

Распознавание образов

4.1. Машина (метод) опорных векторов

(SVM, Support Vector Machine – Вапник В. и др. – 60-80 годы)

Линейно разделимый случай

Имеется множество прецедентов (Ξ, Y) , где $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ – обучающая выборка, а $Y = (y_1, \dots, y_N)$ – множество меток двух классов ω_1 и ω_2 . Требуется по обучающей выборке построить линейную решающую функцию $f(\mathbf{x})$, которая удовлетворяла бы условию $f(\mathbf{x}_i) > 0$ для всех $\mathbf{x}_i \in \omega_1$, $f(\mathbf{x}_i) < 0$ для всех $\mathbf{x}_i \in \omega_2$.

$$y_i = \begin{cases} 1, & \mathbf{x}_i \in \omega_1, \\ -1, & \mathbf{x}_i \in \omega_2. \end{cases}$$

Эту задачу можно переформулировать следующим образом.

Найти линейную решающую функцию $f(\mathbf{x})$, которая удовлетворяла бы условию

$$y_i f(\mathbf{x}_i) > 1 \text{ для всех } \mathbf{x}_i \in \Xi. \quad (4.1)$$

$$y_i ((\mathbf{w}, \mathbf{x}_i) + b) \geq 1, \quad i = 1, \dots, N, \quad (4.2)$$

где \mathbf{w} – вектор весовых коэффициентов, b – некоторое число. Тогда разделяющей два класса гиперплоскостью будет

$$(\mathbf{w}, \mathbf{x}) + b = 0.$$

Все гиперплоскости вида $(\mathbf{w}, \mathbf{x}) + b' = 0$,
где $b' \in (b-1, b+1)$

также будут разделяющими (рис. 4.1).

Расстояние между граничными гиперплоскостями

$(\mathbf{w}, \mathbf{x}) - 1 + b = 0$ и $(\mathbf{w}, \mathbf{x}) + b + 1 = 0$ равно $2/\|\mathbf{w}\|$

$$\left(\frac{1}{\|\mathbf{w}\|} \mathbf{w}, \mathbf{x}\right) + \frac{1}{\|\mathbf{w}\|} (b-1) = 0$$

и $\left(\frac{1}{\|\mathbf{w}\|} \mathbf{w}, \mathbf{x}\right) + \frac{1}{\|\mathbf{w}\|} (b+1) = 0$ – нормальные уравнения этих гиперплоскостей.

$p_1 = \frac{1}{\|\mathbf{w}\|} (b-1)$ и $p_2 = \frac{1}{\|\mathbf{w}\|} (b+1)$ – расстояния от гиперплоскостей до начала координат.

$$p_2 - p_1 = 2/\|\mathbf{w}\|$$

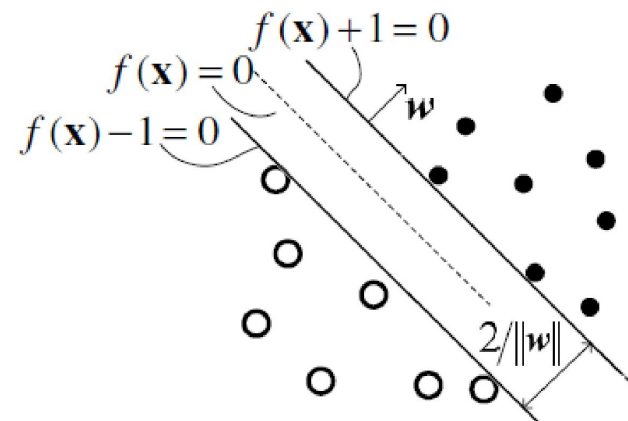


Рис. 4.1

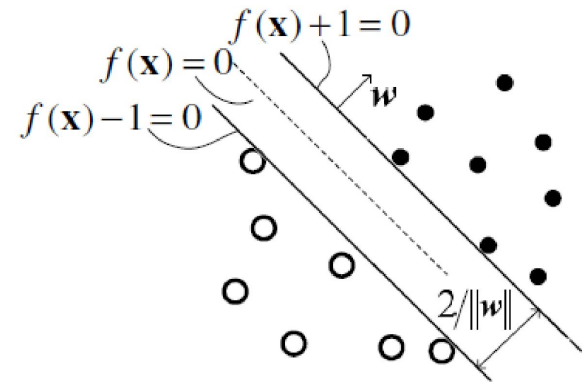


Рис. 4.1

Для надёжного разделения классов необходимо чтобы расстояние между разделяющими гиперплоскостями было как можно большим.

