

4.5 Методы, основанные на голосовании по системам логических 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.

В качестве примера закономерностей могут быть приведены представительные наборы, являющиеся по сути подмножествами признаковых описаний, характерных для одного из распознаваемых классов. Аналогом представительных наборов в задачах с вещественнозначной информацией являются логические закономерности классов. Под логической закономерностью класса K_l понимается область признакового пространства,

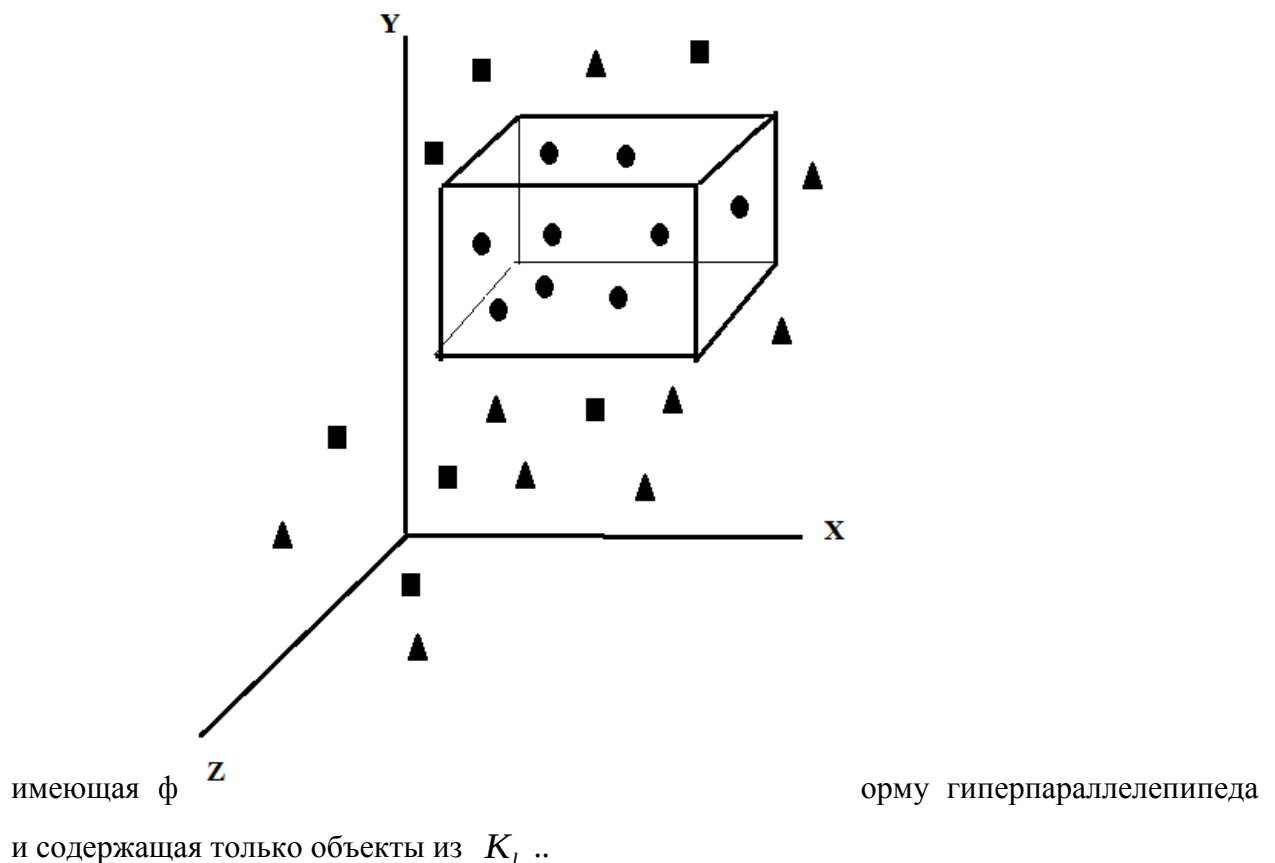


Рис. На рисунке представлена логическая закономерность, содержащая объекты класса K_1 , обозначенные синим кружком, и не содержащая объектов обозначенных по другому классов K_2 и K_3

Математически логическая закономерность класса K_l , которую мы будем обозначать $r(l)$, описывается с помощью наборов предикатов вида

$$P_i[r(l)] = "a_i^{r(l)} \leq x_i \leq b_i^{r(l)}", \quad (1)$$

где $i = 1, \dots, n$.

Напомним, что предикатом называется утверждение, принимающее значения «ИСТИНА» или «ЛОЖЬ» в зависимости от значений входящих в них переменных. Полностью логическая закономерность задаётся конъюнкцией предикатов вида (1):

$$\mathbf{P}[r(l)] = P_1[r(l)] \& \dots \& P_n[r(l)] \quad (2)$$

Очевидно, что множество векторов $(x_1, \dots, x_n)$, для которых, как раз представляет собой гиперпараллелепипед в многомерном пространстве признаков. Не все признаки являются на самом деле существенными для закономерности $r(l)$. Для несущественного признака $x_{i'}$ отрезок $[a_{i'}^{r(l)}, b_{i'}^{r(l)}]$ совпадает с отрезком, из которого принимает значения признак $x_{i'}$.

