Krassimir Markov, Vladimir Ryazanov,
Vitalii Velychko, Levon Aslanyan
(editors)

New Trends in Classification and Data Mining

I T H E A
SOFIA
2010

Krassimir Markov, Vladimir Ryazanov, Vitalii Velychko, Levon Aslanyan (ed.)

New Trends in Classification and Data Mining

ITHEA®

Sofia, Bulgaria, 2010

First edition

Recommended for publication by The Scientific Concil of the Institute of Information Theories and Applications FOI ITHEA

This book maintains articles on actual problems of classification, data mining and forecasting as well as natural language processing:

- new approaches, models, algorithms and methods for classification, forecasting and clusterisation. Classification of non complete and noise data;
- discrete optimization in logic recognition algorithms construction, complexity, asymptotically optimal algorithms, mixed-integer problem of minimization of empirical risk, multi-objective linear integer programming problems;
- questions of complexity of some discrete optimization tasks and corresponding tasks of data analysis and pattern recognition;
- the algebraic approach for pattern recognition - problems of correct classification algorithms construction, logical correctors and resolvability of challenges of classification, construction of optimum algebraic correctors over sets of algorithms of computation of estimations, conditions of correct algorithms existence;
- regressions, restoring of dependences according to training sampling, parametrical approach for piecewise linear dependences restoration, and nonparametric regressions based on collective solution on set of tasks of recognition;
- multi-agent systems in knowledge discovery, collective evolutionary systems, advantages and disadvantages of synthetic data mining methods, intelligent search agent model realizing information extraction on ontological model of data mining methods;
- methods of search of logic regularities sets of classes and extraction of optimal subsets, construction of convex combination of associated predictors that minimizes mean error;
- algorithmic constructions in a model of recognizing the nearest neighbors in binary data sets, discrete isoperimetry problem solutions, logic-combinatorial scheme in high-throughput gene expression data;
- researches in area of neural network classifiers, and applications in finance field;
- text mining, automatic classification of scientific papers, information extraction from natural language texts, semantic text analysis, natural language processing.

It is represented that book articles will be interesting as experts in the field of classifying, data mining and forecasting, and to practical users from medicine, sociology, economy, chemistry, biology, and other areas.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Bulgaria

ISBN 978-954-16-0042-9

C\o Jusautor, Sofia, 2010

RESTORING OF DEPENDENCES ON SAMPLINGS OF PRECEDENTS WITH USAGE OF MODELS OF RECOGNITION

V.V.Ryazanov, Ju.I.Tkachev

Abstract. *Two approaches to solution of the task of restoring of dependence between a vector of independent variables and a dependent scalar according to training sampling are considered. The first (parametrical) approach is grounded on a hypothesis of existence of piecewise linear dependence, and private linear dependences correspond to some intervals of change of the dependent parameter. The second (nonparametric) approach consists in solution of main task as search of collective solution on set of tasks of recognition*

Keywords: *dependence restoring, regression, algorithm of recognition, piecewise linear function, feature, dynamic programming*

ACM Classification Keywords*: A.0 General Literature - Conference proceedings, G.1.2 Approximation: Nonlinear approximation, H.4.2 Types of Systems: Decision support, I.2 Artificial intelligence, I.5 Pattern recognition*

Introduction

The task of restoring of dependence between a vector of variable (features) $\mathbf{x} = (x_1, x_1, \ldots x_n)$, $x_i \in M_i, i = 1,2,\ldots,n$, where M_i are sets of any nature, and a scalar y on sampling $\{(y_i, \mathbf{x}_i)\}_{i=1}^m$ is considered. Assuming existence between them functional link $y = f(\mathbf{x})$, on sampling function from some parametrical class of functions or the algorithm is selected, allowing to calculate for a vector of variables $\mathbf{x}$ appropriate value of the dependent value y. The given task in statistical setting is known as the task of restoring of regression - functions of a conditional expectation. Now there are various parametrical and nonparametric approaches to restoring of regressions [1,2]. Parametrical approaches demand a priori knowledge of analytical sort of functions. Nonparametric approaches use as a rule methods of a frequency estimation and functions of distances. The given approaches have the essential limitations linked to such properties of the real data as heterogeneity of features, various informativity of features, co-ordination of metrics of various features, etc. At the same time, for a case of the discrete value $y \in \{1,2,\ldots,l\}$ (the standard task of recognition [3,4]) the given limitations are not critical. Enumerated above difficulty are successfully overcome, for example, in the logical models of recognition [3-8] which are not demanding solution of additional task of preprocessing of partially contradictory polytypic no presentable data.