Ставится задача нахождения минимума квадратичного функционала $\frac{1}{2} (\mathbf{w}, \mathbf{w})$ в выпуклом многограннике, задаваемом системой неравенств (4.2)

$$\begin{aligned} & \frac{1}{2} (\mathbf{w}, \mathbf{w}) \rightarrow \min; \\ & y_i ((\mathbf{w}, \mathbf{x}_i) + b) \geq 1, \quad i = 1, \dots, N, \end{aligned} \quad (4.2).$$

Решение этой оптимизационной задачи равносильно поиску седловой точки лагранжиана

$$L(\mathbf{w}, b, \lambda) = \frac{1}{2} (\mathbf{w}, \mathbf{w}) - \sum_{i=1}^N \lambda_i (y_i ((\mathbf{w}, \mathbf{x}_i) + b) - 1) \rightarrow \min_{\mathbf{w}, b} \max_{\lambda}$$

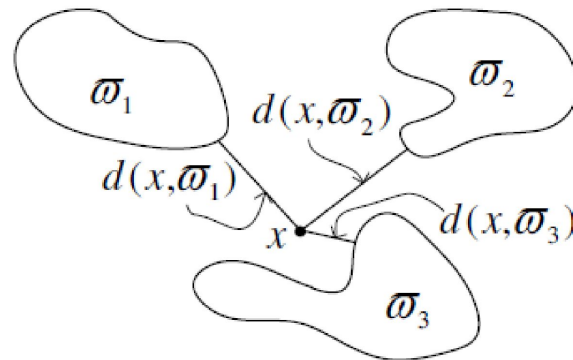
Свойства Метода опорных векторов:

- 1) это наиболее быстрый способ нахождения решающих функций;
- 2) метод сводится к решению задачи квадратичного программирования в выпуклой области, которая всегда имеет единственное решение;
- 3) метод находит разделяющую полосу максимальной ширины, что позволяет в дальнейшем осуществлять более уверенную классификацию;
- 4) метод чувствителен к шумам и стандартизации данных;
- 5) не существует общего подхода к автоматическому выбору ядра (и построению спрямляющего подпространства в целом) в случае линейной неразделимости классов.

Большой класс нейронных сетей можно представить в виде SVM с определёнными ядрами.

4.2. Классификация с помощью функции расстояния

Этот способ предполагает определение функции, оценивающей меру принадлежности предъявленного для классификации образа x классу ω_i , т.е. некоторой функции $d(x, \omega_i)$, которая удовлетворяла бы условию $x \in \omega_i$, если $d(x, \omega_i) \leq d(x, \omega_j)$ для всех $j \neq i$



Чтобы определить меру принадлежности образа x классу ω_i нужно выбрать способ определения меры близости между двумя образами, между образом и классом и, наконец, между двумя классами. Так как каждый образ x характеризуется некоторым вектором признаков \mathbf{x} , то меру близости между образами \mathbf{x} и \mathbf{y} можно задать с помощью меры близости $d(\mathbf{x}, \mathbf{y})$ между векторами-образами \mathbf{x} и \mathbf{y} их признаков.

В качестве такой меры близости чаще всего используют метрику, т.е. такую неотрицательную функцию

$$d : R^n \times R^n \rightarrow R_+,$$

которая удовлетворяет условиям аксиомы метрики.

Пространство R^n с введённой метрикой называют метрическим пространством.

Векторы признаков, между которыми измеряется расстояние, могут иметь разную размерность, разные порядки величин, различные приоритеты.

Поэтому, прежде чем использовать метрики, желательно нормализовать и стандартизировать значения признаков.

Способы определения расстояния между вектором-образом и классом

1. *Определение расстояния до центра класса.*

Этот способ применяется в том случае, когда класс «хорошо описывается» одним эталонным образом – центром класса (качество такого описания может быть измерено величиной дисперсии элементов класса), а цена ошибки неправильной классификации не очень велика. Этот способ согласуется с так называемым «принципом компактности» в распознавании образов, согласно которому векторы-образы одного класса должны быть «компактно» расположены в пространстве признаков.

2. Метод ближайшего соседа.

В этом способе расстояние определяется в соответствии со следующим алгоритмом:

- а) Определяется тот элемент x_s обучающей выборки, который ближе всего к предъявленному образцу x ,
- б) Проверяется условие: если $x_s \in \omega_j$, то считается что и $x \in \omega_j$.

3. определение расстояния до эталонного образа.

Этот способ применяется в том случае, когда класс плохо описывается одним эталонным образом – центром класса. Признаком такой ситуации является большое значение дисперсии элементов класса. Но можно предположить, что элементы класса будут «хорошо группироваться» вокруг нескольких эталонных образов.

Другими словами, каждый класс описывается не одним, а несколькими эталонными образами. В этом случае расстояние между вектором-образом x и эталонным классом ω_i сводится к вычислению расстояния между этим вектором и ближайшим к нему эталонным образом данного класса:

4.3. Метод потенциальных функций

Предположим, что имеется множество прецедентов, т.е. обучающая выборка $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ в пространстве признаков и множество меток $Y = \{y_1, \dots, y_N\}$ принадлежности классам.

Рассмотрим случай двух классов ω_1 и ω_2 . Через X_1 и X_2 обозначим их области предпочтения, а через Ξ_1 и Ξ_2 - прецеденты первого и второго классов соответственно.

