АНАЛИЗ ДАННЫХ И ЗНАНИЙ В ЭКСПЕРТНЫХ СИСТЕМАХ
(Вычислительные системы)
1990 Выпуск 134

УДК 519.237:519.853.4

РЕШЕНИЕ ЗАДАЧИ МНОГООТКЛИКОВОЙ РЕГРЕССИИ ДЛЯ РАЗНОТИПНЫХ ПРИЗНАКОВ В КЛАССЕ ЛОГИЧЕСКИХ РЕШАЮЩИХ ПРАВИЛ

Н.Г.Старцева

Введение

К настоящему времени наиболее широкое распространение получили регрессии с одномерным откликом [1]. Однако реальные объекты, для описания которых привлекается регрессионный анализ, нередко имеют несколько откликов.В связи с этим представляет интерес многооткликовая регрессия. Появились такие модификации многооткликовой регресии, как псевдонезависимая модель, модель в виде системы одновременных (синхронных) уравнений. В первом случае речь идет о ряде стохастически связанных между собой одномерных регрессионных уравнений. Во втором - предполагается, что между разными откликами системы существуют линейные связи.

В работе предлагается другой подход к построению многооткликовой регрессии - рассмотрение ее в классе логических решающих правил. При этом допускается разнотипность переменных факторов, т.е. исходная система факторов может одновременно содержать булевы, номинальные, ранговые, количественные факторы. В дальнейшем признаки (факторы): булевы, номинальные,ранговые будем называть дискретными признаками.

Регрессию с одномерным откликом в классе логических решающих правил рассматривал Г.С.Лбов [2]. Класс логических решающих правил обладает рядом достоинств: на его языке представимы регрессионные модели различной сложности, модели описываются на языке, близком к естественному языку логических суждений, анализируется разнотипная информация.

§1. Постановка задачи многооткликовой регрессии в случае разнотипных переменных

Пусть для описания каждого наблюдения a из некоторой генеральной совокупности Γ используются признаки $X_1, \ldots, X_l$, $Z_1, \ldots, Z_n$, $Y_1, \ldots, Y_k$. Пусть признаки $X_1, \ldots, X_l$, $Z_1, \ldots, Z_n$ - независимые переменные (факторы), а $Y_1, \ldots, Y_k$ - зависимые переменные (отклики). Пусть признаки $X_1, \ldots, X_l$, $Y_1, \ldots, Y_k$ - непрерывные случайные величины, признаки $Z_1, \ldots$ $\ldots, Z_n$ - дискретные случайные величины, для которых определены условные плотности распределения $p(x, y/z)$ и распределения вероятностей $P(z)$ в многомерной области D, где $x =$ $= (X_1(a), \ldots, X_l(a))$, $z = (Z_1(a), \ldots, Z_n(a))$, $y =$ $= (Y_1(a), \ldots, Y_k(a))$; $D = D_x \times D_z \times D_y$, $x \in D_x$, $z \in D_z$,

$$y \in D_y,\ D_x = \prod_{j=1}^{l} D_{x_j},\ D_z = \prod_{j=1}^{n} D_{z_j},\ D_y = \prod_{j=1}^{k} D_{y_j}.$$ Здесь

D_{x_j}, D_{z_j}, D_{y_j} - соответственно области значений признаков X_j, Z_j, Y_j. Рассматривается случай независимых переменных откликов $Y_1, \ldots, Y_k$.

Целью регрессионного анализа является построение правила предсказания, которое позволило бы по описанию x и z предсказать значение y.

Под решающим правилом f из некоторого класса Φ будем понимать отображение $f: D_x \times D_z \rightarrow D_y$.

Прогнозируемый вектор значений признаков $Y_1,\ldots,Y_j,\ldots,Y_k$ с помощью правила f обозначим через $y' = f(x,z) = (y'_1,\ldots,y'_k)$ (прогнозируемые отклики).

