LECTURE 20: LINEAR DISCRIMINANT ANALYSIS

- Objectives:
  - Review maximum likelihood classification
  - Appreciate the importance of weighted distance measures
  - Introduce the concept of discrimination
  - Understand under what conditions linear discriminant analysis is useful

This material can be found in most pattern recognition textbooks. This is the book we recommend:

- R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification* (Second Edition), Wiley Interscience, New York, New York, USA, ISBN: 0-471-05669-3, 2000.

and use in our pattern recognition course. The material in this lecture follows this textbook closely:

J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

Each of these sources contain references to the seminal publications in this area, including our all-time favorite:

- K. Fukunga, *Introduction to Statistical Pattern Recognition*, MacMillan Publishing Company, San Diego, California, USA, ISBN: 0-1226-9851-7, 1990.

ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION

Professor Joseph Picone
Department of Electrical and Computer Engineering
Mississippi State University

email: picone@isip.msstate.edu
phone/fax: 601-325-3149; office: 413 Simrall
URL: http://www.isip.msstate.edu/resources/courses/ece_8463

Modern speech understanding systems merge interdisciplinary technologies from Signal Processing, Pattern Recognition, Natural Language, and Linguistics into a unified statistical framework. These systems, which have applications in a wide range of signal processing problems, represent a revolution in Digital Signal Processing (DSP). Once a field dominated by vector-oriented processors and linear algebra-based mathematics, the current generation of DSP-based systems rely on sophisticated statistical models implemented using a complex software paradigm. Such systems are now capable of understanding continuous speech input for vocabularies of hundreds of thousands of words in operational environments.

In this course, we will explore the core components of modern statistically-based speech recognition systems. We will view speech recognition problem in terms of three tasks: signal modeling, network searching, and language understanding. We will conclude our discussion with an overview of state-of-the-art systems, and a review of available resources to support further research and technology development.

Tar files containing a compilation of all the notes are available. However, these files are large and will require a substantial amount of time to download. A tar file of the html version of the notes is available here. These were generated using wget:

```
wget -np -k -m http://www.isip.msstate.edu/publications/courses/ece_8463/lectures/current
```

A pdf file containing the entire set of lecture notes is available here. These were generated using Adobe Acrobat.

Questions or comments about the material presented here can be directed to help@isip.msstate.edu.

LECTURE 20: LINEAR DISCRIMINANT ANALYSIS

- Objectives:
  - Review maximum likelihood classification
  - Appreciate the importance of weighted distance measures
  - Introduce the concept of discrimination
  - Understand under what conditions linear discriminant analysis is useful

This material can be found in most pattern recognition textbooks. This is the book we recommend:

- R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification* (Second Edition), Wiley Interscience, New York, New York, USA, ISBN: 0-471-05669-3, 2000.

and use in our pattern recognition course. The material in this lecture follows this textbook closely:

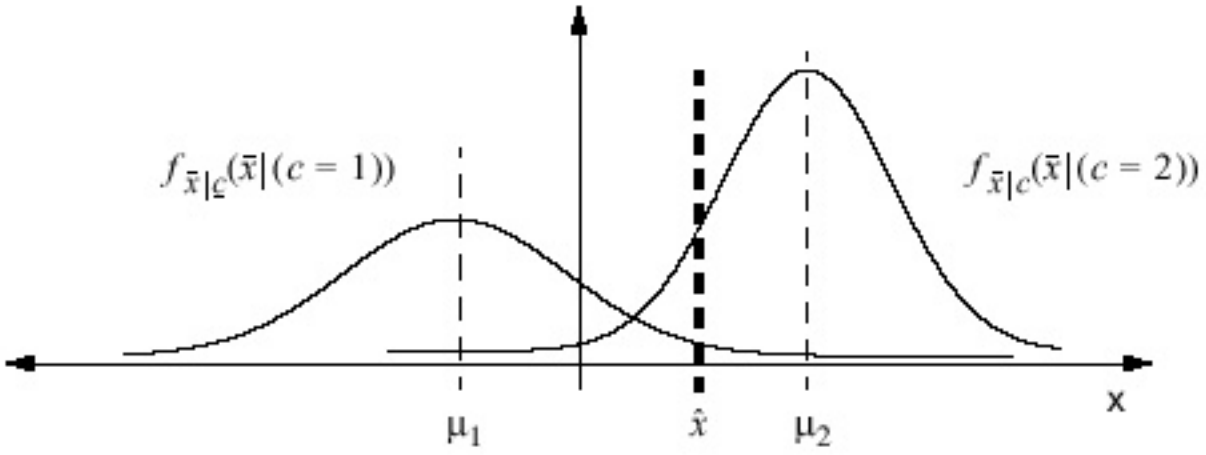
J. Deller, et. al., *Discrete-Time Processing of Speech Signals*, MacMillan Publishing Co., ISBN: 0-7803-5386-2, 2000.

