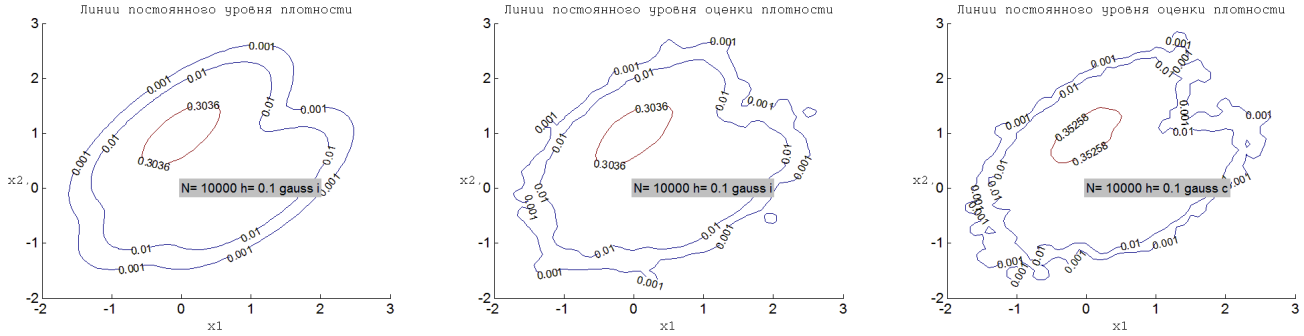
Рис.3. Примеры оценки плотности распределения с использованием ядра в виде оконной функции

На рис.4,5 приведены результаты восстановления многомерной плотности распределения.



а) б) в)

Рис.4. Пример оценки плотности распределения случайного вектора с использованием ядер в виде гауссианы различной формы: а) истинная восстанавливаемая плотность; б) восстановленная ядерная оценка при $\Xi = I$; в) восстановленная ядерная оценка при $\Xi = \tilde{C}$;



а) б) в)

Рис.5. Линии постоянного уровня плотности распределения случайного вектора и его оценок с использованием ядер в виде гауссианы а) для истинной плотности; б) для восстановленной оценки при $\Xi = I$; в) для восстановленной оценки при $\Xi = \tilde{C}$;

Следует также отметить, что здесь подбор параметра, определяющего размер ядра, специально не осуществлялся и для всех экспериментов использовались параметры $h(N) = N^{-r/n}$, $r = 0.5$.

Использование оценок на основе метода Парзена в задачах распознавания

Использование оценок Парзена в алгоритмах распознавания приводит к правилам стандартного вида. Конкретно, для случая двух классов при неизвестных априорных вероятностях решающее правило будет иметь вид

$$\frac{N_1}{N_1 + N_2}\tilde{p}(x/\omega_1) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{N_2}{N_1 + N_2}\tilde{p}(x/\omega_2).$$

При проведении анализа алгоритмов подобного типа условные вероятности ошибок и, в итоге, суммарная вероятность ошибки всех алгоритмов при однократном обучении являются случайными величинами, которые зависят от использованных реализаций обучающих данных.

2.Оценка плотности распределения вероятностей методом k-ближайших соседей

Другой классический метод непараметрической оценки плотности распределения реализует подход, который направлен на сокращение объема вычислений при построении аппроксимирующего ряда с использованием ядер. Данный метод свободен от необходимости подбора ядер и их параметров (размера ячеек) в зависимости объема обучающих данных и их локализации. Одна из возможных идей в этой ситуации состоит в том, чтобы **сделать объем ячеек переменным, т.е. сделать его функцией данных, а не их количества.**

Именно эта идея и реализована в методе k-соседей. Здесь так же, как и в методе Парзена, каждый элемент обучающей выборки $X^N = \{x^{(1)},...,x^{(N)}\},\ x^{(i)} \in R^n$ является центром ячейки.

Однако, **в отличие от метода Парзена**, объем ячейки для этого элемента устанавливается не фиксированным и одинаковым для всех элементов, а переменным, так чтобы вместить в него ровно k ближайших соседей – других элементов обучающей выборки.

