BAYES CLASSIFIER IN MULTIDIMENSIONAL DATA CLASSIFICATION

E. Ocelíková[*], D. Klimešová[**]

** Technical University of Košice, Faculty of Electrical Engineering and Informatics*
Department. of Cybernetics and Artificial Intelligence, Letná 9/B, 041 20 Košice
phone: ++421 55 602 2570 e-mail : Eva.Ocelikova@tuke.sk

*** Czech University of Agriculture, Prague, Faculty of Economics and Management,*
Dept. of Information Engineering Kamycka 129, 165 21 Prague 6 – Suchdol
phone: ++420 2 2438 2272 e-mail: klimesova@pef.czu.cz

and

Czech Academy of Sciences, Institute of Information Theory and Automation
Department. of Image Processing, Pod vodárenskou vĕží 4, 182 08 Prague 8
phone: ++420 2 6605 2247 e-mail: klimes@utia.cas.cz...

Abstract: This paper deals with the classification of objects into the limited number of classes. Objects are characterised by n-features, e.g. n-dimensional vectors describe them. The paper focuses on the Bayes classifier based on the probability principle, with the fixed number of the features during classification process. Bayes classifier, which uses criterion of the minimum error was applied on the set of the multispectral data. They represent real images of the Earth surface obtained from remote Earth sensing. The paper describes experience and results obtained during the classification of extensive set of these multispectral data and analysis of influence of dispersions and mean values of features on classification results.

Keywords: classification**,** Bayes classifier, feature, multispectral data, decision rule.

1 INTRODUCTION

Classification process may be applied in different areas of research and practice, e.g. farms, military, medicine, remote Earth sensing, etc. The classical classification techniques use statistical approach, which typically assumes the normal multidimensional distribution of probability in the experimental data set. Data classification may be supervised and unsupervised [1] [2].

The supervised classification method requires the presence of training data set typically defined by the expert - the teacher. Each class of objects is characterised by the basic statistical parameters (mean values vector, covariance matrix), which are computed from the training set. These parameters guide the discrimination process. The Bayesian classifiers are typical representatives (Bayes classifier, Fisher, Wald sequential) [4].

The unsupervised classification is also known as classification without the teacher. This classification uses in most cases the methods of cluster analysis.

The device, that performs the function of classification, is called ***classifier***. The classifier is system containing several inputs that are transported with signals carrying information about the objects. The system generates information about the competence of objects into particular class on the output. Particular inputs represent features that describe the object.

According to the number of features it is possible to divide classifiers into the following classes:

1. The classifiers with ***variable number of features***, so-called sequential classifiers. This one starts process of classification with valuation of the one feature and if the process is decidable, classification is ended. In inverse case the next feature is added and the process repeats while the process of classification is undecided.
2. The classifiers with ***fixed number of features***. These are characteristic with the exact number of features. This number is fixed during the whole classification process.

According to the other aspects it is possible to divide classification systems into simple and complex, parametric and nonparametric, etc.

This paper mentions problems (as similarity and large dispersion of particular classes), which are possible to arise at the classification by Bayes classifier.

In the following we will look at Bayes classifier.

2 BAYES CLASSIFIER

As mentioned Bayes classifier is the classifier with fixed number of the features. The classified patterns $\boldsymbol{x}$ are described by n-features $\boldsymbol{x} = [x_1, x_2, ..., x_n]^T$ and by pattern space which contains R disjunctive subsets Q_r, $r = 1,...,R$. Bayes classifier uses criterion of the minimum error

$$J = \sum_{r=1}^{R} \int_{Q_r} \sum_{s=1}^{R} \lambda(d_B(\boldsymbol{x}) | \omega_s) p(\boldsymbol{x} | \omega_s) P(\omega_s) d\boldsymbol{x} \qquad (1),$$

where $\lambda(d_B(\boldsymbol{x}) | \omega_s)$ is the loss function and $d_B(\boldsymbol{x})$ is optimal Bayes decision rule which minimises average loss (1). The valuation of criterion in this form is difficult therefore it is transformed to more easily form

$$L_{\boldsymbol{x}}(\omega_r) = \sum_{s=1}^{R} \lambda(\omega_r | \omega_s) p(\boldsymbol{x} | \omega_s) P(\omega_s). \qquad (2)$$