На этапе обучения для каждого класса K_l строится множество логических закономерностей $\tilde{R}_l$. Границы $(a_i^{r(l)}, b_i^{r(l)})$ подбираются таким образом, чтобы равенство $\mathbf{P}[r(l)] = \text{«ИСТИНА»}$ выполнялось бы на максимально большом числе объектов обучающей выборки из класса K_l и равенство $\mathbf{P}[r(l)] = \text{«ЛОЖЬ»}$ выполнялось бы на всех объектах обучающей выборки из класса K_l . Наряду с полными логическими закономерностями, удовлетворяющими последним условиям, используются также частичные логические закономерности, для которых допускается попадание в них небольшой доли объектов чужих классов. Методы построения логических закономерностей подробно излагаются в работе [11], а также книге [9].

Предположим, что нам требуется распознать новый объект s^* . Для каждого класса K_l ищется число закономерностей в $\tilde{R}_l$, для которых $\mathbf{P}[r(l)] = \text{«ИСТИНА»}$. В качестве оценки за класс K_l используется доля таких закономерностей в $\tilde{R}_l$. Классификация s^* производится с помощью стандартного решающего правила. То есть объект относится в тот класс, оценка за который максимальна.

4.6 Метод мультимодельных статистически взвешенных синдромов

Метод мультимодельных статистически взвешенных синдромов является методом распознавания, основанном на принятии коллективных решений по системам синдромов. 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. 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. Пример синдромов, характеризующих разделение двух классов, приведён на рисунке 2

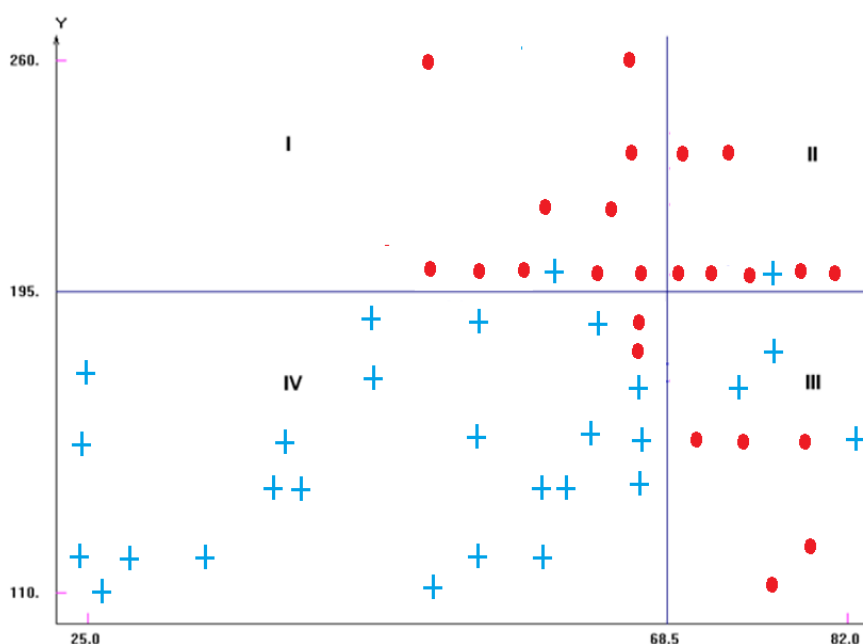


Рис. 2 Внутри синдромов I (верхний слева) и II (верхний справа) преобладают объекты класса K_1 , обозначенного $\bullet$. Внутри синдрома IV преобладают объекты класса K_2 , обозначенные $+$.

Поиск синдромов производится с использованием четырёх семейств разбиений, имеющих различный уровень сложности. Примеры разбиений для каждого из семейств приведены на рисунке 3. Семейство I включает всевозможные разбиения интервалов допустимых значений отдельных признаков на два интервала с помощью одной

границной точки. 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). Семейство IV включает всевозможные разбиения совместных двумерных областей допустимых значений пар признаков на 2 подобласти с помощью прямой граничной линии, произвольно ориентированной относительно координатных осей.

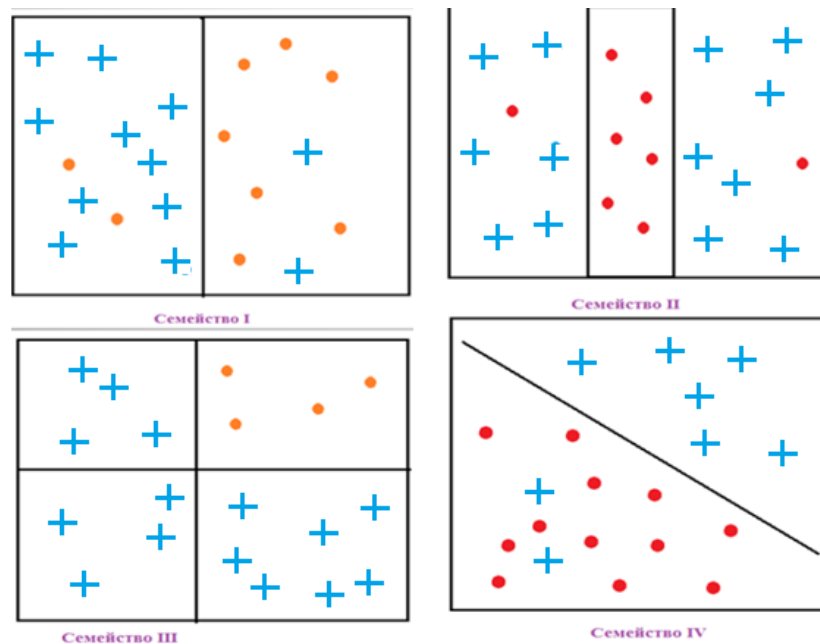


Рис 3. Примеры разбиений для каждого из четырёх семейств, используемых в методе SVC.