При реализации алгоритма оценки плотности распределения в произвольной точке $x \in R^n$ определяется радиус $r = r(x, X^N)$ от этой точки до k-го ближайшего соседа. При определении близости можно использовать любую метрику. Мы дальше будем пользоваться евклидовой метрикой. Тогда оценка плотности распределения в этой точке определяется как

$$\tilde{p}(x) = \frac{k}{NV_n(k, N, X^N)},$$

где $V_n(k, N, X^N)$ – объем множества точек, лежащих в пределах расстояния $r = r(x, X^N)$ от x.

При использования евклидовой метрики этот объем определяется соотношением для гипершара

$$V_n = V_n(k, N, X^N) = \frac{2r^n \pi^{n/2}}{n\Gamma(n/2)},$$

где $\Gamma(n/2)$ – гипергеометрическая функция, которая при любом натуральном m равна

$$\Gamma(m) = (m-1)!, \quad \Gamma(m+1/2) = \frac{\sqrt{\pi}}{2^m}(2m-1)!!.$$

В частности $\Gamma(1/2)=\sqrt{\pi}$, $\Gamma(1)=1$, $\Gamma(2)=1$. Тогда: для $n=1$ очевидно $V_1=2r$; для $n=2$ – $V_2=\pi r^2$; для $n=3$ – $V_3=4\pi r^3/3$.

Если выполняются следующие предельные соотношения

$$\lim_{N\to\infty} k(N)=\infty,\ \lim_{N\to\infty} k(N)/N=0,$$

то для непрерывной плотности распределения $p(x)$ в точке x получаемая оценка является асимптотически несмещенной и состоятельной.

Приводятся следующие рекомендации для выбора параметра k в зависимости от N, обеспечивающие выполнение указанных условий $k=aN^{\gamma},\ 0<\gamma<1$ или $k=b\ln(N+1)$, где $a,\gamma,b-$ подбираемые константы.

Метод k ближайших соседей позволяет получить оценку исключительно просто, что делает его привлекательным в практическом плане.

Однако, при этом предполагается, что плотность распределения в пределах гипершара остается примерно постоянной, что может выполняться при использования достаточно малых радиусов $r=r(x,X^N)$. Следовательно, количество используемых соседей должно быть также достаточно малым, а это приводит к тому, что **точность оценки будет невелика при небольших объемах обучающей выборки** N**.**

Этот недостаток важен, если конечной целью является построение оценки плотности распределения.

Если оценка используется как вспомогательная для построения решающего правила в задаче распознавания, то указанным недостатком можно пренебречь, поскольку решающее правило оказывается не только весьма практичным, но и достаточно эффективным.

На рис.6 приведены графики истинной плотности и ее оценки для показательного распределения вероятностей при различных объемах обучающей выборки $N = 100; 1000; 10000$ и $k = 2N^{\gamma}$, $\gamma = 0.1$. На рис.7 приведены графики истинной плотности и ее оценки для $k = N^{\gamma}$, $\gamma = 0.5$.

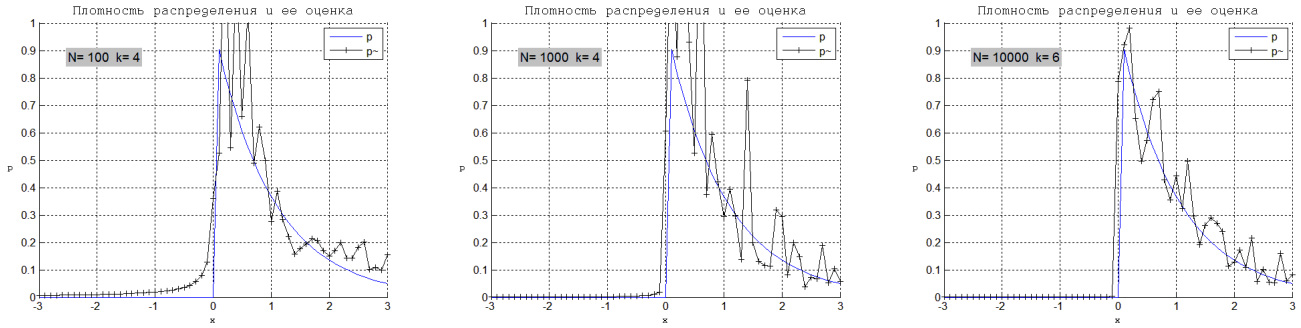


Рис.6. Примеры оценки плотности распределения с использованием настойки $k = 2N^{\gamma}$, $\gamma = 0.1$