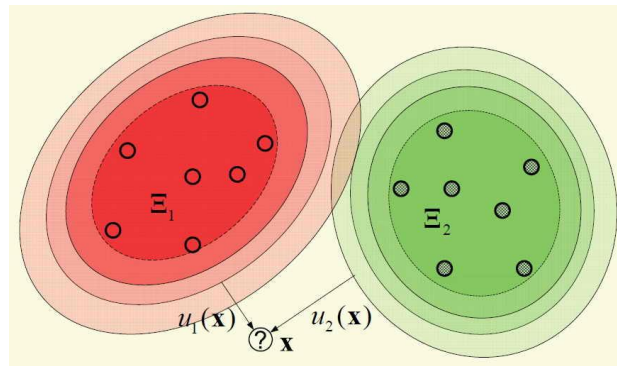
Требуется найти такую решающую (дискриминантную) функцию $d(\mathbf{x})$, чтобы $d(\mathbf{x}) > 0$ для всех $\mathbf{x} \in \Xi_1$ и чтобы $d(\mathbf{x}) < 0$ для всех $\mathbf{x} \in \Xi_2$. Если считать, что

$$y_i = \begin{cases} 1, & \mathbf{x}_i \in \Xi_1, \\ -1, & \mathbf{x}_i \in \Xi_2, \end{cases} \quad i = 1, \dots, N,$$

то требуется найти такую функцию $d(\mathbf{x})$, чтобы $y_i d(\mathbf{x}_i) > 0$ для всех $\mathbf{x}_i \in \Xi$.

Метод потенциальных функций связан с определением так называемой потенциальной функции $u(\mathbf{x}, \mathbf{y})$, т.е. некоторой положительной функции, принимающей тем большие значения, чем «ближе» точки \mathbf{x} и \mathbf{y} друг к другу. Если векторы признаков рассматриваются в метрическом пространстве, то в качестве потенциальной функции $u(\mathbf{x}, \mathbf{y})$ можно выбрать функцию вида $u(\mathbf{x}, \mathbf{y}) = \tilde{u}(d(\mathbf{x}, \mathbf{y}))$, где $d(\mathbf{x}, \mathbf{y})$ – некоторая метрика, а $\tilde{u}(t)$ положительная монотонно убывающая функция. Например

$$\tilde{u}(t) = \frac{1}{1 + \alpha t}, \quad \tilde{u}(t) = e^{-\alpha t}, \quad \alpha > 0. \quad u_i(\mathbf{x}) = \sum_{\mathbf{y} \in \Xi_i} u(\mathbf{x}, \mathbf{y}), \quad i = 1, 2.$$



Лепский, Броневиц, 2009. презентация

Если потенциал $u_1(\mathbf{x}) > u_2(\mathbf{x})$, то $\mathbf{x} \in X_1$. Если же $u_2(\mathbf{x}) > u_1(\mathbf{x})$, то $\mathbf{x} \in X_2$. Таким образом, функция

$$d(\mathbf{x}) = u_1(\mathbf{x}) - u_2(\mathbf{x}) = \sum_{k=1}^N y_k u(\mathbf{x}, \mathbf{x}_k)$$

будет решающей (дискриминантной) функцией.

Решающая функция может и не содержать всех слагаемых и будет иметь вид

$$d(\mathbf{x}) = \sum_{j=1}^N w_j u(\mathbf{x}, \mathbf{x}_j), \quad (1)$$

где $\mathbf{x}_j \in \Xi$ – множество векторов обучающей выборки, w_j – неизвестные коэффициенты.

Функцию (1) можно записать с помощью скалярного произведения:

$$d(\mathbf{x}) = (\mathbf{w}, \mathbf{u}(\mathbf{x})), \quad \text{где } \mathbf{w} = (w_j), \quad \mathbf{u}(\mathbf{x}) = (u(\mathbf{x}, \mathbf{x}_j)).$$

В методе потенциальных функций дискриминантная функция $d(\mathbf{x})$ находится по обучающей выборке $\Xi = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ путём коррекции k -й аппроксимирующей функции $d_k(\mathbf{x})$ с помощью следующей рекуррентной процедуры:

$$d_{k+1}(\mathbf{x}) = d_k(\mathbf{x}) + r_{k+1} u(\mathbf{x}, \mathbf{x}_{k+1}), \quad d_0(\mathbf{x}) \equiv 0, \quad (2)$$

где $\{r_k\}$ – некоторая последовательность, которая должна быть такой, чтобы гарантировать сходимость в некотором смысле $d_k(\mathbf{x})$ к $d(\mathbf{x})$ при $k \rightarrow \infty$.