Определим условную плотность распределения:

$$p(y/x,z) = \frac{p(y,x/z)}{p(x/z)} = \frac{p(y,x/z)}{\int\limits_{D_y} p(y,x/z)dy}.$$

Критерием качества решающего правила f будем считать величину:

$$F_f = M_z M_x M_y\{g(z,x,y)\} =$$

$$= \sum_{z\in D_z} P(z)\int\limits_{D_x} p(x/z)\left\{\int\limits_{D_y} g(z,x,y)p(y/x,z)dy\right\}dx,$$

где

$$g(z,x,y) = \sum_{j=1}^{k}\lambda_j(y_j-y'_j)^2,$$

$$M_y\{g(z,x,y)\} = \int\limits_{D_y} g(z,x,y)p(y/x,z)dy,$$

$$M_x M_y\{g(z,x,y)\} = \int\limits_{D_x} p(x/z)[M_y\{g(z,x,y)\}]dx,$$

$$M_z M_x M_y\{g(z,x,y)\} = \sum_{z\in D_z} P(z)M_x M_y\{g(z,x,y)\},$$

λ_j - весовой коэффициент.

Данный критерий является распространением известного критерия наименьших квадратов в случае многооткликовой регрессии от разнотипных признаков. Так как $Y_1,\ldots,Y_k$ - независимые переменные, то

$$\int\limits_{D_y} g(z,x,y)p(y/x,z)dy = \sum_{j=1}^{k}\int\limits_{D_{y_1}}\ldots\int\limits_{D_{y_j}}\ldots\int\limits_{D_{y_k}} \lambda_j(y_j-y_j')^2 \times$$

$$\times \prod_{j=1}^{k} p(y_j/x,z)dy_1\ldots dy_j\ldots dy_k =$$

$$= \sum_{j=1}^{k}\int\limits_{D_{y_1}} p(y_1/x,z)dy_1\ldots\int\limits_{D_{y_j}} \lambda_j(y_j-y_j')^2 p(y_j/x,z)dy_j\ldots$$

$$\ldots\int\limits_{D_{y_k}} p(y_k/x,z)dy_k = \sum_{j=1}^{k}\lambda_j\int\limits_{D_{y_j}} (y_j-y_j')^2 p(y_j/x,z)dy_j ,$$

так как

$$\int\limits_{D_{y_j}} p(y_j/x,z)dy_j = 1, \quad j = \overline{1,k} .$$

Тогда

$$F_f = \sum_{z\in D_z} P(z)\int\limits_{D_x} p(x/z) \times$$

$$\times\left\{\sum_{j=1}^{k}\lambda_j\int\limits_{D_{y_j}} (y_j-y_j')^2 p(y_j/x,z)dy_j\right\} dx .$$

Известно [2], что оптимальным решающим правилом, обеспечивающим минимум критерия, является следующее правило:

$$f_0(x,z) = y' = (y_1',\ldots, y_j',\ldots, y_k'),$$

где

$$y'_j = M(y_j/x,z) = \int\limits_{D_{y_j}} y_j p(y_j/x,z)dy_j, \quad j = \overline{1,k}.$$

§2. Решение задачи многооткликовой регрессии в классе логических решающих правил

Рассмотрим решение задачи в классе логических решающих правил Φ_M (будем аппроксимировать функцию $y' = f(x,z)$ в классе логических решающих правил, $f \in \Phi_M$).

Опишем класс логических решающих правил. Здесь решающим правилом f задается разбиение области $D_X \times D_Z$ на M подмножеств $\alpha = \{E^1,\dots,E^t,\dots,E^M\}$, т.е. $D_X \times D_Z = \bigcup\limits_{t=1}^{M} E^t$, $E^t \cap E^l = \emptyset$ для $t \neq l$, $2 \le M < \infty$. Отличительной особенностью данного подхода является то, что для описания каждого подмножества E^t используется своя подсистема признаков $\{X_{j_1},\dots,X_{j_{l_t}}, Z_{j_1},\dots,Z_{j_{n_t}}\}$, состоящая из l_t непрерывных признаков и из n_t дискретных. Здесь $E^t = E^t_X \times E^t_Z$, где

$$E^t_X = \tilde{E}^t_X \times \prod_{j \notin I^t_X} D_{X_j}, \quad E^t_Z = \tilde{E}^t_Z \times \prod_{j \notin I^t_Z} D_{Z_j},$$

$$I^t_X = \{j_1,\dots,j_{l_t}\}, \quad I^t_Z = \{j_1,\dots,j_{n_t}\}.$$

Общее число признаков, используемых для описания разбиения α, будет $m = |I_X| + |I_Z|$, где

$$I_X = \bigcup_{t=1}^{M} I^t_X, \quad I_Z = \bigcup_{t=1}^{M} I^t_Z, \quad m \le n+l.$$