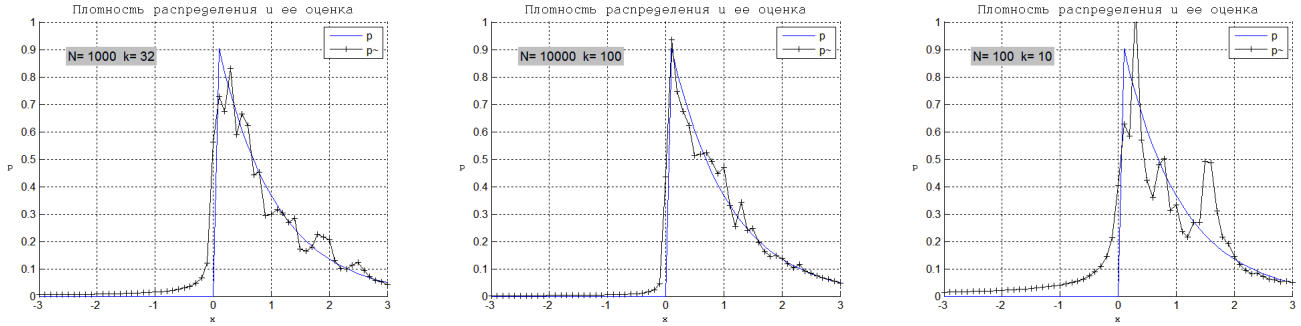
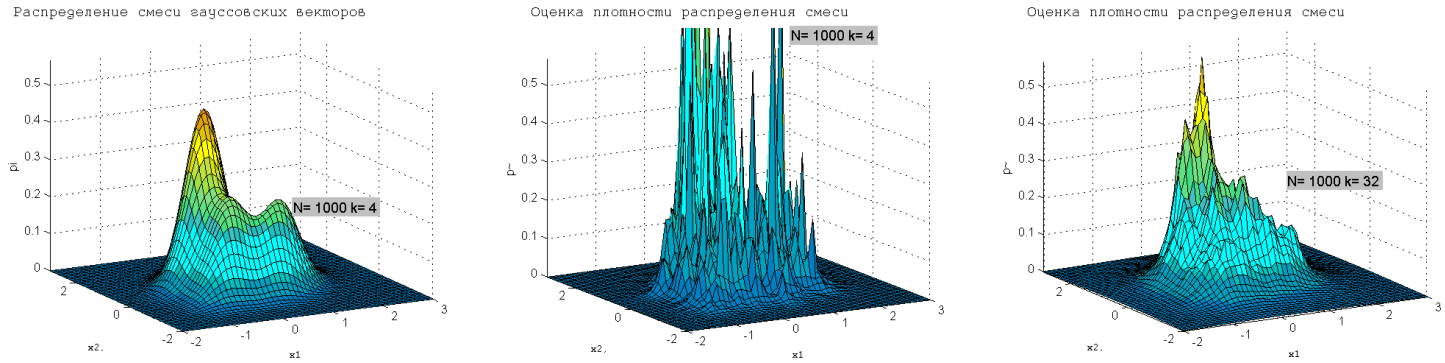


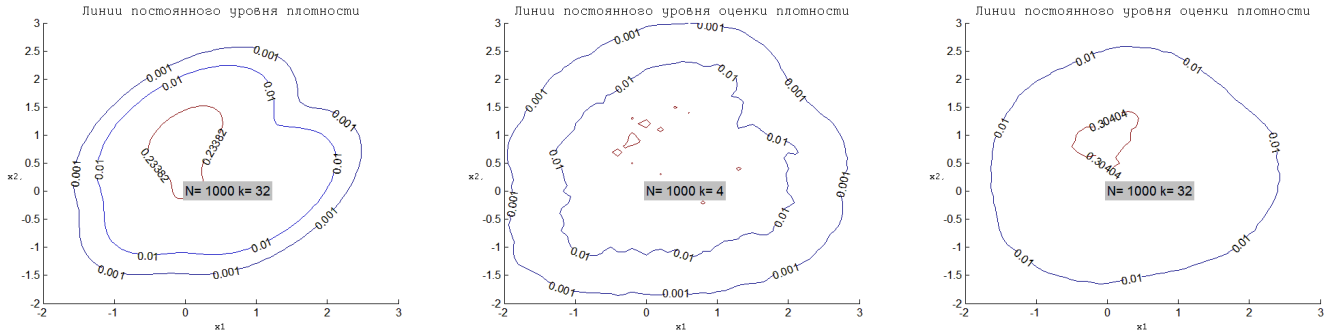
Рис.7. Примеры оценки плотности распределения с использованием настойки $k = N^{\gamma}$, $\gamma = 0.5$

На рис.8,9 приведены результаты для многомерной плотности распределения.



