

# МАТЕМАТИЧЕСКИЕ ОСНОВЫ ТЕОРИИ ПРОГНОЗИРОВАНИЯ

Лектор

Сенько Олег Валентинович

Лекция 10

# Коллективные методы

Одним из способов получения ансамбля является использование алгоритмов, обученных по разным обучающим выборкам, возникающим в результате случайного процесса, лежащего в основе исследуемой задачи. Обычно при решении прикладной задачи в распоряжении исследователя имеется обучающая выборка  $\tilde{S}_t = \{s_1, \dots, s_m\}$  ограниченного объёма. Однако процесс генерации семейства выборок из генеральной совокупности может быть имитирован с помощью процедуры бутстрэп (bootstrap), которая основана на выборках с возвращениями из  $\tilde{S}_t$ .

# Коллективные методы (бэггинг)

В результате получаются выборки  $\tilde{S}_*^{bg}$ , включающие объекты из обучающей выборки  $\tilde{S}_t$ . Однако некоторые объекты могут встречаться более одного раза, а другие объекты отсутствовать. Предположим, что с помощью процедуры бутстрэп получено выборок  $\tilde{S}_*^{bg}$ . Выберем метод, который далее используется  $T$  для обучения отдельных алгоритмов распознавания. В результате у нас будет получено  $T$  алгоритмов  $A_1^{bg}, \dots, A_T^{bg}$ . Для получения коллективного решения может быть использован простейший комитетный метод, относящий объект в тот класс, куда его отнесло большинство алгоритмов.

# Коллективные методы (бэггинг)

Данная процедура носит название бэггинг (bagging), что является сокращением названия **Bootstrap Aggregating**. Процедура бэггинг показывает высокий прирост обобщающей способности по сравнению с алгоритмом, обученным с помощью базового метода по исходной обучающей выборке  $\tilde{S}_t$ , в тех случаях, когда вариационная составляющая ошибки базового метода высока. К таким моделям относятся в частности решающие деревья и нейросетевые методы. При использовании в качестве базового метода решающих деревьев процедура бэггинг приводит к построению ансамблей решающих деревьев (решающих лесов).

# Коллективные методы (бустинг)

Основной идеей алгоритма бустинг является пошаговое наращивание ансамбля алгоритмов. Алгоритм, который присоединяется к ансамблю на  $k$ -ом шаге обучается по выборке, которая формируется из объектов исходной обучающей выборки  $\tilde{S}_t$ . В отличие от метода бэггинг объекты выбираются не равноправно, а исходя из некоторого вероятностного распределения, заданного на выборке  $\tilde{S}_t$ . Данное распределение вычисляется по результатам классификации с помощью ансамбля, полученного на предыдущем шаге.

# Коллективные методы (бустинг)

Приведём схему одного из наиболее популярных вариантов метода бустинг AdaBoost (Adaptive boosting) более подробно.

На первом шаге присваиваем начальные значения весов

$(w_1^1, \dots, w_m^1)$  объектам обучающей выборки. Поскольку веса

имеют вероятностную интерпретацию, то для них соблюдаются

ограничения  $\sum_{j=1}^m w_j^1 = 1$   $w_j^1 \in [0, 1]$  . Обычно начальное

распределение выбирается равномерным

$w_j^1 = \frac{1}{m}$ ,  $j = 1, \dots, m$  . Выбираем число итераций  $T$  .

# Коллективные методы (бустинг)

На  $k$ -ой итерации генерируем выборку  $\tilde{S}_k^{bs}$  из исходной выборки  $\tilde{S}_t$  согласно распределению задаваемому весами  $(w_1^k, \dots, w_m^k)$ .

Обучаем распознающий алгоритм  $A_k^{bs}$  по выборке  $\tilde{S}_k^{bs}$ .