В ходе поиска выбирается разбиение с максимальным значением функционала качества. В различных вариантах метода используется два функционала качества, зависящих от обучающей выборки $\tilde{S}_t$, распознаваемого класса K_l , и разбиения R :

- интегральный $F_i(\tilde{S}_t, K_l, R)$;
- локальный $F_{loc}(\tilde{S}_t, K_l, R)$.

Обозначим через $q_1, \dots, q_r$ элементы некоторого разбиения R . Пусть ν_0^l является долей объектов класса K_l в обучающей выборке $\tilde{S}_t$. ν_i^l - доля объектов K_l среди объектов, описания которых принадлежат элементу q_i , m_i - число объектов, описания которых принадлежат q_i . Интегральный функционал задаётся формулой

$F_i(\tilde{S}_t, K_l, R) = \frac{1}{\nu_0^l(1-\nu_0^l)} \sum_{i=1}^r (\nu_i^l - \nu_0^l)^2 m_i$. В то время как локальный функционал

задаётся формулой $F_{loc}(\tilde{S}_t, K_l, R) = \max_{i=1, \dots, r} \frac{(\nu_i^l - \nu_0^l)^2 m_i}{\nu_0^l(1-\nu_0^l)}$

Метод СВС, впервые предложенный в работе [13] был основан на использовании одномерных семейств разбиений. Позже была предложена модификация СВС –метод мультимодельные статистически взвешенных синдромов (МСВС) [25]. В методе МСВС наряду с одномерными семействами I и II используются также семейства III и IV. Синдромы, задаваемые некоторым оптимальным разбиением R^* включаются в финальный набор, используемый в дальнейшем для распознавания новых объектов, если R^* удовлетворяет специальному критерию. В методе СВС для поиска синдромов используется интегральный функционал $F_i(\tilde{S}_t, K_l, R)$. Для формирования финального набора используется простой критерий: все элементы оптимального разбиения R^* включаются в набор, если величина интегрального функционала $F_i(\tilde{S}_t, K_l, R^*)$ превышает задаваемый пользователем порог δ . Опыт решения прикладных задач показывает, что эффективность распознавания достигается при значениях δ , меняющихся от 2 до 10. Несколько более сложный критерий используется в методе МСВС. Для поиска синдромов используется локальный функционал $F_{loc}(\tilde{S}_t, K_l, R)$. Синдромы оптимального разбиения R^* включаются в финальный набор в случае выполнения неравенства $\kappa F_{loc}(\tilde{S}_t, K_l, R) > \delta$, где величина параметра κ зависит от сложности используемой модели. Эксперименты на прикладных задачах показали, что высокая эффективность достигается при $\kappa = 1$ для простейших разбиений из семейства I и $\kappa = 0.5$ для разбиений из семейства II-IV.

Предположим, что на этапе обучения для класса K_l найдено множество синдромов $\tilde{Q}_l$. Пусть описание $\mathbf{x}^*$ распознаваемого объекта s^* принадлежит синдромам $q_1, \dots, q_r$ из множества $\tilde{Q}_l$. Оценка s^* за класс K_l вычисляется по формуле

$$\Gamma(s^*, K_l) = \frac{\sum_{i=1}^r w_i^l \nu_i^l}{\sum_{i=1}^r w_i^l},$$

где ν_i^l - доля объектов класса K_l в синдроме q_i , w_i^l - вес синдрома при классификации объектов класса K_l , который вычисляется по формуле $w_i^l = \frac{m_i}{m_{i+1}} \frac{1}{\nu_0^l (1 - \nu_0^l)}$, где m_i - число объектов обучающей выборки, попавших в синдром q_i . Данная формула была получена в работе [1] через максимизацию специального функционала, сходного с функционалом правдоподобия.

4.7 Метод опорных векторов.

4.7.1 Линейная разделимость.

Принцип максимизации зазора. Метод опорных векторов является универсальным методом распознавания, позволяющим наряду с линейными реализовывать также нелинейные решающие правила. Исходный вариант метода был предложен для задач с двумя распознаваемыми классами K_1 и K_2 . В случаях, когда объекты разных классов в обучающей выборке линейно разделимы, обычно существует целая совокупность линейных поверхностей, осуществляющих такое разделение. На рисунке 1 представлены двумерные данные, где объекты двух классов могут быть разделены с помощью прямых A, B, C, D. Однако наша интуиция, подсказывает что наилучшей обобщающей способностью должна обладать разделяющая прямая F, одинаково удалённая от групп объектов из разных классов. Однако наша интуиция, подсказывает что наилучшее

й способностью должна обладать разделяющая прямая F, одинаково удалённая от групп объектов из разных классов.