The nonparametric method of restoring of dependence assumes "carrying over" of all marked above problems with features on recognition level. According to training sampling, N tasks of recognition are formed and, respectively, N algorithms of recognition are constructed. N recognition tasks are solved independently for any vector $\mathbf{x}$ of features, and value of the dependent value $y = f(\mathbf{x})$ is calculated as collective solution over recognition tasks.

The parametrical approach assumes that to some segments of a range of the dependent value there correspond the linear dependences from features. The task of restoring of piecewise linear dependence is reduced to solution of the task of dynamic programming variables in which correspond to points of splitting of a range of the dependent value on intervals, and addends of function to be optimized define quality of approximating of dependence in an appropriate interval by linear function.

Results of practical approbation are performed.

1. Restoring of piecewise linear dependences on samplings of precedents

We consider that training sampling $\{(y_i, \mathbf{x}_i)\}_{i=1}^{m}$, in the form of the training table T_{nm} where each string is a vector of values of features and to it appropriate value of the dependent value y is set, $T_{nm} = \left(\begin{array}{cccc|c} x_{11} & x_{12} & \cdots & x_{1n} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2n} & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & y_m \end{array}\right)$. In the present section we consider that $x_i \in R, i = 1,2,...,n$. Without loss of generality we can consider that all $y_i, i = 1,2,...,m$ are various ones and arranged on increase: $y_i < y_{i+1}, i = 1,2,...,m-1$. Let us divide a change interval $[y_1, y_m]$ of $y(\mathbf{x})$ on $l \geq 2$ intervals $\Delta_1 = [y_1, y_{i_1}), \Delta_2 = [y_{i_1}, y_{i_2}), \cdots, \Delta_l = [y_{i_{l-1}}, y_m]$. Any splitting is set by a vector of parameters $z = (z_0, z_1, z_2, ...z_l\ ), z_i \in \{y_1, y_2, ..., y_m\}, z_i < z_{i+1}, i = 0,2,...,l-1, z_0 \equiv y_1, z_l \equiv y_m$.

For any segment Δ_i by data $\{(y_j, \mathbf{x}_j), j = 1,2,...,m,\}$: $y_j \in \Delta_i$ there is by means of the least squares method a function $g_i(\mathbf{x}) = \sum_{t=1}^{n} a_t^i x_t + b^i$ in the best way approximating given subsample. Quality of the present function (and the segment Δ_i) is estimated as $f_i(z_{i-1}, z_i) = \frac{1}{|\Delta_i|} \sum_{j:y_j \in \Delta_i} (y_j - g_i(\mathbf{x}_j))^2$. Then the task of search of the best piecewise linear approximating with l components is formulated in the form of the following task of dynamic programming:

$$\Phi(z_1, z_2, ..., z_{l-1}) = f_1(y_1, z_1) + \sum_{i=2}^{l-1} f_i(z_{i-1}, z_i) + f_l(z_{l-1}, y_m) \qquad (1)$$

$$z_i \in \{y_2, y_3, ..., y_{m-1}\}, z_i < z_{i+1}, i = 1,2,...,l-2, \qquad (2)$$

Solving the dynamic programming task an optimal piecewise linear dependence is found. Thus the subsamples of $\{(y_j, \mathbf{x}_j), j = 1,2,...,m\}$: $y_j \in \Delta_i$, are set which sets of vectors $\mathbf{x}_j$ we consider as the description of some classes. Under descriptions of classes $K_i, i = 1,2,...,l$, some algorithm of recognition A is calculated which is applied to classification of any new objects $\mathbf{x}$.

