Classification of Classifiers

G. Gluhchev[1], M. Savov[1] and O. Boumbarov[2]

Abstract – **In this paper the relationship between three broadly used in practice classifiers – Mahalanobis distance, K nearest neighbors and majority voting, and the optimal classifier in terms of minimum average losses is outlined. Their performance efficiency is experimentally tested on the real problem of signature recognition.**

Keywords **– Mahalanobis distance, K nearest neighbors, Majority voting.**

I. INTRODUCTION

In this paper three of the most popular classifiers are discussed, namely, the Mahalanobis distance based classifier, the K nearest neighbors one and the majority voting. Using the theoretical set-up of the optimal classifier in terms of minimum average losses we make an attempt to show how different classifiers refer to the optimal (Bayesian) one.

The statistical pattern recognition theory assumes that some a priori information is available, including prior probabilities $P(\Omega_1)$ and $P(\Omega_2)$ of the classes, feature density functions $f_1(x)$ and $f_2(x)$, and losses incurred by wrong classification c_{12} and c_{21} respectively. The optimal classifier minimizing the average losses is defined as [1]

$$\begin{aligned} &x \in \Omega_1 \text{ if } P(\Omega_1 \mid x) \geq P(\Omega_2 \mid x) c_{12} / c_{21}, \\ &x \in \Omega_2 \text{ , otherwise} \end{aligned} \tag{1}$$

or

$$\begin{aligned} &x \in \Omega_1 \text{ if } f_1(x) \geq f_2(x) P(\Omega_2) c_{12} / P(\Omega_1) c_{21}, \\ &x \in \Omega_2 \text{ , otherwise} \end{aligned} \tag{2}$$

The condition (1) includes a posteriori probabilities of the classes, while the condition two is based on the maximum likelihood ratio. If $P(\Omega_1 = P(\Omega_2)$ and $c_{12} = c_{21}$ the decision is made according to the maximal a posteriori probability or likelihood ratio. Since these are constant we will assume they are equal.

[1]Georgi Gluhchev and Mladen Savov are with the Institute of Information Technologies of the Bulgarian Academy of Sciences, 2 Acad. G. Bonchev Str., Sofia 1113, Bulgaria, E-mail: {gluhchev@iinf.bas.bg, ihaiha99@abv.bg}

[2]Ognyan Boumbarov is with the Technical University in Sofia, 8 Kl. Ohridski Str., Sofia 1000, Bulgaria

II. MAHALANOBIS DISTANCE

In case of normal distributions the inequality (2) will look as follows

$$e^{-[(x-m_1)^t S_1^{-1}(x-m_1)-(x-m_2)^t S_2^{-1}(x-m_2)]} \geq 1 \tag{3}$$

or after taking a logarithm

$$(x-m_1)^t S_1^{-1}(x-m_1) \geq (x-m_2)^t S_2^{-1}(x-m_2) \tag{4}$$

Eq. (4) is actually a comparison of Mahalanobis distance of x to the centers m_1 and m_2 of the classes, and S_1 and S_2 are their covariance matrices.

III. K - NEAREST NEIGHBORS

When no justified assumptions could be made about the priors and class-conditional distributions, non-parametric classifiers are used. One of the most popular among them is the K- nearest neighbor, where a point x is attached to the class Ω_i, provided the ratio κ_i/K of its κ_i representatives among the K nearest neighbors to K is maximal [1]. It is worth to note that without going by the statistical estimations this empirical classifier evaluates the average a posteriori probability $P(\Omega_i \mid x)$ in a neighborhood of x. Thus, one could conclude that the K – nearest neighbor classifier is an empirical approach to the optimal Bayesian classifier. However, one has to pay attention that this classifier assumes implicitly that the quantity of the training samples corresponds to the prior probabilities of the classes. If this is not the case, the classification error may be too high.

IV. PARZEN WINDOWS

The Parzen window is used for the evaluation of the feature density function in a neighborhood of a point [1]. Therefore, according to inequality (2) the classifier based on Parzen windows could be optimal, as well.

The accuracy of the evaluation depends on the quantity of samples, on the one hand, and on the volume of the neighborhood, on the other hand. This approach resembles to a large extends the K-nearest neighbors one. The difference is that instead of the number K here the volume of the neighborhood is predefined. The advantage consists in its independence from the prior probabilities of the classes, i.e., different size of the training sequences for different classes will not affect the evaluation. Using Parzen windows for the classification actually means that equal prior probabilities are assumed.

V. Majority voting

Often in practice a decision is made depending on the number of votes. Such an approach is applicable to the classification problem provided all the features are treated as independent voters of equal importance in the following way.

The interval $[xmin_{ij}, xmax_{ij}]$ is determined for the i^{th} feature and the j^{th} class. All of the feature values of an unknown object x are tested for belonging to the corresponding interval of the classes. If the test result is positive for a particular class, its score is increased by 1. The winner is determined by the maximal score.

This classifier could be treated as a relative to the above mentioned one, provided a Parzen window of size equal to the interval of the corresponding feature is determined for each class. A maximal number of votes could be assigned to more than one class when this approach is used. For some specific problems like signature verification additional samples from the same class may solve the problem.

VI. Experimental comparison

To test how the above reasoning is supported by the practice, the classifiers have been applied to the real problem of signature authentication. For this signings of 14 volunteers has been captured by a TV camera. Every volunteer submitted 10 signatures that have been used for training. The following 8 features have been measured from each signing: 1) d – signing length as a number of frames, 2) α – hand orientation, 3) β – pen azimuth, 4) γ – pen tilt, 5) $\delta = \alpha - \beta$, 6) r_1 / r_2 - ratio of the distances between the pen center and hand contour, 7) P –perimeter of the polygon defined by the characteristic points of the upper hand contour, 8) A – area of the polygon [3].

The classifiers authentication performance has been evaluated in terms of mean, minimal and maximal error. To do this, 1000 signatures of every volunteer have been simulated using the Matlab's random number generator and the assumption of statistically independent features [2]. The classification results are shown in Table 1.

For the Mahalanobis distance an average classification error of 0,2% was obtained (Table 1, line 2, column 2). An absolute result of 0% errors was obtained for 6 volunteers, while the maximal error of 1% was obtained by one of them.

For the K-nearest neighbors classifier an average error of about 1.02% was obtained when one neighbor was used. For three or five neighbors the average error was slightly higher (Table 1, lines 3 и 4).

For the majority vote about 6% of wrong classifications and a maximal error of 19.6% have been observed (Table 1, line 5).

Table 1. Results from the experimental comparison of the classifiers

Classifier	*Average error %*	*Minimal error %*	*Maximal error %*
Mahalanobis	0.2	0	1
1 neighbor	1.02	0	6.2
3 neighbors	1.05	0	5.7
Majority vote	6.01	0.1	19.6

A separate investigation with Parzen windows has not been carried out due to the small number of the training data.

VII. Conclusion

In this paper three of the most popular classifiers have been analyzed and compared. The relationship between them was outlined, stemming from the assumptions about the available a priori information. It was shown that the Mahalanobis based classifier was quasi optimal in the sense of minimal losses and normal distribution of features. Empirical estimations of the a posteriori probabilities of the classes are obtained when K-nearest neighbor classifier is applied, provided the volume of the training sequences is proportional to the prior probabilities. Similar behavior could be expected if the class density functions are evaluated using Parzen windows.

The majority vote could be thought as a degenerated variance of the above classifiers. The experimental comparison carried out with real data has confirmed the theoretical analysis.

These results could be taken into account when practical classification problems have to be solved.

Acknowledgements

This investigation was supported partly by the Ministry of Education and Sciences in Bulgaria, Contract No BY – TH-202/2006 and BioSecure Network of Excellence, Contract No 507634/2004.

References

[1] R. O. Duda, P. E. Hart, "Pattern Classification and Scene Analysis", Jon Wiley & Sons, N. Y., London, Sydney Toronto, 1999

[2] M. Savov, G. Gluhchev, O. Bumbarov, "Experiments on Signature Identification with "Hand-pen" System Features", Proc. of the Int. Coference "Automatics and Informatics", Sofia, 2006, pp. 139-141

[3] M. Savov, G. Gluhchev. "Signature verification via Hand-Pen motion investigation", Proc. Int. Conf. "Recent Advances in Soft Computing", Canterbury, 2006, pp. 490-495