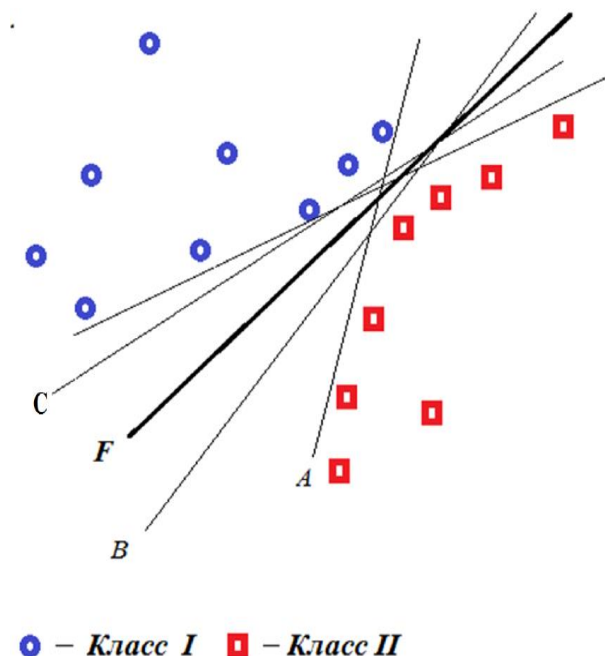


Рис. 1 Иллюстрируются различные варианты разделения классов K_1 и K_2 с помощью линейных границ.

Интуитивные представления об оптимальной разделимости формализует проведение разделяющей гиперплоскости посередине между двумя параллельными гиперплоскостями, каждая из которых отделяет объекты одного из классов. При этом две плоскости строятся таким образом, чтобы «зазор» между ними был бы максимальным.

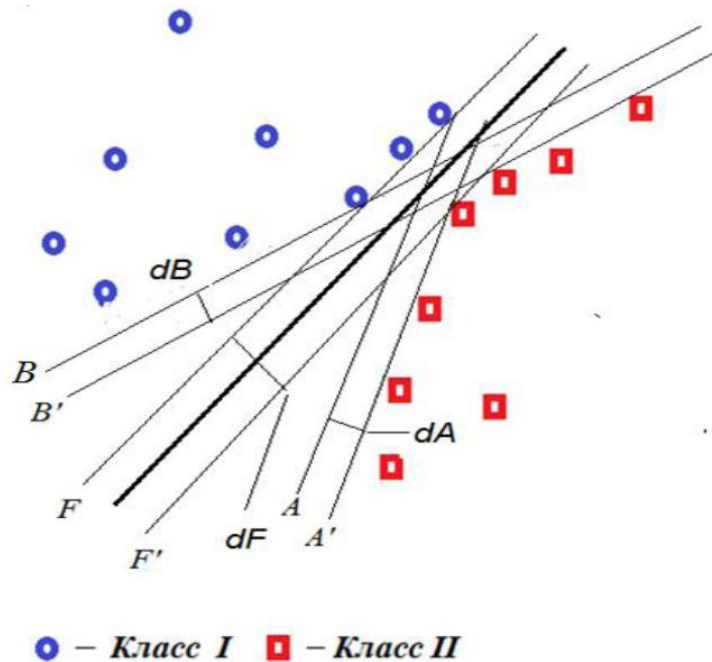


Рис. 1 Иллюстрируются разделение классов K_1 и K_2 с помощью линейных границ с использованием концепции максимального «зазора».

Напомним, что пара параллельных гиперплоскостей P_1 и P_2 в n -мерном пространстве $\mathbf{R}^n$ описывается с помощью уравнений:

$$(P_1) \quad \mathbf{w}\mathbf{x}^t = b_1, \quad (1)$$

$$(P_2) \quad \mathbf{w}\mathbf{x}^t = b_2,$$

где $\mathbf{W}$ является направляющим вектором для гиперплоскостей.

Пусть $\mathbf{z} = \lambda \mathbf{w}$, где λ - некоторое вещественное число. Нетрудно таким образом подобрать λ и b , чтобы система

$$(P_1) \quad \mathbf{z}\mathbf{x}^t = b + 1, \quad (2)$$

$$(P_2) \quad \mathbf{z}\mathbf{x}^t = b - 1,$$

Описывала те же самые гиперплоскости, что и система (1). Пусть точки $\mathbf{x}_1$ и $\mathbf{x}_2$ принадлежат плоскостям P_1 и P_2 соответственно. Расстояние (величина зазора) δ между гиперплоскостями P_1 и P_2 равно проекции разности $(\mathbf{x}_1 - \mathbf{x}_2)$ на направление $\mathbf{z}$, Данная проекция по определению равна $\frac{\mathbf{z}(\mathbf{x}_1^t - \mathbf{x}_2^t)}{|\mathbf{z}|}$. Однако согласно

системе (2) $\frac{\mathbf{z}(\mathbf{x}_1^t - \mathbf{x}_2^t)}{|\mathbf{z}|} = \frac{2}{|\mathbf{z}|}$. Следовательно задача поиска двух максимально удалённых друг от друга параллельных гиперплоскостей, каждая из которых отделяет объекты одного из классов, может быть сведена к оптимизационной задаче с ограничениями.

$$\frac{2}{|\mathbf{z}|} \rightarrow \max \quad (3)$$

$$\mathbf{z}\mathbf{x}_j^t \geq b + 1 \quad \text{при } s_j \in \tilde{S}_t \cap K_1$$

$$\mathbf{z}\mathbf{x}_j^t \leq b - 1 \quad \text{при } s_j \in \tilde{S}_t \cap K_2, \quad j = 1, \dots, m.$$

При этом оптимизация производится по компонентам направляющего вектора $\mathbf{z} = (z_1, \dots, z_n)$ и параметру сдвига b .

Введём обозначение: $\alpha_j = 1$ при $s_j \in \tilde{S}_t \cap K_1$ и $s_j \in \tilde{S}_t \cap K_2$. Учитывая, функция

$\frac{2}{|\mathbf{z}|}$ монотонно возрастает с уменьшением $\sum_{i=1}^n z_i^2$, переходим от задачи (3) к задаче

$$\frac{1}{2} \sum_{i=1}^n z_i^2 \rightarrow \min \quad (4)$$

$$\alpha_j (\mathbf{z}\mathbf{x}_j^t - b) \geq 1, \quad j = 1, \dots, m.$$

Задача (4) относится к хорошо изученному классу задач квадратичного программирования.