В простейшем случае

$$\widetilde{E}_X^t = \prod_{j_\nu \in I_X^t} E_{j_\nu}, \qquad \widetilde{E}_Z^t = \prod_{j_\nu \in I_Z^t} E_{j_\nu},$$

где $E_{j_\nu} \subset D_{X_j}$ или $E_{j_\nu} \subset D_{Z_j}$ и $E_{j_\nu} \cup \overline{E}_{j_\nu} = D_{X_j}$ или $E_{j_\nu} \cup \overline{E}_{j_\nu} = D_{Z_j}$. Очевидно, что такое представление подмножеств $\widetilde{E}_X^t$, $\widetilde{E}_Z^t$, $t=\overline{1,M}$, есть своеобразное ограничение на разбиение множества D.

Класс разбиений, удовлетворяющий этому ограничению, обозначим через Ψ. В дальнейшем будем рассматривать только данный класс разбиений множества D.

Разбиение $\alpha \in \Psi_M$ удобнее всего описывать в виде дерева. Под деревом B принимается корневое дихстомическое дерево, у которого каждой внутренней вершине (узлу) ставится в соответствие некоторый предикат; ветвям, исходящим из внутренней вершины, соответствует истинность или ложность высказывания, получающегося при замене признаков их значениями. Если каждой конечной вершине b^t, $t = \overline{1,M}$, дерева B приписать решение, то такое дерево будем называть деревом решений (логическим решающим правилом). Логическое решающее правило f задается парой $\langle \alpha, r^\alpha \rangle$, $\alpha \in \Psi_M$, $r^\alpha \in R_M$, где R_M - множество возможных решений. Класс логических решающих правил определяется декартовым произведением $\Phi_M = \Psi_M \times R_M$.

Рассматриваются следующие типы предикатов: $J(a, E_j) = [X_j(a) \in E_j]$, где $a \in \Gamma$, $E_j \subset W_j$. Для дискретного признака под W_j понимается любое подмножество из множества различных значений признака X_j; для количественного признака W_j - множество всевозможных подынтервалов некоторого интервала $D_j = [\beta_j, \gamma_j]$.

В статье рассматривается следующая модель регрессионного анализа: для любого разбиения α укажем набор решений $r^{\alpha} = \{y'_1, \ldots, y'_t, \ldots, y'_M\}$, где $y'_t = (y'_{1t}, \ldots, y'_{jt}, \ldots, y'_{kt})$ - некоторый вектор, $y'_{jt} \in [\beta_j, \gamma_j]$, $y'_t = f_t(x,z) = (f_{1t}(x,z), \ldots, f_{jt}(x,z), \ldots, f_{kt}(x,z))$.

Критерий качества предсказания при использовании решающего правила $f \in \Phi_M$ вычисляется следующим образом:

$$F_f = F(\alpha, r^{\alpha}) = \sum_{t=1}^{M} P[(x,z) \in E^t] \cdot F_f^t,$$

$$P[(x,z) \in E^t] = \sum_{z \in E_Z^t} P(z) \int_{E_X^t} p(x/z) dx,$$

$$F_f^t = \sum_{z \in E_Z^t} P(z) \int_{E_X^t} p(x/z) \times$$

$$\times \left\{ \sum_{j=1}^{k} \lambda_j \int_{\beta_j}^{\gamma_j} (y_j - y'_{jt})^2 p(y_j/x,z) dy_j \right\} dx,$$

где $P[(x,z) \in E^t]$ - вероятность того, что точка (x,z) попала в область E^t, F_f^t - значение критерия качества для области E^t.

