

EFFECT OF ESTIMATION AND DIMENSIONALITY
ON THE PERFORMANCE OF TWO-CLASS GAUSSIAN CLASSIFIERS

A Thesis

Submitted to the Faculty of Graduate Studies and Research

in Partial Fulfilment of the Requirement for

the Degree of

Doctor of Philosophy

in the Department of Electrical Engineering

University of Saskatchewan

by

Talaat Salem El-Sheikh

Saskatoon, Saskatchewan

April, 1980

The author claims copyright. Use shall not be made of
the material contained herein without proper acknowledgement,
as indicated on the following page.

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

In the name of Allah, the Benificent, the Merciful

The author has agreed that the Library, University of Saskatchewan, shall make this thesis freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this thesis for scholarly purposes may be granted by the professor or professors who supervised the thesis work recorded herein or, in their absence, by the Head of the Department or the Dean of the College in which the thesis work was done. It is understood that due recognition will be given to the author of this thesis and to the University of Saskatchewan in any use of material in this thesis. Copying or publication or any other use of the thesis for financial gain without approval by the University of Saskatchewan and the author's written permission is prohibited.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Electrical Engineering

University of Saskatchewan

Saskatoon, Canada

ACKNOWLEDGEMENTS

The author would like to express his appreciation to his supervisor Dr. A. G. Wacker for his guidance, assistance and encouragement both in the course of the work and during the preparation of this thesis. The author also wishes to thank the other members of his Advisory Committee for their valuable suggestions and discussions.

Financial assistance was gratefully received from the University of Saskatchewan in the form of a scholarship and from the Department of Electrical Engineering in the form of computer time. Special thanks are due the Department of Electrical Engineering, Cairo University for kindly offering a study leave to the author during his research work.

UNIVERSITY of SASKATCHEWAN

Electrical Engineering Abstract 80A206

EFFECT OF ESTIMATION AND DIMENSIONALITY ON
THE PERFORMANCE OF TWO-CLASS GAUSSIAN CLASSIFIERS

Student: Talaat Salem El-Sheikh

Supervisor: Dr. A. G. Wacker

Ph. D. Thesis Presented to the College of Graduate Studies

April 1980

ABSTRACT

In this thesis the phenomenon observed in "real life", where the probability of correct classification P_{cr} (for classifiers based on estimated distributions) peaks with the dimensionality N , is extensively investigated for two-class multivariate Gaussian classifiers. The case of equal prior class probabilities is the only case considered. Regarding the parameters of the class distributions, two main situations are considered. For the first situation it is assumed that the two class distributions have a common covariance matrix and that this information is known a priori. For the other and more general case, the parameters of the two class distributions are assumed to be different (i.e. the two covariance matrices are unequal and the two mean vectors are unequal). On the other hand, three interesting cases with different degrees of knowledge of the parameters are considered. In the first case all the parameters of the class distributions are assumed known. For the

second case the covariance matrices are assumed known, while all the parameters are assumed unknown in the last and most general case. For the two latter cases, it is assumed that a number of design vectors K are available from each class to be used for estimating the unknown parameters. The classification rule considered is a suboptimal Bayes (minimum error for known distributions) rule in which the unbiased sample mean vectors and unbiased sample covariance matrices are used in place of the true parameters, generally resulting in hyperquadratic decision boundaries.

For the known common covariance matrix case, a basic question investigated is the variation (with N) in the Mahalanobis distance (between the true distributions) required to keep P_{cr} constant. Numerical results are plotted for several cases. Analytical results are also obtained which relate the rate of variation of the Mahalanobis distance with N and the corresponding asymptotic behaviour of P_{cr} . Results for more highly structured problems, involving specific covariance matrices, show that in some cases increasing correlation between the measurements yields higher values of P_{cr} .

For the case of a common covariance matrix, the average probability of correct classification \bar{P}_{cr} is calculated over two different collections of Gaussian problems. The relationship between \bar{P}_{cr} and N is established for these collections for different degrees of knowledge of the parameters.