а) б) в)

Рис.8. Примеры оценки плотности распределения случайного вектора с использованием различных настроек $k(N)$: $k = 2N^{\gamma}$, $\gamma = 0.1$ – б; $k = N^{\gamma}$, $\gamma = 0.5$ – в



а) б) в)

Рис.9. Линии постоянного уровня плотности распределения случайного вектора и его оценок с использованием с использованием различных настроек $k(N)$: $k = 2N^{\gamma}$, $\gamma = 0.1$ – б; $k = N^{\gamma}$, $\gamma = 0.5$ – в

Визуальные анализ представленных результатов показывает, что в целом качество получаемой оценки при малых N существенно хуже, чем оценки по методу Парзена. При увеличении числа используемых соседей точность оценки повышается и становится сравнимой с оценками на основе метода Парзена.

Использование оценок на основе метода k-ближайших соседей в задачах распознавания

Напомним, что оценка плотности распределения вектора $\mathbf{x}$ методом k- соседей на основе обучающей выборки X^N определяется как

$$\tilde{p}(x) = \frac{k}{NV_n(k, N, X^N)}.$$

Пусть X^N состоит из совокупностей обучающих векторов классов $X^{N_i} = \{x^{(i,1)}, ..., x^{(i,N_i)}\}$, $i = \overline{1,M}$. Пусть также в гипершар с объемом $V_n(k, N, X^N)$ попадает k_i образов каждого класса: $0 \le k_i \le k$, $i = \overline{1,M}$.

Тогда вполне естественно представить

$$\tilde{p}(x) = \frac{\sum_{i=1}^{M} k_i}{NV_n(k, N, X^N)} = \sum_{i=1}^{M} \frac{N_i}{N} \frac{k_i}{N_i V_n} = \sum_{i=1}^{N} \tilde{p}(\omega_i)\tilde{p}(x/\omega_i).$$

Отсюда следует вполне очевидная интерпретация

$$\tilde{p}(\omega_i) = \frac{N_i}{N}, \quad \tilde{p}(x/\omega_i) = \frac{k_i}{N_i V_n}.$$

Подставляя полученные выражения в общую структуру решающего правила максимума апостериорной вероятности, получим алгоритм распознавания на основе метода k-ближайших соседей

$$\omega_i : k_i \geq k_j, \quad j = \overline{1,M}, \quad i \neq j.$$

Для случая двух классов решающее правило будет иметь совсем уже тривиальный вид

$$k_1 \underset{\omega_2}{\overset{\omega_1}{\gtrless}} k_2.$$

Несмотря на простейший вид алгоритмов, они имеют весьма высокую эффективность. Показано, что границы суммарной вероятности ошибки $\tilde{E}_s$ определяются неравенствами

$$E_s \leq \tilde{E}_s \leq E_s\left(2 - \frac{M}{M-1}E_s\right) \leq 2E_s,$$

где E_s – вероятность суммарной ошибки, достигаемая при использовании оптимального байесовского решающего правила при известных функциях правдоподобия классов.

Т.е. качество полученного алгоритма по уровню ошибки всего лишь в два раза может быть хуже, чем качество оптимального алгоритма, синтезированного в условиях полной априорной неопределенности.

Для получения конкретных оценок условных вероятностей ошибок алгоритмов распознавания можно использовать метод скользящего контроля или реализовать статистическое имитационное моделирования с тестированием алгоритма на поступающих новых образах.