Под оптимальным правилом $f(M) \in \Phi_M$ понимается правило, при котором

$$F_{f(M)} = F^*(\alpha, r^{\alpha}) = \min_{\alpha \in \Psi_M} \cdot \min_{r^{\alpha} \in R_M} F(\alpha, r^{\alpha}).$$

Очевидно, что $F_{f(M)} \geq F_{f_0}$.

Наилучший набор решений r^{α} для фиксированного разбиения α задается следующим образом: если $(x,z) \in E^t$, то

$$y'_{jt} = \sum_{z \in E^t_z} P(z) \left[\int_{E^t_x} M(y_j/x,z) p(x/z) dx \right], \quad t = \overline{1,M}.$$

При $M = 1$ получим

$$y'_j = \sum_{z \in D_z} P(z) \int_{D_x} M(y_j/x,z) p(x/z) dx,$$

а критерий

$$F_{f(1)} = \sum_{z \in D_z} P(z) \int_{D_x} p(x/z) \times$$

$$\times \left\{ \sum_{j=1}^{k} \lambda_j \int_{\beta_j}^{\gamma_j} (y_j - y'_j)^2 p(y_j/x,z) dy_j \right\} dx .$$

В дальнейшем будем использовать нормированные критерии:

$$\frac{F_{f(M)}}{F_{f(1)}} = F(M); \quad \frac{F_{f_0}}{F_{f(1)}} = F_0 .$$

На практике условные плотности распределения и распределения вероятностей неизвестны, решающее правило строится по подмножеству наблюдений $A \subseteq \Gamma$, которому соответствует таблица данных $V = \{x^i_\nu, z^i_\mu, y^i_j\}$, $i = \overline{1,N}$; $\nu = \overline{1,l}$; $\mu = \overline{1,n}$; $j = \overline{1,k}$, называемая обучающей выборкой, N - объем обучающей выборки.

Для фиксированного разбиения α набор решений $r^{\alpha} = (y'_1, \ldots, y'_t, \ldots, y'_M)$ строится так:

$$y'_{jt} = \bar{y}_{jt} = \sum_{(x^i,z^i)\in E^t} y^i_j / N_t \quad ,$$

где N_t - число точек (x^i,z^i), принадлежащих подмножеству E^t, $y'_t = (y'_{1t},\ldots,y'_{jt},\ldots,y'_{kt})$.

Критерий качества предсказания при фиксированном разбиении α оцениваем следующим образом:

$$\bar{F}(\alpha,r^{\alpha}) = \bar{F}_{\bar{f}(M)} = \frac{1}{N}\left[\sum_{t=1}^{M} \sum_{(x^i,z^i)\in E^t} \sum_{j=1}^{k} \lambda_j (y^i_j - \bar{y}_{jt})^2\right],$$

а нормированный критерий

$$\bar{F}(M) = \frac{\bar{F}_{\bar{f}(M)}}{\bar{F}_{\bar{f}(1)}} ,$$

где

$$\bar{F}_{\bar{f}(1)} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} \lambda_j (y^i_j - \bar{y}_j)^2 , \quad \bar{y}_j = \frac{1}{N}\sum_{i=1}^{N} y^i_j \ .$$

Под наилучшим решающим правилом $\bar{f}^*(M) \in \Phi_M$, определенным по таблице V , будем понимать правило,для которого

$$\bar{F}^*(M) = \min_{\bar{f}(M)\in\Phi_M} \bar{F}(M) .$$

§3. Алгоритм многооткликовой регрессии

Как уже отмечалось в предыдущем параграфе, разбиение α представляется в виде дерева решений, при построении которого осуществляется последовательное "наращивание" вершин дерева в соответствии с принципом присоединения "лучшей" вершины к "лучшей". Подобное построение дерева решений, но для задач дискриминантного анализа рассматривалось в [3].