Решение задачи квадратичного программирования. Важным инструментом исследования экстремальных значений оптимизируемых функций при ограничениях является функция Лагранжа или лагранжиан, который для задачи (4) записывается в виде

$$\mathbf{L}(\mathbf{z}, b, \boldsymbol{\lambda}) = \frac{1}{2} \sum_{i=1}^n z_i^2 - \sum_{j=1}^m \lambda_j [\alpha_j (\mathbf{z} \mathbf{x}_j^t - b) - 1],$$

где $(\lambda_1, \dots, \lambda_m)$ являются неотрицательными вещественными, которые называются множителями Лагранжа.

Из известной теоремы Каруша-Куна-Такера (ККТ) следует, что для точки $(\mathbf{z}^*, b^*)$, в которой функция $\frac{1}{2} \sum_{i=1}^n z_i^2$ достигает своего минимума при ограничениях задачи (4), и некоторого вектора значений неотрицательных множителей Лагранжа $\boldsymbol{\lambda}^* = (\lambda_1^*, \dots, \lambda_m^*)$ соблюдаются условия стационарности лагранжиана $\mathbf{L}(\mathbf{z}, b, \boldsymbol{\lambda})$ по переменным $(\mathbf{z}, b)$.

Также из теоремы ККТ следует необходимость выполнения m равенств, которые носят название условий дополняющей нежёсткости

$$\lambda_j^* [\alpha_j (\mathbf{z}^* \mathbf{x}_j^t - b^*) - 1] = 0, \quad j = 1, \dots, m$$

Условия стационарности заключаются в выполнении n равенств

$$\left. \frac{\partial \mathbf{L}(\mathbf{z}, b, \boldsymbol{\lambda})}{\partial z_i} \right|_{(\mathbf{z}^*, b^*, \boldsymbol{\lambda}^*)} = z_i^* - \sum_{j=1}^m \lambda_j^* \alpha_j x_{ji} = 0, \quad i = 1, \dots, n \quad (5)$$

В векторной форме система (5) принимает вид

$$\mathbf{z}^* - \sum_{j=1}^m \lambda_j^* \alpha_j \mathbf{x}_j = 0.$$

Из условия стационарности также следует выполнение равенства

$$\left. \frac{\partial \mathbf{L}(\mathbf{z}, b, \boldsymbol{\lambda})}{\partial b} \right|_{(\mathbf{z}^*, b^*, \boldsymbol{\lambda}^*)} = \sum_{j=1}^m \lambda_j^* \alpha_j = 0 \quad (6)$$

Условия стационарности (5,6) для лагранжиана $\mathbf{L}(\mathbf{z}, b, \boldsymbol{\lambda})$ являются необходимыми условиями экстремума при ограничениях задачи (4).

Поиск оптимальных значений множителей Лагранжа. Предположим, что $(\mathbf{z}', b', \boldsymbol{\lambda})$ является некоторой точкой, в которой соблюдаются условия стационарности и соблюдаются ограничения задачи (4).

Нетрудно показать, воспользовавшись уравнениями (5,6), что лагранжиан в точке $(\mathbf{z}', b', \boldsymbol{\lambda})$ может быть записан в виде

$$\mathbf{L}(\mathbf{z}', b', \boldsymbol{\lambda}) = g(\boldsymbol{\lambda}) = \sum_{j=1}^m \lambda_j - \frac{1}{2} \sum_{j'=1}^m \sum_{j''=1}^m \lambda_{j'} \lambda_{j''} \alpha_{j'} \alpha_{j''} (\mathbf{x}_{j'} \mathbf{x}_{j''}^t).$$

Отметим, что в силу соблюдения ограничений задачи (4) и неотрицательности множителей Лагранжа в точке $(\mathbf{z}', b', \boldsymbol{\lambda})$ выполняется неравенство

$$g(\boldsymbol{\lambda}) \leq \frac{1}{2} \sum_{i=1}^n (z'_i)^2 \leq \frac{1}{2} \sum_{i=1}^n (z_i^*)^2$$

Из условий дополняющей нежёсткости следует, что в точке $(\mathbf{z}^*, b^*, \boldsymbol{\lambda}^*)$ справедливо равенство

$$\mathbf{L}(\mathbf{z}^*, b^*, \boldsymbol{\lambda}^*) = \frac{1}{2} \sum_{i=1}^n (z_i^*)^2 - \sum_{j=1}^m \lambda_j^* [\alpha_j (\mathbf{z}^* \mathbf{x}_j^t - b^*) - 1] = \frac{1}{2} \sum_{i=1}^n (z_i^*)^2$$

Таким образом максимум $g(\boldsymbol{\lambda})$ равен $\frac{1}{2} \sum_{i=1}^n (z_i^*)^2$ и достигается при $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.