Вычисляем взвешенную ошибку по формуле  $\varepsilon_k = \sum_{j=1}^m w_j^k e_j^k$ ,

где  $e_j^k = 1$ , если алгоритм  $A_k^{bs}$  неправильно классифицировал

объект  $s_j \in \tilde{S}_k^{bs}$  и  $e_j^k = 0$  в противном случае. В том случае,

если  $\varepsilon_k \geq 0.5$  или  $\varepsilon_k = 0$  игнорируем шаг и заново

генерируем выборку  $\tilde{S}_k^{bs}$  исходя из весовых коэффициентов

$$w_j^k = \frac{1}{m}, \quad j = 1, \dots, m$$

# Коллективные методы (бустинг)

В случае если  $\varepsilon_k \in (0, 0.5)$  вычисляем коэффициенты

$\tau_k = \frac{\varepsilon_k}{1 - \varepsilon_k}$  и пересчитываем веса объектов по формуле

$$w_j^{k+1} = \frac{w_j^k (\tau_k)^{1-e_j^k}}{\sum_{j=1}^m w_j^k (\tau_k)^{1-e_j^k}} \quad j = 1, \dots, m$$

Процесс продолжается до тех пор, пока не выполнено  $T$

итераций . В результате мы получаем совокупность

распознающих алгоритмов  $A_1^{bs}, \dots, A_T^{bs}$  .



# Коллективные методы (бустинг)

Предположим, что нам требуется распознать объект  $s^*$ . Оценка объекта  $s^*$  за класс  $K_l$  вычисляется по формуле

$$\Gamma_l(s^*) = \sum_{k=1}^T \ln\left(\frac{1}{\tau_k}\right) \beta_l(s^*) \quad \text{где} \quad \beta_l(s^*) = 1, \text{ если объект } s^* \text{ отнесён}$$

в класс  $K_l$  алгоритмом  $A_k^{bs}$  и  $\beta_l(s^*) = 0$  в противном случае.

Объект  $s^*$  будет отнесён к классу, оценка за который максимальна.

Описанный вариант метода носит название AdaBoost. M1.

# Коллективные методы (бустинг)

Эффективность процедур бустинга подтверждается многочисленными экспериментами на реальных данных. В настоящее время существует большое количество вариантов метода, имеющих разное обоснование.

# Коллективные методы, основанные на голосовании по системам закономерностей

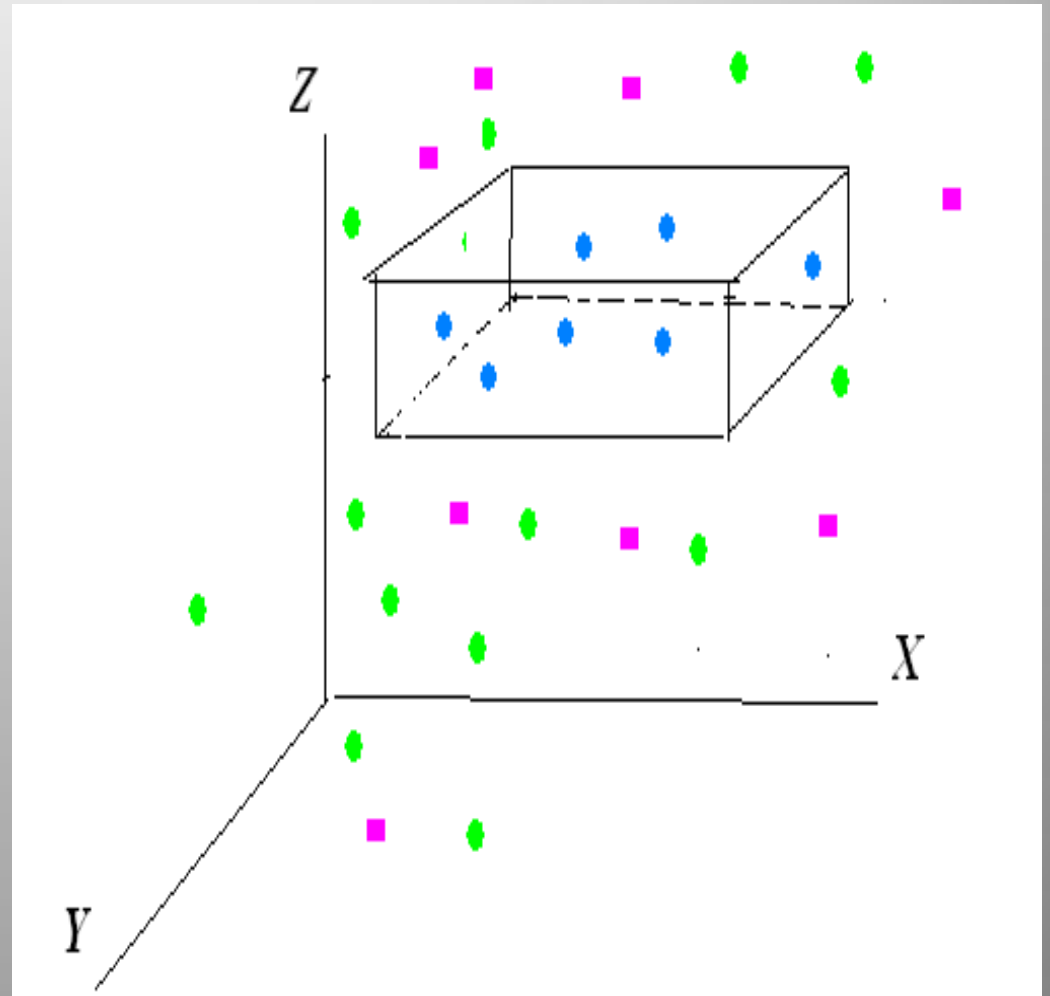
Одним из эффективных подходов к решению задач прогнозирования и распознавания является использование коллективных решений по системам закономерностей. Под закономерностью понимается распознающий или прогностический алгоритм, определённый на некоторой подобласти признакового пространства или связанный с некоторым подмножеством признаков.