The equation (2) represents the loss that will be after the classification of pattern $\boldsymbol{x}$ into the class ω_r, if the pattern does not belong to this class. The optimal decision rule is $L_{\boldsymbol{x}}(d_B(\boldsymbol{x})) = \min L_{\boldsymbol{x}}(\omega_r)$. It is possible to realize on the principle discriminate functions [7]. The form of discriminate function depends on the choice of the loss functions.

If the ***symmetric loss function*** is selected within the form

$$\lambda(\omega_r | \omega_s) = 0 \quad \text{if} \quad r = s$$
$$\lambda(\omega_r | \omega_s) = 1 \quad \text{if} \quad r \neq s \qquad (3)$$

discriminate function is $g_r(\boldsymbol{x}) = p(\boldsymbol{x} | \omega_s) P(\omega_s)$. This loss function states that we lose one unit for only misrecognition and nothing is lost for correct classification.

If the ***diagonal loss function*** is selected within the form

$$\lambda(\omega_r | \omega_s) = -h_i \quad \text{if} \quad r = s$$
$$\lambda(\omega_r | \omega_s) = 0 \quad \text{if} \quad r \neq s \qquad (4)$$

where $h_i > 0$, then discriminate function is $g_r(\boldsymbol{x}) = -h_s p(\boldsymbol{x} | \omega_s) P(\omega_s)$. In this case the loss function states negative loss to a correct decision and no loss to an incorrect decision. The choice of constants h_i is intuitive.

If the patterns $\boldsymbol{x}$ from miscellaneous classes are like statistical set, they can be approximated by normal distribution. The conditional probability density function has the form

$$p(\boldsymbol{x} | \omega) = \frac{1}{2\pi^{R/2}\sqrt{|W|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \mu)^T W^{-1} (\boldsymbol{x} - \mu)\right) \qquad (5)$$

where $|W|$ is determinant of covariance matrix. Covariance matrix W is

$$W = \frac{1}{K} \sum_{i=1}^{K} ((\boldsymbol{x}^i - \boldsymbol{\mu})^T (\boldsymbol{x}^i - \boldsymbol{\mu}), \qquad (6)$$

where K is number of patterns in the class and $\boldsymbol{\mu}$ is the mean vector

$$\boldsymbol{\mu} = \frac{1}{K} \sum_{i=1}^{K} \boldsymbol{x}^i. \qquad (7)$$

If $g_s(\boldsymbol{x}) > g_r(\boldsymbol{x})$ is true then pattern $\boldsymbol{x}$ is nearest to the class ω_s. If there are noises in the input set then classical Bayes classifier assigns this pattern (noise) to the nearest cluster. Augmented Bayes classifier allows solving this problem. This augmented classifier tests Mahalanobis distance D_M by chi-quadrate distribution $\chi^2_\alpha(n)$ with number of features n and with significance level α.

Mahalanobis distance becomes for class ω_r *,where* r=1, 2, …, R

$$D^2_{Mr} = (\boldsymbol{x} - \boldsymbol{\mu}_r)^T W^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_r). \qquad (8)$$

On the base of test the result of classification is:

1. If $D_{Mr}^2 \le \chi_\alpha^2(n)$ then $\boldsymbol{x}$ belongs to class ω_r.
2. If $D_{Mr}^2 > \chi_\alpha^2(n)$ then $\boldsymbol{x}$ belongs to class T_0, where T_0 is *"other"* class.

3 CLASSIFICATION OF IMAGE OF EARTH SURFACE

The testing of augmented Bayes classifier was applied on training set of image data obtained from remote Earth sensing. Data source contained 800 patterns. The patterns had multispectral character and seven dimensional vectors represented them. The patterns were divided into 8 classes:

1. stubble 2. grass 3. asphalt

4.concrete 5. roof 6. water

7. shadow 8. tree.