4.4. Статистический подход к задаче распознавания

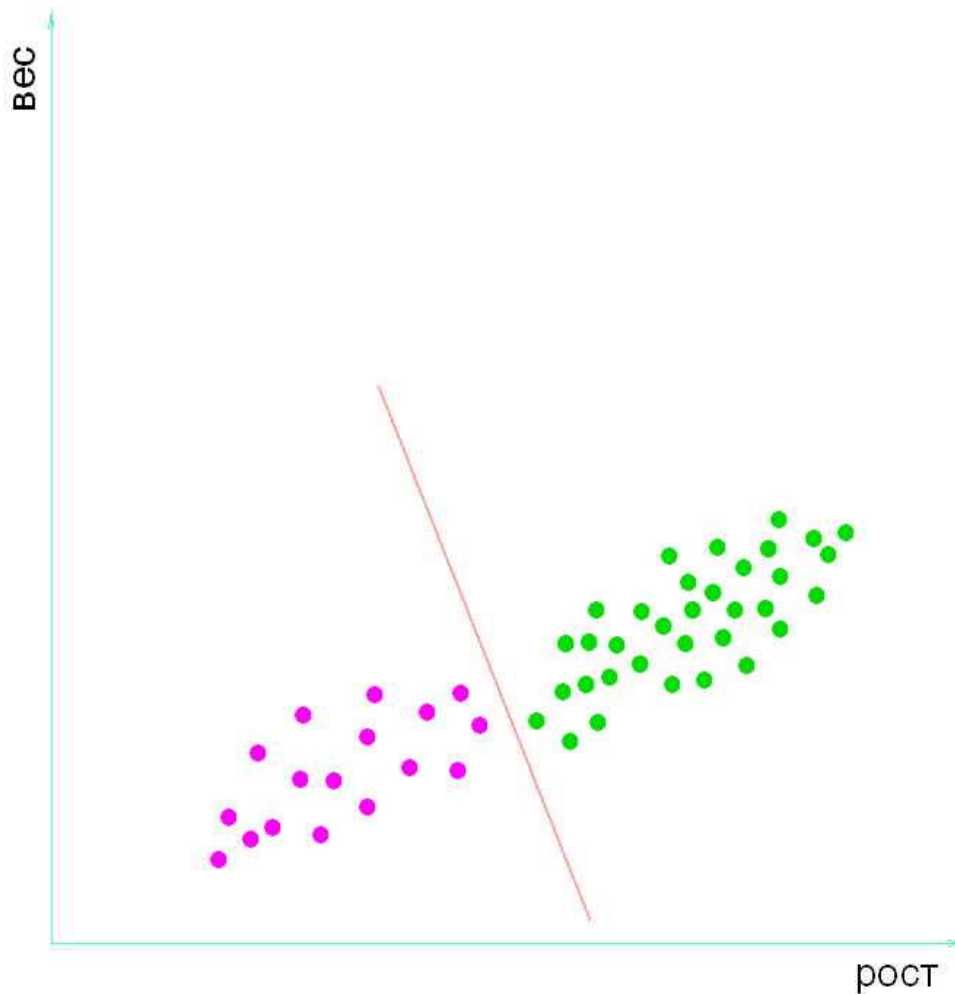


Рис. 4.1. Образы жокеев и баскетболистов в пространстве признаков. Детерминированный подход

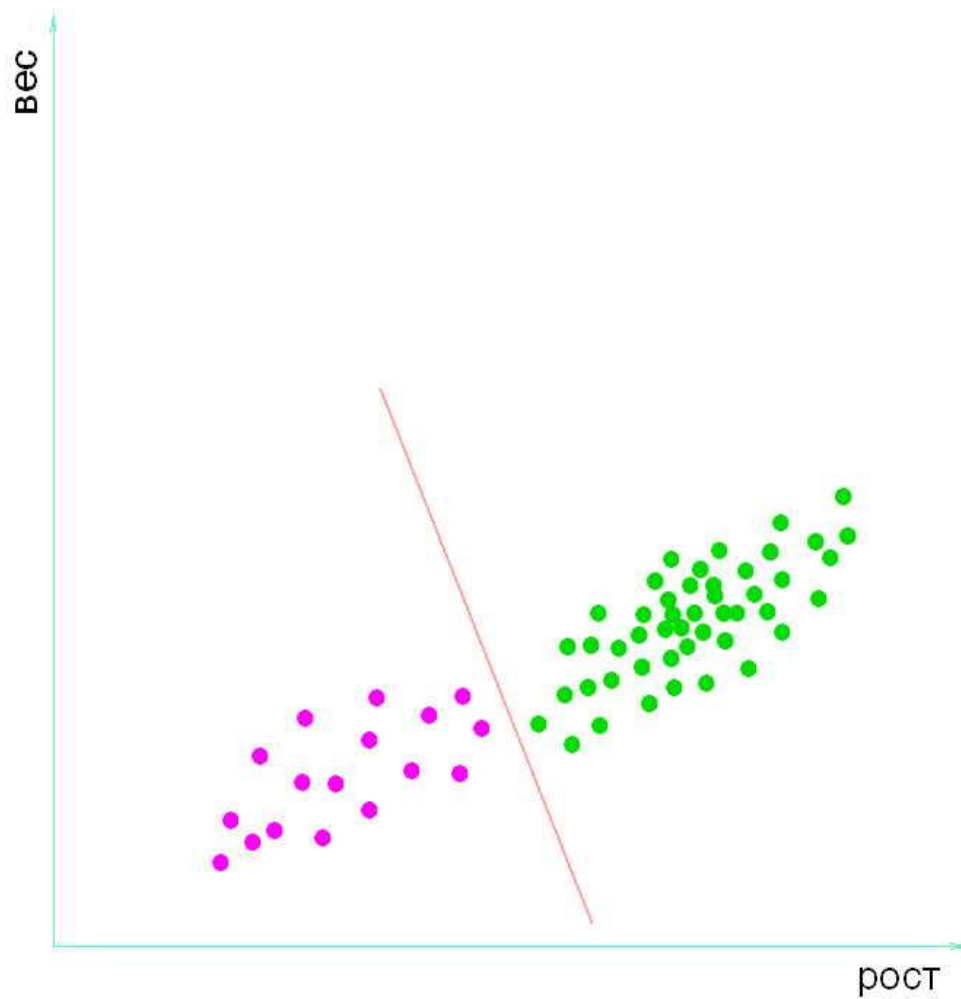


Рис. 4.2 Образы жокеев и баскетболистов в пространстве признаков.