For the general case of unequal covariance matrices (i.e. hyperquadratic decision boundaries), a technique is developed in order

to numerically calculate the conditional probability of correct classification for a particular problem given a particular design sample (i.e. particular estimated parameters). In order to calculate P_{cr} , this conditional probability is averaged over a sufficient number of design samples. The average probability of correct classification \bar{P}_{cr} , for this case, is obtained by simulation over four different collections of Gaussian problems. The relationship between \bar{P}_{cr} and N is established for both the case of completely known and the case of completely unknown parameters.

For the case of completely unknown parameters, peaking in the P_{cr} vs N relationship is always encountered. The value of the optimal dimensionality N_{opt} is usually small. Depending on the specific relation that exists between the consecutive features, N_{opt} varies for the cases considered between $K/5$ and $K/2$. It is noted that if dimensionality is increased beyond about $K/5$, in general peaking or at least saturation occurs. On the other hand, if N is kept below $K/10$, the performance is fairly close to the case of known parameters.

TABLE OF CONTENTS

	page
Copyright	ii
Acknowledgements	iii
Abstract	iv
Table of Contents	vii
List of Figures	ix
List of Tables	xii
Principal Symbols	xiii
1. STATISTICAL PATTERN CLASSIFICATION	1
1.1 General	1
1.2 Peaking Phenomenon	6
1.3 Thesis Outline	10
2. PROBLEM DEFINITION AND GENERAL APPROACH	11
2.1 Motivation	11
2.2 Statistical Model	12
2.2.1 Basic assumptions	12
2.2.2 Performance measure of a classification rule	13
2.2.3 Bayes decision rule	13
2.2.4 Parameter estimation	15
2.2.5 Collection of Problems	20
2.3 General Objective	24
2.4 Summary of Work	24

Table of Contents (con'd):	page
3. PROBABILITY OF CORRECT CLASSIFICATION FOR THE CASE OF A COMMON COVARIANCE MATRIX	27
3.1 Known Parameters	28
3.2 Known Common Covariance Matrix	29
3.2.1 Probability of correct classification	29
3.2.2 Necessary variation of δ_N^2 with N to keep P_{cr} constant.	33
3.2.3 Peaking phenomenon	37
3.2.4 Effect of covariance matrix structure	45
3.3 Unknown Parameters	54
4. AVERAGE CLASSIFICATION ACCURACY OVER COLLECTIONS OF PROBLEMS FOR COMMON COVARIANCE MATRIX CASE	58
4.1 Collection of Problems	58
4.2 Infinite Sample Size	64
4.3 Finite Sample Size	71
5. AVERAGE CLASSIFICATION ACCURACY OVER COLLECTIONS OF PROBLEMS FOR THE CASE OF UNEQUAL PARAMETERS	81
5.1 Calculation of the Conditional Probability of Correct Classification	81
5.2 Collection of Problems	88
5.3 Average Probability of Correct Classification	96
6. CONCLUDING REMARKS AND FUTURE WORK	109
7. REFERENCES	116

LIST OF FIGURES

Fig.		Page
1.1	A two-dimensional pattern space with samples: . samples of class C_1 ; x samples of class C_2 ; o unknown pattern	4
3.1	Relationship between normalized Mahalanobis distance and dimensionality for fixed P_{cr} and for (a) 1 design vector/class, (b) 4 design vectors/class and (c) 10 design vectors/class	35
3.2	Relationship between optimal dimensionality and R (where $\delta_N^2 = \delta_1^2 N^{1/R}$) for (a) 1 design vector/class, (b) 4 design vectors/class and (c) 10 design vectors/class	39
3.3	Relationship between optimal probability of correct classification and R (where $\delta_N^2 = \delta_1^2 N^{1/R}$) for (a) 1 design vector class, (b) 4 design vectors/class and (c) 10 design vectors/class	40
3.4	Relationship between probability of correct classification and dimensionality for 1 design vector/class and unit univariate Mahalanobis distance and $R = 2.1$	41
3.5	Relationship between (a) optimal dimensionality, (b) optimal probability of correct classification, and $(1-\beta)$ where $\delta_{N+1}^2 - \delta_N^2 = \beta^2 (\delta_N^2 - \delta_{N-1}^2)$	44
3.6	Relationship between Mahalanobis distance and dimensionality for different orientations of the vector connecting the two mean vectors	50
3.7	Relationship between optimal dimensionality and correlation coefficient for the equal-variance-equal-covariance matrix	51
4.1	Relationship between average (over possible mean vectors) probability of correct classification and dimensionality for the case of known common covariance matrix and for different number of design vectors	69
4.2	Relationship between average (over possible mean vectors and covariance matrices) probability of correct classification and dimensionality for the case of known common covariance matrix and for different number of design vectors	70