До построения дерева решений в алгоритме предусмотрена нормировка всех значений y_j^i, $i=\overline{1,N}$; $j=\overline{1,k}$. Нормировка проводится следующим образом: $\tilde{y}_j^i = (y_j^i-\bar{y}_j)/s_j$, где

$$s_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_j^i - \bar{y}_j)^2}\,.$$

Нормированные значения $\{\tilde{y}_j^i\}$ используются при вычислении критерия $\bar{F}(M)$. В алгоритме рассматриваются следующие типы предикатов:

1. Для дискретных признаков $J(a,E_j) = [Z_j(a)\in E_j]$, где $a\in \Gamma$, $E_j\in W_j$, $j = \overline{1,n}$. Под W_j понимается множество различных значений признака Z_j или всевозможных объединений двух различных значений.

2. Для количественных признаков $J(a,E_j) = [X_j(a)\in E_j]$, где $a\in \Gamma$, $E_j\in W_j$, $j = \overline{1,1}$. Под W_j понимается множество интервалов типа $[\rho',\rho'']$, $[-\infty,\rho'']$, где $\rho',\rho''\in D'_{X_j}$, D'_{X_j} - множество среднеарифметических значений двух соседних несовпадающих выборочных значений признака X_j. Для сокращения машинного времени среднеарифметическое значение между двумя соседними точками множества D_{X_j} не рассматривается, если для всех $l = \overline{1,k}$ имеем $(y_l^r - y_l^{r+1}) > \varepsilon$, где $\varepsilon = \varepsilon'\cdot s$; $r = \overline{1,N}$;

$$s=\sqrt{\frac{1}{N\cdot k}\sum_{i=1}^{N}\sum_{l=1}^{k}(y_l^i-\bar{y})^2};\quad \bar{y}=\frac{1}{k}\sum_{l=1}^{k}\bar{y}_l;\quad \varepsilon'\in[0,1],$$

ε' - входной параметр, близкий к нулю.

Предположим, что построено дерево с m конечными вершинами, $m = \overline{2,M}$. Этому этапу соответствует некоторое разбиение пространства переменных $\alpha = \{E^1, \ldots, E^t, \ldots, E^m\}$. Далее для каждого подмножества E^t ищется свое, лучшее в смысле критерия $\Delta \overline{F}^t$ высказывание J_t. Множество E^t разбивается на два подмножества E^{t_1} и E^{t_2}, $E^{t_1} \cup E^{t_2} = E^t$.

Критерием оценки качества вершины b^t дерева решений является величина:

$$\Delta\overline{F}^t = \overline{F}(m) - \overline{F}(m+1) =$$

$$= \frac{1}{N}\Bigg[\sum_{(x^i, z^i) \in E^t} \sum_{j=1}^{k} \lambda_j (y_j^i - \overline{y}_{jt})^2 -$$

$$- \sum_{(x^i, z^i) \in E^{t_1}} \sum_{j=1}^{k} \lambda_j (y_j^i - \overline{y}_{jt_1})^2 -$$

$$- \sum_{(x^i, z^i) \in E^{t_2}} \sum_{j=1}^{k} \lambda_j (y_j^i - \overline{y}_{jt_2})^2 \Bigg] \Big/ \overline{F}_{\overline{f}(1)},$$

где

$$\overline{y}_{jt_1} = \frac{1}{N_{t_1}} \sum_{(x^i, z^i) \in E^{t_1}} y_j^i ,$$

$$\overline{y}_{jt_2} = \frac{1}{N_{t_2}} \sum_{(x^i, z^i) \in E^{t_2}} y_j^i ,$$

N_{t_1}, N_{t_2} - число точек (x^i, z^i), принадлежащих соответственно подмножествам E^{t_1}, E^{t_2}.

Находим

$$\Delta\bar{F}^{t^*} = \max_{t=\overline{1,m}} \Delta\bar{F}^{t}.$$

Дальнейшее ветвление дерева проводим из вершины b^{t^*} и получаем дерево с $(m+1)$ конечной вершиной. Этому этапу соответствует некоторое разбиение пространства переменных $\alpha =$

$= \{E^1, \ldots, E^{t-1}, E^{t^*_1}, E^{t^*_2}, E^{t+1}, \ldots, E^m\}$.