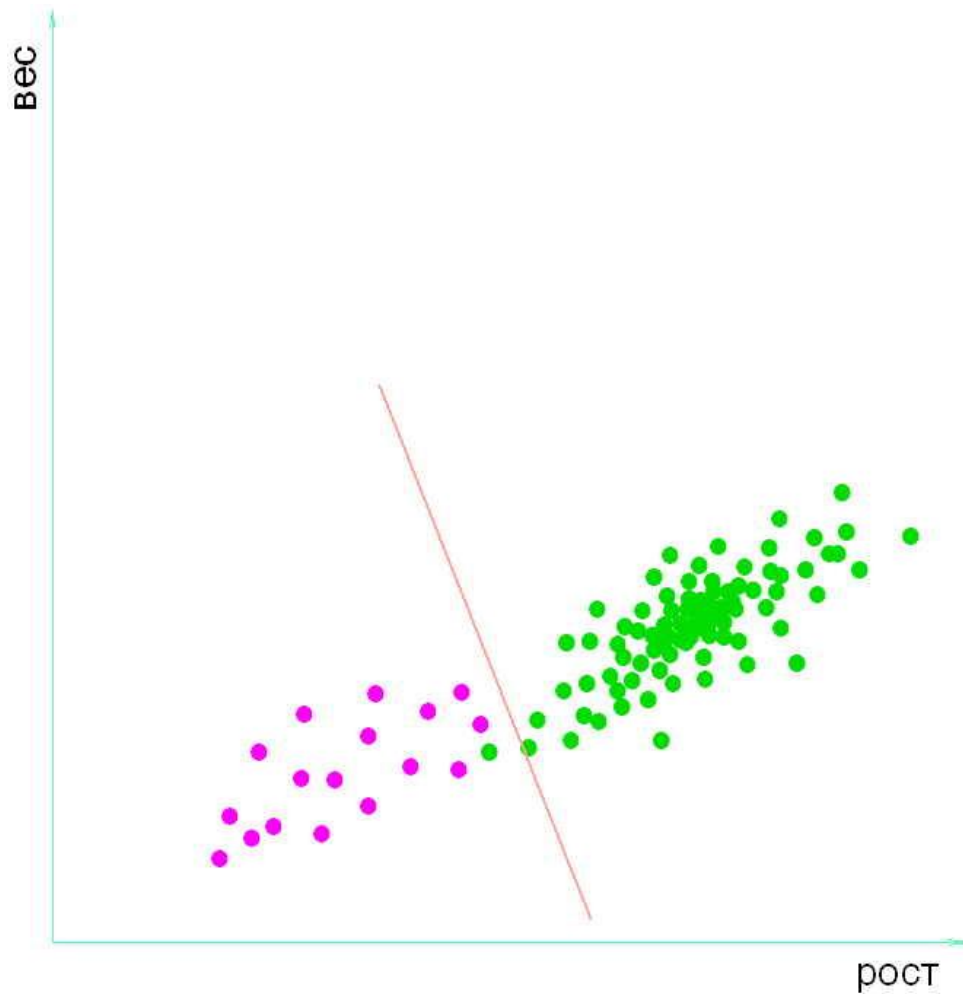


Рис. 4.3 Образы жокеев и баскетболистов в пространстве признаков. Статистический подход

При большом количестве объектов относительную частоту появления образа в классе можно оценить, построив гистограмму по объектам, принадлежащим заданному классу.

При большом количестве образов в k -м классе эта частота стремится к вероятности $p_k(\mathbf{x})$ появления образа \mathbf{x} в данном классе, и гистограмму можно рассматривать как дискретную аппроксимацию функции плотности распределения $f_k(\mathbf{x})$.

В основе статистических методов классификации лежит предположение, что функция плотности вероятности $f(\mathbf{x})$ для любого из выделяемых классов отлична от нуля на всем множестве значений признаков. То есть любой вектор \mathbf{x} может появиться в любом из классов, но с разной вероятностью.

