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

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

In the name of Allah, the Beneficent, the Merciful

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

Table of Contents (con'd):

LIST OF FIGURES

List of Figures (con'd) page

List of Figures (con'd) page

LIST OF TABLES

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