# Коллективные методы, основанные на голосовании по системам закономерностей

В качестве примера закономерностей могут быть приведены представительные наборы, являющиеся по сути подмножествами признаков описаний, характерных для одного из распознаваемых классов. Аналогом представительных наборов в задачах с вещественнозначной информацией являются логические закономерности классов.

# Методы, основанные на голосовании по системам логических закономерностей

Под логической закономерностью класса  $K_l$  понимается область признакового пространства, имеющая форму гиперпараллелепипеда и содержащая только объекты  $K_l$ .



# Методы, основанные на голосовании по системам логических закономерностей

Логическая закономерность  $\mathbf{r}$  задаётся с помощью конъюнкций предикатов вида

$$P_i^r(s) = "b_i^{rl} \leq x_i(s) \leq b_i^{rr} "$$

При этом для характеризующего закономерность класса  $K_l$  предиката

$$\mathbf{P}^r(s) = P_1^r(s) \& \dots \& P_n^r(s)$$

Должны выполняться следующие условия

$$1) \quad \exists s_{j^*} \in K_l \cap \tilde{S}_t, \quad \mathbf{P}^r(s_{j^*}) = 1$$

# Методы, основанные на голосовании по системам логических закономерностей

$$2) \quad \forall s_{j^*} \in \tilde{S}_t, \quad s_{j^*} \notin K_l, \quad \mathbf{P}^r(s_{j^*}) = 0$$

3)  $\mathbf{P}^r(s)$  доставляет экстремум некоторому функционалу качества  $\varphi(\mathbf{P})$ , заданному на множестве всевозможных предикатов, удовлетворяющих условиям 1), 2).

На практике используются такие функционалы качества  $\varphi(\mathbf{P})$

как число объектов из  $\tilde{S}_t \cap K_l$ , для которых  $\mathbf{P}^r(s) = 1$ , доля

объектов  $\tilde{S}_t \cap K_l$ , для которых  $\mathbf{P}^r(s) = 1$ .

# Методы, основанные на голосовании по системам логических закономерностей

Наряду с полными логическими закономерностями, для которых выполняются все условия 1) – 3), используются также частичные логические закономерности, для которых допускаются некоторые нарушения условия 2). То есть допускается существование небольшой доли нарушений условия 2) для тех объектов, для которых выполняется условие  $\mathbf{P}^r(s) = 1$

На этапе обучения для каждого из классов  $K_l$  ищется множество логических закономерностей  $\tilde{R}_l$  .