Each of these sources contain references to the seminal publications in this area, including our all-time favorite:

- K. Fukunga, *Introduction to Statistical Pattern Recognition*, MacMillan Publishing Company, San Diego, California, USA, ISBN: 0-1226-9851-7, 1990.

MAXIMUM LIKELIHOOD CLASSIFICATION

Consider the problem of assigning a measurement to one of two sets:



What is the best criterion for making a decision?

Ideally, we would select the class for which the conditional probability is highest:

$$c^* = \underset{c}{\text{argmax}} \; P((c=\hat{c})|(\bar{x}=\hat{\bar{x}}))$$

However, we can't estimate this probability directly from the training data. Hence, we consider:

$$c^* = \underset{c}{\text{argmax}} \; P((\bar{x}=\hat{\bar{x}})\big|(c=\hat{c}))$$

By definition

$$P((c=\hat{c})|(\bar{x}=\hat{\bar{x}})) = \frac{P((c=\hat{c}),(\bar{x}=\hat{\bar{x}}))}{P(\bar{x}=\hat{\bar{x}})}$$

and

$$P((\bar{x}=\hat{\bar{x}})\big|(c=\hat{c})) = \frac{P((c=\hat{c}),(\hat{x}=\hat{\bar{x}}))}{P(c=\hat{c})}$$

from which we have

$$P((c=\hat{c})\big|(\bar{x}=\hat{\bar{x}})) = \frac{P((\bar{x}=\hat{\bar{x}})\big|(c=\hat{c}))P(c=\hat{c})}{P(\bar{x}=\hat{\bar{x}})}$$

Clearly, the choice of c that maximizes the right side also maximizes the left side. Therefore,

$$c^* = \underset{c}{\text{argmax}} \left[P((\bar{x} = \hat{\bar{x}}) \middle| (c = \hat{c}))\right]$$

$$= \underset{c}{\text{argmx}} [P((\bar{x} = \hat{\bar{x}}) \middle| (c = \hat{c}))P(c = \hat{c})]$$

if the class probabilities are equal,

$$c^* = \underset{c}{\text{argmx}} [P((\bar{x} = \hat{\bar{x}}) \middle| (c = \hat{c}))]$$

A quantity *related* to the probability of an event which is used to make a decision about the occurrence of that event is often called a *likelihood measure*.

A decision rule that maximizes a likelihood is called a maximum likelihood decision.

In a case where the number of outcomes is not finite, we can use an analogous continuous distribution. It is common to assume a multivariate Gaussian distribution:

$$f_{\bar{x}|c}(x_1,\ldots,x_N \middle| c) = f_{\bar{x}|c}(\hat{\bar{x}} \middle| \hat{c})$$

$$= \frac{1}{\sqrt{2\pi \left| C_{\bar{x}|c} \right|}} \exp\left\{ -\frac{1}{2} (\hat{\bar{x}} - \bar{\mu}_{\bar{x}|c})^T \underline{C}_{\bar{x}|c}^{-1} (\hat{\bar{x}} - \bar{\mu}_{\hat{\bar{x}}|c}) \right\}$$

We can elect to maximize the log, $\ln[f_{\bar{x}|c}(\bar{x}|c)]$ rather than the likelihood (we refer to this as the log likelihood). This gives the decision rule:

$$c^* = \underset{c}{\text{argmin}} \left[(\hat{\bar{x}} - \bar{\mu}_{\bar{x}|c})^T \underline{C}_{\bar{x}|c}^{-1} (\hat{\bar{x}} - \bar{\mu}_{\hat{\bar{x}}|c}) + \ln\left\{ \left| \underline{C}_{\bar{x}|c}^{-1} \right| \right\} \right]$$

(Note that the maximization became a minimization.)

