

# New Bounds and Approximations for the Error of Linear Classifiers

Luis Rueda

School of Computer Science, University of Windsor  
401 Sunset Avenue, Windsor, ON, N9B 3P4, Canada  
lrueda@uwindsor.ca

**Abstract.** In this paper, we derive lower and upper bounds for the probability of error for a linear classifier, where the random vectors representing the underlying classes obey the *multivariate* normal distribution. The expression of the error is derived in the *one-dimensional space*, independently of the dimensionality of the original problem. Based on the two bounds, we propose an approximating expression for the error of a generic linear classifier. In particular, we derive the corresponding bounds and the expression for approximating the error of Fisher's classifier. Our empirical results on synthetic data, including up to five-hundred-dimensional featured samples, show that the computations for the error are extremely fast and quite accurate; the approximation differs from the *actual* error by at most  $\varepsilon = 0.0184340683$ .

## 1 Introduction

Assessing the performance of classifiers is a fundamental problem in pattern recognition, for which various approaches have been proposed in the literature. The main idea is to measure the discriminability of the classifier by means of its *misclassification rate* or *error rate*. The error rate or classification error, in general, measured as the *probability of error*, provides a quite useful insight about the quality of a classifier. We consider the classical problem of deriving the *true* error rate for a linear classifier, which we presently refer to as the *classification error* or *probability of error*. We deal with two classes,  $\omega_1$  and  $\omega_2$ , whose *a priori* probabilities are  $P(\omega_1)$  and  $P(\omega_2)$  respectively, and which are represented by two normally distributed  $d$ -dimensional random vectors,  $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and  $\mathbf{x}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ , respectively. In order to derive a linear classification scheme, the aim is to find a linear function of the form:

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0 = 0, \quad (1)$$

that classifies an unknown object, represented by a real-valued feature vector,  $\mathbf{x} = [x_1, \dots, x_d]^t$ , into the respective class, where  $\mathbf{w}$  is a  $d$ -dimensional *weight* vector, such that  $\mathbf{w} \neq \mathbf{0}_d$ , and  $w_0$  is a *threshold* weight.

The problem of estimating the classification error has been studied for various cases, including an asymptotic formula for the expected error of the pseudo-Fisher classifier for the case in which the dimensionality of the problem is relatively larger than the size of the training dataset [8]. Bounds for more generic scenarios have been derived for

linear classifiers that use kernel functions [5]. In the case of the Bayesian (quadratic) classifier for normal distributions, it is well-known that bounds on the classification exist, namely Chernoff's and Battacharyya's bounds [3, 4, 12], and the approximation method introduced by Lee *et al.* [7]. These bounds are applicable to the *optimal* classifier, which is linear only for equal covariance matrices, and that they are not tight enough for the majority of the cases.

## 2 Bounds and Approximations for the Error

We assume that we are dealing with the case in which  $\boldsymbol{\mu}_1$  is on the "negative side" of the classifier, i.e.  $g(\boldsymbol{\mu}_1) < 0$ , which implies that  $g(\boldsymbol{\mu}_2) > 0$ . It is, thus, easy to see that to evaluate the opposite case, i.e. when  $g(\boldsymbol{\mu}_1) > 0$  and  $g(\boldsymbol{\mu}_2) < 0$ , it suffices to rename the classes in such a way that the new class  $\omega_1$  satisfies  $g(\boldsymbol{\mu}_1) < 0$ .

