NPS-EC-01-005

NAVAL POSTGRADUATE SCHOOL
Monterey, California

INVESTIGATION OF FEATURE DIMENSION REDUCTION SCHEMES FOR CLASSIFICATION APPLICATIONS

by

M. P. Fargues

June 1, 2001

Prepared for: Center for Reconnaissance Research, Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
Monterey, California

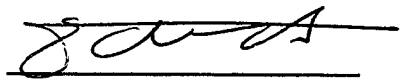
RADM D. Ellison
Superintendent

R. Elster
Provost

This report was sponsored by the Naval Postgraduate School Center for Reconnaissance Research.

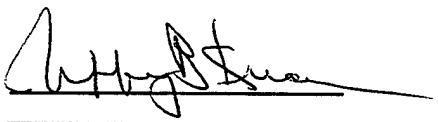
Approved for public release; distribution is unlimited.

The report was prepared by:



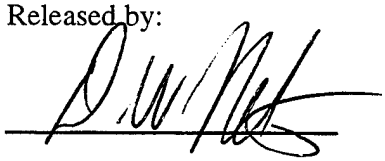
MONIQUE P. FARGUES
Department of Electrical and
Computer Engineering

Reviewed by:



JEFFREY B. KNORR
Chairman
Department of Electrical and
Computer Engineering

Released by:



DAVID W. NETZER
Associate Provost and
Dean of Research

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1, 2001	Final Report, Sep 1999-Dec2000

4. TITLE AND SUBTITLE
Investigation of Feature Dimension Reduction Schemes for Classification Applications

5. FUNDING NUMBERS
MIPR#A448195

6. AUTHOR(S)
M.P. Fargues

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Electrical and Computer Engineering
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-EC-01-005

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Center for Reconnaissance Research
Naval Postgraduate School
Monterey, CA 93943

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this report are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (Maximum 200 words)

Extracting relevant features that allow for class discrimination is the first critical step in classification applications. However, this step often leads to high-dimensional feature spaces, which requires large datasets to create viable classification schemes. As a result, there is a strong incentive to reduce the feature space dimension. Two classical types of approaches to reduce feature dimension exist: Principal Component Analysis (PCA)-based or discriminant-based approaches. The main difference between the two types lies in the criterion selected; PCA-based schemes seek a projection direction which bests represents the data in a norm sense, while discriminant-based schemes seek a projection that best separates the class data.

This study presents a comparison of three discriminant-based feature dimension reduction schemes: the Mean Separator Neural Network (MSNN), the Mahalanobis-based Dimension Reduction scheme (MBDR), and the kernel-based Generalized Discriminant Analysis (GDA) approach. PCA is included for comparison purposes as it is also widely used in classification applications. All four feature dimension reduction schemes are implemented and evaluated by applying the transformed features to a basic minimum distance classifier. Three classification datasets commonly used in statistics for benchmarking purposes are selected to compare the schemes and results discussed. Results show the kernel-based generalized discriminant analysis approach to lead to consistently higher classification performances than the other schemes considered in the study for the data investigated.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Feature dimension, classification, kernel-based scheme	33
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

Classification applications require the extraction of class discriminative information. However, this step often leads to high-dimensional feature spaces, which requires large datasets to create viable classification schemes. This study presents follow-on work to those of Duzenli [DUZ98] and San Pedro [SAN00], and considers two discriminant-based feature dimension reduction schemes for classification applications. The two feature reduction schemes considered are the Mahalanobis-based dimension reduction (MBDR) scheme recently proposed by Brunzell [BRE99], and the kernel-based generalized discriminant analysis approach (GDA) proposed by Baudat & Anouar [BAN00]. The GDA is part of a new breed of kernel-based algorithms that are currently being considered by the research community to develop new learning techniques, as they can be used to derive nonlinear generalizations of currently known algorithms. Finally, the classical PCA and the MSNN proposed earlier in [DUZ98] are included in this study for comparison purposes.

The four feature dimension reduction schemes considered were implemented in MATLAB and evaluated by applying the transformed features to a basic minimum distance classifier. Performances are evaluated by applying these schemes to three datasets commonly used in statistics for benchmarking purposes. Results show overall best results to be obtained for the GDA for the datasets considered. Results also show there is no consistent second best feature reduction scheme among the MSNN, the MBDR, and the PCA, as performances for these three schemes are data dependent.

I. INTRODUCTION

A. BACKGROUND

This work presented in this report is part of a larger scale study conducted during 1999 and 2000 where we investigated various feature extraction and dimension reduction schemes, and their application to the classification of digital modulation types. The overall study was divided in three separate phases.

- The first phase of the overall study investigated extensions to the MSNN approach originally derived in 1998 [DUZ98] to include variance information in the optimization criterion. Results obtained with synthetic data and basic communication schemes were presented in San Pedro [SAN00]. Results showed no significant improvements over the original MSNN for the data investigated.
- The second phase of the overall study, which this report specifically focuses on, investigated two new feature dimension reduction schemes and their resulting performances on benchmarking datasets.
- The third phase of the overall study investigated the application of a selected few higher-order statistic parameters to the classification of digital modulation schemes of types [2,4,8]-PSK, [2,4,8]-FSK, and [16,64,256]-QAM in low SNR levels and multipath propagation channel environments. A hierarchical tree-based classifier was proposed and its performances studied over various types of propagation channels [FAH01, HAT01].

B. OBJECTIVES

Extracting relevant features that allow for class discrimination is the first critical step in classification applications. However, this step often leads to high-dimensional feature spaces,

which requires large (and potentially not available) datasets to create viable classification schemes. In addition, some of the features may carry little useful information or be correlated with others resulting in redundancies in the feature space. As a result, there is a strong incentive to reduce the feature space dimension. Two classical types of approaches to reduce feature dimension exist: Principal Component Analysis (PCA)-based or discriminant-based approaches. The main difference between the two types lies in the criterion selected; PCA-based schemes seek a projection direction which bests represents the data in a norm sense, while discriminant-based schemes seek a projection that best separates the class data [DHS01]. We proposed in earlier work a simple discriminant-based feature dimension reduction scheme called the Mean Separator Neural Network (MSNN). The MSNN belongs to the class of projection pursuit algorithms, where the goal is to find a projection direction that emphasizes class discrimination [BIS95]. Results showed the MSNN scheme to have very good performances for the underwater data considered during this earlier study [DUZ98, DFA98, FDU98]. The MSNN approach can be viewed as a one-layer neural network (NN) implementation where the goal is to find the projection index, i.e., the weight vector, which maximizes the absolute difference between the means of the projected class data. As a result, it suffers of the same drawback as that present in numerous other NN implementations: the iterative procedure is not insured to converge to the global minimum due to the nonlinear activation function present in the optimization criterion. While the "local minima" issue was shown not to be a problem for the data investigated in our earlier study, it motivated this follow-on work where we investigate two alternate discriminant-based dimension reduction schemes which do not exhibit such a behavior.

The two feature reduction schemes considered are the Mahalanobis-based dimension reduction (MBDR) recently proposed by Brunzell [BRU97], and the kernel-based generalized

discriminant analysis (GDA) proposed by Baudat & Anouar [BAN00]. In addition, we benchmark these two schemes against the classical PCA approach, and the MSNN scheme.

Chapter II briefly reviews the PCA approach, as applied to classification applications. Chapters III and IV present the Mahalanobis-based dimension reduction approach and kernel-based generalized discriminant schemes respectively. The basic MSNN scheme is described in Chapter V. The four feature dimension reduction schemes considered in this study are implemented in MATLAB and evaluated by applying the transformed features to a basic minimum distance classifier. Three classification datasets commonly used in statistics for benchmarking purposes are selected to compare the schemes and results discussed in Chapter VI. Finally, Chapter VII presents conclusions.