Contingent tables Table 1 and Table 2 show the evaluation of obtained classification results.

Symmetric loss function was used for realization of the tests. The tests were made with the significance level $\alpha = 0{,}05$ and $\alpha = 0{,}01$. Contingent Table 1. contains the results of the classification with significance level $\alpha = 0{,}05$. Less satisfactory results were obtained in 4th and 5th classes. Similarity of mean values and large dispersions of the patterns involve this, (see Figure 1 and Figure 2). The best results were obtained in 2nd and 6th classes with 97% successful classification.

The results of classification were similar when significance level $\alpha = 0{,}01$, (see contingent Table 2). Correct classification increased by the choice of significance level $\alpha = 0{,}01$ about 2% in 5th class, about 3% in 3rd and 4th classes. The best results were obtained in 6th class with 100% successful classification. In this class a little dispersion of the patterns was observed.

Explanation of symbols (Table 1 and Table 2)

9 - the class “other“

N_1 - the number of patterns of the original classes

N_1% - the number of correctly classified patterns in percentage

E_1% - error of the first kind in percentage; it means error which arises after the assignment of patterns from certain original class to another classes after classification

N_2 - the number of patterns classified into certain class after classification

E_2% - error of the second kind in percentage; it means error which arises after the assignment of patterns from all original classes into one certain class after classification.

Table 1. Contingent table - results of Bayes classification with $\alpha = 0{,}05$

Original classes	Classes after the classification									N_1	N_1%	E_1%
	1	2	3	4	5	6	7	8	9			
1	95	5	0	0	0	0	0	0	0	100	95	5
2	0	97	2	0	0	0	0	0	1	99	97	3
3	0	0	95	0	2	0	0	0	3	97	97	3
4	0	0	5	85	6	0	0	0	4	96	88	12
5	0	1	0	13	80	0	3	0	3	97	82	18
6	0	0	0	0	0	97	0	0	3	97	100	0
7	0	0	0	0	3	0	90	5	2	98	91	9
8	0	0	0	0	0	0	8	92	0	100	92	8
N_2	95	103	102	98	91	97	101	97	16	784		
E_2%	0	6	7	14	13	0	11	6				

Table 2. Contingent table- results of Bayes classification with $\alpha = 0{,}01$

Original classes	Classes after the classification									N_1	$N_1\%$	$E_1\%$
	1	2	3	4	5	6	7	8	9			
1	95	5	0	0	0	0	0	0	0	100	95	5
2	0	98	2	0	0	0	0	0	0	100	98	2
3	0	0	98	0	2	0	0	0	0	100	98	2
4	0	0	5	88	6	0	0	0	1	99	88	12
5	0	1	0	13	82	0	3	0	1	99	82	18
6	0	0	0	0	0	100	0	0	0	100	100	0
7	0	0	0	0	3	0	91	6	0	100	91	9
8	0	0	0	0	0	0	8	92	0	100	92	8
N_2	95	104	105	101	93	100	102	98	2	798		
$E_2\%$	0	6	7	13	12	0	11	7				