Given a linear classifier of the form of (1), and assuming that the space is divided into two regions,  $\mathcal{R}_1$  and  $\mathcal{R}_2$ , which represent the areas in which an object is assigned to  $\omega_1$  and  $\omega_2$  respectively, the probability of error,  $\text{Pr}[\text{error}]$ , is calculated as follows [3]:

$$\text{Pr}[\text{error}] = \int_{\mathcal{R}_2} p_{x_1}(\mathbf{x}|\omega_1)P(\omega_1)d\mathbf{x} + \int_{\mathcal{R}_1} p_{x_2}(\mathbf{x}|\omega_2)P(\omega_2)d\mathbf{x}, \quad (2)$$

where  $p_{x_i}(\mathbf{x}|\omega_i)$  is the probability of  $\mathbf{x}$  given  $\omega_i$ ,  $\mathcal{R}_1$  is the region determined by  $g(\mathbf{x}) < 0$ , and  $\mathcal{R}_2$  is the region determined by  $g(\mathbf{x}) > 0$ . Note that  $g(\mathbf{x}) = 0$  is excluded from (2) since  $\int_{g(\mathbf{x})=0} p_{x_i}(\mathbf{x}|\omega_i) = 0$ .

The expression given in (2) is quite involved due to the fact that it invokes integrating multivariate normal distributions. Fortunately, when dealing with normal distributions, simpler expressions that involve  $N(0, 1)$  random variables can be used (see [11]). An elegant way of writing the integrals of (2) in terms of an  $N(0, 1)$  random variable,  $x$ , is as follows [4]:

$$\text{Pr}[\text{error}] = P(\omega_1) \int_{-\infty}^{a_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + P(\omega_2) \int_{-\infty}^{a_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad (3)$$

where  $x$  is an  $N(0, 1)$  random variable, and

$$a_1 = (\mathbf{w}^t \boldsymbol{\Sigma}_1 \mathbf{w})^{-\frac{1}{2}} (w_0 + \mathbf{w}^t \boldsymbol{\mu}_1), \text{ and} \quad (4)$$

$$a_2 = -(\mathbf{w}^t \boldsymbol{\Sigma}_2 \mathbf{w})^{-\frac{1}{2}} (w_0 + \mathbf{w}^t \boldsymbol{\mu}_2). \quad (5)$$

Although the multi-dimensional problem is reduced to its equivalent in the one-dimensional space, the analytic form of the integral for the univariate normal distribution density function is still not possible, and thus an algebraic analysis is only possible by means of bounds. To derive the expressions for the lower and upper bounds, we use a well-known inequality for the cumulative normal distribution function [6].

We know that  $a_1 = (\mathbf{w}^t \boldsymbol{\Sigma}_1 \mathbf{w})^{-\frac{1}{2}} (w_0 + \mathbf{w}^t \boldsymbol{\mu}_1)$ . Since  $g(\boldsymbol{\mu}_1) < 0$ , it then follows that  $\mathbf{w}^t \boldsymbol{\mu}_1 + w_0 < 0$ . Also,  $\boldsymbol{\Sigma}_1$  is positive definite, which implies that  $(\mathbf{w}^t \boldsymbol{\Sigma}_1 \mathbf{w})^{-\frac{1}{2}} > 0$ . Thus, it follows that  $a_1 < 0$ . Also, we know that  $a_2 = -(\mathbf{w}^t \boldsymbol{\Sigma}_2 \mathbf{w})^{-\frac{1}{2}} (w_0 + \mathbf{w}^t \boldsymbol{\mu}_2)$ .

Since  $g(\boldsymbol{\mu}_2) = \mathbf{w}^t \boldsymbol{\mu}_2 + w_0 > 0$ , and  $\boldsymbol{\Sigma}_1$  is positive definite, which implies that  $(\mathbf{w}^t \boldsymbol{\Sigma}_2 \mathbf{w})^{-\frac{1}{2}} > 0$ , it follows that  $a_2 < 0$ .