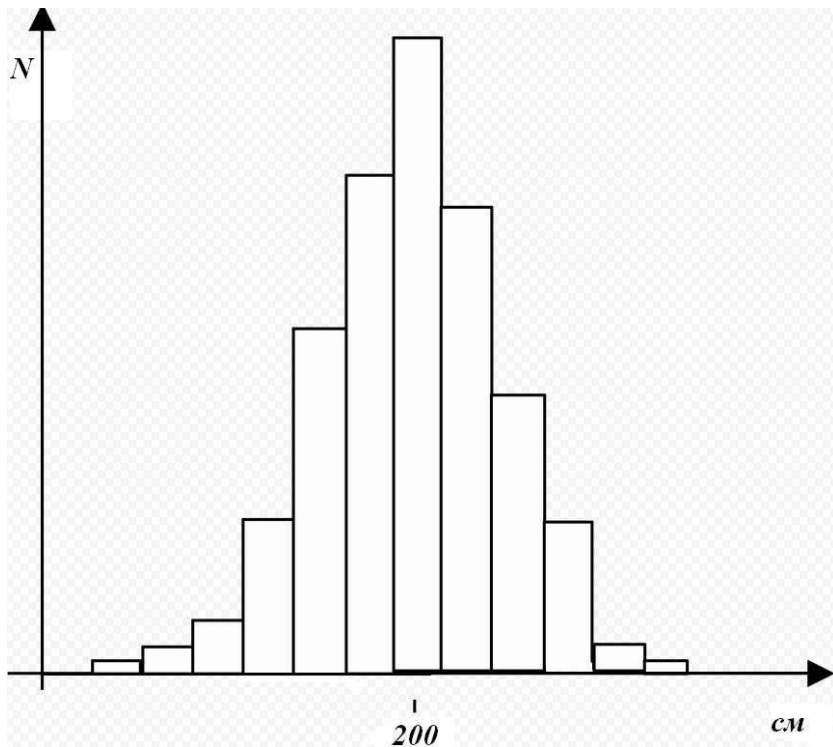


Рис. 4.4 Гистограмма

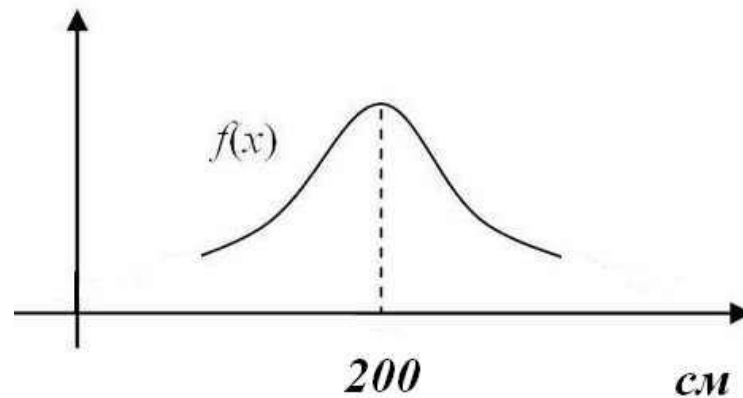


Рис. 4.5 Функция плотности распределения вероятностей

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$$

μ — математическое ожидание (среднее значение),
 σ — среднеквадратическое отклонение (σ^2 — дисперсия)
распределения.