We can define a distance measure based on this as:

$$d_{ml}(\bar{x}, \bar{\mu}_{\bar{x}|c}) = (\hat{\bar{x}} - \bar{\mu}_{\bar{x}|c})^T \underline{C}_{\bar{x}|c}^{-1} (\hat{\bar{x}} - \bar{\mu}_{\hat{\bar{x}}|c}) + \ln\left\{ \left| \underline{C}_{\bar{x}|c}^{-1} \right| \right\}$$

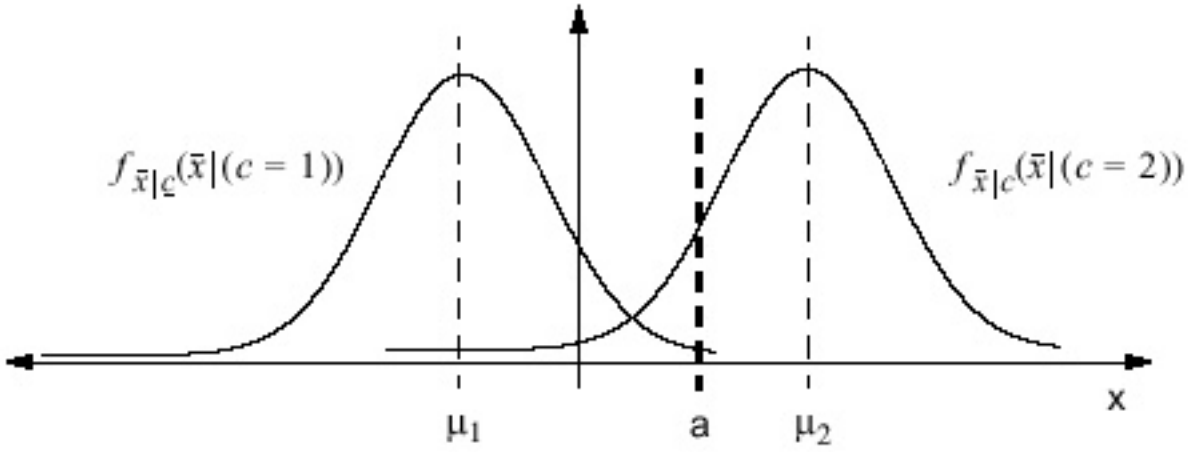
Note that the distance is conditioned on each class mean and covariance. This is why "generic" distance comparisons are a joke.

If the mean and covariance are the same across all classes, this expression simplifies to:

$$d_M(\bar{x}, \bar{\mu}_{\bar{x}|c}) = (\hat{\bar{x}} - \bar{\mu}_{\bar{x}|c})^T \underline{C}_{\bar{x}|c}^{-1} (\hat{\bar{x}} - \bar{\mu}_{\hat{\bar{x}}|c})$$

This is frequently called the *Mahalanobis distance*. But this is nothing more than a weighted Euclidean distance.

This result has a relatively simple geometric interpretation for the case of a single random variable with classes of equal variances:



The decision rule involves setting a threshold:

$$a = \left(\frac{\mu_1 + \mu_2}{2}\right) + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P(c=2)}{P(c=1)}\right)$$

and,

$$\begin{array}{lll} if & x < a & x \in (c=1) \\ else & & x \in (c=2) \end{array}$$

If the variances are not equal, the threshold shifts towards the distribution with the smaller variance.

What is an example of an application where the classes are not equiprobable?

TWO-DIMENSIONAL GAUSSIAN DISTRIBUTIONS

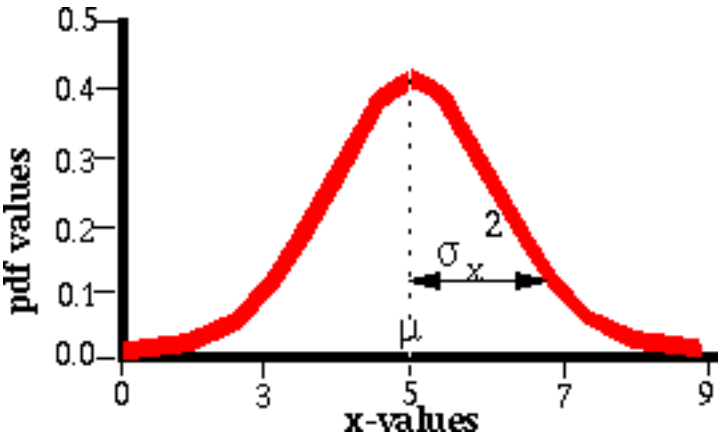
The equation of multivariate Gaussian distribution is as follows:

$$z = p(x,y) = \frac{1}{(2\pi)^{d/2}\sqrt{|C|}} e^{-\frac{1}{2}((x,y)-\mu)C^{-1}((x,y)-\mu)^T}$$

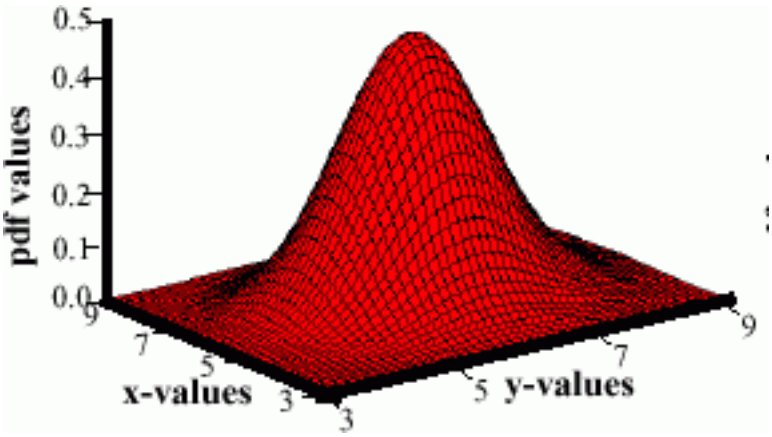
$$C = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix}$$

where is the covariance matrix and *d* is the dimension of the distribution.