Since  $a_i < 0$  for  $a = 1, 2$ , the inequality given in [6] can be expressed, in terms of  $a_i$ , as follows:

$$\frac{2e^{-\frac{a_i^2}{2}}}{(-a_i + \sqrt{a_i^2 + 4})\sqrt{2\pi}} \leq \int_{-\infty}^{a_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \leq \frac{2e^{-\frac{a_i^2}{2}}}{(-a_i + \sqrt{a_i^2 + 2})\sqrt{2\pi}}. \quad (6)$$

Pre-multiplying by  $P(\omega_i)$ , and adding the corresponding terms of (6) for  $i = 1, 2$ , we obtain the following inequality:

$$\begin{aligned} \sqrt{\frac{2}{\pi}} \left( \frac{P(\omega_1)e^{-\frac{a_1^2}{2}}}{-a_1 + \sqrt{a_1^2 + 4}} + \frac{P(\omega_2)e^{-\frac{a_2^2}{2}}}{-a_2 + \sqrt{a_2^2 + 4}} \right) &\leq \text{Pr}[\text{error}] \leq \\ &\sqrt{\frac{2}{\pi}} \left( \frac{P(\omega_1)e^{-\frac{a_1^2}{2}}}{-a_1 + \sqrt{a_1^2 + 2}} + \frac{P(\omega_2)e^{-\frac{a_2^2}{2}}}{-a_2 + \sqrt{a_2^2 + 2}} \right), \end{aligned} \quad (7)$$

where  $a_1$  and  $a_2$  are obtained as in (4) and (5) respectively.

The bounds for the classification error are important to obtain a fair assessment about the classifier, without computing the error using numeric integration methods. One should note, however, that these bounds are not tight enough for values of  $a_i$  close to 0, while being asymptotically accurate for  $a_i < -1$ . An alternative for this is to use the following approximating function for the probability of error:

$$\begin{aligned} \text{Pr}[\text{error}] \cong \frac{1}{\sqrt{2\pi}} \left[ P(\omega_1)e^{-\frac{a_1^2}{2}} \left( \frac{1}{-a_1 + \sqrt{a_1^2 + 4}} + \frac{1}{-a_1 + \sqrt{a_1^2 + 2}} \right) \right. \\ \left. + P(\omega_2)e^{-\frac{a_2^2}{2}} \left( \frac{1}{-a_2 + \sqrt{a_2^2 + 4}} + \frac{1}{-a_2 + \sqrt{a_2^2 + 2}} \right) \right]. \end{aligned} \quad (8)$$

Thus, taking the average between the lower and the upper bounds as in (8) appears to be a very good approximation for the actual probability of error. To assess this, we first analyze the relationship between the integral and the average of the two bounds in (6). We are tempted to call the result given below *Lemma*, but unfortunately, one of the steps cannot be proved algebraically, and hence we leave it as a proposition. The complete proof of this result can be found in [9].

**Proposition 1.** For all  $a_i < 0$

$$|g(a_i)| = \left| \int_{-\infty}^{a_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - \frac{e^{-\frac{a_i^2}{2}}}{\sqrt{2\pi}} \left( \frac{1}{-a_i + \sqrt{a_i^2 + 4}} + \frac{1}{-a_i + \sqrt{a_i^2 + 2}} \right) \right| \leq \varepsilon, \quad (9)$$

where  $\varepsilon = 0.0184340683$ . □

Using this result, we now state the relationship between the approximation function in (8) and the actual probability of error, which is computed as in (3).

**Theorem 1.** *The approximation given in (8) differs from the actual error in (3) in at most  $\varepsilon = 0.0184340683$ .*

*Proof (Sketch).* Since  $P(\omega_1) \geq 0$  and  $P(\omega_2) \geq 0$ , using the inequality given in (9), we can write:

$$|P(\omega_1)g(a_i)| \leq P(\omega_1)\varepsilon, \text{ and} \tag{10}$$

$$|P(\omega_2)g(a_i)| \leq P(\omega_2)\varepsilon = [1 - P(\omega_1)]\varepsilon. \tag{11}$$

Adding (10) and (11), and rearranging, the result follows. The complete proof can be found in [9]. □

The result of Theorem 1 is quite important in our analysis, since it shows that the approximation for the error is very accurate, e.g. it differs from the actual error in at most two digits. This relationship is corroborated in the empirical results discussed in Section 4.