Таким образом, оптимальные значения неотрицательных множителей Лагранжа $(\lambda_1^*, \dots, \lambda_m^*)$ могут быть найдены как решение оптимизационной задачи, которая называется квадратичного программирования, двойственной по отношению к задаче (4):

$$\sum_{j=1}^m \lambda_j - \frac{1}{2} \sum_{j'=1}^m \sum_{j''=1}^m \lambda_{j'} \lambda_{j''} \alpha_{j'} \alpha_{j''} (\mathbf{x}_{j'} \mathbf{x}_{j''}^t) \rightarrow \max \quad (7)$$

$$\sum_{j=1}^m \lambda_j \alpha_j = 0$$

$$\lambda_j \geq 0, j = 1, \dots, m$$

Пусть $(\hat{\lambda}_1, \dots, \hat{\lambda}_m)$ - решение задачи (7) Направляющий вектор оптимальной разделяющей гиперплоскости находится по формуле $\mathbf{z}^* = \sum_{j=1}^m \hat{\lambda}_j \alpha_j \mathbf{x}_j$.

То есть направляющий вектор разделяющей гиперплоскости является линейной комбинацией векторных описаний объектов обучающей выборки, для которых значения соответствующих оптимальных множителей Лагранжа отличны от 0. Такие векторные описания принято называть опорными векторами. Пусть

$$J_0 = \{j = 1, \dots, m \mid [\alpha_j (\mathbf{z}^* \mathbf{x}_j^t - b) - 1] \neq 0\}$$

Из условий дополняющей нежёсткости видно, при $j \in J_0$ обязательно должно выполняться равенство $\hat{\lambda}_j = 0$. Поэтому векторное описание $\mathbf{x}_j$ соответствующего объекта обучающей выборки является опорным вектором, если j не принадлежит J_0 . Оценка параметра сдвига $\hat{b}$ находится из ограничения, соответствующего произвольному опорному вектору.

Распознавание новых объектов. Классификация нового распознаваемого объекта s с описанием $\mathbf{x}$ вычисляется согласно знаку выражения

$$g(\mathbf{x}) = \sum_{j=1}^m \hat{\lambda}_j \alpha_j (\mathbf{x}_j \mathbf{x}^t) - \hat{b}$$

Объект s относится к классу K_1 , если $g(\mathbf{x}) > 0$ и объект s относится к классу K_2 в противном случае.

4.7.2 Случай отсутствия линейной разделимости

Существенным недостатком рассмотренного варианта метода опорных векторов является требование линейной разделимости классов. Однако данный недостаток может быть легко преодолен с помощью следующей модификации, основанной на использовании дополнительного вектора неотрицательных переменных $(\zeta_1, \dots, \zeta_m)$.

Требования об отделимости классов из задачи (3) заменяются более мягкими требованиями:

$$\mathbf{z} \mathbf{x}_j^t \geq b + 1 - \zeta_j \quad \text{при} \quad s_j \in \tilde{S}_t \cap K_1$$

$$\mathbf{z}\mathbf{x}_j^t \leq b - 1 + \zeta_j \quad \text{при } s_j \in \tilde{S}_t \cap K_2, \quad j = 1, \dots, m.$$

При этом выдвигается требование минимальности суммы $\sum_{j=1}^m \zeta_j$. Поиск оптимальных параметров разделяющей гиперплоскости при отсутствии линейной разделимости таким образом сводится к решению задачи квадратично программирования

$$\frac{1}{2} \sum_{i=1}^n z_i^2 + C \sum_{j=1}^m \zeta_j \rightarrow \min$$

$$\alpha_j(\mathbf{z}\mathbf{x}_j^t - b) \geq 1 - \zeta_j, \quad \zeta_j \geq 0, \quad j = 1, \dots, m,$$

Положительная константа C является открытым параметром алгоритма. Иными словами оптимальное значение C подбирается пользователем.

Пусть $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$ - вектор множителей Лагранжа, соответствующих ограничениям $\alpha_j(\mathbf{z}\mathbf{x}_j^t - b) \geq 1 - \zeta_j$;

$\eta^* = (\eta_1^*, \dots, \eta_m^*)$ - вектор множителей Лагранжа, соответствующих ограничениям $\zeta_j \geq 0$, $j = 1, \dots, m$;

Из теоремы ККТ следует, что для точки $(\mathbf{z}^*, b^*, \zeta^*)$, в которой функция $\frac{1}{2} \sum_{i=1}^n z_i^2 + C \sum_{j=1}^m \zeta_j$ достигает своего минимума при ограничениях задачи (4), и некоторых векторов значений неотрицательных множителей Лагранжа λ^* и η^* соблюдаются условия стационарности лагранжиана

$$\mathbf{L}(\mathbf{z}, \zeta, b, \lambda, \eta) = \frac{1}{2} \sum_{i=1}^n z_i^2 + C \sum_{j=1}^m \zeta_j - \sum_{j=1}^m \lambda_j [\alpha_j(\mathbf{z}\mathbf{x}_j^t - b) - 1 - \zeta_j] + \sum_{j=1}^m \eta_j \zeta_j$$

по переменным $(\mathbf{z}, \eta, b)$.