1D Gaussian distribution

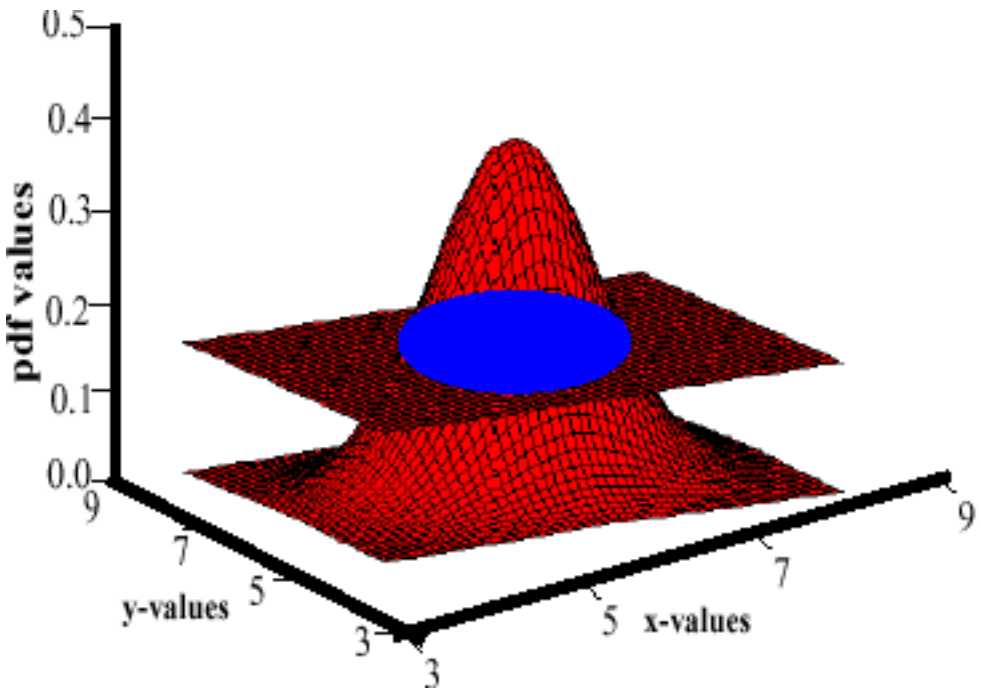


2D Gaussian distribution

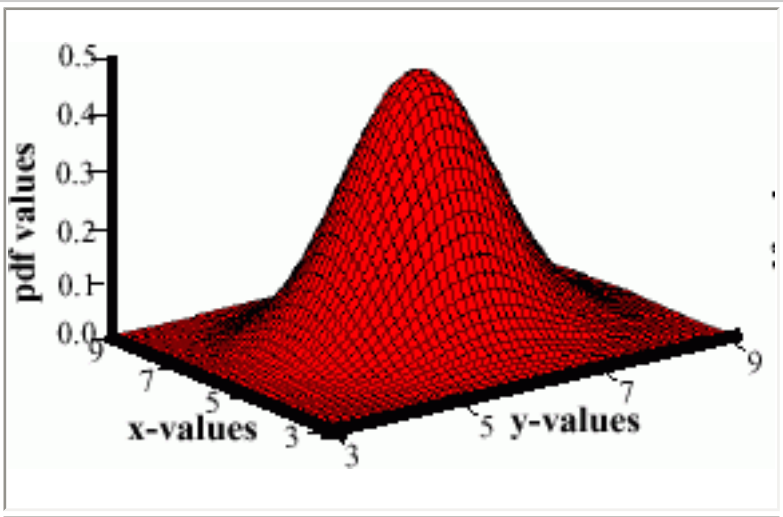
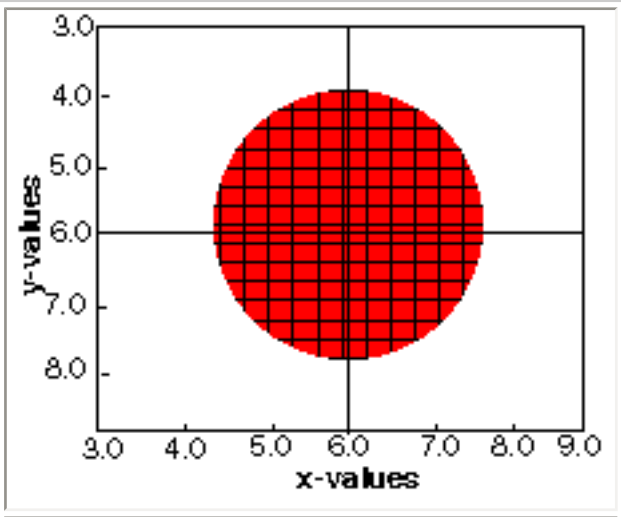
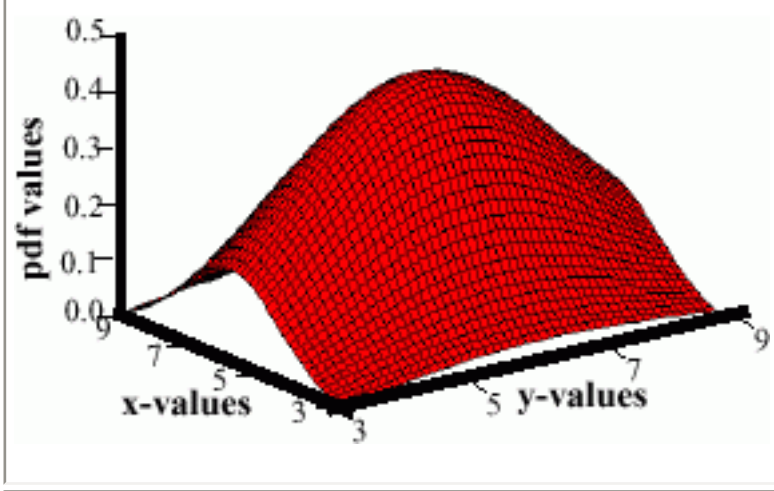