На рис.10 и 11 а,б представлены результаты использования алгоритмов с восстановлением плотности по методам Парзена и k –соседей для случая $n=2$ и для различных видов восстанавливаемых плотностей распределения двух классов

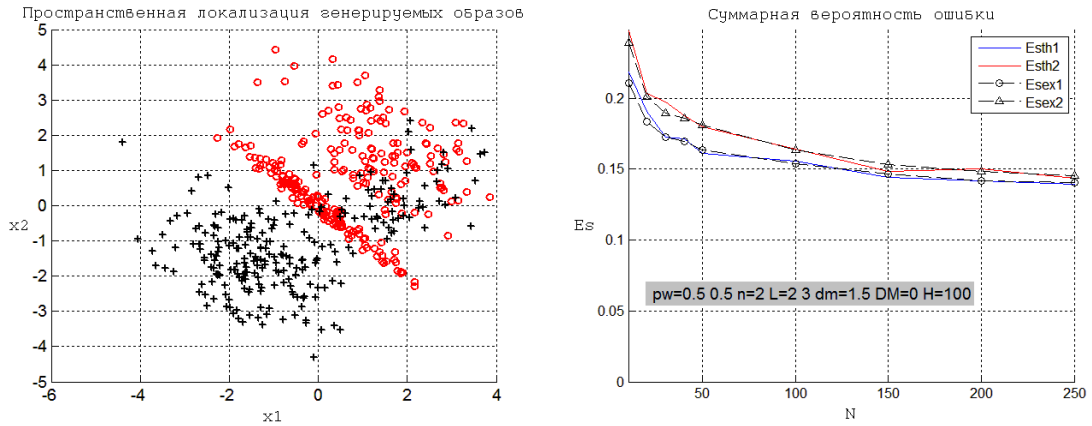


Рис.10. Области локализации и зависимости для суммарной вероятности ошибки от объема обучающих выборок при усреднении по $H=100$ актам обучения

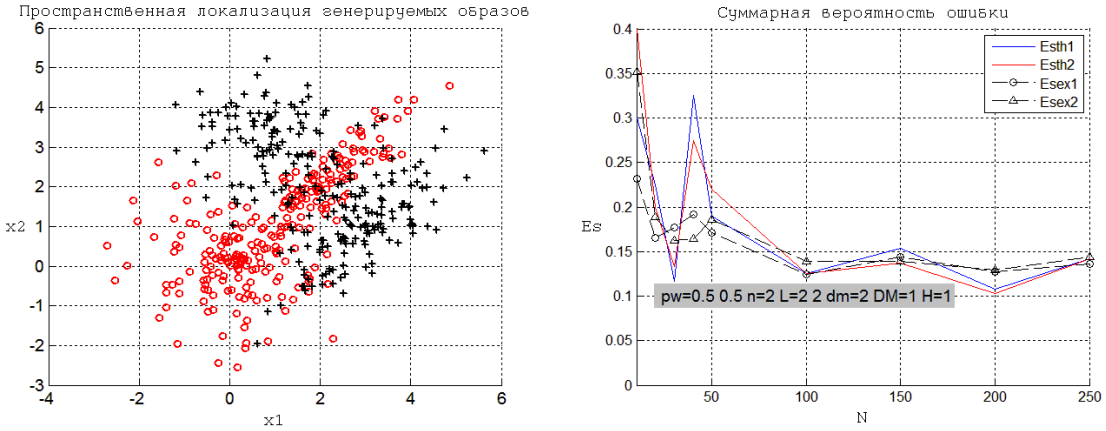


Рис.11. Области локализации классов и зависимости для суммарной вероятности ошибки от объема обучающих выборок при обучении по одной $H=1$ выборке

3. **Нелокальные методы оценивания плотности распределения вероятностей. Построение гистограммы**

Данный класс методов отличается от ранее рассмотренных тем, что оценка в каждой точке получается на основе аппроксимации плотности распределения вероятностей по всем полученным обучающим данным. При этом нет необходимости хранить всю обучающую выборку, т.е. при реализации такого подхода обеспечивается определенное сжатие данных. К наиболее известным методам нелокального оценивания относятся гистограммный метод.

Этот достаточно простой метод непараметрической оценки плотности распределения реализует подход, основанный на построении гистограмм по обучающей выборки $X^N = \{x^{(1)},...,x^{(N)}\},\ x^{(i)} \in R^n$. Принципиальное отличие данного метода от ранее рассмотренных состоит в том, что он не требует для получения оценки в каждой точке x хранения всей обучающей выборки, т.е. фактически реализует определенное сжатие обучающих данных.