Данные условия записываются в виде

$$\left. \frac{\mathbf{L}(\mathbf{z}, \zeta, b, \lambda, \eta)}{\partial z_i} \right|_{(\mathbf{z}^*, b^*, \zeta^*)} = z_i^* - \sum_{j=1}^m \lambda_j^* \alpha_j x_{ji} = 0, \quad i = 1, \dots, n;$$

$$\left. \frac{\partial \mathbf{L}(\mathbf{z}, b, \lambda)}{\partial b} \right|_{(\mathbf{z}^*, b^*, \lambda^*)} = \sum_{j=1}^m \lambda_j^* \alpha_j = 0;$$

$$\left. \frac{\partial \mathbf{L}(\mathbf{z}, b, \lambda)}{\partial \zeta_j} \right|_{(\mathbf{z}^*, b^*, \lambda^*)} = C - \lambda_j - \eta_j = 0, \quad j = 1, \dots, m$$

Также из теоремы ККТ следует необходимость выполнения m равенств, которые носят название условий дополняющей нежёсткости

$$\lambda_j^* [\alpha_j (\mathbf{z}^* \mathbf{x}_j^t - b) - 1 + \zeta_j] = 0, \quad \eta_j \zeta_j = 0, \quad j = 1, \dots, m.$$

Оптимальные значения множителей $(\lambda_1^*, \dots, \lambda_m^*)$ могут быть найдены как решение двойственной задачи квадратичного программирования.

$$\sum_{j=1}^m \lambda_j - \frac{1}{2} \sum_{j'=1}^m \sum_{j''=1}^m \lambda_{j'} \lambda_{j''} \alpha_{j'} \alpha_{j''} (\mathbf{x}_{j'} \mathbf{x}_{j''}^t) \rightarrow \max \quad (7)$$

$$\sum_{j=1}^m \lambda_j \alpha_j = 0$$

$$C \geq \lambda_j \geq 0, \quad j = 1, \dots, m$$

Как и в случае линейной разделимости направляющий вектор оптимальной разделяющей гиперплоскости находится по формуле $\mathbf{z}^* = \sum_{j=1}^m \lambda_j^* \alpha_j \mathbf{x}_j$. Из условий $C - \lambda_j - \eta_j = 0$ и $\zeta_j \eta_j = 0$ следует что $\eta_j > 0$ и $\zeta_j = 0$ при $0 < \lambda_j < C$.

Также как и в случае существования линейной разделимости параметра сдвига $\hat{b}$ находится из ограничения, соответствующего произвольному опорному вектору $\mathbf{x}_j$. Действительно, из условий дополняющей нежёсткости и из следующего из них равенства $\zeta_j = 0$ следует выполнение равенства $\alpha_j (\mathbf{z}^* \mathbf{x}_j^t - \hat{b}) - 1 = 0$, эквивалентного равенству $\hat{b} = \mathbf{z}^* \mathbf{x}_j^t - \alpha_j$.

Распознавание нового объекта S производится по его описанию $\mathbf{x}$ также как и в случае линейно разделимых классов с помощью решающего правила (8) по величине распознающей функции $g(\mathbf{x})$.

4.7.3 Построение оптимальных нелинейных разделяющих поверхностей с помощью метода опорных векторов.

Предположим что в исходном признаковом пространстве эффективное линейное разделение отсутствует. Однако может существовать такое евклидово пространство H_y и такое отображение Φ из области пространства $\mathbf{R}^n$ содержащей описания распознаваемых объектов, в пространство H_y , что образы объектов обучающей выборки из классов K_1 и K_2 оказываются разделимыми с помощью некоторой гиперплоскости P_y . Пусть $\{\mathbf{y}_1, \dots, \mathbf{y}_m\}$ - образы в пространстве H_y векторов описаний объектов обучающей выборки $\{\mathbf{y}_1, \dots, \mathbf{y}_m\}$.

Линейная разделимость означает существование решения аналога задачи квадратичного программирования (4) для пространства H_y , которое сводится к решению двойственной задачи

$$\sum_{j=1}^m \lambda_j - \frac{1}{2} \sum_{j'=1}^m \sum_{j''=1}^m \lambda_{j'} \lambda_{j''} \alpha_{j'} \alpha_{j''} (\mathbf{y}_{j'} \mathbf{y}_{j''}^t) \rightarrow \max$$

$$\sum_{j=1}^m \lambda_j \alpha_j = 0$$

$$\lambda_j \geq 0, j = 1, \dots, m$$

Отметим, что необходимость полного восстановления преобразования $\Phi(\mathbf{x})$ для поиска всех коэффициентов задачи квадратичного программирования (13) отсутствует. Достаточно найти функцию, связывающую скалярное произведение $(\mathbf{y}_{j'} \mathbf{y}_{j''}^t)$ с векторами $\mathbf{x}_{j'}$ и $\mathbf{x}_{j''}$, где $\mathbf{y}_{j'} = \Phi(\mathbf{x}_{j'})$ и $\mathbf{y}_{j''} = \Phi(\mathbf{x}_{j''})$.

Такую функцию мы далее будем называть потенциальной и обозначать $K(\mathbf{x}_{j'}, \mathbf{x}_{j''})$. Можно подобрать потенциальную функцию таким образом, чтобы решение (13) было оптимальным. При этом поиск оптимальной потенциальной функции может производиться внутри некоторого заранее заданного семейства. Например, потенциальную функцию можно задать с помощью простого сдвига $K(\mathbf{x}_{j'}, \mathbf{x}_{j''}) = \mathbf{x}_{j'}^t \mathbf{x}_{j''} + \theta$. Решение, полученное путём замены скалярных произведений на потенциальные функции, может

рассматриваться как построении линейной разделяющей поверхности в трансформированном пространстве, если удаётся доказать существование отображения $\Phi(\mathbf{x})$, для которого при произвольных $\mathbf{x}'$ и $\mathbf{x}''$ из $\mathbf{R}^n$ выполняется равенство

$$K(\mathbf{x}', \mathbf{x}'') = F(\mathbf{x}') F^t(\mathbf{x}'')$$

Существование преобразования $\Phi(\mathbf{x})$, для которого выполняется равенство (15), было показано для неотрицательных симметричных потенциальных функций вида

$$K(\mathbf{x}', \mathbf{x}'') = [\mathbf{x}'(\mathbf{x}'')^t]^d + c,$$

где d - целое число, c - вещественная константа.