Finally, the task of $y = f(\mathbf{x})$ calculation for any $\mathbf{x}$ is solved in a two stages: object $\mathbf{x}$ classification $A: \mathbf{x} \rightarrow K_i$ is carried out, further $y = f(\mathbf{x}) = g_i(\mathbf{x}) = \sum_{t=1}^{n} a_t^i x_t + b^i$ is calculated.

2. The nonparametric method of regression based on construction of collective solutions on set of tasks of recognition

In the present section we consider that $x_i \in M_i, i = 1,2,...,n$, M_i are sets of any nature. We will take number $v \leq m$ and $v+1$ points from a segment $R = [y_1, y_m]$: $d_0 = y_1$, $d_0 < d_1 < \cdots < d_{v-1} < d_v = y_m$. We receive set R splitting into v intervals $\Delta_1 = [d_0, d_1), \Delta_2 = [d_1, d_2), \cdots, \Delta_v = [d_{v-1}, d_v]$, $\mathbf{\Delta} = \{\Delta_1, ..., \Delta_v\}$. We will put $c_k = \dfrac{d_{k-1} + d_k}{2}$ - interval $\Delta_k, k = 1,2,...,v$ centre.

Let's take number $2 \leq l \leq v$ and we will define N segment R splittings into l intervals having put to each splitting in correspondence a vector $\mathbf{k}_i = (k_i^{(1)}, k_i^{(2)}, ..., k_i^{(l-1)}), i = 1,2,..., N, k_i^{(j)} < k_i^{(j+1)} < n$ with integer positive components. The vector data sets labels intervals in appropriate splitting. Intervals $\Delta_1, ..., \Delta_{k_i^{(1)}}$ flagged by a label «1», intervals $\Delta_{k_i^{(1)}+1}, ..., \Delta_{k_i^{(2)}}$ - a label «2», intervals $\Delta_{k_i^{(l-2)}+1}, ..., \Delta_{k_i^{(l-1)}}$ - a label «l-*1*», intervals $\Delta_{k_i^{(l-1)}+1}, ..., \Delta_n$ - a label «l». Each splitting of a segment R defines set $\mathbf{M} = M_1 \times \cdots \times M_n$ splitting into l subsets (classes) $K_1^i, ..., K_l^i$: $\mathbf{M} = \bigcup_{t=1}^{l} K_t^i, \nu \neq \mu \Rightarrow K_\nu^i \cap K_\mu^i = \varnothing$. Splitting $\mathbf{K}^i = \{K_1^i, ..., K_l^i\}$ is set by a rule: the object $\mathbf{x}$ belongs to a class K_j^i if and only if $y = f(\mathbf{x}) \in \Delta_k$ and the interval Δ_k is flagged by a label « j ». Each splitting $\mathbf{K}^i$ defines the standard task of recognition Z_i with l classes.

Let A_i - some algorithm of solution of the task Z_i, carrying any object $\mathbf{x}$ to one of classes $K_{a_i}^i, a_i \in \{1,2,...,l\}$.

Let's consider the direct product $\mathbf{K}^1 \times \cdots \times \mathbf{K}^l \times \mathbf{\Delta}$ as set of simple events. Event $(K_{a_1}^1, ..., K_{a_N}^N, \Delta_j)$ means reference of the object $\mathbf{x}$ by algorithm A_1 in a class $K_{a_1}^1$, …, by algorithm A_N in a class $K_{a_N}^N$, thus $y = f(\mathbf{x}) \in \Delta_j$. Probability of such event we will designate as $\mathrm{P}(a_1, ..., a_N, \Delta_j)$.

According to the formula of Bayes, we have

$$\mathrm{P}(\Delta_j | a_1, ..., a_N) = \frac{\mathrm{P}(a_1, ..., a_N, \Delta_j)}{\mathrm{P}(a_1, ..., a_N)} = \frac{\mathrm{P}(\Delta_j)}{\mathrm{P}(a_1, ..., a_N)} \mathrm{P}(a_1, ..., a_N | \Delta_j) \tag{3}$$

If algorithms are statistically independent, we have $\mathrm{P}(a_1, ..., a_N | \Delta_j) = \prod_{i=1}^{N} \mathrm{P}(K_{a_i}^i | \Delta_j)$, $\mathrm{P}(a_1, ..., a_N) = \prod_{i=1}^{N} \mathrm{P}(K_{a_i}^i)$ and the formula (3) becomes (4).