Число конечных вершин m увеличиваем до тех пор, пока не выполнится одно из трех условий:

1. $\bar{F}(m) \le \delta$ (δ - некоторый входной параметр, близкий к 0, $\delta \in [0,1]$).

2. Число конечных вершин m не превысит некоторое число M (M - входной параметр).

3. Все вершины b^t делению не подлежат. Вершина b^t не подлежит делению, если не существует ни одного разбиения множества E^t на два подмножества E^{t_1} и E^{t_2} таких, что $N_{t_1} > \mu$ и $N_{t_2} > \mu$, где μ - минимально допустимое число наблюдений в конечной вершине (μ - входной параметр).

Конечной вершине b^t, $t = \overline{1,M}$, дерева решений ставится в соответствие следующий набор решений:

1) вектор средних $(\bar{y}_{1t}, \ldots, \bar{y}_{jt}, \ldots, \bar{y}_{kt})$, где

$$\bar{y}_{jt} = \sum_{(x^i, z^i)\in E^t} y^i_j / N_t ;$$

2) вектор оценок стандартных уклонений $(s_{1t}, \ldots$
$\ldots, s_{jt}, \ldots, s_{kt})$, где

$$s_{jt} = \frac{1}{N_t} \sum_{(x^i, z^i)\in E^t} (y^i_j - \bar{y}_{jt})^2 ;$$

3) вектор наибольших значений $(\hat{y}_{1t}, \dots, \hat{y}_{jt}, \dots, \hat{y}_{kt})$, где

$$\hat{y}_{jt} = \max_{(x^i, z^i) \in E^t} y_j^i ;$$

4) вектор наименьших значений $(\check{y}_{1t}, \dots, \check{y}_{jt}, \dots, \check{y}_{kt})$, где

$$\check{y}_{jt} = \min_{(x^i, z^i) \in E^t} y_j^i .$$

Необходимо отметить, что для формирования векторов решений используются первоначальные ненормированные значения откликов y_j^i .

Пусть решающее правило построено. Для прогноза значений $y_1, \dots, y_k$ наблюдения a в предикаты, находящиеся в узлах дерева, подставляются конкретные значения признаков, и достигается некоторая конечная вершина b^t , которой соответствует вектор средних $(\bar{y}_{1t}, \dots, \bar{y}_{kt})$. Компоненты этого вектора и являются прогнозируемыми значениями $y_1, \dots, y_k$ для рассматриваемого наблюдения a .

Заметим, что если в таблице V для некоторого наблюдения a отсутствует значение признака X_j или Z_j ,то при выборе предиката $J(a, E_j)$ это наблюдение не рассматривается. Если же для прогнозируемого наблюдения a пропущено значение признака X_j или Z_j , то при определении вершины b^t рассматривается среднее значение для этого признака.

В заключение отметим, что алгоритм реализован на языке FORTRAN для машин типа ЕС. Работа алгоритма продемонстрирована на тестовом примере (см.приложение 1). В приложении 1 описаны обучающая выборка V и прогнозируемая выборка. В таблицах имеются пропуски (пропуск выделен чертой). В приложении 2 описаны результаты работы программы.

Литература

1. ДРЕЙПЕР Н., СМИТ Г. Прикладной регрессионный анализ: Пер. с англ. /Под ред. Ю.П.Адлера и В.Г.Горского. - М.: Финансы и статистика, 1986. Т.2.

2. LBOV G.S. The logical decision rules for automatic discovery of Knowledge in expert system data base //Int. J. of pattern recognition and artificial intelligence. - Singapore, New Jersey, Hong Kong, 1989. - Vol. 3, N 1.

3. ЛБОВ Г.С., СТАРЦЕВА Н.Г. Алгоритм многоклассового распознавания,основанный на логических решающих функциях. - Новосибирск.–1985. - Вып. 111: Вычислительные системы. - С. 3-10.