II. PRINCIPAL COMPONENT ANALYSIS

A. INTRODUCTION

Principal Component Analysis (PCA) is one possible approach to reduce the dimensionality of the class features under consideration. The method projects high-dimensional data vectors onto a lower dimensional space by using a projection which bests represents the data in a mean square sense, i.e., leads to projected data vectors which preserve most of the energy contained in the original data [DHS01, BIS95]. This linear dimension reduction scheme uses the Karhunen-Loeve transformation which represents a given data vector as a linear combination of the eigenvectors obtained from the data covariance matrix. As a result, lower dimensional data vectors may be obtained by projecting the high-dimensional data vectors onto a user-specified number of eigenvectors associated with the largest eigenvalues of the data covariance matrix. PCA is widely used in engineering applications such as for example in compression as it preserves most of the original overall data information, and in statistics where it can be applied to decorrelate data prior to processing, etc.... However, the PCA projection criterion is not necessarily well designed for classification applications where the goal is to best discriminate between classes, not preserve most of the energy in a lower dimensional class feature space. Nevertheless it is a classical tool applied extensively, and we will use it in our comparison of the various dimension reduction schemes considered in this study.

B. DESCRIPTION

The PCA maps an ensemble of P N-dimensional vectors $X=[\underline{x}_1,....,\underline{x}_P]$ onto an ensemble of P D-dimensional vectors $Y=[\underline{y}_1, ...,\underline{y}_P]$ where D<N using a linear transformation which can be represented by the rectangular matrix A so that:

$$\underline{y}_i = A^H \underline{x}_i, \quad i = 1,\ldots,P, \tag{2.1}$$

where A has orthogonal column vectors. For PCA, the matrix A is selected as the P*D matrix containing the D eigenvectors associated with the larger eigenvalues of the data covariance matrix X^HX. With such a choice of transformation matrix A, the transformed data vectors y_i have uncorrelated components as:

$$YY^H = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_D \end{bmatrix},$$

where λ_i, $i=1,\ldots D$, are the eigenvectors of the data covariance matrix X^HX. The concept of PCA is illustrated next by considering three classes of two-dimensional data, as shown in Figures II-1 and II-2, where the data dimension is to be reduced to one. The transformation matrix A is of dimension 2*1, and the projected data sets lie on a line. Figures II-1 & II-2 show that the PCA projection direction preserves most of the signal energy but also generates projected data with significant amount of overlap between two of the projected class data.

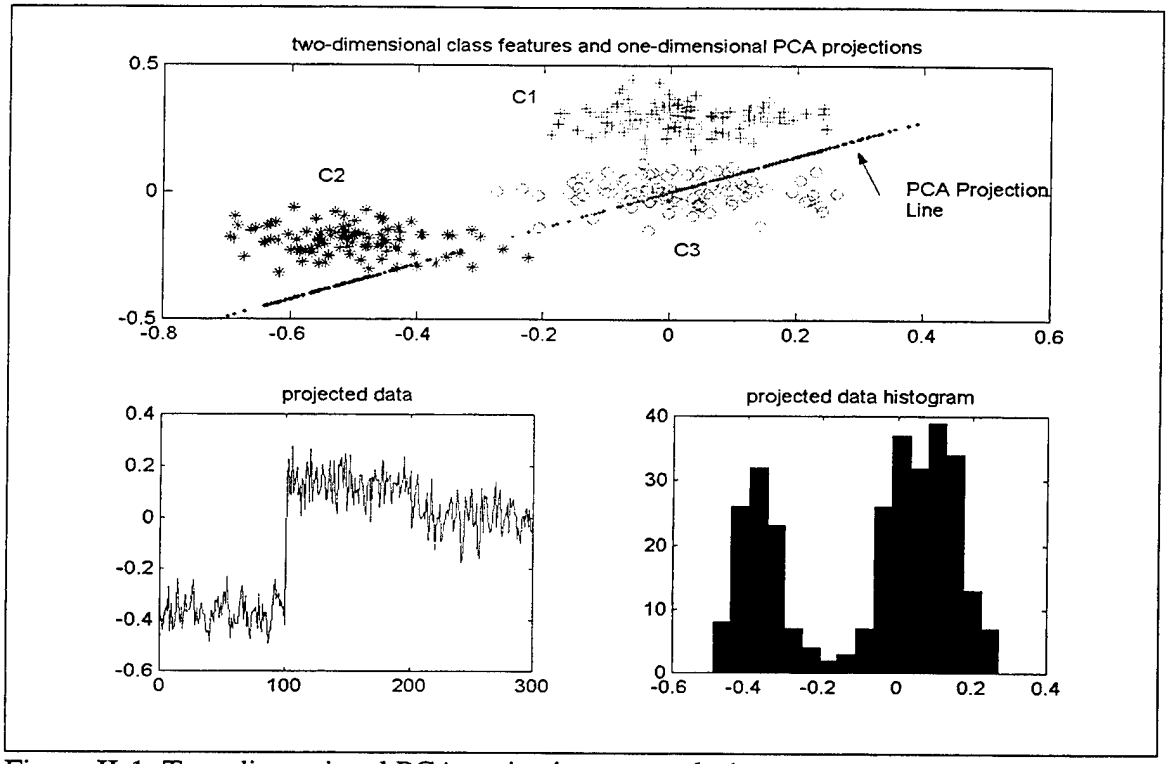


Figure II-1. Two-dimensional PCA projection, example 1.

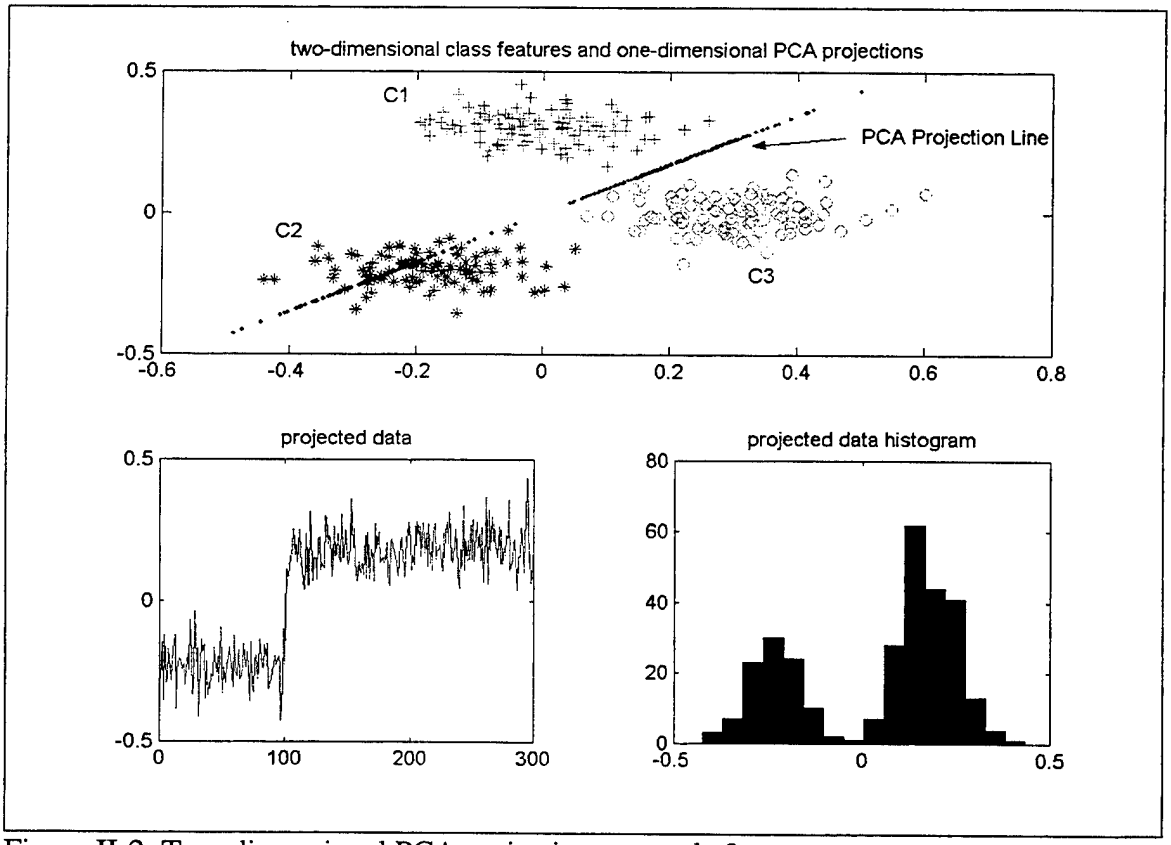


Figure II-2. Two-dimensional PCA projection, example 2.