List of Figures (con'd)

page

- 4.3 Relationship between average squared Mahalanobis distance and dimensionality for fixed average (over possible mean vectors) probability of correct classification and for the case of known common covariance matrix 74
- 4.4 Relationship between average squared Mahalanobis distance and dimensionality for fixed average (over possible mean vectors) probability of correct classification and for the case of unknown parameters 75
- 4.5 Relationship between average (over possible mean vectors) probability of correct classification and dimensionality for the case of unknown parameters 76
- 4.6 Relationship between optimal dimensionality and β for the case of known common covariance matrix where β is the ratio between contributions of successive features to the average squared Mahalanobis distance 78
- 4.7 Relationship between optimal dimensionality and β for the case of unknown parameters where β is the ratio between contributions of successive features to the average squared Mahalanobis distance 79
- 4.8 Relationship between average (over possible mean vectors and covariance matrices) probability of correct classification and dimensionality for the case of unknown parameters and for different number of design vectors 80
- 5.1 Relationship between average probability of correct classification and dimensionality for the case of identical features (i.e. collection Ω_1) 99
- 5.2 Relationship between average probability of correct classification and dimensionality for the case of identically distributed features (i.e. collection Ω_2) 100
- 5.3 Relationship between average probability of correct classification and dimensionality for the case of proportional features (i.e. collection Ω_3) 101
- 5.4 Relationship between average probability of correct classification and dimensionality for the case of proportionally distributed features (i.e. collection Ω_4) 102

- 5.5 Histogram indicating the distribution of the probability of correct classification for the case of identical features (i.e. collection Ω_1) and for (a) 10 design vectors/class, (b) 20 design vectors/class (c) infinite design vectors/class 105
- 5.6 Histogram indicating the distribution of the probability of correct classification for the case of identically distributed features (i.e. collection Ω_2) and for (a) 10 design vectors/class, (b) 20 design vectors/class (c) infinite design vectors/class 106
- 5.7 Histogram indicating the distribution of the probability of correct classification for the case of proportional features (i.e. collection Ω_3) and for (a) 10 design vectors/class (b) 40 design vectors/class (c) infinite design vectors/class 107
- 5.8 Histogram indicating the distribution of the probability of correct classification for the case of proportionality distributed features (i.e. collection Ω_4) and for (a) 10 design vectors/class (b) 40 design vectors/class (c) infinite design vectors/class 108

LIST OF TABLES

Table		page
3.1	Variation of $\alpha(N)$ with N (for $\alpha(N)$ see Equation (3.33))	49
5.1	Number of samples and number of problems for the different collections of problems	97

LIST OF PRINCIPAL SYMBOLS

B	-	Eigenvector matrix
C	-	Class
K	-	Design sample size
M	-	Mean vector
\hat{M}	-	Estimated mean vector
N	-	Dimensionality
N_{opt}	-	Optimum dimensionality
P_{cr}	-	Probability of correct classification
\hat{P}_{cr}	-	Estimated probability of correct classification
\bar{P}_{cr}	-	Average probability of correct classification
X	-	Observation vector
Σ	-	Covariance matrix
$\hat{\Sigma}$	-	Estimated covariance matrix
ρ	-	Correlation Coefficient
λ	-	Eigenvalue
Λ	-	Eigenvalue matrix
μ	-	Transformed difference-of-means vector
Ω	-	Parameter space
δ_N	-	Mahalanobis distance
ϕ	-	Characteristic function