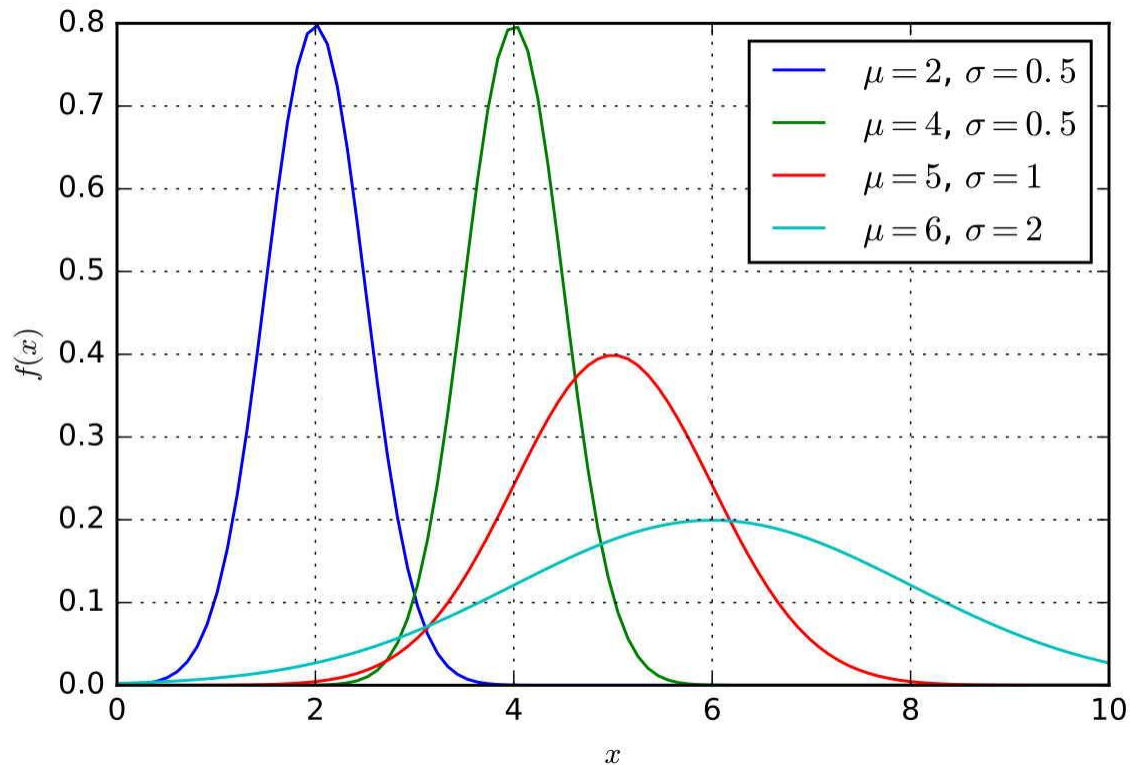


Рис. 4. 6. Нормальное распределение

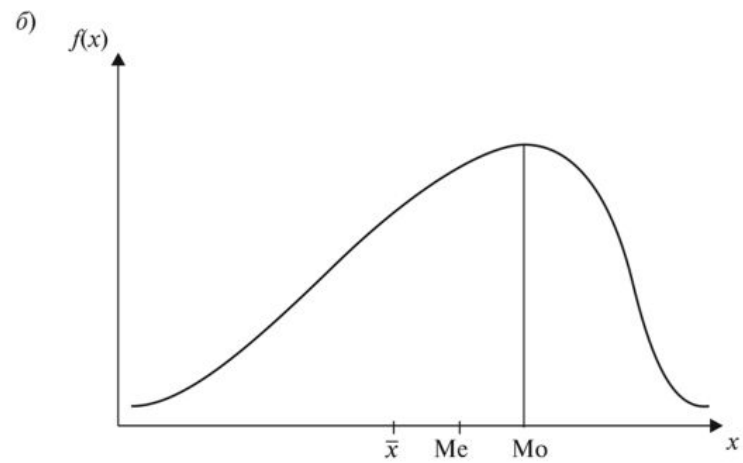
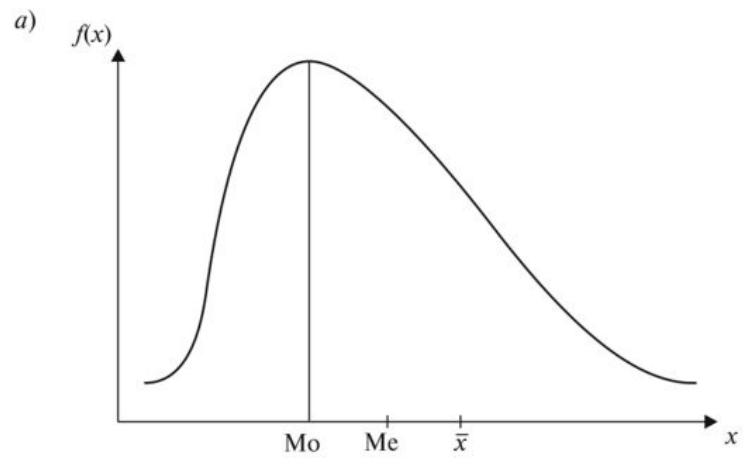


Рис. 4.7 Несимметричные распределения

В основе статистических методов классификации лежит предположение, что функция плотности вероятности $f(x)$ для любого из выделяемых классов отлична от нуля на всем множестве значений признаков.

То есть любой вектор x может появиться в любом из классов, но с разной вероятностью.

Поскольку мы неизбежно вынуждены строго определить границы между классами в пространстве X , то всегда существует вероятность, что какое-то количество точек из любого класса попадет в другие. Эта ошибка называется *ошибкой первого рода* (α).

С другой стороны, в этот класс может попасть какое-то количество точек из других классов. Такая ошибка называется *ошибкой второго рода* (β).

Полная вероятность ошибки выделения каждого класса, таким образом, есть $p(\alpha) + p(\beta)$.

При большом количестве образов мы можем попробовать минимизировать среднюю вероятность ошибки $p(\alpha) + p(\beta)$ при многократном принятии решения.

Критерий Байеса обеспечивает минимальную в среднем вероятность ошибки распознавания при многократном принятии решения.

С этой целью вводится понятие «риска», то есть платы за каждую ошибку, и определяется условие, соответствующее минимуму среднего риска.

Для пары классов X_i и X_j оно имеет следующий вид:

$$\frac{P_i f_i(\mathbf{x})}{P_j f_j(\mathbf{x})} = 1 \quad (4.1)$$

или в логарифмической форме

$$\ln \frac{P_i f_i(\mathbf{x})}{P_j f_j(\mathbf{x})} = 0. \quad (4.2)$$

Поскольку мы ищем минимум среднего риска при многократном принятии решения, в эти выражения, кроме функций плотности вероятности, входят также *априорные вероятности* P_i и P_j ,

Можно сказать, что эти вероятности характеризуют частоту появления каждого из классов на анализируемом множестве образов, которая пропорциональна доле площади под этим классом.

Второе выражение (4.2) соответствует уже знакомому нам виду разделяющей функции для пары классов: $d_{ij}(\mathbf{x})=0$. То есть, если $P_i f_i(\mathbf{x}) \geq P_j f_j(\mathbf{x})$, то мы принимаем решение в пользу класса X_i , в противном случае принимаем решение в пользу класса X_j .

Отношение (4.1) называется *отношением правдоподобия*, а функции, стоящие в числителе и знаменателе, — *функциями правдоподобия*.

По-другому можно сказать, что на каждом интервале мы принимаем решение в пользу того класса, полная вероятность которого в рамках заданного множества образов при конкретном значении \mathbf{x} максимальна.

Такое правило принятия решения называют *принцип максимума правдоподобия*.

Ясно, что функция правдоподобия представляет собой решающую функцию.

