report for

Comparison of Pattern Recognition Techniques

submitted to fulfill the semester project requirement for

EE 8993: Pattern Recognition

May 5, 1999

submitted to:

Dr. N.H. Younan

Department of Electrical and Computer Engineering
413 Simrall, Hardy Rd.
Mississippi State University
Box 9571
MS State, MS 39762

submitted by:

Yu Zeng

Mississippi State University
Box 9671
Mississippi State, Mississippi 39762
Tel: 601-325-8335
Fax: 601-325-8192
email: zeng@isip.msstate.edu

1 ABSTRACT

Principal Components Analysis (PCA) is commonly used in many fields including feature extraction and data compression. Independent Components Analysis (ICA) is a new technique that has been demonstrated to offer improved performance for problems such as blind source separation where higher-order statistics of the data are important. The goal of this project is to investigate the theoretical relationship between PCA and ICA, and demonstrate the nature of this relationship on several classes of problems.

2 INTRODUCTION

2.1 Principal Components Analysis

Principal component analysis [1] is commonly used for feature extraction. According to this technique, the first principal component of a data set is the direction along which there is largest variance over all samples. The basic idea of this technique is that the direction along which there is maximum variation is most likely to contain the information about the class discrimination. Mapping the given data set into this space will make the difference between classes more significant.

PCA maps the data set using an orthonormal matrix $Y = \Phi^T X$ to the feature space, where $X = [x_1, x_2, \ldots, x_n]^T$ is the original data set, $Y = [y_1, y_2, \ldots, y_n]^T$ is the output data set. The transformation matrix Φ can be achieve by $\Sigma\Phi = \Phi\Lambda$, where Φ and Λ are the eigenvector matrix and the eigenvalue matrix of the covariance matrix Σ of the source data X. Since Φ is an orthonormal matrix, in the transformed space, we have

$$\|Y\|^2 = Y^T Y = X^T \Phi \Phi^T X = X^T X = \|X\|$$

The Euclidean distance can be used as the measure of the distance between two data.

Only the mean and the covariance will be used to find out the transformation matrix.

2.2 Independent Components Analysis

Independent Component Analysis is very similar to PCA. Given a data set $X = [x_1, x_2, \ldots, x_n]^T$, ICA will find an $n \times n$ matrix W, which will minimize the average mutual information of the output matrix $Y = [y_1, y_2, \ldots, y_n]^T$. The difference is transformation matrix is not required to be orthogonal, as shown in figure 1. The computation of this transform involves higher-order statistics of the data, so it's well used for problems such as blind source separation where higher-order statistics of the data are important.

2.3 Algorithm for ICA

My approach to find the ICA transformation follows that described in [2]. Since a closed-form solution is not tractable, I used a stochastic gradient ascent algorithm. The idea is pick a nonlinear

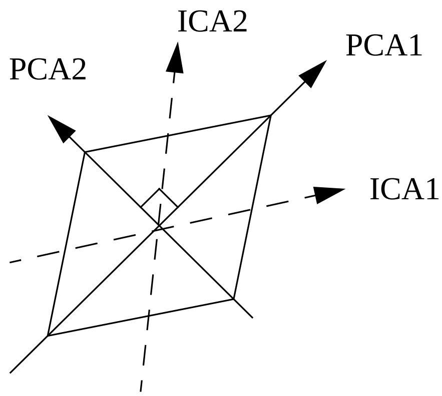

Figure 1. A comparison between transformations found by PCA and ICA when the data is uniformly distributed within the diamond-shaped region.

function $g(u)$, which has the same covariance matrix as the underlying independent components, maximizing the joint entropy of $g(u)$ will lead to ICA solution.

The specific form of the stochastic gradient ascent that is used involves a learning rule that changes weights according to the gradient of the entropy [2]:

$$\Delta W \propto \frac{\partial H(y)}{\partial W} W^T W = (I + \hat{y} u^T) W$$

in which $\hat{y}_i = \frac{\partial}{\partial y_i} \frac{\partial y_i}{\partial u_i} = \frac{\partial}{\partial u_i} In \frac{\partial y_i}{\partial u_i}$ is computed as $\hat{y}_i = 1 - 2y_i$, and $y_i = (1 + e^{-u_i})^{-1}$.

3 EXPERIMENTS

3.1 Data

As showing in figure 2, the original data contains two classes. If use the 1-NNR classification rule, the data to the left hand of the separating hyperplan will be classified into class 1, and the data to the right hand of the separating hyperplan will be classified into class 2.