# Методы, основанные на голосовании по системам логических закономерностей

Предположим, что нам требуется распознать новый объект

$s^*$ . Для каждого из классов  $K_l$  ищется число закономерностей из  $\tilde{R}_l$  для которых  $\mathbf{P}^r(s^*) = 1$ . При этом доля таких закономерностей считается оценкой за класс  $K_l$ . Для классификации используется стандартное решающее правило, т.е. объект относится в класс, оценка за который максимальна.

# Методы, основанные на голосовании по системам логических закономерностей

Поиск оптимальной системы логических закономерностей

производится по набору случайно выбранных из обучающей

выборки эталонных объектов (опорных эталонов). Для каждого

из опорного эталона  $S_i$  поиск оптимальных границ

$(b_1^{rl}, b_1^{rr}), \dots, (b_n^{rl}, b_n^{rr})$

осуществляется сначала на некоторой

неравномерной сетке пространства, которая задается с

помощью разбиения интервала значений каждого из

признаков. После нахождения оптимальной закономерности на

заданной сетке, поиск продолжается на заданной в окрестности

этого оптимального решения, но уже на более мелкой сетке.

# Методы, основанные на голосовании по системам логических закономерностей

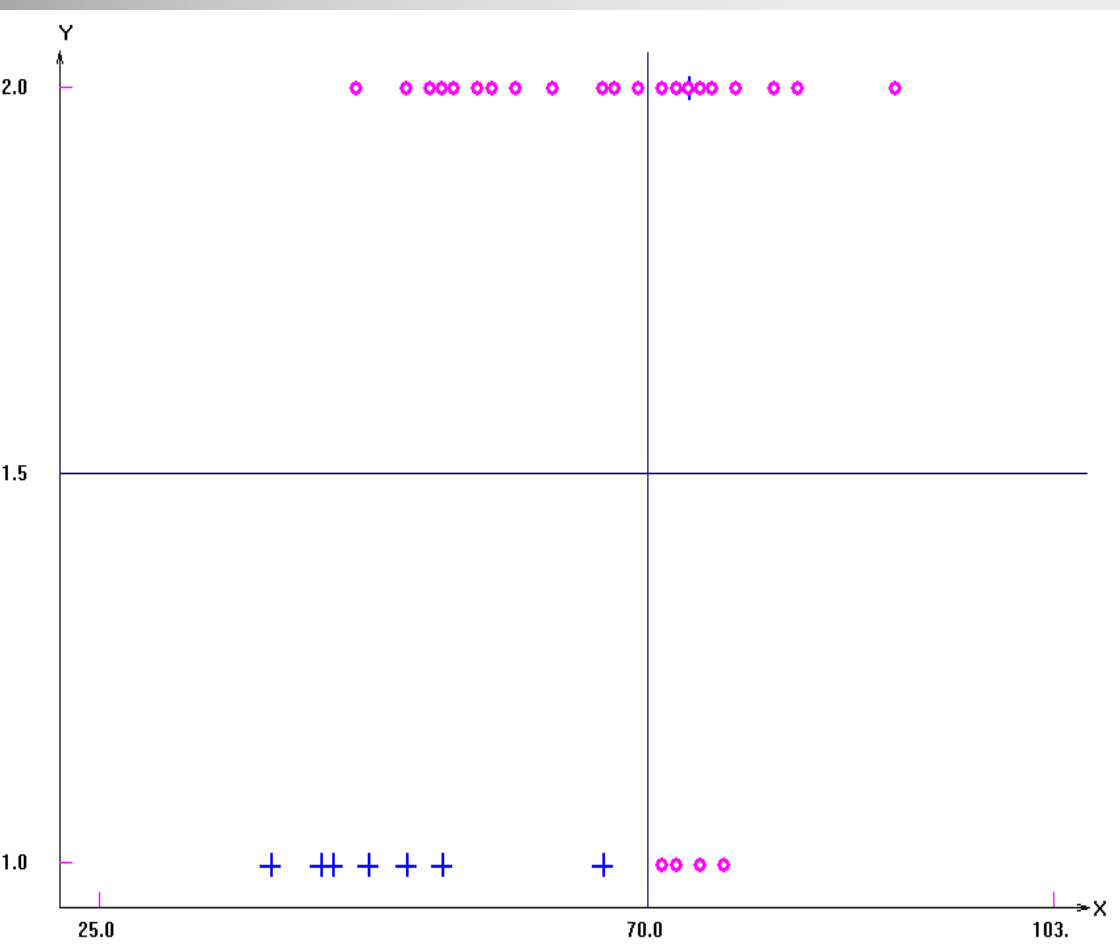
Процесс заканчивается, если при переходе к более мелкой сетке не удастся найти ЛЗ с более *высоким* критерия качества  $\varphi(\mathbf{P})$ .