Поступила в ред.-изд.отд.
24 января 1989 года

Приложение 1

Контрольный пример

Обучающая выборка					
№ объекта	Признаки				
	X_1	X_2	Z_1	Y_1	Y_2
1	10	0.8	1	1	0.1
2	11	0.5	2	2	0.2
3	12	0.7	4	1	0.1
4	16	0.1	5	2	0.3
5	10	0.4	1	1	0.2
6	12	0.6	2	3	0.15
7	13	0.5	4	1	0.25
8	-	0.6	3	4	0.17
9	14	0.7	1	1	0.19
10	15	0.2	6	3	0.4
11	20	0.9	5	2	0.4
12	8	0.1	1	6	0.4
13	9	0.2	1	5	0.2
14	7	0.3	1	6	0.4
15	9	0.1	2	5	0.3
16	8	0.2	3	3	0.6
17	5	0.3	2	6	0.1
18	4	0.1	3	3	0.7
19	6	0.2	4	7	0.6
Прогнозируемая выборка					
1	10	0.7	1		
2	12	0.6	2		
3	8	0.3	2		
4	7	0.2	3		
5	-	0.2	4		
6	15	0.4	6		

Номер вершины NF(K5,3)=1

S(1,1)=3,26315	D(1,1)=1,96929	SMOL(1,1)=1,0	BIG(1,1)=7.0
S(1,2)=0,30316	D(1,2)=0,17529	SMOL(1,2)=0,1	BIG(1,2)=0,7
F(1,1)=18,31398	F(1,2)=0,51805		

Номер вершины NF(K5,3)=2

S(2,1)=5,12499	D(2,1)=1,36359	SMOL(2,1)=3,0	BIG(2,1)=7,0
S(2,2)=0,41250	D(2,2)=0,19645	SMOL(2,2)=0,1	BIG(2,2)=0,7
F(2,1)=2,28345	F(2,2)=0,39284		

Номер вершины NF(K5,3)=3

S(3,1)=1,70000	D(3,1)=0,78102	SMOL(3,1)=1,0	BIG(3,1)=3,0
S(3,2)=0,22900	D(3,2)=0,10329	SMOL(3,2)=0,1	BIG(3,2)=0,4
F(3,1)=2,47449	F(3,2)=0,45293		

Номер вершины NF(K5,3)=6

S(4,1)=1,00000	D(4,1)=0,0	SMOL(4,1)=1,0	BIG(4,1)=1,0
S(4,2)=0,16800	D(4,2)=0,05913	SMOL(4,2)=0,1	BIG(4,2)=0,25
F(4,1)=0,0	F(4,2)=0,0		

Номер вершины NF(K5,3)=7

S(5,1)=2,40000	D(5,1)=0,48990	SMOL(5,1)=2,0	BIG(5,1)=3,0
S(5,2)=0,29000	D(5,2)=0,10198	SMOL(5,2)=0,15	BIG(5,2)=0,4
F(5,1)=0,0	F(5,2)=0,0		

Номер вершины NF(K5,3)=4

S(6,1)=5,50000	D(6,1)=1,5	SMOL(6,1)=3,0	BIG(6,1)=7,0
S(6,2)=0,50000	D(6,2)=0,1	SMOL(6,2)=0,4	BIG(6,2)=0,6
F(6,1)=0,0	F(6,2)=0,0		

Номер вершины NF(K5,3)=5

S(7,1)=4,75000	D(7,1)=1,08972	SMOL(7,1)=6,0	BIG(7,1)=6,0
S(7,2)=0,32500	D(7,2)=0,22776	SMOL(7,2)=0,1	BIG(7,2)=0,7
F(7,1)=0,0	F(7,2)=0,0		

Пояснения к таблице:

S(K5,J) - среднее значение J-го признака в вершине NF(K5,3)

D(K5,J) - стандартное отклонение J-го признака в вершине NF(K5,3)

SMOL(K5,J) - наименьшее значение J-го признака в вершине NF(K5,3)

BIG(K5,J) - наибольшее значение J-го признака в вершине NF(K5,3)

F(K5,1) - значение критерия вершины NF(5,3)