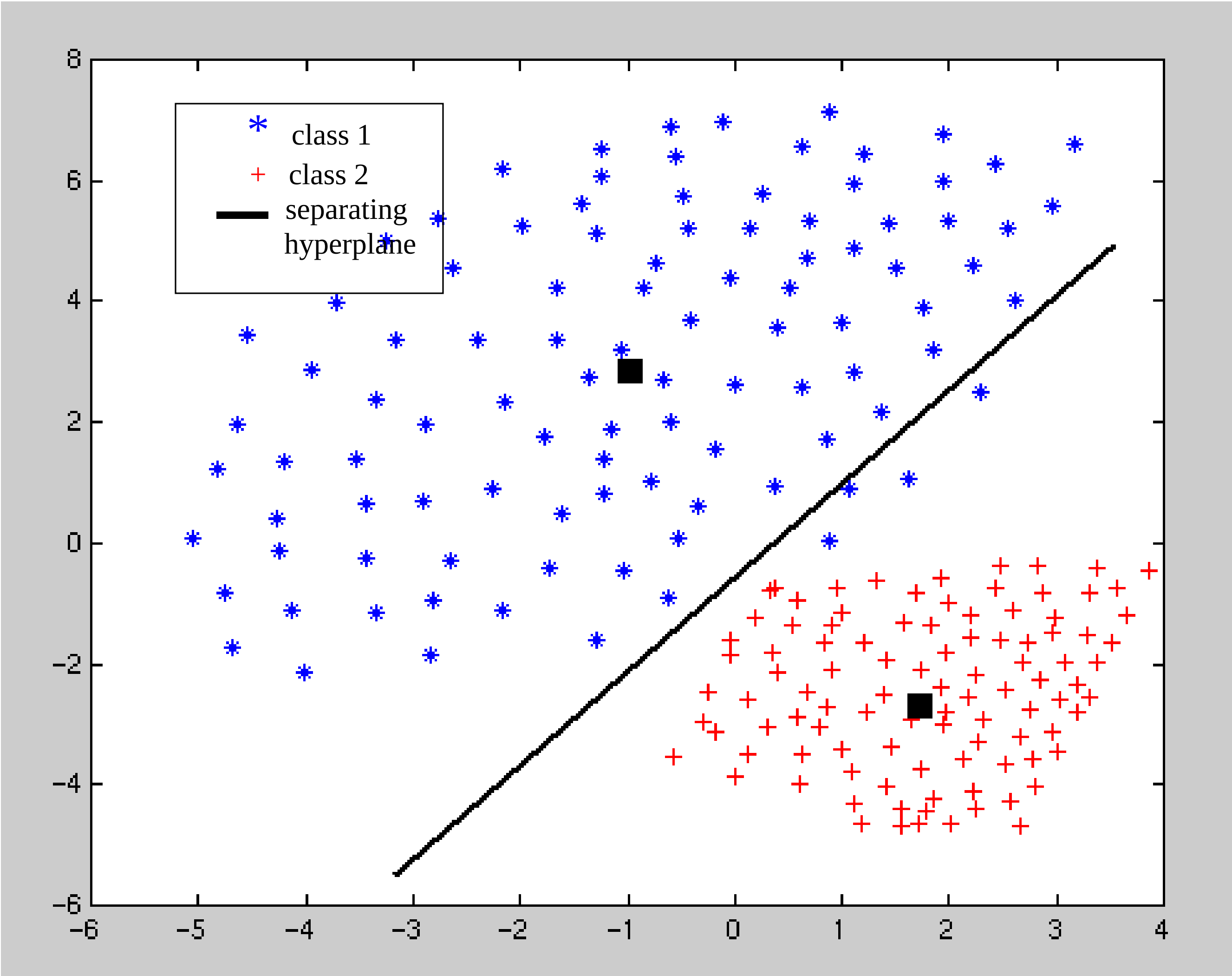

Figure 2. The original data

Using PCA, I computed the transformation matrix Φ_1 and Φ_2 for data set 1 and data set 2 respectively. For each given test data point x, I transformed it to y_1 and y_2 separately. Then, in the transformed space, I computed the Euclidean distance between the test data point and the mean value of each class d_1 and d_2. Using the 1-NNR rule, x belongs to the class which results in the smaller distance. The result of PCA is showing in figure 3:

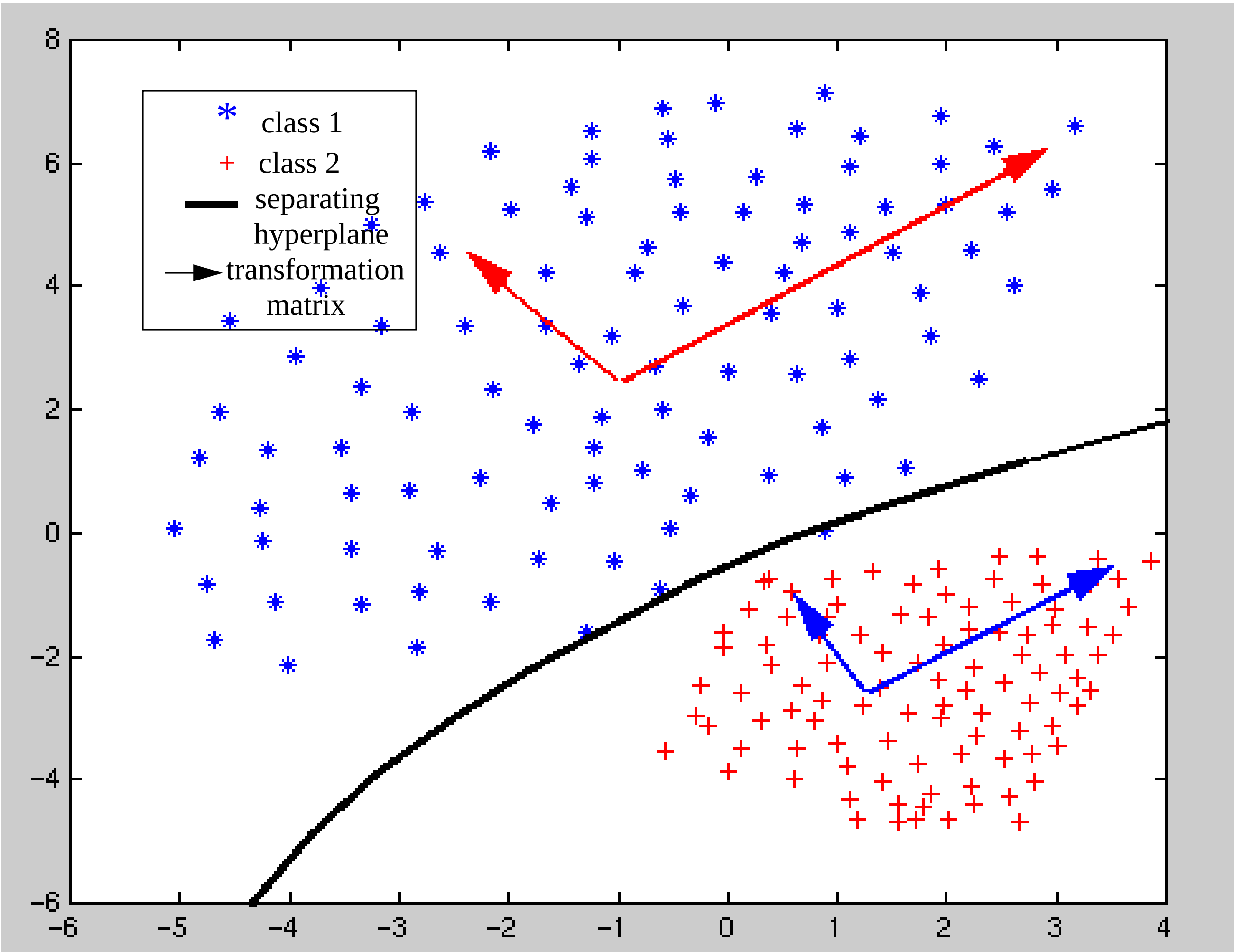

Figure 3. The result using PCA

Using ICA, I followed the same procedure as what I did using PCA. The result is showing in figure 4:

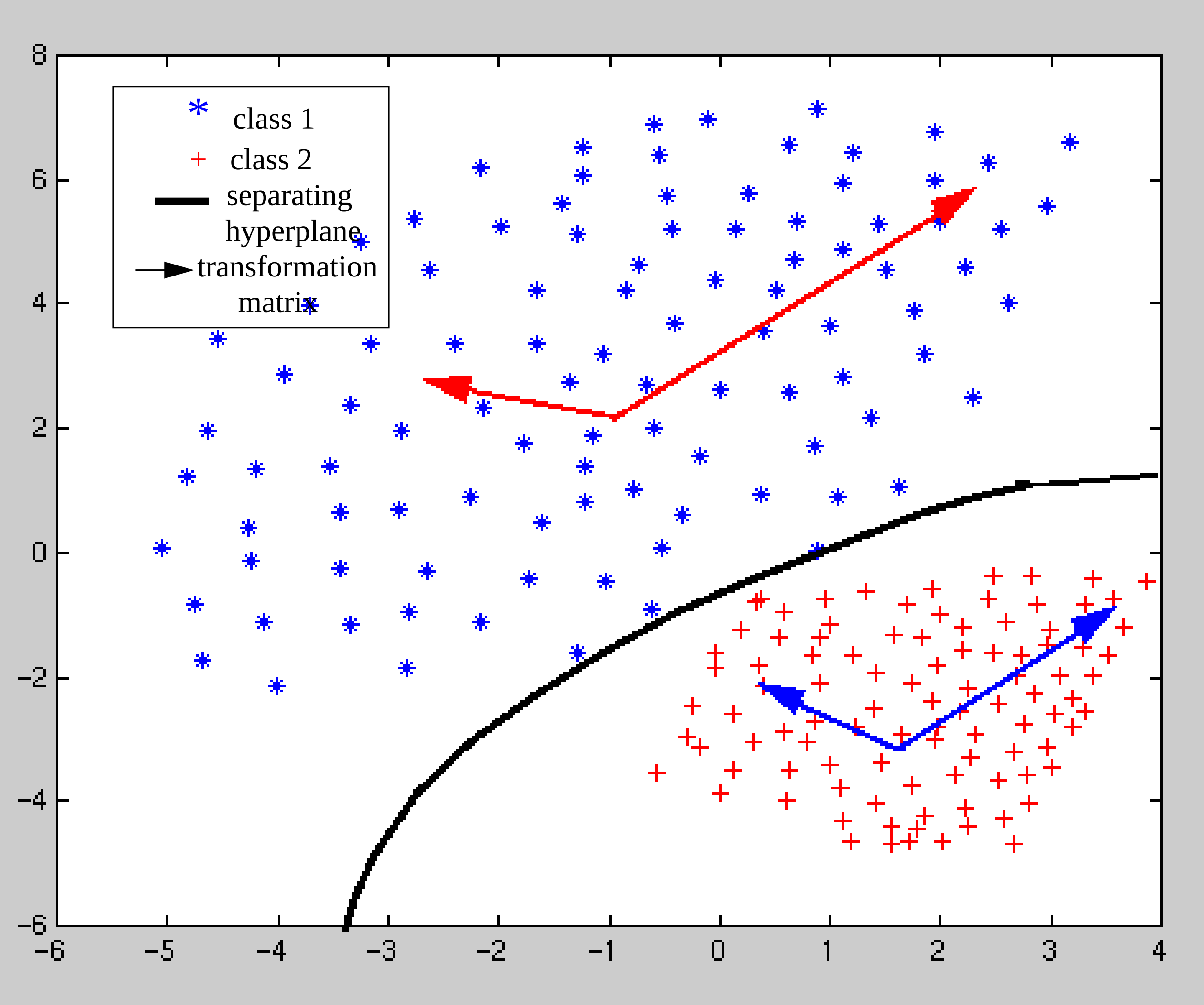

Figure 4. Result using ICA

4 SUMMARY

Both PCA and ICA can improve the performance of data discrimination problem than 1-NNR rule. Ideally, since the transformation matrix that ICA find is not required to be orthogonal, so ICA will give a better result than PCA do. Since ICA uses all order of statistics, it is more useful than PCA while the higher-order statistics of data are important.

5 REFERENCE

[1] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, San Diego, California, USA, 1990.

[2] A. J. Bell and T. J. Sejnowski, “An Information-Maximization Approach to Blind Separation and Blind Deconvolution,” Technical Report No. INC-9501, Institute for Neural Computing, University of California at San Diego, San Diego, California, USA, February 1995.

[3] A. Hyvarinen, “Fast and Robust Fixed-Point Algorithms for Independent Components Analysis”, Laboratory of Computer and Information Science, http://www.cis.hut.fi/~aapo/pub.html