III. MAHALANOBIS-BASED DIMENSION REDUCTION

A. INTRODUCTION

As shown earlier, PCA may not be well suited to reduce feature dimension in classification applications where the main goal is to preserve class discrimination. Fisher's linear discriminant (LDA) introduced by Fisher is better suited as it seeks a projection direction which best discriminates between the classes considered [FUK90, DHS01]. Fisher's discriminant was initially derived for the two-class problem and extended later to the more than two-class problem. The Fisher projection index for the 2-class problem is derived as the direction that maximizes the following ratio:

$$J(\underline{w}) = \frac{\underline{w}^t S_B \underline{w}}{\underline{w}^t S_w \underline{w}}. \tag{3.1}$$

The matrices S_B and S_W respectively represent the between-class and within-class scatter matrices defined as:

$$S_B = (\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T, \quad S_w = \sum_{i=1}^{2} \Sigma_i,$$

where $\underline{m}_1$ and $\underline{m}_2$ respectively represent class-specific means for classes C_1 and C_2 and $\Sigma_i, i = 1, \cdots, 2,$ are the class-specific data covariance matrices for the two classes under consideration. As a result the projection criterion aims at maximizing a ratio of the separation between projected class data and the projected class-specific data variance information, thereby preserving discrimination information between the two classes considered. It can be shown that the criterion function $J(\underline{w})$ may be maximized by finding the projection vector $\underline{w}$ which satisfies the following generalized eigenvalue problem [DHS01, FUK90]:

$$S_B \underline{w} = \lambda S_w \underline{w},$$

which leads to

$$\underline{w} = S_w^{-1}(\underline{m}_1 - \underline{m}_2). \tag{3.2}$$

The Fisher Linear Discriminant can be extended to a higher number of classes (called the Multiple Discriminant Analysis (MDA) approach) by generalizing between-class and within-class scatter matrices to the more than two classes problem [DHS01, FUK90].

The feature dimension reduction proposed by Brunzell [BRU97, BRE99] follows the same basic concept as that present in the MDA; that is to find a linear projection that preserves the separation between classes. However, Brunzell proposes to accomplish the task by defining a pairwise Mahalanobis class distance measure and stacking all possible pairwise Mahalanobis-based distances into a transformation matrix, so the name Mahalanobis-based Distance Reduction (MBDR) approach. The MBDR approach and the Fisher Linear discriminant are identical for the two-class problem and the difference between the two schemes lies in the generalization to the more-than-two-classes problem, where the MBDR scheme preserves the pairwise approach while the MDA does not. Brunzell showed that his proposed transformation preserves the separation between classes. Performance evaluations of the MBDR feature dimension reduction scheme were conducted by applying the proposed scheme to seven datasets widely used in classification benchmarking, where the data dimensions are reduced to two and classification performances obtained with a basic quadratic classifier computed. Brunzell showed that classification performances obtained using the MBDR scheme are as good or better than the basic and variants of the Fisher LDA approach on the benchmarking data sets considered. As a result, we will consider the MBDR approach and not the basic LDA implementation in our classifier performance comparisons.

B. DESCRIPTION

The Mahalanobis-Based Dimension Reduction (MBDR) transformation matrix proposed by Brunzell is defined as:

$$U = \left[C_{1,2}^{-1} m_{1,2}, \cdots, C_{i,j}^{-1} m_{i,j}, \cdots, C_{c-1,c}^{-1} m_{c-1,c} \right], \; for \; 1 \le i \le j \le c, \tag{3.3}$$

where $C_{i,j}$ are pairwise covariance matrices defined as $C_{i,j} = \Sigma_i + \Sigma_j$, $m_{i,j}$ are pairwise class-specific mean vector differences defined as $\underline{m}_{i,j} = \underline{m}_i - \underline{m}_j$, and c represents the total number of classes. The feature dimension reduction scheme is applied by computing the SVD of the matrix U and selecting as transformation matrix that which contains the first k singular vectors associated with the k largest singular values of U.

Applying the MBDR matrix to the data considered in Figures II-1 & II-2 leads to Figure II-3 & II-4. Results show the projection direction much better suited to preserve class discrimination than PCA is, as expected from similarities to the LDA approach.

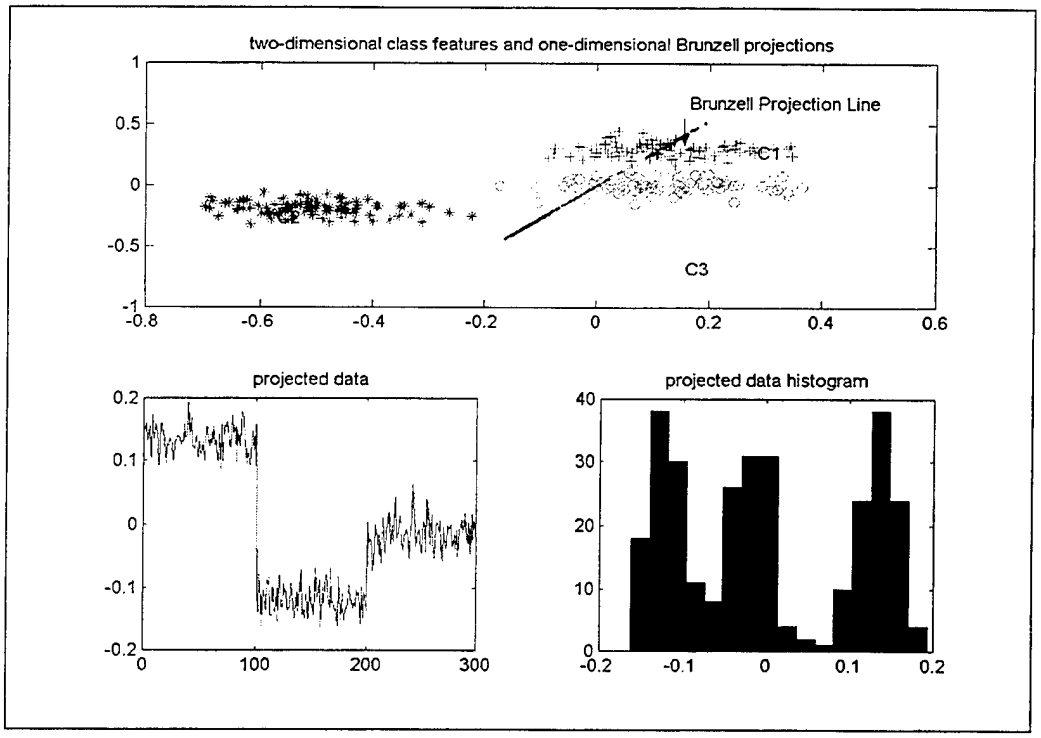


Figure II-3. Two-dimensional MBDR projection, example 1.

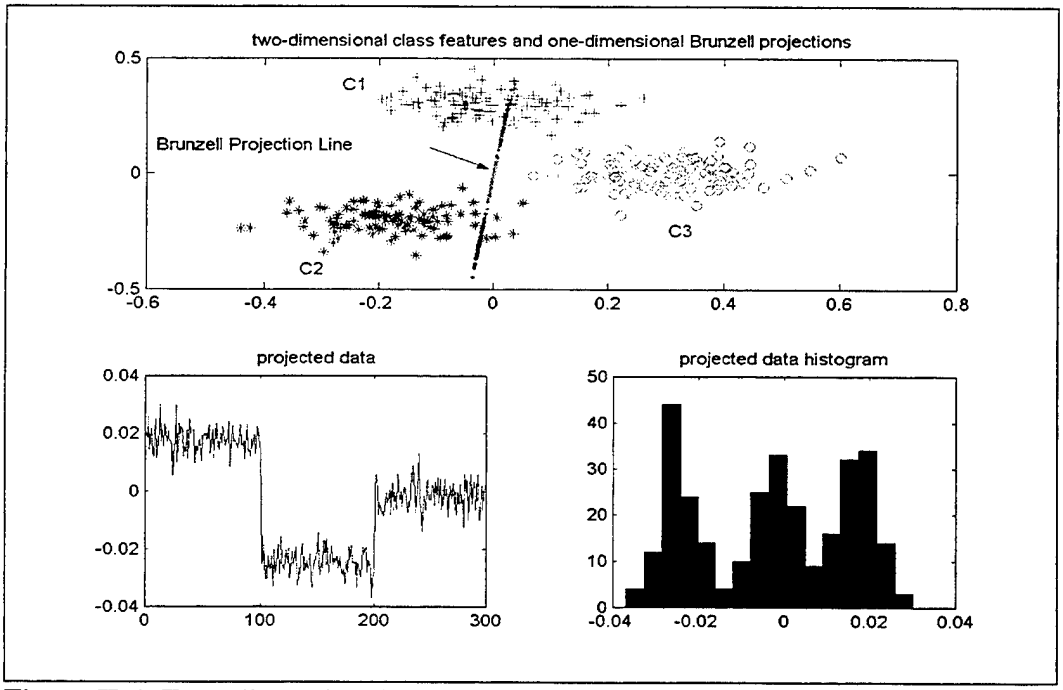


Figure II-4. Two-dimensional MBDR projection, example 2.