$$\mathrm{P}(\Delta_j | a_1, ..., a_N) = \frac{\mathrm{P}(\Delta_j)}{\prod_{i=1}^{N} \mathrm{P}(K_{a_i}^i)} \prod_{i=1}^{N} \mathrm{P}(K_{a_i}^i | \Delta_j) \tag{4}$$

Let's enter notations $p_k = \mathrm{P}(\Delta_k | a_1, ..., a_N), k = 1, ..., v$.

Function $F:(p_1,...,p_v)\to R$ where $p_1,...,p_v$ are received according to (3) is named as the Bayesian corrector. Function $F:(p_1,...,p_v)\to R$ where $p_1,...,p_v$ are received according to (4) is named as the naive Bayesian corrector. We will consider later the naive Bayesian corrector.

Note. We will notice that in cases when $y=f(\mathbf{x})\in\{1,2,...,l\}$, the value y means a class label at classification of the object $\mathbf{x}$ and the primary goal is the recognition task. Here, all splittings $\mathbf{K}^i$ are coincide ones. For the set of collection of recognition algorithms $A_i, i=1,2,...,N$, the model of the naive Bayesian corrector $F:(p_1,...,p_v)\to\{1,2,...,l\}$ in the recognition task is known (see, for example, [10]).

Definition 1. «As the answer on an average» of the Bayesian corrector for the object $\mathbf{x}$, we will name the value $\tilde{y}=\sum_{k=1}^{v} p_k c_k$.

Definition 2. «As the answer on a maximum» of the Bayesian corrector for the object $\mathbf{x}$, we will name the value $\tilde{y}=c_k$, where $k=\arg\max_j p_j$.

Let's describe the common algorithm of restoring of dependence $y=f(\mathbf{x})$ according to the training sampling, based on usage of Bayesian model, and algorithm of calculation of value of the dependent value y for any $\mathbf{x}$.

Algorithm of restoring of dependence $y=f(\mathbf{x})$.

1. Splitting of a segment $R=[y_1,y_m]$ into intervals $\Delta_k, k=1,2,...,v$.
2. The calculation of splittings $\mathbf{K}^i, i=1,...,N,$ setting of tasks of recognition $Z_i, i=1,...,N,$ a choice of algorithms of recognition $A_i, i=1,...,N,$ and their training.
3. Calculation of estimations of probabilities $\mathrm{P}(K_j^i|\Delta_k), \mathrm{P}(\Delta_k), \mathrm{P}(K_j^i), i=1,...,N, j=1,...,l, k=1,...,v.$

Algorithm of calculation of value of the dependent value.

1. Classification of the object $\mathbf{x}$ by algorithms $A_i, i=1,...,N.$
2. Calculation of values of probabilities $p_1,...,p_v$ according to (4).
3. Calculation of $\tilde{y}=\sum_{k=1}^{v} p_k c_k$ or $\tilde{y}=c_k$, $k=\arg\max_j p_j$.

Practical implementation of the model of restoring of dependence set stated above and its application demands a concrete definition of all resulted parameters of model and algorithms of recognition. Here can be used any algorithms of recognition on precedents, and for calculation of estimations of probabilities – approaches and methods of mathematical statistics. In the present paper, the use of special collection of logical algorithms of recognition (test algorithms, algorithms of voting by representative sets, algorithms of voting by systems of logical regularities) and heuristic estimation of probabilities is considered and proved. The common feature of the given algorithms is faultless recognition of objects of no contradictory training sampling.

3. Restoring of dependences on the basis of application of collections of logical algorithms of recognition

Let's put $l=2$. For simplicity we consider that all values y_i in training sampling are various and $y_i < y_{i+1}, i=1,...,m-1$. We will consider next two ways of construction of intervals Δ_k.