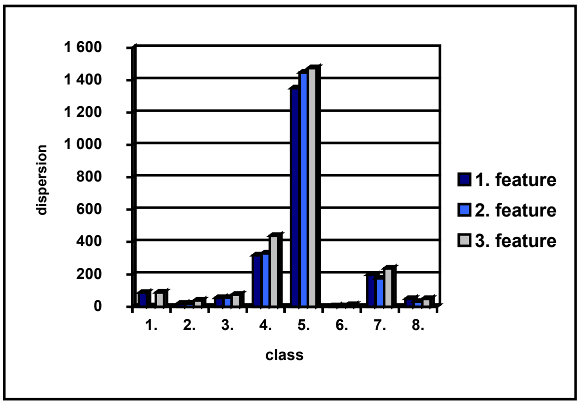

Fig. 1. Dispersions of features in the classes

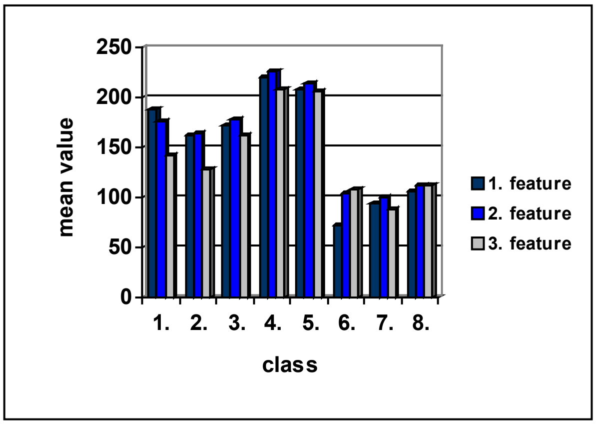

Fig. 2. Mean values of features in the classes

CONCLUSION

Bayes classifier absolutely successfully recognized the patterns from the class in which dispersion of the patterns was small (6th class). The data from all classes were classified overall successfully except the 4th and 5th classes. These classes were classified less successfully because they had large dispersions and very similar mean values of features. The advantage of Bayes classifier its classification rate because it always calculates R discriminate equations only in the decision process (R is the number of classes). For successful usage of Bayes classifier, a priori information about number of patterns, multiplicity of patterns in the various classes, mean values of features are necessary and the condition of semi definition covariance matrix must be valid. Bayes classifier in the case did not classify the data when some of the mentioned conditions were absent.

ACKNOWLEDGMENT

The work has been supported by the grant agency VEGA project No. 1/2185/05 and project SR-CzR No. 008. This support is very gratefully acknowledgement.

6 REFERENCES

Devijaver, P. E., Kittler, J. (1982). Pattern recognition: a statistical approach. Prentice-Hall, Inc., London.

Fukunaga, K. (1972). Introduction to statistical pattern recognition. Academic Press, New York.

Kharin Yu. S. (1997). Statistical techniques: Optimality and robustness, 1st International Workshop on STRIP '97, Prague, June, pp. 135-140.

Landryová, L., Gondek, H. and Fecko, P. (2004). Research for Mining Optimization and Biogas Utilization with Minimizing its Escape into the Atmosphere In: 21st Annual International Coal - Energy and the Environment Pittsburgh Coal Conference. Osaka, Japan, International Convention Centre, pp. 6, poster 12/2.

Leondes, C. T. (1998). Image Processing and Pattern Recognition.. Academic Press, New York.

Ocelíková, E., Madarász, L. (2002). Contribution to Creation of Macrosituations of Complex System. *Journal of Advanced Computional Intelligence,* Vol. 6, No.2, pp. 79-83.

Ocelíková, E., Krištof, J. (2000). Statistical Classification Techniques In: Proc. of the 34th Spring International Conference "Modelling and Simulation of Systems MOSIS", May 2-4, Rožnov pod Radhoštem, pp. 111-116.

Ocelíková, E., Klimešová, D. (2004). Contribution to the Classification of Multidimensional Data. In: Proceedings of 2nd Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence, SAMI 2004, Herľany, Slovakia, January 16-17, pp. 71-76.

Ocelíková, E., Zolotová, I., Landryová, L. (2003). Risk Classes and Cluster Analysis. *Sborník vìdeckých prací VŠB-TU Ostrava,* r. XLIX, è. 1, èást 2, øada strojní, s. 17-20.

Ocelíková, E., Zolotová, I., Klimešová, D.. (2002): Multikriteriálna klasifikácia zdravotného stavu.. *Lékař a technika,* ročník 33, No. 6, pp.187-191.

Raudys, S. (1997). Intrinsic Dimensionality and Small Sample Properties of Statistical Classifier. 1st International Workshop on STRIP '97, Prague, June, pp. 85-90.

Wilson, d. R., Martinez, T. R. (1997). Improved Heterogeneous Distance Functions. *Journal of Artificial Intelligence Research,* No. 6, pp. 1-34.