IV. KERNEL-BASED GENERALIZED DISCRIMINANT ANALYSIS

A. INTRODUCTION

LDA is a classical scheme well matched to classification applications as it preserves class discriminations. However, it may fail when the problem under consideration contains non-separable class information. A significant amount of research has been conducted recently in the area of kernel-based approaches to address non-separable class problems. The main idea behind kernel-based methods is to nonlinearly transform the input feature space into a higher-dimensional space in which the transformed features are separable. Nonlinear transformations are nothing new on themselves, however, most of the earlier ones involve computations in the transformed space for the resulting classification set-up. The main advantage behind the kernel-based generalized discriminant analysis approach is the fact that all computations may be carried out in the original space by expressing the nonlinear transformation in terms of dot products only. Such a reformulation of the problem leads to the computation of a class separating hyperplane with maximum margin without explicitly carrying the transformation of the features. It also leads to a nonlinear decision boundary in the original feature space. Such nonlinear transformations have been known for sometimes but not taken advantage of until Vapnick presented the support vector machines (SVM) approach [VAP95, CHS00]. Since then, several nonlinear generalizations of algorithms have been proposed; kernel-PCA [SSM99, SSM98, TRC01, MSS99], kernel-based denoising [MSS99], kernel-based LDA [MRW99a, MRW99b, BAN00], etc... Applications can be found in image processing [EPP00, CHV99], pattern recognition [GSO00, MAE99, HAE99], text categorization [TKO99], speech processing [NBR00], time series prediction [MSR97], radar imagery [LCB00], etc... and results have shown in some cases a significant improvement in classifier performances over more established

methods. Our study is restricted to the nonlinear generalization of LDA called the generalized discriminant analysis (GDA) only.

B. DESCRIPTION

The GDA is an extension of the LDA where the LDA criterion is defined in the transformed space. However, computations are carried out in the original feature space by reformulating the GDA criterion in terms of dot products of the nonlinear transformation operation. Recall that the basic LDA projection index is defined as the direction that maximizes the following ratio:

$$J(\underline{w}) = \frac{\underline{w}^t S_B \underline{w}}{\underline{w}^t S_W \underline{w}}, \tag{4.1}$$

where S_B and S_W respectively represent the between-class and within-class scatter matrices. Assume that we have N classes with n_l samples per class C_l, i.e., $\sum_{i=1}^{N} n_i = M$, where M is the total data sample size. Further, assume that $\underline{x}$ is nonlinearly transformed into a different space with a mapping ϕ:

$$\begin{aligned} \phi: \quad & X \rightarrow F \\ & \underline{x} \rightarrow \phi(\underline{x}). \end{aligned}$$

The covariance matrix of the transformed data $\phi(\underline{x})$ is given by:

$$V = \frac{1}{M} \sum_{i=1}^{M} \phi(\underline{x}_i)\phi^t(\underline{x}_i), \tag{4.2}$$

assuming the transformed data $\phi(\underline{x})$ is zero-mean. The data can be centered by following the procedure presented by Baudat & Anouar if it is not centered originally [BAN00, Appendix C]. Using class indices, (4.2) may be rewritten as:

$$V = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{n_i} \phi(\underline{x}_{ik})\phi^t(\underline{x}_{ik}), \tag{4.3}$$

where $\phi(\underline{x}_{ik})$ is the element k of class i.

Next, the covariance matrix of the class centers may be written as:

$$B = \frac{1}{M}\sum_{k=1}^{M} n_k \bar{\phi}_k \bar{\phi}_k^t, \tag{4.4}$$

where $\bar{\phi}_k$ represents the mean value for class C_k defined as:

$$\bar{\phi}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \phi(\underline{x}_{ki}),$$

with $\underline{x}_{ki}$ representing the i^{th} sample of class C_k. The key behind the GDA approach is to express the LDA criterion given by Eq. (4.1) in the transformed space as:

$$J(\underline{v}) = \frac{\underline{v}^t B \underline{v}}{\underline{v}^t V \underline{v}}, \tag{4.5}$$

where B and V are defined in Eqs. (4.3) & (4.4). The criterion is maximized when $\underline{v}$ is selected as the eigenvector associated with the maximum generalized eigenvalue associated with *(B,V)* [DHS01]. Note that the eigenvectors $\underline{v}$ may be written in terms of the elements in the transformed space F. Thus,

$$\underline{v} = \sum_{p=1}^{N}\sum_{q=1}^{n_p} \alpha_{pq}\phi(\underline{x}_{pq}). \tag{4.6}$$

Replacing $\underline{v}$ by its expansion given in Eq. (4.6) into Eq. (4.5), Baudat and Anouar show that the projection index *J(.)* may be rewritten as [BAN00]:

$$J(\underline{\alpha}) = \frac{\underline{\alpha}^t KWK\underline{\alpha}}{\underline{\alpha}^t KK\underline{\alpha}}. \tag{4.7}$$

The matrix K is of dimension M*M and is defined on the class elements by the blocks K_{pq} each of dimensions $n_p * n_p$. Each block matrix K_{pq} is composed of dot products in the transformed feature space F. Thus:

$$K = (K_{pq})_{\substack{p=1,\cdots,N \\ q=1,\cdots,N}}, \; with \; K_{pq} = (k_{ij})_{\substack{i=1,\cdots,n_p \\ j=1,\cdots,n_q}}, \tag{4.8}$$

where for given classes p and q the elements k_{ij} are defined in terms of dot products of the nonlinear transformation, i.e.,

$$(k_{ij})_{pq} = \phi^t(\underline{x}_{pi})\phi(\underline{x}_{qj}).$$

The matrix W is a block diagonal matrix of dimension M*M where each block W_l, $l=1,..,N$ is of dimension $n_l * n_l$ and defined as:

$$W_l = \frac{1}{n_l}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}, \; l = 1, \cdots N. \tag{4.9}$$

Baudat and Anouar show that the above generalized eigenvector problem may be simplified and reformulated as [BAN00]:

$$\lambda \underline{\beta} = P^t W P \underline{\beta}. \tag{4.10}$$

Therefore, the GDA problem becomes to find the eigenvector $\underline{\beta}$ defined in terms of the eigenvector $\underline{\alpha}$ as $\underline{\beta} = \Gamma P^t \underline{\alpha}$, where P and Γ are the eigenvector and eigenvalue matrices of K respectively. The eigenvector $\underline{\alpha}$ may be computed back from $\underline{\beta}$ by the transformation $\underline{\alpha} = \Gamma^{-1} P \underline{\beta}$. One of the potential drawbacks in the GDA is the computational load involved in computing the matrix inverse Γ^{-1}, as Γ is of dimension M*M, where M is the dataset size. However, computationally efficient alternatives have been reported in [MMR01, LRO01, KMW]. Our implementation computes the inverse Γ^{-1} with a reduced-rank pseudo-inverse to avoid ill-conditioning problems.