### 3 Error Analysis for Fisher’s Classifier

We consider the two-class case for Fisher’s classifier – for the general case see [3, 13]. Suppose then that we deal with two classes,  $\omega_1$  and  $\omega_2$ , which are represented by two normally distributed  $d$ -dimensional random vectors,  $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and  $\mathbf{x}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  respectively, where  $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ , and whose *a priori* probabilities are  $P(\omega_1)$  and  $P(\omega_2)$  respectively. The solution for the vector  $\mathbf{w}$  that maximizes the *class separability* is given by:

$$\mathbf{w} = \mathbf{S}_W^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \tag{12}$$

where  $\mathbf{S}_W = \frac{1}{2}(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$ .

To complete the linear classifier, the threshold  $w_0$  has to be obtained. A simple approach (suggested in [12]) is to assume that the distributions in the original space have identical covariance matrices, and take the independent term of the optimal quadratic or Bayesian classifier, which results in:

$$w_0 = -\frac{1}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t \mathbf{S}_W^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - \log \frac{P(\omega_1)}{P(\omega_2)}. \tag{13}$$

Once we have derived the corresponding linear classifier by means of vector  $\mathbf{w}$  and the corresponding threshold, we obtain the boundaries for the integrals in (3). The algebraic expression for the error is stated in the following theorem, whose proof can be found in [9]. A special condition is shown in the corollary given thereafter.

**Theorem 2.** *Let  $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  and  $\mathbf{x}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  be two normally distributed random vectors representing two classes,  $\omega_1$  and  $\omega_2$ , whose *a priori* probabilities are  $P(\omega_1)$  and  $P(\omega_2)$  respectively, and  $g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$  be Fisher’s classifier; where  $\mathbf{w}$  and  $w_0$  are obtained as in (12) and (13) respectively.*

If  $r^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) > \max \left\{ \log \frac{P(\omega_1)}{P(\omega_2)}, \log \frac{P(\omega_2)}{P(\omega_1)} \right\}$ , then:

$$Pr[\text{error}(F)] = P(\omega_1) \int_{-\infty}^{b_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + P(\omega_2) \int_{-\infty}^{b_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad (14)$$

where  $x$  is an  $N(0, 1)$  random variable, and

$$b_1 = \frac{1}{2} \left[ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} \boldsymbol{\Sigma}_1 (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \right]^{-\frac{1}{2}} \left[ -(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - \log \frac{P(\omega_1)}{P(\omega_2)} \right], \quad (15)$$

$$b_2 = -\frac{1}{2} \left[ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} \boldsymbol{\Sigma}_2 (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \right]^{-\frac{1}{2}} \left[ (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^t (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - \log \frac{P(\omega_1)}{P(\omega_2)} \right]. \quad (16)$$

□

**Corollary 1.** Under the conditions of Theorem 2, and assuming that  $P(\omega_1)=P(\omega_2)=0.5$ , the probability of error of Fisher’s classifier can always be computed as in (14). □

The algebraic expression for the error given in Theorem 2 shows that the classification error for Fisher’s classifier can be derived directly from the parameters of the distributions, i.e. without finding the corresponding classifier. It is important to note, however, that the threshold in Fisher’s classifier, and in general, for any classifier, can be obtained in other ways too (see [10]). Also, using the algebraic analysis of the probability of error discussed in the previous subsection, bounds and approximations for the error for Fisher’s classifier can be found (see [9]).

### 4 Empirical Analysis

To test the accuracy and computational efficiency of the error analysis discussed in this paper, we performed a few simulations on synthetic data, which involve two normally distributed classes and a linear classifier. The two classes,  $\omega_1$  and  $\omega_2$ , are then fully specified by their parameters,  $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1$  and  $\boldsymbol{\Sigma}_2$ . We assume that the *a priori* probabilities for the two classes are equal, i.e.  $P(\omega_1) = P(\omega_2) = 0.5$ .