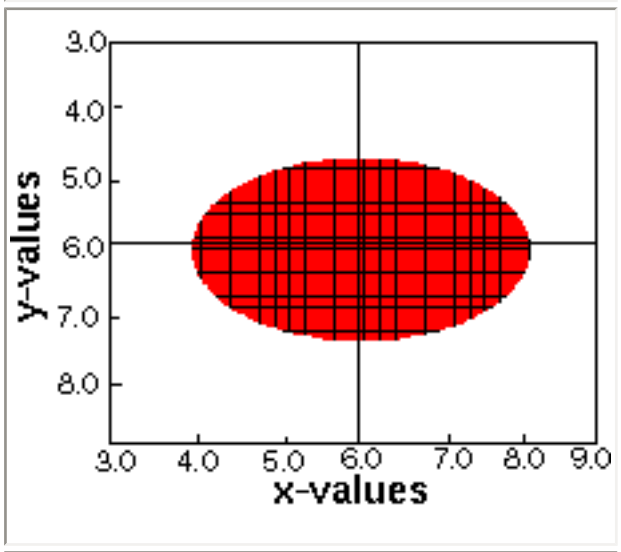
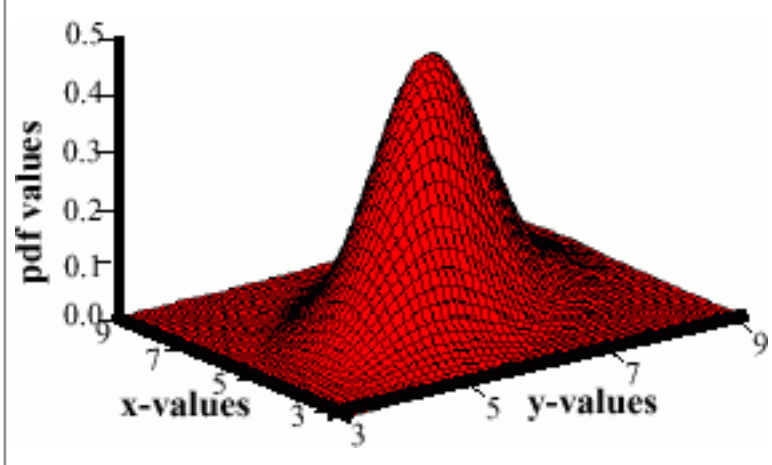
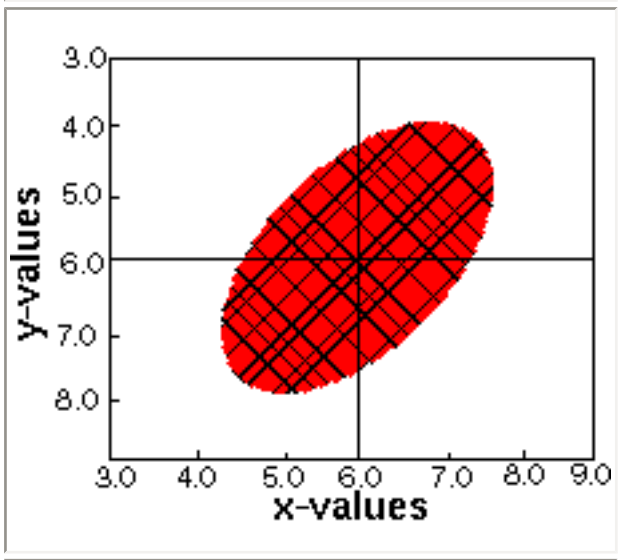
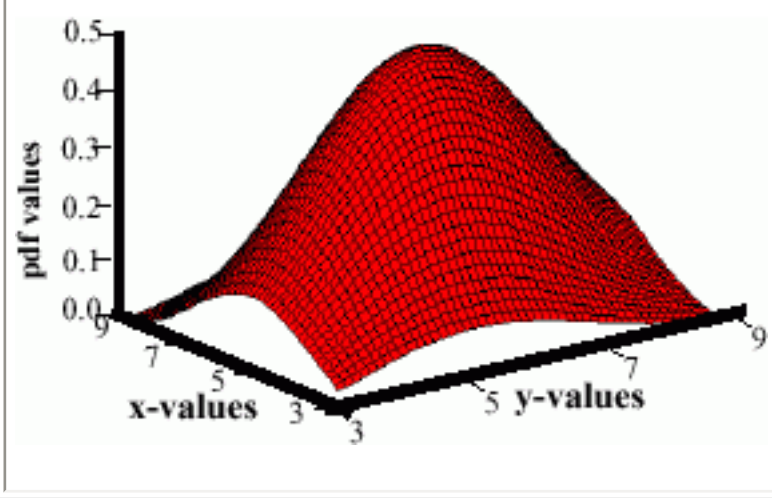
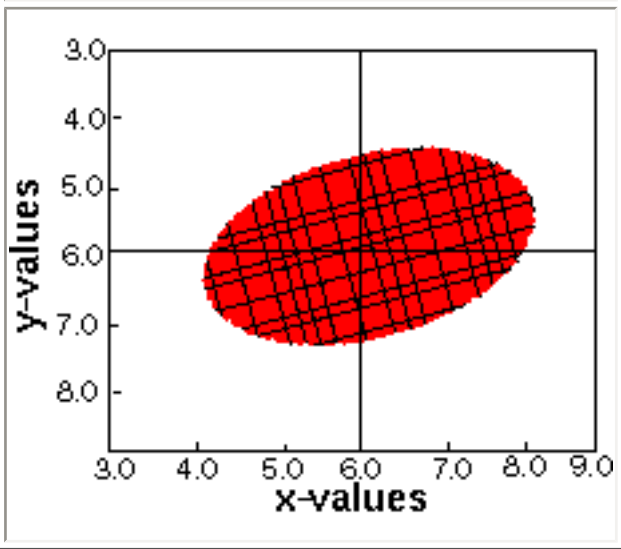