Transformations with Gaussian and polynomial kernels have been used extensively in kernel-based implementations [CHS00, HEA99, MMR01, BUR98]. We selected the Gaussian kernel $k(\underline{x}, \underline{y}) = \exp(-\|\underline{x} - \underline{y}\|^2 / c)$, with variable spread c in this study and implemented the GDA using MATLAB. One important issue is the specific selection of the spread that affects the classification performance, and Muller et. al. address the model selection issue in their tutorial

[MMR01]. However, an automated selection of the spread c was beyond the scope of this study, and c was determined by trial and error in our simulations.

V. MEAN SEPARATOR NEURAL NETWORK

The Mean Separator Neural Network (MSNN) proposed by Duzenli & Fargues belongs to the class of projection pursuit algorithms [DUZ98, DFA98, DUF98]. The basic MSNN implementation is defined to differentiate between two classes $\{\underline{x}_I\}$ and $\{\underline{y}_I\}$. It iteratively looks for a one-dimensional nonlinear projection direction of the feature space that maximizes the mean difference of projected class data means, for a user-specified nonlinear activation function $\Phi(.)$. As a result, the mean difference criterion *MD(.)* to be maximized is defined as:

$$MD(\underline{w}) = -\left(E\left[\Phi(\underline{w}^t \underline{x})\right] - E\left[\Phi(\underline{w}^t \underline{y})\right]\right)^2, \qquad (5.1)$$

where $\underline{w}$ is a column weight vector. The scheme can be viewed as a one-layer back-propagation neural network (BPNN) implementation with one processing element. The MSNN was implemented using the nonlinear *logsig* function for activation function and gradient descent with variable learning rate [DUZ98]. The scheme was extended to classify more than two classes by reformulating the overall problems as a set of pairwise sub-problems [DUZ98]. Results showed the MSNN to lead to similar or better classification performances than more computational expensive BPNNs, and significantly higher classification performances than obtained with classification trees on the data investigated. Further details regarding these comparisons may be found in Duzenli [DUZ98]. Note that the MSNN implementation suffers of the same drawback as that present in other BPNN implementations: the iterative procedure is not insured to converge to the global minimum as a result of the nonlinear activation function $\Phi(.)$ used in the projection criterion definition. Therefore, we run the MSNN a few times with different initial conditions and selected the weight vector $\underline{w}$ leading to the best training performances in our simulations.

Extensions to the basic MSNN algorithm were considered by San Pedro who investigated the following projection criterion that takes into account both mean differences and variance of the projected data:

$$MD_2(\underline{w}) = -\frac{\left(E\left[\Phi(\underline{w}^t \underline{x})\right] - E\left[\Phi(\underline{w}^t \underline{y})\right]\right)^2}{\mathrm{var}\left[\Phi(\underline{w}^t \underline{x})\right] + \mathrm{var}\left[\Phi(\underline{w}^t \underline{y})\right]}. \qquad (5.2)$$

In this case, the goal becomes to maximize a ratio of the projected class means over the projected class variances. The criterion $MD_2(.)$ may be viewed as a nonlinear implementation of the pairwise Fisher Linear Discriminant. However, this approach cannot be solved using eigen-properties any longer, due to the nonlinear activation function $\Phi(.)$, and an iterative procedure is required to maximize the projection criterion $MD_2(.)$. Various stopping criteria and slightly modified versions of the projection criterion and data set-up were also investigated in San Pedro [SAN00]. However, results showed no significant classification performance improvements of the extensions with respect to the basic MSNN implementation on the data investigated for the nonlinear transformations considered. Therefore, we considered only the basic MSNN implementation with pairwise coupling in this benchmarking study.

VI. CLASSIFIER PERFORMANCES COMPARISON

A. DATA DESCRIPTION

The MSNN, PCA, MBDR, and kernel-based GDA approaches were implemented in MATLAB and applied to the following three classification problems, commonly used in statistics for benchmarking purposes, to evaluate the performances of the feature dimension reduction algorithms. All datasets were obtained from [MLD] and further details describing the feature characteristics and statistics of each dataset can found there.

1. *Iris data:* One of the typical benchmarking data sets selected to investigate the performance of a classifier when dealing with nonlinearly separable data is the IRIS dataset [MLD]. This dataset has three classes with four-dimensional features, where two of the classes are not linearly separable, while the third class is linearly separable from the other two. Twenty-five trials per class were selected for training and for testing respectively.
2. *Handwritten Digits data*: This dataset contains attributes representing normalized bitmaps of handwritten digits from a preprinted form. The dataset had 10 classes and 64 features normalized in the range [0,16]. 87 trials per class were selected for training and for testing respectively.
3. *Spam E-mail data:* This dataset contains attributes indicating whether a specific e-mail can be considered as spam or non-spam e-mail. The dataset has two classes (spam and non-spam type) and 57 features per trial. Most of the features indicate whether a particular word or character was frequently occurring in the e-mail, and further details regarding each individual feature can be found in [MLD]. 227 trials per class were selected for training and for testing respectively.

B. CLASSIFIER SET-UP

Once the feature dimensions are reduced to a desired user-selected size, classification of the data is obtained by applying the basic minimum distance classifier described next. First, the training dataset is used to obtain the mean values for class-specific transformed feature vectors. Such class-specific feature vectors are selected to represent each class and are called class-specific mean feature vectors. During testing, unlabelled feature vectors are compared against each class-specific mean feature vectors, and class decision made by selecting the class which leads to the smaller distance between the unlabelled feature vector and all class-specific mean feature vectors.

We varied the size of the projection, i.e., the size of the reduced dimension features, for PCA and MBDR schemes to evaluate the sensitivity of the feature reduction algorithm to the dimensionality of the projection. Such a variation is not possible for the MSNN algorithm, as it implements a fixed one-dimensional projection. However, we run the MSNN algorithm several times for each training dataset starting the iteration with different random initial values each time in an effort to mitigate the local minima issue discussed earlier, and selected the weights leading to the best training dataset classification performances.

C. RESULTS

Figures VI-1 to VI-3 present the overall classification results obtained with the various dimension reduction schemes followed by the minimum distance classifier. Overall classification performances both for training and testing sets are showed to evaluate any potential generalization issues. Corresponding confusion matrices are included in the Appendix.

1. IRIS Data

Figure VI-1 presents the overall classification performance obtained for the IRIS dataset. Recall that this dataset has a relatively low-dimensional feature dimension to start with, and that it was selected because two of the classes (C_2 and C_3) are not linearly separable, while the third class (C_1) is linearly separable from the other two. Results show the GDA approach is successful in separating the two nonlinearly separable classes while the MSNN is not. Two different implementations of the GDA with slightly different overall classification performances are shown: Kernel-1 and Kernel 2.

Kernel-1: spread value c equal to 1.5 and reduced rank for the pseudo inverse of Γ equal to 75 (full matrix size),

Kernel-2: spread value c equal to 1 and reduced rank for the pseudo inverse of Γ equal to 20 (by visual inspection of the eigenvalue spread for Γ).

Simulations showed that large variations in the spread value may result in significant classification performance differences (when the spread is selected too large or too small for the data under investigation). Results also showed the specific selection of the reduced rank value might have some impact on the classification performances. However, no extensive study was conducted, and further study is required to validate these findings. The results presented here show some slight differences due to small variations in the spread and the pseudo rank of Γ.

The PCA, LDA and MBDR approaches based on a two-dimensional projection of the features show a few classification errors for data in class C_2 and C_3 resulting in 8% overall classification errors.

2. Handwritten Digits Data

Figure VI-2 presents the overall classification performance obtained for the handwritten digit dataset. Recall that this dataset has a relatively high-dimensional feature

dimension to start with (60 features). Results show the best performance is obtained for the kernel-based implementation, followed by PCA (when the projection dimension is larger than 10), the MSNN, and finally the MBDR approach. A few comments are in order.