1. We will take $v=m, N=v-1$. We will put $d_0 = y_1, d_1 = \frac{y_1+y_2}{2}, \cdots, d_{m-1} = \frac{y_{m-1}+y_m}{2}, d_m = y_m$. For algorithm A_i intervals $\Delta_1,...,\Delta_i$ are marked by a label «1», the others - «2».
2. The minimum value $\varepsilon = y_{i+1} - y_i, i=1,...,m-1$, is founded. We will put $v = 2m-2, N = v-1 = 2m-3$, and $d_0 = y_1 - \frac{\varepsilon}{2}, d_1 = y_1 + \frac{\varepsilon}{2}, d_2 = y_2 - \frac{\varepsilon}{2}, d_3 = y_2 + \frac{\varepsilon}{2}, \cdots,$

$d_{2i} = y_{i+1} - \frac{\varepsilon}{2}, d_{2i+1} = y_{i+1} + \frac{\varepsilon}{2}, \cdots, d_{v-1} = y_m - \frac{\varepsilon}{2}, d_v = y_m + \frac{\varepsilon}{2}$. For algorithm of recognition $A_i, i=1,...,N$, we will mark intervals $\Delta_1,...,\Delta_i$ by label «1», the others – by «2».

As frequency estimation $\mathrm{P}(K_j^i), i=1,...,N, j=1,...,l,$ we will name a share of the objects belonging to a class K_j^i in the task Z_i. As frequency estimations $\mathrm{P}(K_j^i|\Delta_k), i=1,...,N, j=1,...,l, k=1,...,v$, we will name the ratio $\frac{m_{ij}^{(k)}}{m_{ij}}$, where $m_{ij} = \left|\{\mathbf{x}_t : \mathbf{x}_t \in K_j^i\}\right|$, $m_{ij}^{(k)} = \left|\{\mathbf{x}_t : \mathbf{x}_t \in K_j^i, y_t \in \Delta_k\}\right|$. As a frequency estimation $\mathrm{P}(\Delta_k), k=1,...,v,$ we will name the value $\frac{\left|\{\mathbf{x}_t, t=1,...,m : y_t \in \Delta_k\}\right|}{m}$.

Definition 3. The model of restoring of dependence is named correct if for training sampling $\{(y_i, \mathbf{x}_i)\}_{i=1}^m$ takes place $\tilde{y}_i = y_i, i=1,...,m$.

Definition 4. The Bayesian corrector over logical algorithms of recognition with frequency estimations of probabilities is named as model A_1 at usage of the second way of construction of intervals and «the answer on a maximum».

Theorem 1. The model A_1 at the no contradictory training information is correct.

At practical calculation of values $\mathrm{P}(K_j^i|\Delta_k)$ there are situations, when $p_k = \mathrm{P}(\Delta_k|a_1,...,a_N) = 0, k=1,...,v$. Let d is a natural number. We will consider $\mathrm{P}'(K_j^i|\Delta_k) = \sum_{t=\max(k-d,1)}^{\min(k+d,n)} w_{t-k} \mathrm{P}(K_j^i|\Delta_k)$, and $\tilde{\mathrm{P}}(K_j^i|\Delta_k) = \frac{\mathrm{P}'(K_j^i|\Delta_k)}{\sum_{t=1}^{v} \frac{\mathrm{P}(\Delta_t)}{\mathrm{P}(K_j^i)} \mathrm{P}'(K_j^i|\Delta_t)}$. Here d is named as width of a window of smoothing, and nonnegative values $w_{-d},...,w_d$ are smoothing scales. It is visible that, $\tilde{\mathrm{P}}(K_j^i|\Delta_k) \geq 0$, $\sum_{k=1}^{v} \mathrm{P}(\Delta_k)\tilde{\mathrm{P}}(K_j^i|\Delta_k) = \mathrm{P}(K_j^i)$, i.e. $\tilde{\mathrm{P}}(K_j^i|\Delta_k)$ are conditional probabilities formally.

Replacement process $\mathrm{P}(K_j^i|\Delta_k) \to \widetilde{\mathrm{P}}(K_j^i|\Delta_k)$ we will name as smoothing. In addition, we will set superimpose limitation of symmetry $w_{-t} = w_t, t = 1,...,d$, and character of decrease $w_0 > 2w_1 \geq 4w_2 \geq \cdots \geq 2^d w_d$. We will name модель A_1 as model A_2 at usage of procedure of smoothing with exponential function of scales. The Theorem 2 is correct.