Существование преобразования $\Phi(\mathbf{x})$ с выполнением равенство (15) доказано также для ядерных функции типа гауссианы

$$K(\mathbf{x}', \mathbf{x}'') = \frac{1}{\sqrt{2\pi d}} e^{-\frac{(\mathbf{x}' - \mathbf{x}'')^2}{2d}},$$

где σ - вещественная неотрицательная константа (размер ядра). Поскольку в общем случае преобразование является нелинейным, то преобразом в пространстве $\mathbf{R}^n$ линейной разделяющей гиперплоскости, существующей в пространстве H_y , может оказаться нелинейная поверхность.

Для большого числа прикладных задач линейная разделимость является недостижимой. Поэтому выбор ядерной функции может производиться из требования о минимальности числа ошибок в смысле задачи квадратичного программирования (9). На практике подбор ядерных функций и их параметров производится исходя из требования достижения максимальной обобщающей способности, которая оценивается с помощью скользящего контроля или оценок на контрольной выборке. Опыт решения прикладных задач показывает, что высокая эффективность распознавания достигается при выборе в качестве ядерной функции гауссианы.

Прототипом метода опорных векторов явился метод «Обобщенный портрет», разработанный В.Н.Вапником и А.Я.Червоненкисом [1]. В современном варианте метод был предложен в работе [21]. Подробное описание метода появилось в работе [18] в 1998 году. В настоящее время метод опорных векторов является одним из наиболее распространённым в мире средством решения задач распознавания, высокая эффективность которого подтверждается практикой. В связи с этим были предложены подходы, использующие основные принципы метода опорных векторов для решения задач регрессионного анализа.

Литература

- [1] Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). М.: Наука. 1974. - 416 с.
- [3] Воронцов К.В. (Курс лекций). www.machinelearning.ru
- [4] Докукин А.А., Сенько О.В. [Оптимальные выпуклые корректирующие процедуры в задачах высокой размерности Журнал вычислительной математики и математической физики](#). 2011. Т. 51. № 9. С. 1751-1760.
- [5] А.М. Дубров, В.С. Мхитарян, Л.И. Трошин Многомерные статистические методы: Учебник, - М.: Финансы и статистика, 2000, - 352с.
- [6] Дмитриев А.Н., Журавлев Ю.И., Кренделев Ф.П., О математических принципах классификации предметов и явлений. Сб. "Дискретный анализ". Вып. 7. Новосибирск, ИМ СО АН СССР. 1966. С. 3-11.
- [7] Донской В.И. Алгоритмические модели обучения классификации: обоснование, сравнение, выбор. –Симферополь, «ДИАЙПИ», 2014,-227 с.
- [8] Дюкова Е.В. Алгоритмы распознавания типа “Кора”: сложность реализации и метрические свойства// Распознавание, классификация, прогноз (матем. методы и их применение). М.: Наука, 1989. Вып.2. С. 99-125.
- [9] Журавлев Ю.И., Никифоров В.В. Алгоритмы распознавания, основанные на вычислении оценок // Кибернетика. 1971. №3. С. 1-11.
- [10] Журавлев Ю.И., ИЗБРАННЫЕ НАУЧНЫЕ ТРУДЫ. - М.: Издательство Магистр, 1998. - 420 с.
- [11] Журавлев Ю.И., Рязанов В.В., Сенько О.В. «Распознавание». Математические методы. Программная система. Практические применения. — М.: Фазис, 2006. 159 с.

[

[12] Кузнецов В.А., Сенько О.В., Кузнецова А.В. и др. Распознавание нечетких систем по методу статистически взвешенных синдромов и его применение для иммуногематологической характеристики нормы и хронической патологии. // Химическая физика. 1996. Т.15. N 1. С.81-100.

[13] Лбов Г.С., Старцева Н.С. Логические решающие правила и вопросы статистической устойчивости решений. Новосибирск: Изд-во Ин-та математики, 1999, 212 с.

[14] Мерков А.Б. Распознавание образов: Введение в методы статистического обучения. . М.: Едиториал УРСС, 2011. — 256 с.

[15] А.С. Потапов. Распознавание образов и машинное восприятие. Общий подход на основе принципа минимальной длины описаний. –Спб.: Политехника, 2007, -548 с.

[16] Рязанов В.В. Логические закономерности в задачах распознавания (параметрический подход)//Ж. вычисл. матем. и матем. физ., 2007, [том 47, номер 10](#), страницы 1793–1808

[17] L. Breiman Bagging predictors. Machine learning, 24, 123-140, 1996.

[18] Chris. J.C. Burges A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers, Boston. Manufactured in The Netherlands. Appeared in: Data Mining and Knowledge Discovery 2, 121-167, 1998.

[19] T. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization. Machine Learning, 40(2):139–157, 2000.\

[20] Kuncheva L.I. Combining Pattern Classifiers. Methods and Algorithms. — Wiley Interscience, New Jersey, 2004.

[21] [Cortes, C.](#); Vapnik, V. (1995). "Support-vector networks". // [Machine Learning](#) **20** (3): 273

[22] Senko O.B., Kuznetsova A.B. A recognition method based on collective decision making using systems of regularities of various types. Pattern Recognition and Image Analysis. 2010. V. 20. № 2. P. 152–162.