SUPPORT REGION: A CONVENIENT VISUALIZATION TOOL

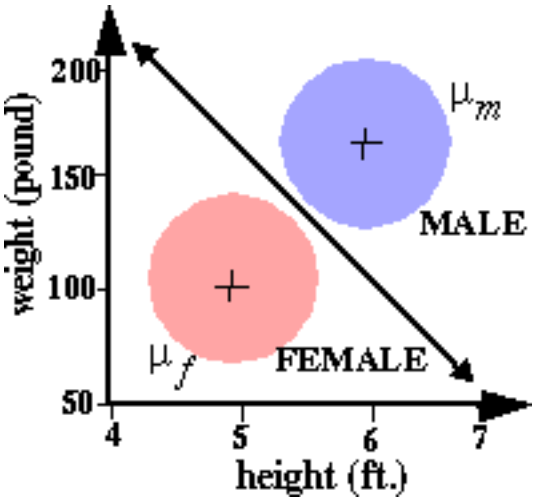
- A support region is the surface defined by the intersection of a Gaussian distribution with a plane.
- The shape of a support region is elliptical and depends on the covariance matrix of the original data.
- A convenient visualization tool in pattern recognition.



Some examples demonstrating the relationship between the covariance matrix and the 2D Gaussian distribution are shown below:

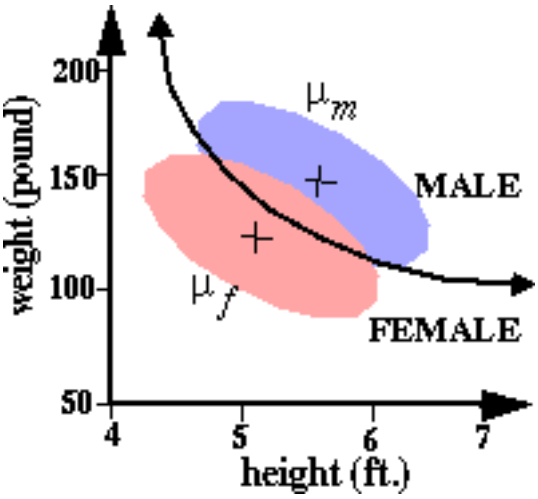
Identity: $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$		
Unequal Variances: $C = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$		
Nonzero off-diagonal: $C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$		
Unconstrained: $C = \begin{bmatrix} 5 & 0.5 \\ 0.5 & 2 \end{bmatrix}$		

. **Case I: Distributions with equal variances**



. **Decision surface boundary is a line separating the two distributions (general case is a hyperplane).**