- Задача поиска оптимальной ЛЗ на каждой сетке сводится к поиску максимальной совместной подсистемы некоторой системы неравенств. Далее все вычисления повторяются для  $k$  случайно выбранных «опорных» эталонов класса  $K_j$ , а все найденные логические закономерности объединяются в одно множество  $\tilde{\mathbf{P}}_j$ .

# Метод статистически взвешенных синдромов (СВС)

Коллективные решения в методе СВС принимаются по информации о принадлежности векторного  $\mathbf{x}^*$  описания распознаваемого объекта “синдромам” из некоторой заранее заданной системы  $\tilde{Q}$ . Под синдромом понимается такая область признакового пространства, в которой содержание объектов одного из классов, отличается от содержания объектов этого класса в обучающей выборке или по крайней мере в одной из соседних областях.

# Метод статистически взвешенных синдромов (СВС)



Пример синдромов, характеризующих разделение объектов из классов  $K_1^{(+)}$  и  $K_2^{(0)}$  приведён на рисунке 2.

Видно, описания объектов из  $K_1$  сосредоточены главным образом в нижнем левом квадранте «синдроме».

Рисунок 2

# Метод статистически взвешенных синдромов (СВС)

Синдромы ищутся для каждого из распознаваемых классов с помощью построения оптимальных разбиений интервалов допустимых значений единичных признаков или совместных двумерных областей допустимых значений пар признаков. При этом поиск производится внутри нескольких семейств разбиений, имеющих различный уровень сложности. В ходе поиска выбирается разбиение с максимальным значением функционала качества. Используется два функционала качества: интегральный  $F_s(K_j, \tilde{S}_t, R)$  и локальный.  $F_L(K_j, \tilde{S}_t, R)$

# Метод статистически взвешенных синдромов (СВС)

Обозначим через  $r_1, \dots, r_t$  элементы некоторого разбиения  $R$ .

Пусть  $v_0^j$  - доля объектов класса  $K_j$  в обучающей выборке  $\tilde{S}_t$ ,

$v_i^j$  - доля объектов  $K_j$ , описания которых принадлежат

элементу  $r_i$ ,  $m_i$  - число объектов, описания которых

принадлежат элементу  $r_i$ .

Интегральный функционал определяется формулой

$$F_s(K_j, \tilde{S}_t, R) = \sum_{i=1}^t (v_0^j - v_i^j)^2 m_i$$

# Метод статистически взвешенных синдромов (СВС)

Локальный функционал качества задаётся в виде

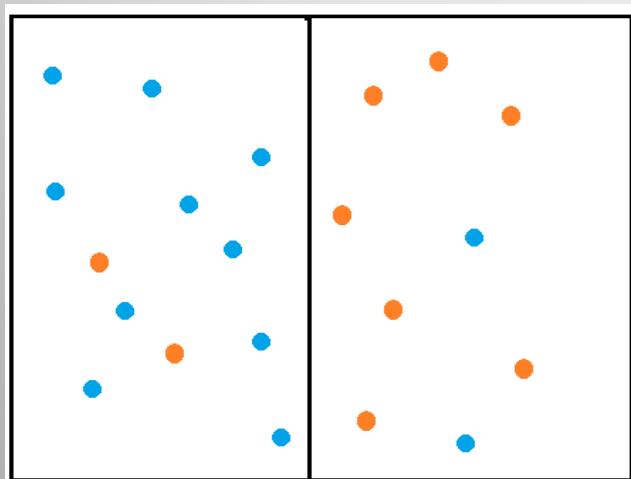
$$F_L(K_j, \tilde{S}_t, R) = \max_{i \in \{1, \dots, t\}} (v_0^j - v_i^j)^2 m_i$$