Для случая одномерного признака и трех классов правило принятия решения по принципу максимума правдоподобия показано на рис. 4.8.

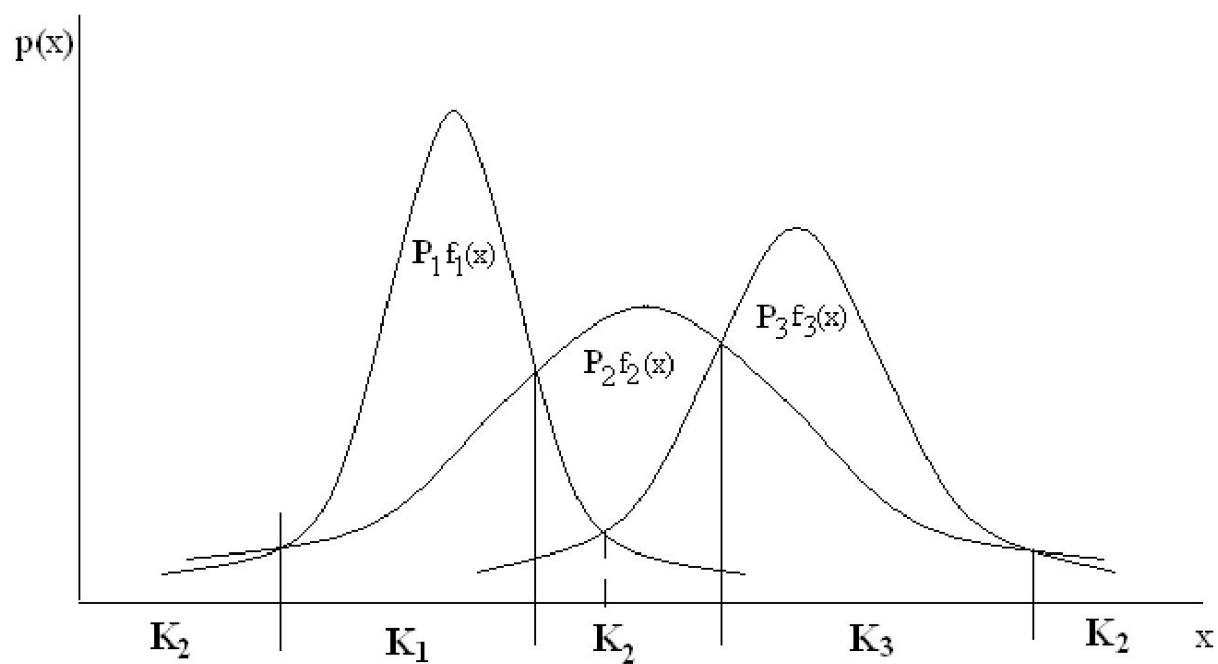


Рис. 4.8.

Классификация без обучения(кластеризация)

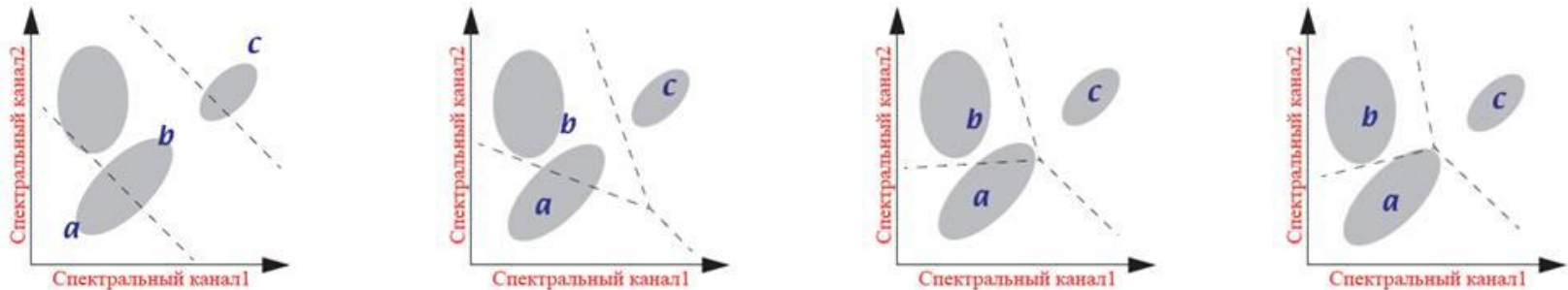
Наиболее распространенные методы классификации без обучения:

ISODATA;

К-Средних.

Метод классификации без обучения ISODATA (Итерационная самоорганизующаяся методика анализа данных - Iterative Self-Organizing Data Analysis Technique).

ISODATA это процесс, который основан на кластерном анализе.



- 1) Всё пространство признаков делится на n равных множеств, для каждого из них определяется среднее значение;
- 2) Первая итерация кластеризации: в пространстве для каждой точки рассчитывается расстояние до средних значений, и каждую точку относят в определенный кластер. В один кластер попадают те точки, между которыми меньше расстояния в пространстве признаков;
- 3) Расчет новых средних значений для полученных классов;
- 4) Следующая итерация с новыми значениями средних, и уточнение границ классов, при этом число классов может меняться.