One of the tests involves the analysis of the *actual* classification error, as well as the bounds and the approximation introduced in this paper. To conduct the test, we generated random parameters for  $d$ -dimensional classes, where  $d = 20, 40, \dots, 200$ . The linear classifier used to test our method is the traditional Fisher’s classifier, where the threshold is obtained as in (13). The mean vectors for the two classes were generated randomly from a uniform distribution specified by the intervals  $[0, 0.2]$  for  $\boldsymbol{\mu}_1$ , and  $[0.2, 0.4]$  for  $\boldsymbol{\mu}_2$ . The covariance matrices were generated by invoking the *random correlation* method for generating *positive semidefinite* matrices [2]. The empirical results

**Table 1.** Comparison of the actual value, bounds and approximation for the error.

Dim.	$a_1$	$a_2$	Pr[error]	Lower bnd.	Upper bnd.	Approx.	Difference
20	-0.42678	-0.48947	0.32351	0.28622	0.36949	0.32785	0.00434
40	-0.81392	-0.78878	0.21148	0.19581	0.23851	0.21716	0.00567
60	-1.07009	-0.88201	0.16559	0.15554	0.18523	0.17039	0.00480
80	-1.14236	-1.04401	0.13745	0.13039	0.15274	0.14157	0.00412
100	-1.39993	-1.48340	0.07488	0.07240	0.08175	0.07707	0.00220
120	-1.97384	-1.89796	0.02653	0.02604	0.02836	0.02720	0.00068
140	-2.12477	-2.02669	0.01908	0.01878	0.02030	0.01954	0.00046
160	-2.34827	-2.17967	0.01204	0.01188	0.01274	0.01231	0.00027
180	-3.04408	-2.50596	0.00364	0.00360	0.00381	0.00371	0.00007
200	-3.46395	-2.99247	0.00083	0.00082	0.00086	0.00084	0.00001

obtained from our simulations are shown in Table 1. The second and third columns contain the boundaries for the two integrals, computed as in (4) and (5) respectively. The fourth column corresponds to the probability of error for Fisher's classifier, obtained as in (3), where the integrals were computed numerically by invoking the near-minimax Chebyshev approximations for the error function [1]. The fifth and sixth columns contain the lower and upper bounds for the classification error, which were computed as in (7). The seventh column represents the approximation of the classification error, obtained as in (8), and the last column corresponds to the difference (in absolute value) between the *actual* probability of error, the fourth column, and the *approximation* of the error, the seventh column. The results from the table show that the lower and upper bounds for the error are quite *loose* for large values of the classification error. Conversely, they are very *tight* for small values of the classification error. This is observed in the last row for dimension 200, in which the bounds are found to be very close to each other. A similar behavior is observed when analyzing the difference between the actual error and the approximation. The approximation differs from the actual error in nearly  $10^{-5}$  for dimensions 180 and 200. Observe also, that in all cases, even in the case of dimension 20 in which the error is large, the approximation of the error differs from the actual value in less than  $\varepsilon = 0.0184340683$ , and hence achieving at least *two digits* of precision.

To experimentally analyze the computational efficiency for computing the classification error, we conducted simulations on test suites involving dimensions  $d = 50, 100, \dots, 500$ . The results obtained are depicted in Table 2. The second and third columns contain the average for the probability of error, and the average of the difference between the actual value and the approximation. The fourth column contains the CPU time (in seconds) for computing the actual probability of error, the bounds, and the approximation for each of the experiments. The methods were tested using Matlab in an Intel 2.0Ghz workstation running Windows XP.

The results in the table show that our computational method is extremely fast for dimension 50, performing the computations in less than *a tenth* of a second. As the dimension of the feature space increases, the running times also increase. We also observe that the average differences between the actual error and the approximation are very small, e.g. below  $\varepsilon = 0.0184340683$ .