F(K5,2) - значение критерия дерева на шаге K5

J = 1,K; K5 = 1, KWT

Результаты прогноза

Количество объектов на контроле M1 = 6

Объект номер 1 = 1 попал в вершину 6

SY(1,1)=1,00000 DY(1,1)=0,0 BIGY(1,1)=1.0000 SMOLY(1,1)=1.0000
SY(1,2)=0,16800 DY(1,2)=0,05913 BIGY(1,2)=0,25 SMOLY(1,2)=0,1000

Объект номер 1 = 2 попал в вершину 7

SY(2,1)=2,40000 DY(2,1)=0.48990 BIGY(2,1)=3,0000 SMOLY(2,1)=2,0000
SY(2,2)=2,9000 DY(2,2)=0,10198 BIGY(2,2)=0,4000 SMOLY(2,2)=0,1500

Объект номер 1 = 3 попал в вершину 4

SY(3,1)=5,50000 DY(3,1)=1,50000 BIGY(3,1)=7,0000 SMOLY(3,1)=3,0000
SY(3,2)=0,50000 DY(3,2)=0,10000 BIGY(3,2)=6,000 SMOLY(3,2)=0,4000

Объект номер 1 = 4 попал в вершину 4

SY(4,1)=5,50000 DY(4,1)=1,50000 BIGY(4,1)=7,0000 SMOLY(4,1)=3,0000
SY(4,2)=0,50000 DY(4,2)=0,10000 BIGY(4,2)=0,6000 SMOLY(4,2)=0,4000

Объект номер 1 = 5 попал в вершину 6

SY(5,1)=1,00000 DY(5,1)=0,0 BIGY(5,1)=1,0000 SMOLY(5,1)=1,0000
SY(5,2)=0,16800 DY(5,2)=0,05913 BIGY(5,2)=0,2500 SMOLY(5,2)=0,1000

Объект номер 1 = 6 попал в вершину 7

SY(6,1)=2,40000 DY(6,1)=0,48990 BIGY(6,1)=3,0000 SMOLY(6,1)=2,0000
SY(6,2)=0,29000 DY(6,2)=0,10198 BIGY(6,2)=0,4000 SMOLY(6,2)=0,1500

Пояснения к таблице:

SY(I,K1) - среднее значение для I-го объекта K1 компоненты
DY(I,K1) - стандартное отклонение для I-го объекта K1 компоненты
BIGY(I,K1) - наибольшее значение для I-го объекта K1 компоненты
SMOLY(I,K1) - наименьшее значение для I-го объекта K1 компоненты
I = 1,M1 K1 = 1,K

Приложение 2

Исходные данные

Количество объектов M = 19
Количество признаков факторов N = 3
Количество признаков откликов K = 2
Параметр KBET = 3, DEL = 0,005, EPS = 0,17569

Количество ветвей KWT = 7
Конечное значение критерия ED = 0,39284
Код пропуска B = 99,9

Решающее правило

K5	NF(K5,3)	NF(K5,1)	NF(K5,2)	NJF(1,K5)	NJF(2,K5)	APF(1,K5)	APF(2,K5)	NF(K5,4)	NF(K5,5)
1	1	2	3	1	1	-99999,0	9,5	1	19
2	2	4	5	1	1	5,5	8,5	1	8
3	3	6	7	0	3	1,0	4,0	1	10
4	6	0	0	0	0	0,0	0,0	2	5
5	7	0	0	0	0	0,0	0,0	2	5
6	4	0	0	0	0	0,0	0,0	2	4
7	5	0	0	0	0	0,0	0,0	2	4

Пояснения к таблице:

K5 – номер;
NF(K5,3) – номер вершины исходящей;
NF(K5,1) – номер вершины, в которую надо идти, если высказывание – истина;
NF(K5,2) – если ложно;
NJF(1,K5) – индекс высказывания;
NJF(2,K5) – номер признака, по которому строится высказывание;
APF(J,K5) – J = 1,2 – пороги;
NF(K5,4) указывает конечность вершины;
NF(K5,5) указывает число объектов в вершине.