- MSNN performances vary from run to run due to the local minimum issues inherent in this algorithm, and two different runs are shown here: MSNN-t1 and MSNN-t2, where the difference lies in the random initial values selected during the training phase. This result also further highlights the fact that the MSNN should be run a few times on a given training data, and the version leading to the best performances selected in an effort to minimize this drawback.
- The PCA feature dimension reduction process clearly degrades the discrimination quality of the class features, as the classification performances degrade with decreasing feature size (projected features of dimension 2, 10, 20, 30 are shown here). Simulations showed classification performances to be identical for dimensions 30 to 40. This result also highlights a well-known problem of PCA when applied to classification applications; that is the dimension reduction criterion is not necessarily designed to preserve class discrimination information.
- The MBDR scheme also degrades the discrimination quality of the class features, as classification performances decrease with decreasing feature sizes (two-, and four-, and ten-dimensional projections are reported here). Simulations showed performances to be identical for projection sizes between 4 to 8.
- Simulations showed the kernel-based implementation (using a four-dimensional projection) clearly leads to the best classification performances of all the schemes considered for this dataset.

3. SPAM E-mail Data

Figure VI-3 presents the overall classification performance obtained for the SPAM e-mail dataset. Recall that this dataset has a relatively high-dimensional feature dimension to start with (57 features) and only two classes. Results show the best overall classification performance is obtained for the kernel-based implementation, followed by the MBDR scheme, the MSNN implementation, and finally by various implementations of the PCA (where 2, 10, 20, and 30-dimensional projections were investigated). A few comments are in order.

- A one-dimensional projection for the MBDR approach was selected as only as only one eigenvalue of the matrix U defined earlier in Eq. (3.3) was non zero.
- Simulations showed the PCA feature reduction scheme has the worst classification performances of all schemes considered, and that no improvements are observed by increasing the transformed feature space dimension from 10 to 40.
- The best overall classification performance was obtained with the kernel-based classifier followed by the MSNN implementation and the MBDR scheme.

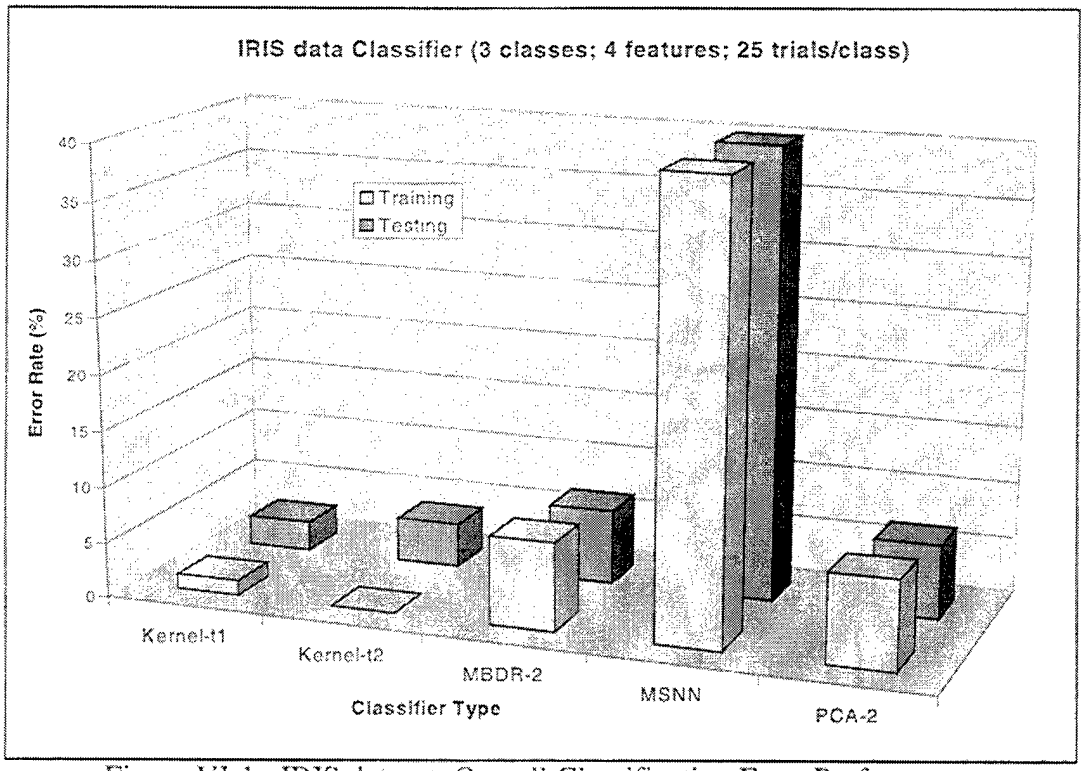


Figure VI-1. IRIS dataset; Overall Classification Error Performance

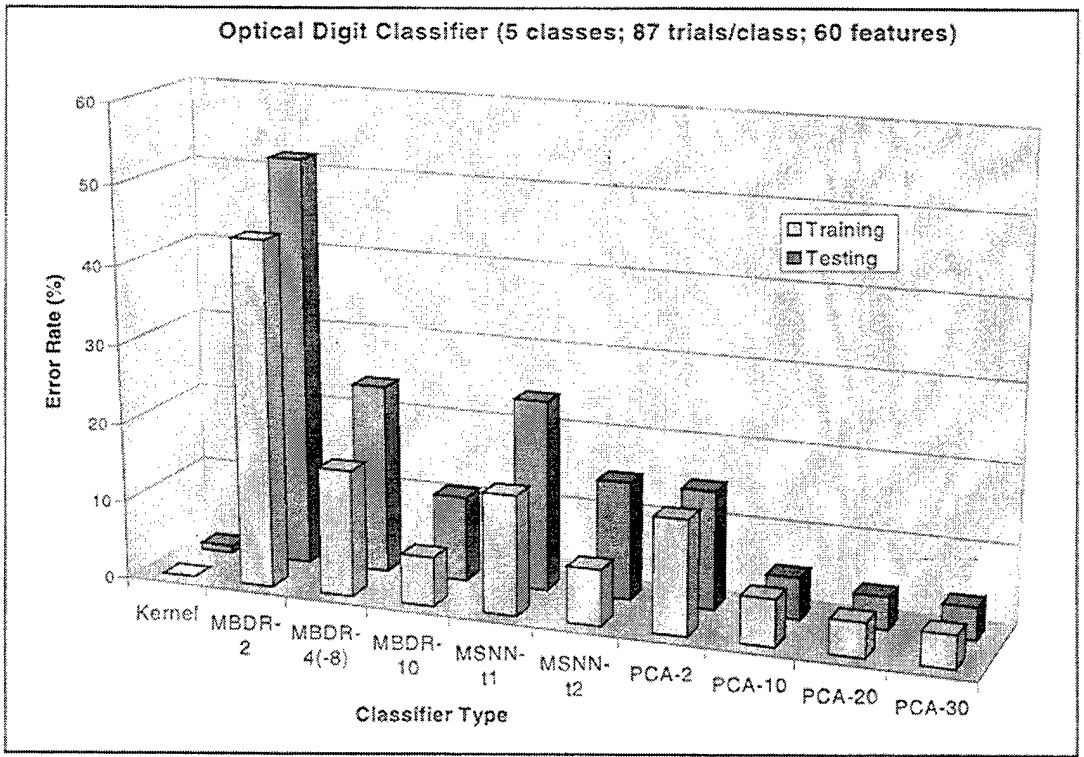


Figure VI-2. Handwritten Digits dataset; Overall Classification Error Performance