Поиск разбиений с максимальным значением одного из функционалов производится в рамках одного из четырёх семейств. Примеры разбиений для каждого из семейств приведены на рисунке.

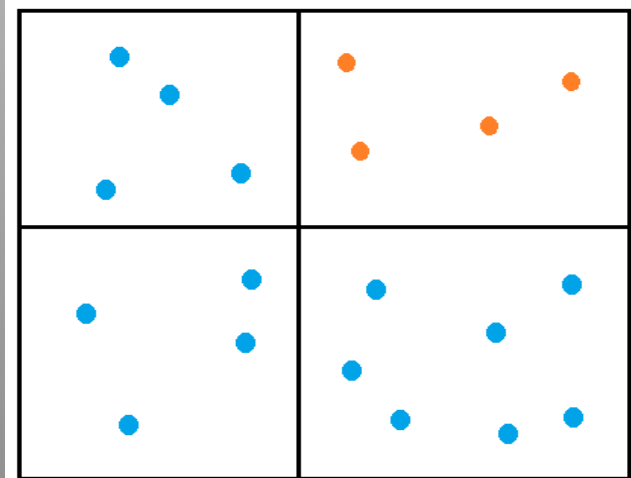
Семейство I включает всевозможные разбиения интервалов допустимых значений отдельных признаков на два интервала с помощью одной граничной точки.



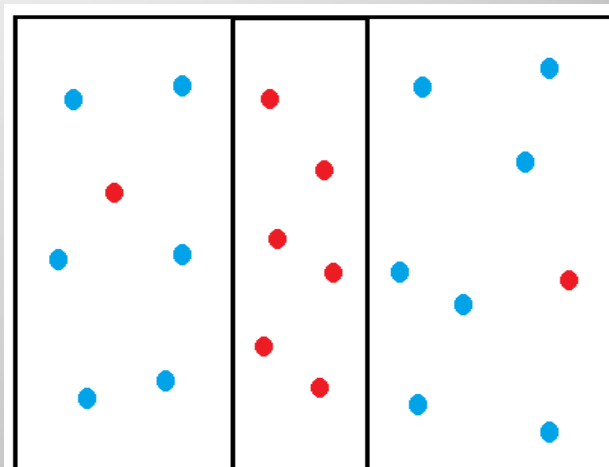
# Метод статистически взвешенных синдромов (СВС)



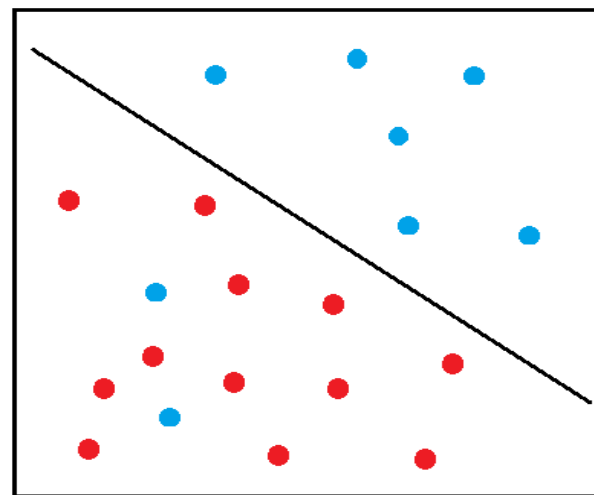
Семейство I



Семейство III



Семейство II



Семейство IV

# Метод статистически взвешенных синдромов (СВС)

Семейство II включает всевозможные разбиения интервалов допустимых значений отдельных признаков на 3 интервала с помощью двух граничных точек.