Theorem 2. The model A_2 at the no contradictory training information is correct one

4. Experimental matching of models of restoring on the model data

It is spent three sorts of experiments for matching of the nonparametric method of restoring of dependences offered in the present article with linear regression and the nonlinear regression grounded on nuclear smoothing.

1. Matching on elementary functions. For the first experiment the elementary nonlinear one-dimensional dependences (a parabola and a sinusoid) are used. On fig. 1, 2 results of matching of the restored dependences with true on some segment on an example of one-dimensional dependences $y = x^2$ and $y = \sin(x)$ are resulted. In the task "parabola" training sampling was under construction by a rule $x_i = 2i, y_i = (x_i)^2, i = 1,...,50$, and control sampling – by a rule $x_i = i, y_i = (x_i)^2, i = 1,...,100$. In the task "sinusoid" training sampling was under construction by a rule $x_i = 2i, y_i = \sin(\frac{2\pi}{100}x_i), i = 1,...,50$, and control – by a rule $x_i = i, y_i = \sin(\frac{2\pi}{100}x_i), i = 1,...,100$. In figures 1,2 plots of dependences and results of their restoring are shown, in table 1 norms of vectors of errors of each of algorithms on the given tasks are presented.

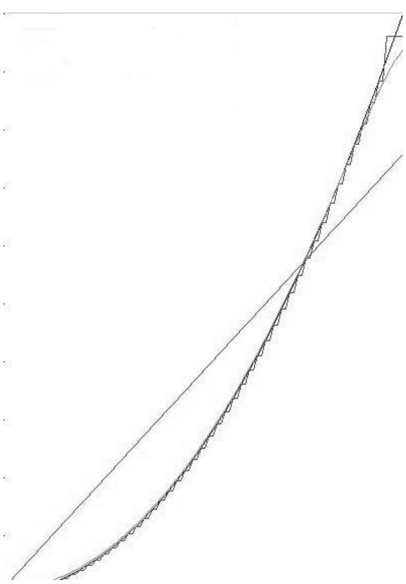

Figure 1. The model problem "parabola"

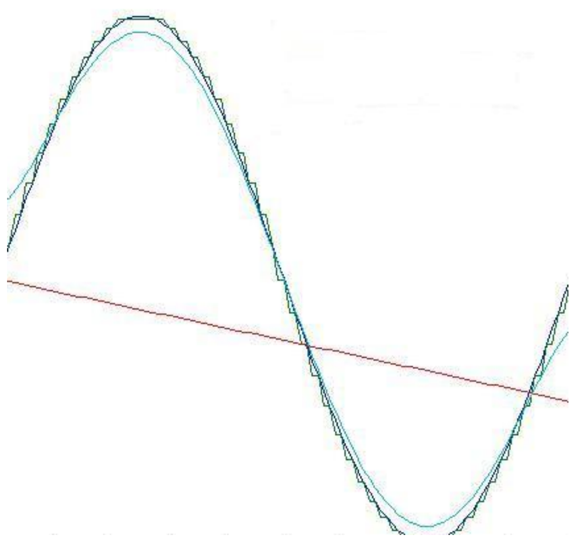

Figure 2. The model problem " sinusoid "

Table 1.

Task	Bayesian corrector	Linear regression	Nuclear smoothing
Parabola	0.31	6.52	0.64
Sinusoid	810	11321	1138

2. Matching on the data with noise. There were considered 10 various two-dimensional normal distributions. Means and dispersions of distributions were selected according to the uniform law of distribution. According to each of normal laws of distribution samplings of 12 points which have been united in one $\{(x_i^{(1)}, x_i^{(2)})\}_{i=1}^{120}, x_i^{(1)}, x_i^{(2)} \in R$ are received. The target variable y coincides to within the sign with density of appropriate normal distribution. For the first 5 normal samples value y was the density, for the others – density with the sign a minus. Variables $x^{(1)}, x^{(2)}$ have been added noise $x^{(3)}, \cdots, x^{(49)}$, subordinates to the uniform law of distribution. The sampling $\{(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(49)}, y)\}_{i=1}^{120}$ generated as a result has been divided casually into training and forecasting samplings of 90 and 30 points, accordingly.