**Table 2.** Results for the running times and probabilities of error for different simulations on normally distributed classes whose dimensions range from 20 to 200.

Dim.	Avg. error	Avg. Diff.	Time
50	0.195821	0.005383	0.052680
100	0.114588	0.003450	0.157020
150	0.068887	0.002006	0.622500
200	0.044076	0.001213	1.411930
250	0.027769	0.000714	2.664730
300	0.017695	0.000424	5.195770
350	0.012043	0.000271	8.302140
400	0.008192	0.000174	13.113550
450	0.005368	0.000108	18.286700
500	0.003665	0.000070	26.335270

## 5 Conclusions

In this paper, we derive lower and upper bounds, and an expression that approximates the probability of error, which are obtained from the parameters of the distributions. The result can be used for *any* linear classifier, even though the underlying distributions are not normal. We have shown that the approximations differ from the *actual* error in at most  $\varepsilon = 0.0184340683$ .

Our empirical results on synthetic, *higher*-dimensional data, show that the bounds are very tight for small values of the classification error. Also, the approximation has been empirically shown to be very precise in the estimation of the error, which is approximated by a factor of at most  $\varepsilon = 0.0184340683$ . It has been shown to work efficiently for spaces involving up to *five hundred* features.

Many directions for future work exist, including the generalization of this model for more than two classes, the generalization of the model for piecewise linear classifiers that deals with more than one hyperplane. The latter is a difficult problem, since the linear transformations have to be applied simultaneously for all hyperplanes, leading a yet multivariate integration problem. Another problem that we are investigating is that of comparing the classification error for linear classifiers that use different thresholding methods.

## Acknowledgments

The author's research is partially supported by NSERC, the Natural Sciences and Engineering Council of Canada.

## References

1. W. Cody. A Portable FORTRAN Package of Special Function Routines and Test Drivers. *ACM Transactions on Mathematical Software*, 19:22–32, 1993.
2. P. Davies and N. Higham. Numerically Stable Generation of Correlation Matrices and their Factors. Technical Report 354, Manchester, England, 1999.
3. R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley and Sons, Inc., New York, NY, 2nd edition, 2000.
4. K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 1990.
5. R. Herbrich and T. Graepel. A PAC-Bayesian Margin Bound for Linear Classifiers. *IEEE Transactions on Information Theory*, 48(12):3140–3150, 2002.
6. M. Kendall and A. Stuart. *Kendall's Advanced Theory of Statistics, Volume I: Distribution Theory*. Edward Arnold, sixth edition, 1998.
7. C. Lee and E. Choi. Bayes Error Evaluation of the Gaussian ML Classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1471–1475, May 2000.
8. S. Raudys and R. Duin. Expected Classification Error of the Fisher Linear Classifier with Pseudo-inverse Covariance Matrix. *Pattern Recognition Letters*, 19:385–392, 1999.
9. L. Rueda. A One-dimensional Analysis for the Probability of Error of Linear Classifiers for Normally Distributed Classes. *Submitted for Publication*, 2004. Electronically available at <http://davinci.newcs.uwindsor.ca/~lrueda/papers/ErrorEstJnl.pdf>.
10. L. Rueda. An Efficient Approach to Compute the Threshold for Multi-dimensional Linear Classifiers. *Pattern Recognition*, 37(4):811–826, April 2004.
11. N. Vaswani. A Linear Classifier for Gaussian Class Conditional Distributions with Unequal Covariance Matrices. In *Proceedings of the 16th International Conference on Pattern Recognition*, volume 2, pages 60–63, Quebec, Canada, 2002.
12. A. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, N.York, second edition, 2002.
13. Y. Xu, Y. Yang, and Z. Jin. A Novel Method for Fisher Discriminant Analysis. *Pattern Recognition*, 37(2):381–384, February 2004.