Семейство III включает всевозможные разбиения совместных двумерных областей допустимых значений пар признаков на 4 подобласти с помощью двух граничных точек ( по одной точке для каждого из двух признаков)

# Метод статистически взвешенных синдромов (СВС)

Семейство  $\mathcal{IV}$  включает всевозможные разбиения совместных двумерных областей допустимых значений пар признаков на 2 подобласти с помощью прямой граничной линии, произвольно ориентированной относительно координатных осей.

Найденные оптимальные разбиения используются для построения систем синдромов, если соответствующая им максимальная величина функционала качества превосходит некоторое заранее заданное пользователем пороговое значение  $\delta$ .

# Метод статистически взвешенных синдромов (СВС)

Причём величина порога зависит от сложности модели разбиений.

Порог является минимальным для простейшей одномерной модели I. Для моделей II-IV величина порога домножается на величину  $K$ , задаваемую пользователем, что позволяет регулировать влияние эффекта переобучения.

Одномерные разбиения, найденные внутри семейств I и II могут быть используются при построении не только одномерных, но также и двумерных синдромов.

# Метод статистически взвешенных синдромов (СВС)

Предположим, что на этапе обучения для класса  $K_j$  найдена система синдромов  $\tilde{Q}_j$ . Предположим, что описание  $\mathbf{x}^*$  распознаваемого объекта  $s^*$  принадлежит синдромам  $q_1, \dots, q_r$  из системы  $\tilde{Q}_j$ . Оценка за класс  $K_j$  вычисляется по формуле

$$\Gamma_j(s^*) = \frac{\sum_{i=1}^r w_i v_i^j}{\sum_{i=1}^r w_i}, \text{ где } v_i^j \text{ - доля класса } K_j \text{ в синдроме } q_i,$$

$w_i$  - вес синдрома.

# Метод статистически взвешенных синдромов (СВС)

Вес синдрома вычисляется по формуле

$$w_i = \frac{m_i}{m_i + 1} \frac{1}{v_i^j (1 - v_i^j)}, \text{ где } m_i - \text{число объектов}$$

обучающей выборки с описанием, принадлежащем  $q_i$ .

# Метод комитетов

Метод комитетов представляет собой реализацию подхода к решению задач распознавания, объединяющего принципы линейного разделения классов и вычисления коллективных решений.

Рассмотрим задачу распознавания с двумя классами  $K_1$  и  $K_2$ . Пусть  $f_1(\mathbf{x}), \dots, f_r(\mathbf{x})$  - набор линейных функций вида

$$f_i(\mathbf{x}) = a_{1i}x_1 + \dots + a_{ni}x_n + a_{0i}, \quad \text{где } \mathbf{x} = (x_1, \dots, x_n) \text{ - вектор признаков,}$$

$(a_{0i}, a_{1i}, \dots, a_{ni})$  - вектор вещественных параметров.

# Метод комитетов

Предположим, что для классификации произвольного объекта  $s^*$  с описанием  $\mathbf{x}^*$  используется следующее решающее правило:

объект  $s^*$  относится в класс  $K_1$ , если  $\sum_{i=1}^r \text{sign}[f_i(\mathbf{x})] > 0$  ;

,

объект  $s^*$  относится в класс  $K_2$ , если  $\sum_{i=1}^r \text{sign}[f_i(\mathbf{x})] > 0$  ; (1)

в случае, если  $\sum_{i=1}^r \text{sign}[f_i(\mathbf{x})] = 0$  происходит отказ от распознавания.



# Метод комитетов

Набор функций  $f_1(\mathbf{x}), \dots, f_r(\mathbf{x})$  называется комитетом, если решающее правило (1) правильно классифицирует объекты обучающей выборки.

Распознающий алгоритм, основанный, на решающем правиле (1) потенциально позволяет производить распознавание линейно неразделимых классов, реализуя кусочно-линейную разделяющую поверхность. Обучение сводится к поиску оптимальных (минимальных по числу функций) комитетов. Теоретически показано существование комитета для непротиворечивых данных.