Рассмотрим теоретическое обоснование данного метода и практические рекомендации по его применению. Пусть случайный вектор $\mathbf{x}$ имеет распределение $f_{\mathbf{x}}(x) = p(x)$. Вероятность попадания этого вектора в некоторую область Ω равна $P_\Omega = \int\limits_\Omega p(x)dx$. При этом величину P_Ω можно рассматривать как усредненное по области Ω значение плотности распределения вероятности. Для выборки X^N вероятность попадания в указанную область ее элементов k раз определяется законом Бернулли

$$P(\mathbf{k}=k)=C_N^k P_\Omega^k (1-P_\Omega)^{N-k}.$$

При этом математическое ожидание и дисперсия случайной величины $\mathbf{k}$, как уже ранее отмечалось в аналогичных задачах, будут равны

$$M[\mathbf{k}]=NP_\Omega, \quad D[\mathbf{k}]=NP_\Omega(1-P_\Omega).$$

Отсюда следует, что

$$P_\Omega = M[\mathbf{k}]/N.$$

Показано, что максимально правдоподобной оценкой вероятности P_Ω является величина

$$\tilde{P}_\Omega = k/N,$$

при этом вид выражений для математического ожидания и дисперсии $\mathbf{k}$ гарантирует несмещённость и состоятельность этой оценки. Из теоремы о среднем также следует, что

$$P_\Omega = \int_\Omega p(x)dx = p(x_0)V_\Omega,$$

где $x_0 \in \Omega$ – внутренняя точка области Ω, а V_Ω – объем этой области. Отсюда можно окончательно предложить использовать следующую оценку плотности, имеющую постоянное значение внутри области Ω

$$\tilde{p}(x)=p(x_0)=\tilde{P}_\Omega/V_\Omega = k/NV_\Omega, \quad x\in\Omega.$$

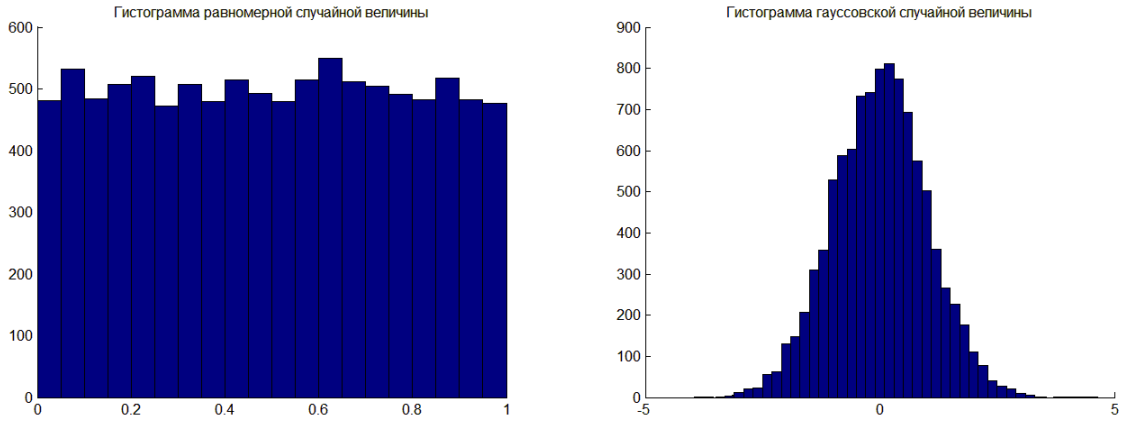
С учетом этого для построения классической гистограммы общую область U_X, в которой гарантированно размещаются элементы выборки X^N разбивают на непересекающиеся подобласти-ячейки $\Omega_j,\ j=\overline{1,M},\ \Omega_j \cup \Omega_s = \varnothing,\ j\neq s$.

Тогда внутри каждой из них используется оценка

$$\tilde{p}(x) = k_j / NV_{\Omega_j}, \quad x \in \Omega_j,$$

где k_j – количество элементов обучающей выборки X^N, попавших в Ω_j. Необходимо отметить, что выбор размера ячеек в данном случае влияет на сходимость (состоятельность) оценки, поскольку нужно обеспечить увеличение количества попадающих в каждую ячейку наблюдений при росте N. В простейшем случае ячейки выбираются одинакового размера.

Примеры гистограмм (ненормированных) для выборки из 10000 реализаций равномерной и гауссовской случайных величин, представлены на рис.12



а) б)

Рис.12. Гистограммы равномерной (а) и гауссовской случайных величин (б)