3. Matching on tasks with discrete features. Sampling of 200 3D objects according to a following rule is generated: - $x_i^{(1)}, x_i^{(2)}$ implementations of a random variable $\xi = 0,...,10$ with probabilities $\frac{1}{11}$, $x_i^{(3)} = i, i = 1,...,200$. It was considered the dependence $y = \begin{cases} x^{(3)}, & x^{(1)} \geq x^{(2)}, \\ -x^{(3)} & x^{(1)} < x^{(2)}. \end{cases}$ The generated sampling was considered divided casually on two on 100 objects (training and forecasting).

In Table 2 norms of vectors of errors of the algorithms, averaged on 10 independent experiments are resulted.

Table 2.

The task	Bayesian corrector	Linear regression	Nuclear smoothing
Data with noise	3.625	4.489	4.150
Discrete features	552.1	850.7	606.2

Results of experiments show a high accuracy of the Bayesian corrector, its ability successfully to handle data with noise and the different type data at rather small training samplings.

5. Conclusion

In a method of restoring of piecewise linear dependences it was supposed that the number of linear components is set. Clearly, that the criterion $\Phi(z_1, z_2, ..., z_{l-1})$ monotonously does not increase with growth of number of components and, therefore, cannot be used for definition of l. The true number of components can be found by exhaustive search of a small number l with calculation of $\Phi(z_1, z_2, ..., z_{l-1})$ in a mode of the cross-validation.

The Bayesian corrector for calculation of values of the dependent value is not the unique way grounded on problem solving of recognition. Here other approaches are possible which are developed now by authors.

Aknowledgement

The study was supported by Russian Foundation for Basic Research (projects № 08-01-00636, 09-01-00409, 09-01-12060-ofi_м), and by target program № 14 of Russian Academy of Sciences

Bibliography

[1] Drejper H, Smith G. Applied regression analysis. M.:Pablishing house Williams, 2007.

[2] Hardle B. Applied nonparametric regression. M, The World, 1993.

[3] Zhuravlev Ju.I. Correct algebras over sets of not correct (heuristic) algorithms. I. Cybernetics. 1977. N4. pp. 5-17., II. Cybernetics, N6, 1977, III. Cybernetics. 1978. N2. pp. 35-43.

[4] Zhuravlev Ju.I. About the algebraic approach to solving of recognition or classification problems. Cybernetics problems. M: The Science, 1978. 33. pp.5-68.

[5] Dmitriev A.N., Zhuravlev Ju.I., Krendelev F.P., About mathematical principles of classification of subjects and the phenomena. The transactions "The Discrete analysis". Issue 7. Novosibirsk, IM SB AS USSR. 1966. pp. 3-11

[6] Baskakova L.V., Zhuravlev Ju.I. Model of recognising algorithms with representative sets and systems of basic sets//Zhurn. vichisl. matem. and matem. phys. 1981. Vol.21, № 5. pp.1264-1275

[7] Ryazanov V.V. Logical regularities in recognition tasks (the parametrical approach)// Zhurn. vichisl. matem. and matem. phys. 2007. Vol.47, № 10. pp.1793-1808

[8] Zhuravlev Ju.I., Ryazanov V.V., Senko O.V. Pattern recognition. Mathematical methods. The program system. Applications. - Moscow: Phasys publisher, 2006, 176 p.

[9] Sigal I.H., Ivanov A.P. Introduction in applied discrete programming: models and computing algorithms. The manual. M: PHYSMATLIT, 2002, 240 p.

[10] P.Domingos and M.Pazzani. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29:103-130, 1997.

Authors information

Vladimir Ryazanov *– Professor, Heard of Department; Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, Russia,* 119991 Moscow, Vavilov's street, 40; *e-mail: rvv@ccas.ru*

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence

Yury Tkachev *– The post-graduate student; Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, Russia, 119991 Moscow, Vavilov's street, 40; e-mail:* tkachevy@gmail.com

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence