INFORMATICA, 2000, Vol. 11, No. 4, 381–396

Generalization Error of Randomized Linear Zero Empirical Error Classifier: Simple Asymptotics for Centered Data Case

Valdas DIČIŪNAS*
Department of Computer Science, Vilnius University
Naugarduko 24, 2600 Vilnius, Lithuania
e-mail: valdas.diciunas@maf.vu.lt

Šarūnas RAUDYS
Institute of Mathematics and Informatics
Akademijos 4, 2600 Vilnius, Lithuania
e-mail: raudys@das.mii.lt

Received: December 2000

Abstract. An estimation of the generalization performance of classifier is one of most important problems in pattern clasification and neural network training theory. In this paper we estimate the generalization error (mean expected probability of classification) for randomized linear zero empirical error (RLZEE) classifier which was considered by Raudys, Dičiūnas and Basalykas. Instead of "non-explicit" asymptotics of a generalization error of RLZEE classifier for centered multivariate spherically Gaussian classes proposed by Basalykas *et al.* (1996) we obtain an "explicit" and more simple asymptotics. We also present the numerical simulations illustrating our theoretical results and comparing them with each other and previously obtained results.

Key words: randomized linear classifier, generalization error, Gaussian classes, probability of misclassification.

1. Introduction

An estimation of the generalization performance of classifier is one of most important problems in pattern clasification and neural network training theory.

A significant part of theoretical results belongs to analysis of a standard linear discriminant function in a case when true pattern classes are Gaussian. Sitgreaves (1961) derived the first exact but very complicated for calculations formula for the expected classification error of the standard Fisher linear discriminant function (DF). John (1961) represented the linear discriminant function with known covariance matrix as a difference of two independent chi-square variables and expressed the expected error in a form

*Part of this work was done while being at the Institute of Mathematics and Informatics, Vilnius, Lithuania.

of infinite sum. Raudys (1967) used this result and derived the first simple asymptotic formula for the expected probability of misclassification (PMC) of the Euclidean distance classifier when the sample size $N \to \infty$. Best known asymptotic expansion for the expected classification error of the Fisher linear DF for the case when covariance matrix is known belongs to Okamoto (1963). Later two simple formulae for the expected error of the standard Fisher linear DF were obtained by Deev (1970, 1972) and Raudys (1972). An experimental comparison of these results was made by Wyman *et al.* (1990).

A large group of linear classification algorithms does not rely on the assumption about normality of the data. For a linear rule which minimizes an empirical error a number of error bounds and approximate asymptotic formulae have been derived (see, e.g., Amari, 1993; Amari *et al.*, 1992; Haussler *et al.*, 1994; Meir, 1995; Vapnik, 1982). In this type of investigations, no assumptions about the distribution of pattern classes is made. Therefore, the obtained estimates are either very pessimistic and unrealistic upper error bounds or depend on some unknown coefficients.

In this paper we consider a randomized linear zero empirical error (RLZEE) classifier and seek for a simple analytical expression for the generalization error when the pattern classes are sphericallly Gaussian and a Mahalanob distance between pattern classes is known. The RLZEE classifier shortly can be defined as follows. A training set

$$L = \{\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \ldots, \mathbf{X}_N^{(1)},\ \mathbf{X}_1^{(2)}, \mathbf{X}_2^{(2)}, \ldots, \mathbf{X}_N^{(2)}\}$$

of p-dimensional ($p \geqslant 2$) observation vectors from two different classes π_1 and π_2 is given. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be centers (means) of classes π_1 and π_2, respectively. We call the data (i.e., classes π_1 and π_2) *centered* if $\mathbf{C}_1 + \mathbf{C}_2 = \mathbf{0}$ and *non-centered* otherwise. According to some a priori distribution we randomly generate coefficients $a_0, a_1, \ldots, a_p \in \mathbb{R}$ and verify if a hyperplane

$$a_1x_1 + \cdots + a_px_p + a_0 = 0$$

discriminates given vectors without errors, i.e., if it satisfies the following (ZEE) condition:

$$\begin{cases} \mathbf{A}^{\mathrm{T}}\mathbf{X}_j^1 + a_0 > 0 & \forall \mathbf{X}_j^1 \in L \cap \pi_1, \\ \mathbf{A}^{\mathrm{T}}\mathbf{X}_j^2 + a_0 \leqslant 0 & \forall \mathbf{X}_j^2 \in L \cap \pi_2, \end{cases} \tag{ZEE}$$

where $\mathbf{A}^{\mathrm{T}} = (a_1, \ldots, a_p)$. Suppose that we successfully found such a hyperplane $(a_0, \mathbf{A})$ (in opposite case we generate new coefficients $a_0, a_1, \ldots, a_p$ and so on). The question is how well this hyperplane discriminates *unknown* data from classes π_1 and π_2, i.e., we are interested in an expected *probability of misclassification* (PMC) of this classifier. In neural networks literature the expected PMC usually is called *generalization error* because it shows how well the given classifier generalizes. In this paper we will use both these terms in arbitrary way.

The RLZEE classifier since 1993 has been intensively studied by Raudys *et al.* (see, for example, Basalykas *et al.*, 1996; Raudys, 1993; Raudys, 1997; Raudys and Dičiūnas, 1996). The main results of these studies were:

(1) The following exact formula for an expected PMC (which we also will call a *generalization error*) in a case of centered p-variate ($p \geqslant 2$) spherically Gaussian classes π_1, π_2 (Basalykas, 1996; Dičiūnas, 2001; Raudys, 1993) was obtained:

$$\mathrm{MEP}_N = \frac{\int_{-\infty}^{\infty}\int_{-1}^{1} \mathrm{P}(\mathrm{MC} \mid u, w)\mathrm{P}(\mathrm{ZEE} \mid u, w) f(u, w)\,\mathrm{d}u\,\mathrm{d}w}{\int_{-\infty}^{\infty}\int_{-1}^{1} \mathrm{P}(\mathrm{ZEE} \mid u, w) f(u, w)\,\mathrm{d}u\,\mathrm{d}w}, \tag{1}$$

where the density of random variables u and w is

$$f(u, w) = \frac{p-1}{2\pi}\frac{(1-u^2)^{(p-3)/2}}{(1+w^2)^{(p+1)/2}}, \quad u \in (-1, 1),\ w \in \mathbb{R}, \tag{2}$$

$$\mathrm{P}(\mathrm{MC} \mid u, w) = \frac{1}{2}\left[\Phi\left(-\frac{u\delta}{2} + w\right) + \Phi\left(-\frac{u\delta}{2} - w\right)\right], \tag{3}$$

$$\mathrm{P}(\mathrm{ZEE} \mid u, w) = \left[\Phi\left(\frac{u\delta}{2} + w\right)\right]^N \left[\Phi\left(\frac{u\delta}{2} - w\right)\right]^N. \tag{4}$$

In (3), (4) Φ is the normal distribution function

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\,\mathrm{d}t, \quad \text{where } \phi(t) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-t^2/2}, \tag{5}$$

N is a half of the length of the training set, and δ is the distance between the classes π_1 and π_2.

(2) The following asymptotics of expected PMC for the same classes as in (1) was derived (Basalykas, 1996):

$$\mathrm{MEP}_N = \Phi\left(-\frac{\delta u_0}{2}\right) + \frac{\delta}{8N}\phi\left(-\frac{\delta u_0}{2}\right)(A_1 + A_2 + A_3 + A_4) + \mathrm{O}\left(\frac{1}{N^2}\right),$$

where u_0 is a root of some integral equation (see (14)) and A_i are terms depending in some complicated form on p, N, δ and u_0 (see Section 2).

(3) The following simple "thumb" asymptotics of expected PMC was constructed heuristically on the ground of the table of the values of exact formula (Raudys, 1997):

$$\mathrm{MEP}_N \approx \Phi\left(-\frac{\delta}{2}\frac{1}{\sqrt{1 + (1.6 + 0.18\delta)(p/N)^{1.8-\delta/5}}}\right). \tag{6}$$

Above mentioned results, however, contain some negative features: (a) the exact formula is the ratio of two double integrals of some "non-regular" functions, and therefore requires long and careful calculations, and (b) the asymptotical formula is "non-explicit" (it contains a root of integral equation which has to be solved in numerical way).

In this paper we obtain an "explicit" asymptotics in a case of *centered* data. In some cases (for example, when $p = \text{const}$, $\delta \geqslant 1$ and N is big enough) this formula becomes very simple; it has not above mentioned negative features (a), (b) and can be used instead of above mentioned results (1)–(3). We also present the numerical simulations illustrating our theoretical results and comparing them with each other and previously obtained results.

2. An Asymptotical Expansion of Generalization Error

As it was already shortly mentioned in Section 1, approximately integrating (1) Basalykas *et al.* (1996) obtained the following analytical expression for generalization error of RLZEE classifier:

$$\mathrm{MEP}_N = \Phi\left(-\frac{\delta u_0}{2}\right) + \frac{\delta}{8N}\phi\left(-\frac{\delta u_0}{2}\right)(A_1 + A_2 + A_3 + A_4) + \mathrm{O}\left(\frac{1}{N^2}\right), \quad (7)$$

where

$$A_1 = \frac{u_0}{\beta_1(u_0)}, \quad (8)$$

$$A_2 = -\frac{\delta^2 u_0}{4Z''(u_0)}, \quad (9)$$

$$A_3 = -\frac{Z'''(u_0)}{(Z''(u_0))^2}, \quad (10)$$

$$A_4 = -\frac{{\beta_1}'(u_0)}{Z''(u_0)\beta_1(u_0)}, \quad (11)$$

for

$$\beta_1(u) = \lambda + \frac{2}{N} + \frac{\delta u}{2}\psi + \psi^2, \quad \psi = \psi(u) = \frac{\phi(\delta u/2)}{\Phi(\delta u/2)}, \quad (12)$$

$$Z(u) = \ln\Phi\left(\frac{\delta u}{2}\right) + \frac{\lambda}{2}\ln\left(1-u^2\right), \quad \lambda = \frac{p-3}{2N}, \quad (13)$$

u_0 $(0 < u_0 < 1)$ is a solution of equation

$$\delta\phi\left(\frac{\delta u}{2}\right)\left(1-u^2\right) = 2\lambda u\Phi\left(\frac{\delta u}{2}\right). \quad (14)$$

and functions Φ and ϕ represent normal distribution and density, respectively (see (5)).

The aim of this section is to obtain an asymptotical solution of equation (14), which allows to derive a simple asymptotical expansion for generalization error. Complete derivation consists of the following steps:

1. Derivation of asymptotical expression of u_0 for $\lambda = (p-3)/(2N) \to 0$.
2. Asymptotical expression of auxiliary terms $\beta_1(u_0)$, $\beta_1'(u_0)$ and $1/Z''(u_0)$.
3. Inserting of auxiliary terms into expressions for A_1, A_2, A_3 and A_4.
4. Asymptotical expression of $\phi(-\delta u_0/2)$ and $\Phi(-\delta u_0/2)$.
5. Final derivation of asymptotics of MEP_N.

2.1. *Asymptotics of u_0 for $\lambda \to 0$*

Equation (14) is equivalent to equation

$$u^2 + 2\frac{\Phi(\delta u/2)}{\delta\phi(\delta u/2)}\lambda u - 1 = 0. \tag{15}$$

For $\lambda \to 0$ its solution $u_0(\lambda) \to 1$. Therefore, let us approximate u_0 by some polynomial depending on λ:

$$u_0 = 1 + a\lambda + b\lambda^2 + c\lambda^3 + \mathrm{O}\big(\lambda^4\big). \tag{16}$$

REMARK 1. Though $\lambda = (p-3)/(2N) \to 0$, the number of features p may increase. For example, for $p = \mathrm{O}(\sqrt{N})$ λ is of order $1/\sqrt{N}$, and according to (16) and estimates presented below we will obtain generalization error with accuracy $\mathrm{O}(1/N^2)$.

Below we use Taylor expansions of functions Φ and ϕ near the point $x_0 = \delta/2$. It is easy to verify that

$$\Phi'(x) = \phi(x),$$
$$\Phi''(x) = f'(x) = -x\phi(x),$$
$$\Phi'''(x) = f''(x) = \big(x^2 - 1\big)\phi(x).$$

Then

$$\Phi(x) = \Phi_\delta + \phi_\delta\left(x - \frac{\delta}{2}\right) - \frac{\delta\phi_\delta}{4}\left(x - \frac{\delta}{2}\right)^2 + \frac{(\delta^2-4)\phi_\delta}{24}\left(x - \frac{\delta}{2}\right)^3 + \mathrm{o}\left(\left(x - \frac{\delta}{2}\right)^3\right), \tag{17}$$

$$\phi(x) = \phi_\delta - \frac{\delta}{2}\phi_\delta\left(x - \frac{\delta}{2}\right) + \frac{(\delta^2-4)\phi_\delta}{8}\left(x - \frac{\delta}{2}\right)^2 + \mathrm{o}\left(\left(x - \frac{\delta}{2}\right)^2\right), \tag{18}$$

where

$$\Phi_\delta \overset{\mathrm{def}}{=} \Phi\left(\frac{\delta}{2}\right) \quad \text{and} \quad \phi_\delta \overset{\mathrm{def}}{=} \phi\left(\frac{\delta}{2}\right). \tag{19}$$

Inserting (16) into (17) and (18) we obtain

$$\begin{aligned}\Phi\left(\frac{\delta u_0}{2}\right) &= \Phi_\delta + \frac{\delta}{2}\phi_\delta\big(a\lambda + b\lambda^2 + \mathrm{o}\big(\lambda^2\big)\big) - \frac{\delta^3}{16}\phi_\delta\big(a\lambda + \mathrm{o}(\lambda)\big)^2 + \mathrm{o}\big(\lambda^2\big)\\ &= \Phi_\delta + \alpha\lambda + \beta\lambda^2 + \mathrm{o}\big(\lambda^2\big),\end{aligned} \tag{20}$$

for

$$\alpha = \frac{\delta\phi_\delta a}{2} \quad \text{and} \quad \beta = \frac{\delta\phi_\delta(8b - \delta^2 a^2)}{16}, \tag{21}$$

and

$$\begin{aligned}\phi\left(\frac{\delta u_0}{2}\right) &= \phi_\delta - \frac{\delta^2}{4}\phi_\delta\big(a\lambda + b\lambda^2 + \mathrm{o}\big(\lambda^2\big)\big) + \frac{(\delta^2-4)\delta^2\phi_\delta}{32}\big(a\lambda + \mathrm{o}(\lambda)\big)^2\\ &\quad + \mathrm{o}\big(\lambda^2\big) = \phi_\delta + \mu\lambda + \nu\lambda^2 + \mathrm{o}\big(\lambda^2\big),\end{aligned} \tag{22}$$

for

$$\mu = -\frac{\delta^2\phi_\delta a}{4} \quad \text{and} \quad \nu = \frac{\delta^2\phi_\delta((\delta^2-4)a^2 - 8b)}{32}. \tag{23}$$

After inserting (16), (20) and (22) into (15) and simple algebra we have

$$\begin{aligned}&2a\phi_\delta\lambda + 2a\mu\lambda^2 + 2a\nu\lambda^3 + a^2\phi_\delta\lambda^2 + a^2\mu\lambda^3 + 2b\phi_\delta\lambda^2 + 2b\mu\lambda^3 + 2ab\phi_\delta\lambda^3\\ &\quad + 2c\phi_\delta\lambda^3 + \frac{2}{\delta}\big(\Phi_\delta\lambda + a\Phi_\delta\lambda^2 + b\Phi_\delta\lambda^3 + \alpha\lambda^2 + a\alpha\lambda^3 + \beta\lambda^3\big) + \mathrm{o}\big(\lambda^3\big) = 0.\end{aligned}$$

Equalizing the coefficients of terms with λ, λ^2 and λ^3 we obtain a system of equations:

$$\begin{cases} 2\delta\phi_\delta a + 2\Phi_\delta = 0,\\ 2\delta\mu a + \delta\phi_\delta a^2 + 2\delta\phi_\delta b + 2a\Phi_\delta + 2\alpha = 0,\\ 2\delta\nu a + \delta\mu a^2 + 2\delta\mu b + 2\delta\phi_\delta ab + 2\delta\phi_\delta c + 2b\Phi_\delta + 2\alpha a + 2\beta = 0. \end{cases} \tag{24}$$

Let us denote

$$h_\delta \overset{\text{def}}{=} \frac{\Phi(\delta/2)}{\delta\phi(\delta/2)} \quad \left(= \frac{\Phi_\delta}{\delta\phi_\delta}\right). \tag{25}$$

It is easy to verify that system of equations (24) has the following solution:

$$\begin{aligned} a &= -h_\delta,\\ b &= \frac{h_\delta}{2}\big(1 + h_\delta + \tfrac{1}{2}\delta^2 h_\delta\big),\\ c &= -\frac{h_\delta}{32}\big(3\delta^4 h_\delta^2 + 16\delta^2 h_\delta^2 + 10\delta^2 h_\delta + 24h_\delta + 8\big). \end{aligned} \tag{26}$$

Below instead of (16) and (26) we also will use more simple expressions

$$u_0 = 1 - h_\delta \lambda + \mathrm{o}(\lambda) \quad \text{and} \quad u_0^2 = 1 - 2h_\delta \lambda + \mathrm{o}(\lambda). \tag{27}$$

2.2. *Asymptotics of* $\beta_1(u_0)$, $\beta_1'(u_0)$ *and* $1/Z''(u_0)$

Let us return to equations (8)–(13). Since

$$\psi' \overset{\text{def}}{=} \psi'(u) = -\frac{\delta}{2}\psi\left(\frac{\delta u}{2} + \psi\right) \quad \text{and} \quad Z'(u) = \frac{\delta}{2}\psi - \frac{\lambda u}{(1-u^2)}, \tag{28}$$

therefore

$$\beta_1{}'(u) = \frac{\delta}{2}\psi\left(1 - \frac{\delta^2 u^2}{4} - \frac{3\delta u}{2}\psi - 2\psi^2\right), \tag{29}$$

$$Z''(u) = -\frac{\delta^2\psi}{4}\left(\frac{\delta u}{2} + \psi\right) - \frac{\lambda(1+u^2)}{(1-u^2)^2}, \tag{30}$$

$$Z'''(u) = -\frac{\delta^3\psi}{8}\left(1 - \frac{\delta^2 u^2}{4} - \frac{3\delta u}{2}\psi - 2\psi^2\right) - \frac{2\lambda u(3+u^2)}{(1-u^2)^3}. \tag{31}$$

Equations (20)–(23), (25) and (26) yield

$$\Phi\left(\frac{\delta u_0}{2}\right) = \Phi_\delta - \frac{\Phi_\delta}{2}\lambda + \mathrm{o}(\lambda), \tag{32}$$

$$\phi\left(\frac{\delta u_0}{2}\right) = \phi_\delta + \frac{\delta\Phi_\delta}{4}\lambda + \mathrm{o}(\lambda). \tag{33}$$

Hence,

$$\begin{aligned}
\psi_0 \overset{\text{def}}{=} \psi(u_0) &= \frac{\phi_\delta}{\Phi_\delta}\frac{1 + \frac{\delta\Phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)}{1 - \frac{\lambda}{2} + \mathrm{o}(\lambda)} \\
&= \frac{\phi_\delta}{\Phi_\delta}\left(1 + \frac{\delta\Phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)\right)\left(1 + \frac{\lambda}{2} + \mathrm{o}(\lambda)\right) \\
&= \frac{\phi_\delta}{\Phi_\delta}\left(1 + \frac{\delta\Phi_\delta + 2\phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)\right),
\end{aligned} \tag{34}$$

and

$$\psi_0^2 = \frac{\phi_\delta^2}{\Phi_\delta^2}\left(1 + \frac{\delta\Phi_\delta + 2\phi_\delta}{2\phi_\delta}\lambda + \mathrm{o}(\lambda)\right). \tag{35}$$

Inserting (34), (35) into equations (12) and (29)–(31) we obtain

$$\begin{aligned}
\beta_1(u_0) &= \lambda + \mathrm{O}\left(\frac{1}{N}\right) + \frac{\delta}{2}\big(1 - h_\delta\lambda + \mathrm{o}(\lambda)\big)\frac{\phi_\delta}{\Phi_\delta}\left(1 + \frac{\delta\Phi_\delta + 2\phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)\right) \\
&\quad + \frac{\phi_\delta^2}{\Phi_\delta^2}\left(1 + \frac{\delta\Phi_\delta + 2\phi_\delta}{2\phi_\delta}\lambda + \mathrm{o}(\lambda)\right) \\
&= \frac{\delta\phi_\delta}{2\Phi_\delta} + \frac{\phi_\delta^2}{\Phi_\delta^2} + \left(\frac{\delta(\delta\Phi_\delta + 2\phi_\delta)}{8\Phi_\delta} + \frac{1}{2}\right)\lambda + \mathrm{O}\left(\frac{1}{N}\right) + \mathrm{o}(\lambda) \\
&= \frac{\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}{2\Phi_\delta^2} \\
&\quad \times \left(1 + \frac{\delta^2\Phi_\delta^2 + 4\Phi_\delta^2 + 6\delta\Phi_\delta\phi_\delta + 8\phi_\delta^2}{4\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}\lambda + \mathrm{O}\left(\frac{1}{N}\right) + \mathrm{o}(\lambda)\right), \qquad (36) \\
\beta_1'(u_0) &= \frac{\delta\phi_\delta}{2\Phi_\delta}\big(1 + \mathrm{o}(1)\big) \\
&\quad \times \left[1 - \frac{\delta^2}{4}\big(1 + \mathrm{o}(1)\big) - \frac{3\delta}{2}\big(1 + \mathrm{o}(1)\big)\frac{\phi_\delta}{\Phi_\delta}\big(1 + \mathrm{o}(1)\big) - \frac{2\phi_\delta^2}{\Phi_\delta^2}\big(1 + \mathrm{o}(1)\big)\right] \\
&= \frac{\delta\phi_\delta(4\Phi_\delta^2 - \delta^2\Phi_\delta^2 - 6\delta\Phi_\delta\phi_\delta - 8\phi_\delta^2)}{8\Phi_\delta^3} + \mathrm{o}(1), \qquad (37)
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{Z''(u_0)} &= \frac{(1 - u_0^2)^2}{-\frac{\delta^2\psi_0}{4}\big(\frac{\delta u_0}{2} + \psi_0\big)(1 - u_0^2)^2 - \lambda(1 + u_0^2)} \\
&= \frac{\lambda(4h_\delta^2\lambda + \mathrm{o}(\lambda))}{-\frac{\delta^2\psi_0}{4}\big(\frac{\delta u_0}{2} + \psi_0\big)\lambda(4h_\delta^2\lambda + \mathrm{o}(\lambda)) - \lambda(2 + \mathrm{o}(1))} \\
&= \frac{4h_\delta^2\lambda + \mathrm{o}(\lambda)}{-2 + \mathrm{o}(1)} = -2h_\delta^2\lambda + \mathrm{o}(\lambda). \qquad (38)
\end{aligned}$$

2.3. *Asymptotics of* A_1*,* A_2*,* A_3 *and* A_4

From (27) and (36) we obtain

$$\begin{aligned}
A_1 &= \left(1 - \frac{\Phi_\delta}{\delta\phi_\delta}\lambda + \mathrm{o}(\lambda)\right)\frac{2\Phi_\delta^2}{\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)} \\
&\quad \times \left(1 - \frac{\delta^2\Phi_\delta^2 + 4\Phi_\delta^2 + 6\delta\Phi_\delta\phi_\delta + 8\phi_\delta^2}{4\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}\lambda + \mathrm{O}\left(\frac{1}{N^2}\right) + \mathrm{o}(\lambda)\right) \\
&= \frac{2\Phi_\delta^2}{\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}\left(1 - \frac{\delta^3\Phi_\delta^2 + 8\delta\Phi_\delta^2 + 6\delta^2\Phi_\delta\phi_\delta + 8\Phi_\delta\phi_\delta + 8\delta\phi_\delta^2}{4\delta\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}\lambda \right. \\
&\qquad \left. + \mathrm{O}\left(\frac{1}{N^2}\right) + \mathrm{o}(\lambda)\right).
\end{aligned}$$

Equations (9), (27) and (38) yield

$$A_2 = -\frac{\delta^2}{4}\big(1+\mathrm{o}(1)\big)\big(-2h_\delta^2\lambda + \mathrm{o}(\lambda)\big) = \frac{\Phi_\delta^2}{2\phi_\delta^2}\lambda + \mathrm{o}(\lambda). \tag{39}$$

Inserting (30) and (31) into (10) we have

$$\begin{aligned} A_3 \;=\; & \left[\frac{\delta^3\psi_0}{8}\left(1 - \frac{u_0^2\delta^2}{4} - \frac{3u_0\delta}{2}\psi_0 - 2\psi_0^2\right)\big(1-u_0^2\big)^3 + 2\lambda u_0\big(3+u_0^2\big)\right] \\ & \times \big(1-u_0^2\big)\left[\frac{\delta^2\psi_0}{4}\left(\frac{u_0\delta}{2} + \psi_0\right)\big(1-u_0^2\big)^2 + \lambda\big(1+u_0^2\big)\right]^{-2}. \end{aligned} \tag{40}$$

According to (16) and (26),

$$u_0 = 1 - h_\delta\lambda + \frac{h_\delta}{2}\left(1 + h_\delta + \frac{1}{2}\delta^2 h_\delta\right)\lambda^2 + \mathrm{o}\big(\lambda^2\big), \tag{41}$$

$$u_0^2 = 1 - 2h_\delta\lambda + h_\delta\Big(1 + 2h_\delta + \frac{1}{2}\delta^2 h_\delta\big)\lambda^2 + \mathrm{o}\big(\lambda^2\big). \tag{42}$$

Therefore,

$$1 - u_0^2 = 2h_\delta\lambda\left(1 - \left(\frac{1}{2} + h_\delta + \frac{1}{4}\delta^2 h_\delta\right)\lambda + \mathrm{o}(\lambda)\right). \tag{43}$$

Inserting (34), (35) and (41)–(43) into (40) we obtain

$$\begin{aligned} A_3 \;=\; & \Big[2\lambda\big(1 - h_\delta\lambda + \mathrm{o}(\lambda)\big)\big(4 - 2h_\delta\lambda + \mathrm{o}(\lambda)\big) \\ & \times 2h_\delta\lambda\left(1 - \left(\frac{1}{2} + h_\delta + \frac{1}{4}\delta^2 h_\delta\right)\lambda + \mathrm{o}(\lambda)\right) + \mathrm{o}\big(\lambda^2\big)\Big] \\ & \times \left[\frac{\delta}{4h_\delta}\big(1+\mathrm{o}(1)\big)\left(\frac{\delta}{2}\big(1+\mathrm{o}(1)\big) + \frac{1}{\delta h_\delta}\big(1+\mathrm{o}(1)\big)\right)\right. \\ & \left.\times\, 4h_\delta^2\big(1+\mathrm{o}(1)\big)\lambda + 2 - 2h_\delta\lambda + \mathrm{o}(\lambda)\right]^{-2}\lambda^{-2} \\ =\; & \left[16h_\delta\left(1 - \frac{3}{2}h_\delta\lambda + \mathrm{o}(\lambda)\right)\left(1 - \left(\frac{1}{2} + h_\delta + \frac{1}{4}\delta^2 h_\delta\right)\lambda + \mathrm{o}(\lambda)\right) + \mathrm{o}(\lambda)\right] \\ & \times \frac{1}{4}\left[1 + \left(\frac{1}{2} + \frac{1}{4}\delta^2 h_\delta - h_\delta\right)\lambda + \mathrm{o}(\lambda)\right]^{-2} \\ =\; & 4h_\delta\left[1 - \left(\frac{1}{2} + \frac{5}{2}h_\delta + \frac{1}{4}\delta^2 h_\delta\right)\lambda + \mathrm{o}(\lambda)\right]\left[1 - \left(1 + \frac{1}{2}\delta^2 h_\delta - 2h_\delta\right)\lambda + \mathrm{o}(\lambda)\right] \\ =\; & \frac{4\Phi_\delta}{\delta\phi_\delta}\left(1 - \frac{3\delta^2\Phi_\delta + 2\Phi_\delta + 6\delta\phi_\delta}{4\delta\phi_\delta}\lambda + \mathrm{o}(\lambda)\right). \end{aligned}$$

Finally, let us insert (36)–(38) into (11):

$$\begin{aligned} A_4 &= \left(2\frac{\Phi_\delta^2}{\delta^2\phi_\delta^2}\lambda + \mathrm{o}(\lambda)\right)\left(\frac{\delta\phi_\delta(4\Phi_\delta^2 - \delta^2\Phi_\delta^2 - 6\Phi_\delta\phi_\delta - 8\phi_\delta^2)}{8\Phi_\delta^3} + \mathrm{o}(1)\right) \\ &\quad\times \frac{2\Phi_\delta^2}{\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)}\big(1+\mathrm{o}(1)\big) \\ &= \frac{\Phi_\delta(4\Phi_\delta^2 - \delta^2\Phi_\delta^2 - 6\delta\Phi_\delta\phi_\delta - 8\phi_\delta^2)}{2\delta\phi_\delta^2(\delta\Phi_\delta + 2\phi_\delta)}\lambda + \mathrm{o}(\lambda). \end{aligned}$$

2.4. *Asymptotics of* $\phi(-\delta u_0/2)$ *and* $\Phi(-\delta u_0/2)$

Obviously, $\phi(-x) = \phi(x)$. Hence, equation (33) yields

$$\phi\left(-\frac{\delta u_0}{2}\right) = \phi_\delta\left(1 + \frac{\delta\Phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)\right). \tag{44}$$

It remains to find an asymptotics of $\Phi(-\delta u_0/2)$. Taylor expansion near the point $x_0 = -\delta/2$ gives

$$\begin{aligned} \Phi\left(-\frac{\delta u_0}{2}\right) &= \Phi\left(-\frac{\delta}{2}\right) + \phi_\delta\frac{\delta}{2}(1-u_0) + \frac{\delta\phi_\delta}{4}\frac{\delta^2}{4}(1-u_0)^2 \\ &\quad + \frac{(\delta^2-4)\phi_\delta}{24}\frac{\delta^3}{8}(1-u_0)^3 + \mathrm{o}\big((1-u_0)^3\big) \\ &= \Phi\left(-\frac{\delta}{2}\right) + \frac{\delta\phi_\delta}{2}\left[h_\delta\lambda - b\lambda^2 - c\lambda^3 + \mathrm{o}\big(\lambda^3\big)\right] \\ &\quad + \frac{\delta^3\phi_\delta}{16}\left[h_\delta^2\lambda^2 - 2h_\delta b\lambda^3 + \mathrm{o}\big(\lambda^3\big)\right] + \frac{\delta^3(\delta^2-4)\phi_\delta h_\delta^3}{192}\lambda^3 + \mathrm{o}\big(\lambda^3\big), \end{aligned}$$

where b and c are given by (26). Substituting (26) into above equation after little algebra we obtain

$$\Phi\left(-\frac{\delta u_0}{2}\right) = \Phi\left(-\frac{\delta}{2}\right) + C_1\lambda + C_2\lambda^2 + C_3\lambda^3 + \mathrm{o}\big(\lambda^3\big), \tag{45}$$

where

$$\begin{aligned} C_1 &= \frac{1}{2}\Phi_\delta, \\ C_2 &= -\frac{\Phi_\delta}{16\delta\phi_\delta}\big(\delta^2\Phi_\delta + 4\Phi_\delta + 4\delta\phi_\delta\big), \\ C_3 &= \frac{\Phi_\delta}{96\delta\phi_\delta^2}\big(2\delta^3\Phi_\delta^2 + 16\delta\Phi_\delta^2 + 9\delta^2\Phi_\delta\phi_\delta + 36\Phi_\delta\phi_\delta + 12\delta\phi_\delta^2\big). \end{aligned} \tag{46}$$

2.5. *Asymptotics of* MEP_N

Using the results of Sections 2.3 and 2.4 we can rewrite equation (7) as

$$\begin{aligned}
\mathrm{MEP}_N &= \Phi\left(-\frac{\delta}{2}\right) + \frac{\Phi_\delta}{2}\lambda + C_2\lambda^2 + C_3\lambda^3 + \mathrm{o}\big(\lambda^3\big) + \frac{\delta\phi_\delta}{8N}\left[1 + \frac{\delta\Phi_\delta}{4\phi_\delta}\lambda + \mathrm{o}(\lambda)\right] \\
&\quad \times \left[\frac{2\Phi_\delta^2}{\phi_\delta(\delta\Phi_\delta + 2\phi_\delta)} + \frac{4\Phi_\delta}{\delta\phi_\delta} + (B_1 + B_2 + B_3 + B_4)\lambda + \mathrm{o}(\lambda)\right] \\
&\quad + \mathrm{O}\left(\frac{1}{N^2}\right) \\
&= \Phi\left(-\frac{\delta}{2}\right) + \frac{\Phi_\delta}{2}\lambda + C_2\lambda^2 + C_3\lambda^3 + \mathrm{o}\big(\lambda^3\big) + \frac{\Phi_\delta(3\delta\Phi_\delta + 4\phi_\delta)}{4(\delta\Phi_\delta + 2\phi_\delta)}\frac{1}{N} \\
&\quad + \frac{\delta\phi_\delta}{8}(B_1 + B_2 + B_3 + B_4 + B_5)\frac{\lambda}{N} + \mathrm{o}\left(\frac{\lambda}{N}\right) + \mathrm{O}\left(\frac{1}{N^2}\right), \qquad (47)
\end{aligned}$$

where

$$\begin{aligned}
B_1 &= -\frac{\Phi_\delta^2(\delta^3\Phi_\delta^2 + 8\delta\Phi_\delta^2 + 6\delta^2\Phi_\delta\phi_\delta + 8\Phi_\delta\phi_\delta + 8\delta\phi_\delta^2)}{2\delta\phi_\delta^2(\delta\Phi_\delta + 2\phi_\delta)^2}, \\
B_2 &= \frac{\Phi_\delta^2}{2\phi_\delta^2}, \\
B_3 &= -\frac{\Phi_\delta(3\delta^2\Phi_\delta + 2\Phi_\delta + 6\delta\phi_\delta)}{\delta^2\phi_\delta^2}, \\
B_4 &= \frac{\Phi_\delta(4\Phi_\delta^2 - \delta^2\Phi_\delta^2 - 6\delta\Phi_\delta\phi_\delta - 8\phi_\delta^2)}{2\delta\phi_\delta^2(\delta\Phi_\delta + 2\phi_\delta)}, \\
B_5 &= \frac{\Phi_\delta^2(3\delta\Phi_\delta + 4\phi_\delta)}{2\phi_\delta^2(\delta\Phi_\delta + 2\phi_\delta)},
\end{aligned}$$

C_2 and C_3 are given by (46) and $\lambda = (p-3)/(2N) \to 0$.

For $p = \mathrm{const}$, $\lambda = \mathrm{O}(1/N)$. In this case we obtain more simple asymptotical expression for MEP_N:

$$\mathrm{MEP}_N = \Phi\left(-\frac{\delta}{2}\right) + \left(\frac{\Phi_\delta(p-3)}{4} + \frac{\Phi_\delta(3\delta\Phi_\delta + 4\phi_\delta)}{4(\delta\Phi_\delta + 2\phi_\delta)}\right)\frac{1}{N} + \mathrm{O}\left(\frac{1}{N^2}\right). \qquad (48)$$

Since

$$\frac{3\delta\Phi_\delta + 4\phi_\delta}{\delta\Phi_\delta + 2\phi_\delta} = 3 - \frac{2\phi_\delta}{\delta\Phi_\delta + 2\phi_\delta},$$

therefore we obtain

$$\mathrm{MEP}_N \sim \Phi\left(-\frac{\delta}{2}\right) + \frac{\Phi_\delta}{4}\frac{p}{N} - \frac{\Phi_\delta\phi_\delta}{2(\delta\Phi_\delta + 2\phi_\delta)}\frac{1}{N}. \qquad (49)$$

Let

$$t(\delta) = \frac{f_\delta}{2(\delta\Phi_\delta + 2f_\delta)}.$$

It is easy to show that $t(\delta) \downarrow 0$ ($\delta \to \infty$) and $\max_\delta t(\delta) = t(0) = 0.25$. Since $p \geqslant 4$, we have $(\Phi_\delta/4)(p/N) \geqslant \Phi_\delta/N$, therefore the last term in the right-hand side of (49) is very small comparing with the first two terms (at least, for $\delta \geqslant 1$, since $t(1) \approx 0.1261$), and we can rewrite (49) in an extremely simple form:

$$\mathrm{MEP}_N \sim \Phi\left(-\frac{\delta}{2}\right) + \frac{1}{4}\Phi\left(\frac{\delta}{2}\right)\frac{p}{N}. \tag{50}$$

3. Numerical Simulations

Figures 1 and 2 compare the following expressions of the generalization error for $\delta = 2$ and $\delta = 4$, respectively:

(i) an asymptotic formula (7) of Basalykas *et al.* (1996);
(ii) our asymptotic formula (47);
(iii) our short asymptotics (50);
(iv) an empirical formula (6) of Raudys (1997).

The exact values of MEP_N obtained using (i)–(iv) for $\delta = 2, 4$ and $p = 10$ and 50, are presented in Tables 1 and 2, respectively. Since there was demonstrated in (Basalykas *et al.*, 1996) that the formula of Basalykas *et al.* is very accurate, here we consider its values as the exact ones. In (Basalykas *et al.*, 1996) was also mentioned that the MEP_N mainly depends on a ratio p/N. Therefore, though our derivation of formulas (47) and (50) is valid for $p \geqslant 4$, these formulas together with Tables 1 and 2 in obvious way can be used for the case $p \leqslant 3$.

Figures 1 and 2 show that for the small N values an empirical formula of Raudys can be useful while for the bigger N values our short asymptotics is sufficiently accurate. We also can see that for the bigger δ one needs to take the bigger N to make our formula accurate. Indeed, one can notice that in equations (16), (20), (22) and (32) the factors C_k of λ^k ($\lambda = (p-3)/(2N)$) increase together with k increasing. Moving from $C_k\lambda^k$ to $C_{k+1}\lambda^{k+1}$ this increasing is of the order δ/f_δ, therefore these equations are valid for such δ that

$$\frac{\delta}{f(\delta/2)} \cdot \frac{p-3}{2N} = \mathrm{o}(1),$$

i.e.,

$$N \gg \frac{\delta p}{2f(\delta/2)}.$$

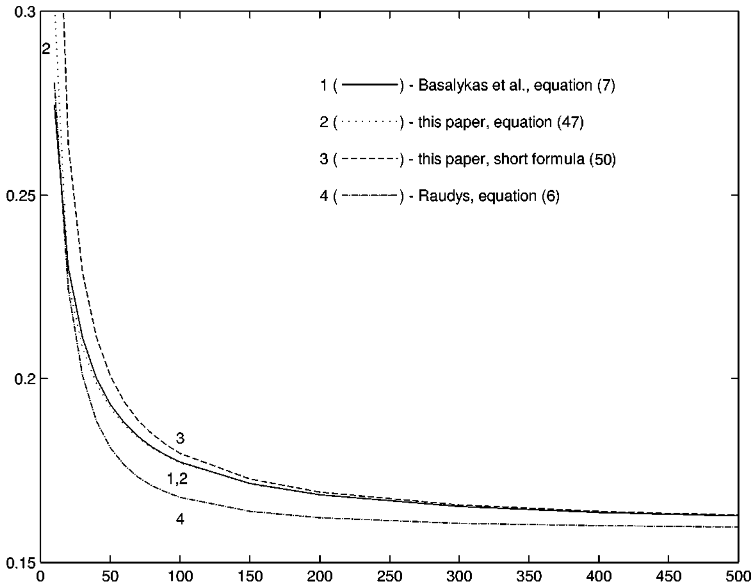

Fig. 1. Generalization error versus length N of learning set for $p = 10$ and $\delta = 2$.

The numerical simulations show that for $\delta \geqslant 1$ starting from $N_0 = \delta p/f(\delta/2)$ our short asymptotics (50) is sufficiently accurate (for $n \geqslant N_0$). By example, if $p = 10$, for $\delta = 2$ and $\delta = 4$ we have $N_0 = 83$ and 741, respectively (compare with Figs. 1, 2 and Table 1).

Table 1

Generalization error values obtained by different authors ($p = 10$). B means Basalykas *et al.*, D (Dičiūnas) – this paper, equation (47), S – short asymptotics (50), R – Raudys

N	10	20	50	100	200	500	1000	2000	10000
				$\delta = 2$, Bayes' error $\mathrm{MEP}_\infty = 0.1587$					
B	0.2745	0.2303	0.1930	0.1773	0.1684	0.1627	0.1607	0.1597	0.1589
D	0.3026	0.2260	0.1922	0.1772	0.1684	0.1627	0.1607	0.1597	0.1589
S	0.3690	0.2638	0.2007	0.1797	0.1692	0.1629	0.1608	0.1597	0.1589
R	0.2805	0.2244	0.1812	0.1677	0.1622	0.1596	0.1590	0.1588	0.1587
				$\delta = 4$, Bayes' error $\mathrm{MEP}_\infty = 0.0228$					
B	0.1189	0.0822	0.0530	0.0403	0.0327	0.0272	0.0251	0.0239	0.0230
D	*	*	0.0715	0.0399	0.0322	0.0271	0.0250	0.0239	0.0230
S	0.2671	0.1449	0.0716	0.0472	0.0350	0.0276	0.0252	0.0240	0.0230
R	0.1362	0.0868	0.0492	0.0358	0.0292	0.0253	0.0240	0.0234	0.0229

* – unapplicapable value.

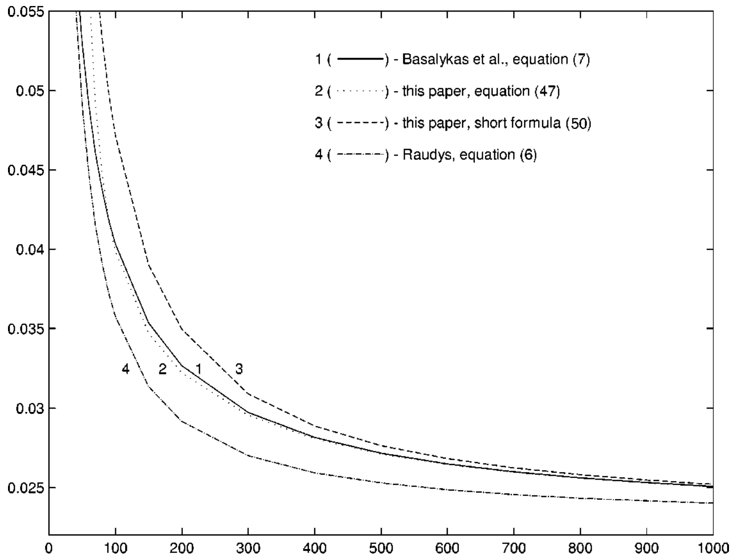

Fig. 2. Generalization error versus length N of learning set for $p = 10$ and $\delta = 4$.

Table 2

Generalization error values obtained by different authors ($p = 50$)

N	10	20	50	100	200	500	1000	2000	10000
				$\delta = 2$, Bayes' error $\text{MEP}_\infty = 0.1587$					
B	0.4039	0.3493	0.2753	0.2313	0.2008	0.1776	0.1686	0.1637	0.1597
D	*	*	*	0.2528	0.2023	0.1776	0.1686	0.1637	0.1597
S	*	*	0.3690	0.2638	0.2112	0.1797	0.1692	0.1639	0.1597
R	0.4108	0.3624	0.2805	0.2244	0.1885	0.1677	0.1622	0.1600	0.1588
				$\delta = 4$, Bayes' error $\text{MEP}_\infty = 0.0228$					
B	0.4039	0.3493	0.2753	0.2313	0.2008	0.1776	0.1686	0.1637	0.1597
D	*	*	*	0.2528	0.2023	0.1776	0.1686	0.1637	0.1597
S	*	*	0.3690	0.2638	0.2112	0.1797	0.1692	0.1639	0.1597
R	0.4108	0.3624	0.2805	0.2244	0.1885	0.1677	0.1622	0.1600	0.1588

* – unapplicapable value.

References

Amari, S. (1993). A universal theorem on learning curves. *Neural Networks*, **6**(1), 61–66.

Amari, S., N. Fujita, S. Shinomoto (1992). Four types of learning curves. *Neural Computation*, **4**(6), 5–18.

Basalykas, A., V. Dičiūnas, Š. Raudys (1996). On expected probability of misclassification of linear zero empirical error classifier. *Informatica*, **7**(2), 137–154; (1997). Letter to Editors. *Informatica*, **8**(2), 310–311.

Deev, A.D. (1970). Representation of statistics of discriminant analysis and asymptotic expansions in dimensionalities comparable with sample size. *Reports of Academy of Sciences of the USSR*, **195**(4), 756–762 (in Russian).

Deev, A.D. (1972). Asymptotic expansions for distributions of statistics W, M, and W* in discriminant analysis. In Yu.N. Blagoveshenskij (Ed.), *Statistical Methods of Classification*, Vol. 31, Moscow Univ. Press, Moscow, pp. 6–57 (in Russian).

Dičiūnas, V. Generalization error of randomized linear zero empirical error classifier: I. Non-centered data case. *Informatica*, to appear in 2001.

Haussler, D., M. Kearns, H.S. Seung, N. Tishby (1994). Rigorous learning curves from statistical mechanics. In *Proceedings of the 7th Annual ACM Workshop on Computational Learning Theory*, Morgan Kaufman, San Mateo, CA, pp. 76–87.

John, S. (1961). Errors in discrimination. *Ann. Math. Statist.*, **32**, 1125–1144.

Meir, R. (1995). Empirical risk minimization versus maximum-likelihood estimation: a case study. *Neural Computation*, **1**, 144–157.

Okamoto, M. (1963). An asymptotic expansion for the distribution of linear discriminant function. *Ann. Math. Statist.*, **34**, 1286–1301; (1968). Erratum. *Ann. Math. Statist.*, **39**, 1358–1359.

Raudys, Š. (1967). On determining training sample size of linear classifier. *Computing Systems*, **28**, 79–87 (in Russian).

Raudys, Š. (1972). On the amount of a priori information in designing the classification algorithm. *Engineering Cybernetics*, Proc. Acad. of Sciences of the USSR, **4**, pp. 168–174 (in Russian).

Raudys, Š. (1993). On shape of pattern error function, initializations and intrinsic dimensionality in ANN classifier design. *Informatica*, **4**(3–4), 360–383.

Raudys, Š. (1997). On dimensionality, sample size, and classification error of nonparametric linear classification algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**(6), 667–671.

Raudys Š., V. Dičiūnas (1994). Generalization error of linear margin classifier. In *Proc. of 6th Microcomputer School on Neural Networks and Applications*, Sedmihorky, Czech Republic, pp. 159–164.

Raudys Š., V. Dičiūnas (1996). Expected error of minimum empirical error and maximal margin classifiers. In: *Proc. of 13th Internat. Conf. on Pattern Recognition*, Vienna, Austria, August 25–29, Vol. 2, IEEE Computer Society Press, Los Alamitos, CA, pp. 875–879.

Raudys Š., V. Pikelis (1980). On dimensionality, sample size, classification error, and complexity of classification algorithm in pattern recognition. *IEEE Transactions on PAMI*, **2**(3), 242–252.

Sitgreaves, R. (1961). Some results on the distribution of the W-classification statistics. In *Studies in Item Selection and Prediction*, Stanford Univ. Press., California, pp. 241–261.

Vapnik, V.N. (1982). *Estimation of Dependencies Based on Empirical Data*. Springer, New York.

Wyman, F., D.M. Young, D.W. Turner (1990). A comparison of asymptotic error rate expansions for the sample linear discriminant function. *Pattern Recognition*, **23**, 775–783.

V. Dičiūnas received a mathematician's diploma from Moscow State University in 1982. He is a Senior Assistant Professor at the Department of Computer Science of Vilnius University. His research interests include artificial neural networks, statistical classification methods, and complexity theory.

Š. Raudys received the M.S. degree in electrical and computer engineering from Kaunas University of Technology in 1963, and the candidate of Sciences and Doctor of Sciences degrees from the Institute of Mathematics and Cybernetics, Academy of Sciences, Lithuania, in 1969 and 1978, respectively. He is a Head of the Department of Data Analysis of the Institute of Mathematics and Informatics. His research interests include statistical pattern recognition, artificial neural nets, expert systems, machine learning, and data analysis methods. He is an Associate Editor of international journals: Pattern Recognition, Pattern Recognition and Image Analysis, Pattern Analysis and Applications, Informatica. He has been invited as a speaker and/or a member of Program Committees of about 30 International Conferences. He is an author of the book *Statistical and Neural Classifiers; An Integrated Approach* (Springer, 2001).

Atsitiktinio nulinės empirinės klaidos tiesinio klasifikatoriaus vidutinė tikėtina klaida: paprasta asimptotika centruotų duomenų atveju

Valdas DIČIŪNAS, Šarūnas RAUDYS

Šiame straipsnyje gauta atsitiktinio nulinės empirinės klaidos tiesinio klasifikatoriaus vidutinės tikėtinos klaidos (VTK) išreikštinė asimptotika centruotų duomenų atveju. Taip pat pasiūlyta kita labai paprasta asimptotinė VTK formulė ir išnagrinėtos salygos, prie kurių ši formulė tampa pakankamai tiksli. Gautų formulių tikslumas palygintas su kitų autorių pasiūlytų formulių tikslumu.