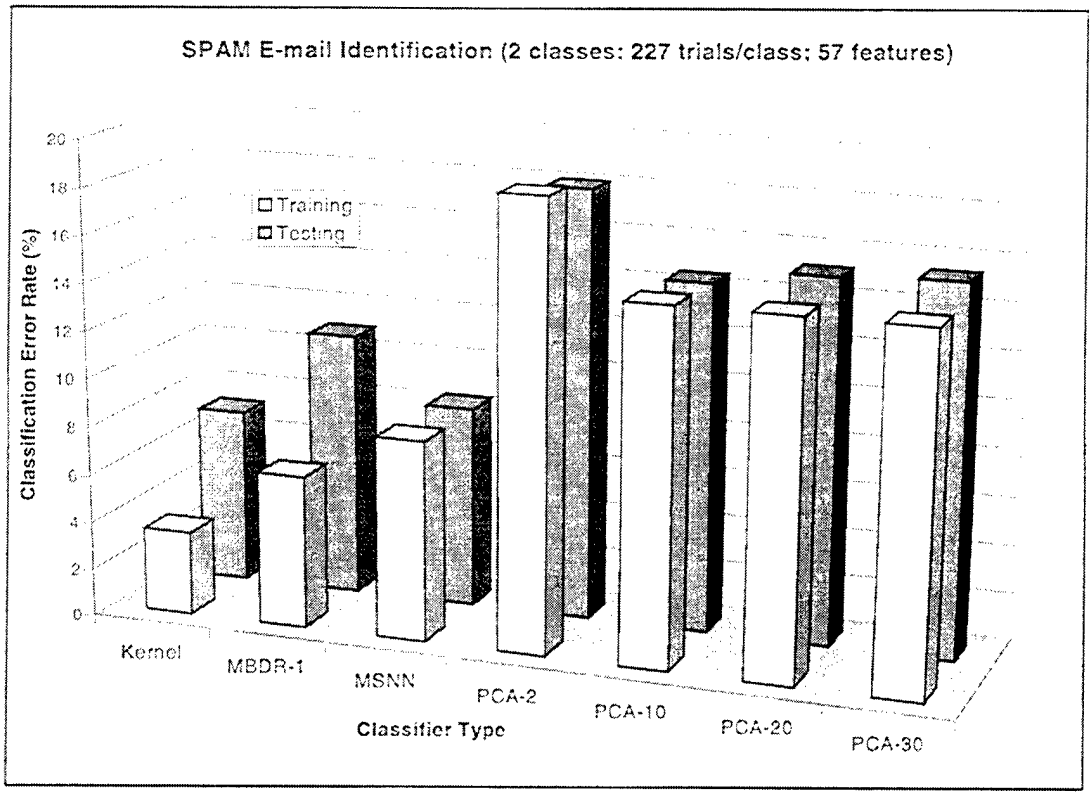


Figure VI-3. SPAM E-mail dataset; Overall Classification Error Performance

VII. CONCLUSIONS

Classification applications require the extraction of class discriminative information. However, this step often leads to high-dimensional feature spaces, which requires large datasets to create viable classification schemes. This study presents follow-on work to [DUZ98, SAN00] and considers two discriminant-based feature dimension reduction schemes for classification applications. The two feature reduction schemes considered are the Mahalanobis-based dimension reduction (MBDR) scheme recently proposed by Brunzell, and the kernel-based generalized discriminant analysis approach (GDA) proposed by Baudat & Anouar. The GDA is part of a new breed of kernel-based algorithms that are currently being considered by the research community to develop new learning techniques, as they can be used to derive nonlinear generalizations of currently known algorithms. Finally, the classical PCA and the MSNN proposed earlier in [DUZ98] are included in this study for comparison purposes.

The four feature dimension reduction schemes considered were implemented in MATLAB and evaluated by applying the transformed features to a basic minimum distance classifier. Performances are evaluated by applying these schemes to three datasets commonly used in statistics for benchmarking purposes. Results show overall best results to be obtained for the GDA for the datasets considered. Results also show there is no consistent second best feature reduction scheme among the MSNN, the MBDR, and the PCA, as performances for these three schemes are data dependent.

Note that our investigation of the generalized discriminant approach (GDA) remains preliminary in nature as our study was restricted to the Gaussian kernel case only, and issues regarding the specific selection of a kernel type were not addressed. In addition, we did not consider issues regarding the specific selection of the spread factor for the Gaussian kernel.

Further investigations addressing these two issues would be needed to complete the study of the GDA behavior. Nevertheless, results are very promising as they show best overall results for the datasets considered to be obtained with the GDA. However, the GDA is also potentially the most computationally intensive of the four schemes considered, depending on the size of the data considered.

Finally, investigating the applicability of the GDA approach to the classification of digital modulation types, and comparing the resulting performances to those obtained using the higher-order statistics based hierarchical approach discussed in [HAT01, FAH01] is left for further study.

APPENDIX

This Appendix contains confusion matrices obtained for training and testing sets for the three datasets [MDL] selected to benchmark the feature reduction schemes considered in this study

1) Digit Data: 5 classes identified by 60 features per trial, 87 trials in training and testing datasets.
2) IRIS Data: 3 classes identified by 4 features per trial, 25 trials in training and testing datasets.
3) Spam e-mail Data: 2 classes identified by 57 features per trial, 227 trials in training and testing datasets.

DIGIT DATA

CLASSIFICATION SCHEMES

	Training Data Results	Testing Data Results

GDA
Kparam=1000;
Tol=.001
(4–dim projection)

Error rate 0%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	87	0	0	0	0
	2	0	87	0	0	0
	3	0	0	87	0	0
	4	0	0	0	87	0
	5	0	0	0	0	87

Error rate 0.91%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	87	0	0	0	0
	2	0	87	1	0	0
	3	0	0	85	0	0
	4	0	0	0	87	1
	5	0	0	1	0	86

MBDR
(2-dim projection)

Error rate 44%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	84	2	30	18	0
	2	0	36	14	4	18
	3	0	16	24	29	0
	4	3	2	12	36	0
	5	0	31	7	0	69

Error rate 52%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	78	03	44	20	0
	2	1	26	6	1	31
	3	1	16	15	32	0
	4	7	4	13	33	0
	5	0	38	9	1	56

MBDR
(4-dim projection)
to
(8-dim projection)

Error rate 16%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	87	0	0	1	0
	2	0	72	0	1	5
	3	0	1	68	20	0
	4	0	10	19	66	1
	5	1	3	0	0	80

Error rate 24%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	86	0	0	1	0
	2	0	72	7	0	15
	3	0	1	63	7	0
	4	0	11	17	79	2
	5	1	3	0	0	70

MBDR
(10-dim projection)

Error rate 6.21%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	87	0	2	0	1
	2	0	78	0	0	12
	3	0	9	85	0	3
	4	0	0	0	87	0
	5	0	0	0	0	71

Error rate 10.80%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	86	0	5	1	0
	2	0	67	1	1	12
	3	0	19	77	0	2
	4	0	0	0	85	0
	5	1	1	4	0	73

PCA
(2-dim projection)

Error rate 14.48%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	86	1	0	0	1
	2	0	72	0	1	5
	3	0	1	68	20	0
	4	0	10	19	66	1
	5	1	3	0	0	80

Error rate 14.94%		Labeled data belonging to Class:				
		1	2	3	4	5
Data identified as Class: →	1	86	0	0	1	0
	2	0	72	7	0	15
	3	0	1	63	7	0
	4	0	11	17	79	2
	5	1	3	0	0	70

IRIS DATA

CLASSIFICATION SCHEMES

Training Data Results

Testing Data Results

GDA
Kparam=1; Tol=.001;
ng=20
(3–dim projection)

Error rate: 1.33%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	24	1
	3	0	0	25

Error rate: 2.67%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	24	1
	3	0	1	24

GDA
Kparam=1.5; Tol=.001;
ng=75
(3–dim projection)

Error rate: 0%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	25	0
	3	0	0	25

Error rate: 4%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	24	1
	3	0	2	23

MBDR
(2-dim projection)

Error rate: 8%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	21	2
	3	0	4	23

Error rate: 6.67%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	23	3
	3	0	2	22

PCA
(2-dim projection)

Error rate: 8%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	22	3
	3	0	3	22

Error rate: 6.67%	Labeled data belonging to Class:			
		1	2	3
Data identified as Class: →	1	25	0	0
	2	0	24	4
	3	0	1	21

MSNN

Except for a few isolated trials, no separation is obtained between classes 2 and 3

SPAM DATA

CLASSIFICATION SCHEMES

Training Data Results

Testing Data Results

GDA
Kparam=2; Tol=.001; ng=200
(2–dim projection)

Error rate: 3.52%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	219	8
	2	8	219

Error rate: 7.27%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	212	15
	2	18	209

MBDR
(1-dim projection)
(only 1 non zero eigenvalue)

Error rate: 6.39%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	206	8
	2	21	219

Error rate: 11.01%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	202	25
	2	25	202

PCA
(1-dim projection)

No separation

MSNN

Error rate: 8.37%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	206	27
	2	21	200

Error rate: 8.37%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	206	27
	2	21	200

PCA
(2-dim projection)

Error rate: 18.7%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	174	32
	2	53	195

Error rate: 18.06%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	177	32
	2	50	195

PCA
(10-dim projection)

Error rate: 14.75%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	187	27
	2	40	200

Error rate: 14.53%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	187	26
	2	40	201

PCA
(20-dim projection)

Error rate: 14.76%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	187	27
	2	40	200

Error rate: 15.2%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	186	28
	2	41	199

PCA
(30-dim projection)

Error rate: 14.76%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	187	27
	2	40	200

Error rate: 15.41%	Labeled data belonging to Class:		
		1	2
Data identified as Class: →	1	186	29
	2	41	198

REFERENCES

[BAN00] G. Baudat & F. Anouar, "Generalized discriminant analysis using a kernel approac Neural Computations, Vol. 12, No. 1, 2000, pp. 2385-2404.

[BIS95] C. Bishop, *Neural Networks for Pattern Recognition,* Oxford Press, 1995.

[BRE99] H. Brunzell & J. Eriksson, " Feature reduction for classification of multidimensional data," Draft, ECE Department, Ohio State University, Nov. 1999.

[BRU97] H. Brunzell, "Extraction of features for classification of impulse radar measurements," SPIE Proceedings, Automatic Object Recognition VII, Vol. 3069, Apr. 1997, pp. 321-330.

[BUR98] C. Burges, "A tutorial on support vector machines for pattern recognition," Knowledge Discovery and Data Mining, Vol. 2, No. 2, 1998.

[CHS00] N. Christianini & J. Shawe-Taylor, *An Introduction to Support Vector Machines and other Kernel-based Learning Methods*, Cambridge Press, 2000.

[CHV99] O. Chapelle, P. Haffner & V. Vapnik, "SVMs for histogram-based image classification" IEEE Transaction on Neural Networks, Vol. 9, 1999.

[DFA98] O. Duzenli & M. Fargues, "Wavelet-based feature extraction methods for classification applications," Proceedings of the 9th IEEE SSAP Workshop, Sept. 1998.

[DHS01] R. Duda, P. Hart & D. Stork, *Pattern Classification,* Wiley Interscience, 2001.

[DUZ98] O. Duzenli, *Classification of Underwater Signals Using Wavelet-Based Decompositions,* MSEE Thesis, Naval Postgraduate School, June 1998.

[EPP00] T. Evgeniou, M. Pontil, C. Papageorgiou, & T. Poggio, "Image representations for object detection using kernel classifiers," Proceedings ACCV, 2000.

[FAH01] M. Fargues & G Hatzichristos, *A Hierarchical Approach to the Classification of Digital Modulation Types in Multipath Environments,* Technical Report, NPS-EC-01-004, Naval Postgraduate School, May 2001.

[FDU98] M. Fargues & O. Duzenli, "Dimension reduction issues in classification applications," Proceedings of the 32nd Asilomar Conference on Signals, Systems, and Computers, Nov. 1998.

[FUK90] K. Fukunaga, *Statistical Pattern Recognition*, 2nd Ed., Academic Press, 1990.

[GSO00] I. Guyon & D. Stork, "Linear discriminant and support vector classifiers," In A.J. Smola, P.L. Bartlett, B. Schölkopf, and D. Schuurmans, editors, Advances in Large Margin Classifiers, Cambridge, MA, 2000, MIT Press, pp. 147-169.

[HAT01] G. Hatzichristos, *Classification of Digital Modulation Types in Multipath Environments,* Electrical Engineer's Thesis, Naval Postgraduate School, March 2001.

[HEA99] "Support vector machines," M. Hearst Ed., IEEE Journal on Intelligent Systems, 1999, pp. 18-28.

[JOA98] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," Proceedings of the European Conference on Machine Learning, 1998, pp. 137-148.

[KMW] *Kernel Machines Webpage* [Online], available: www.kernel-machines.org, last accessed 5/1/01.

[LCB00] C-C. Lim, H-G. Chew & R. Bogner, "Target detection in radar imagery using support vector machines with training size biasing," 2000.

[LRO01] A. Ruiz & P. Lopez-de-Teruel, "Nonlinear kernel-based statistical pattern analysis,"

IEEE Trans. on Neural Networks, Vol. 12, No. 1, Jan. 2001, pp. 16-32.

[MAE99] E. Maeda, "Multi-category classification by kernel-based nonlinear subspace method," IEEE ICASSP'99 Proceedings, 1999.

[MCL92] G. McLachan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, 1992.

[MLD] *Machine learning database repository*, [Online], available: ftp://ftp.ics.uci.edu/pub/machine-learning-databases, last accessed 5/1/01.

[MMR01] K-R. Muller, S. Mika, G. Ratsch, K. Tsuda & B. Scholkopf, "An introduction to kernel-based learning algorithms," IEEE Trans. on Neural Networks, Vol. 12, No. 2, March 2001, pp. 181-201.

[MRS99] S. Mika, G. Rätsch, B. Schölkopf, A. Smola, J. Weston & K-R. Müller, "Invariant feature extraction and classification in kernel spaces," Advances in Neural Information Processing Systems 12, Cambridge, MA, 1999.

[MRW99a] S. Mika, G. Ratsch & J. Weston, "Fisher discriminant analysis with Kernels," Neural Networks for Signal Processing IX, 1999, pp. 41-48.

[MRW99b] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, A. Smola & K-R. Muller, " Invariant feature extraction and classification in kernel spaces," In Neural Networks for Signal Processing IX, 1999, pp. 41-48.

[MSR97] K-R. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen & V. Vapnik, "Predicting time series with support vector machines," Advances in Kernel Methods --- Support Vector Learning, Cambridge, MA. MIT Press, 1999, pp. 243-254.

[MSS99] S. Mika, B. Schölkopf, A. Smola, K-R. Müller, M. Scholz, & G. Rätsch, "Kernel PCA and de-noising in feature spaces," In Advances in Neural Information Processing Systems Vol. 11, Cambridge, MA, 1999. MIT Press, pp. 536-542.

[NBR00] P. Niyogi, C. Burges, P. Ramesh, "Distinctive feature detection using support vector machines," IEEE Proc. ICASSP'00, 2000.

[SAN00] M. San Pedro, *Signal Classification using the Mean Separator Neural Network,* MSEE Thesis, Naval Postgrduate School, March 2000.

[SBS99] B. Schölkopf, C. J. C. Burges, & A. J. Smola. *Advances in Kernel Methods --- Support Vector Learning*. MIT Press, Cambridge, MA, 1999.

[SMB99] B. Scholkopf, S. Mika, J. Burges, P. Knirsch, K-R. Muller, G. Ratsch & J. Smola, "Input space versus feature space in kernel-based methods," IEEE Trans. on Neural Networks, Vol. 10, No. 5, Sept. 1999, pp. 1000-1017.

[SSM98] B. Schölkopf, A. Smola, & K-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, Vol. 10, 1998, pp. 1229-1319.

[SSM99] B. Schölkopf, A. Smola, & K-R. Müller, "Kernel principal component analysis," In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, Advances in Kernel Methods - SV Learning, MIT Press, Cambridge, MA, 1999, pp.327-352.

[TCR01] L. Trejo, A. Cichocki, R. Rosipal & M. Girolami, "Kernel PCA for feature extraction and de-noising in non-linear regression," Neural Computing & Applications, 2001.

[TKO00] S. Tong & D. Koller, "Support vector machine active learning with applications to text classification," Proceedings of the 2000 ICML Conference, 2000.

[VAP95] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, N.Y., 1995.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center.....2
   8725 John J. Kingman Rd., STE 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, Ca 93943-5121

3. Chairman, Code EC 1
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, Ca 93943-5121

4. Prof. Monique P. Fargues, Code EC/Fa 3
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, Ca 93943-5121

5. Prof. John P. Powers, Code EC/Po 1
   Center for Reconnaissance Research
   Department of Electrical and Computer Engineering
   Naval Postgraduate School
   Monterey, Ca 93943-5121

6. Research Office, Code 09
   Naval Postgraduate School
   Monterey, CA 93943.....1