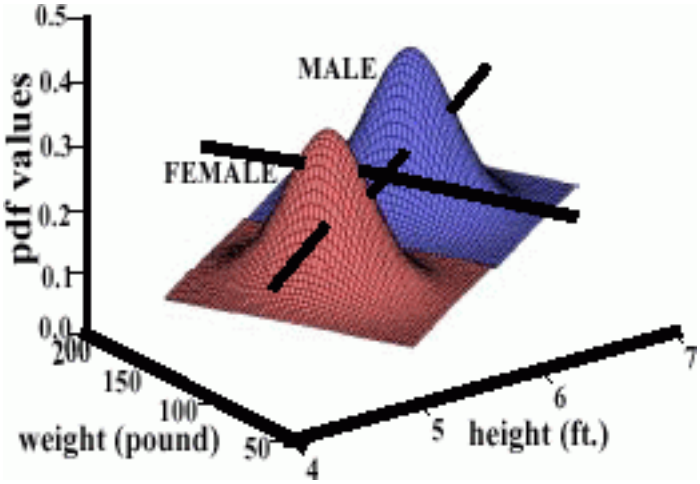
. **Case II: Distributions with unequal variances**



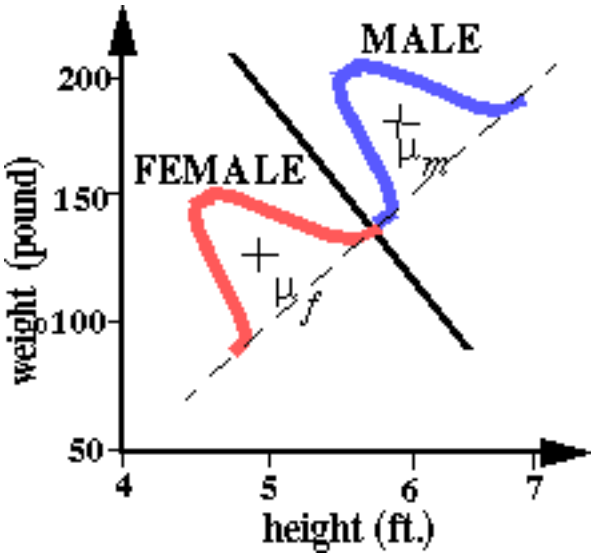
. **Direction of greatest variance is not the direction of discrimination.**

GOAL: MAXIMIZE SEPARABILITY

Principal Component Analysis: Transform features to a new space in which the features are uncorrelated.



Linear Discriminant Analysis: Projection of *d*-dimensional data onto a line; Dimensionality reduction by mapping ***L*** distributions to ***(L-1)***-dimensional subspace; maximize class separability.



OPTIMIZATION CRITERIA BASED ON SCATTER

- Within-class scatter defines the scatter of samples around their respective means.

$$S_w = \sum_{i=1}^{L} P_i E\left[(X-\mu_i)(X-\mu_i)^T\right]$$

- Between-class scatter defines scatter of expected vectors around the global mean.

$$S_b = \sum_{i=1}^{L} P_i\left[(\mu_i-\mu_0)(\mu_i-\mu_0)^T\right]$$

- Mixture-class scatter is the overall scatter obtained from the covariance of all samples:

$$S_m = E\left[(X-\mu_0)(X-\mu_0)^T\right] = S_w + S_b$$

where μ_0 is the overall mean.

- Optimizing criteria used to obtain the transforms is a combination of within-class scatter, between-class scatter and overall scatter.

$$criterion = \frac{tr(S_b)}{tr(S_w)} \qquad criterion = S_b - \mu(S_w - c)$$

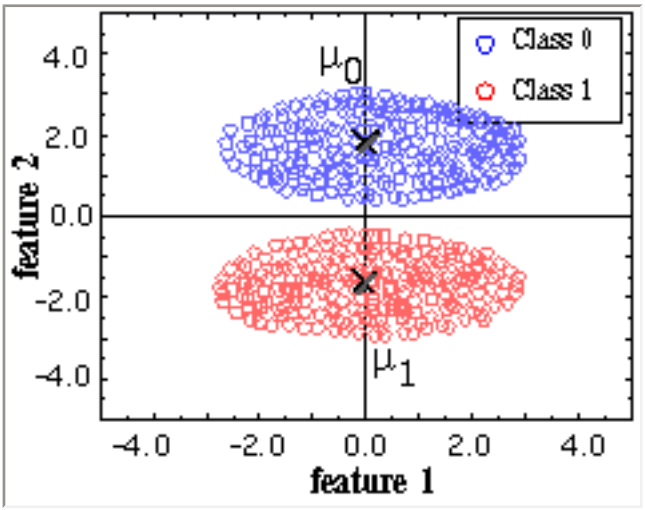
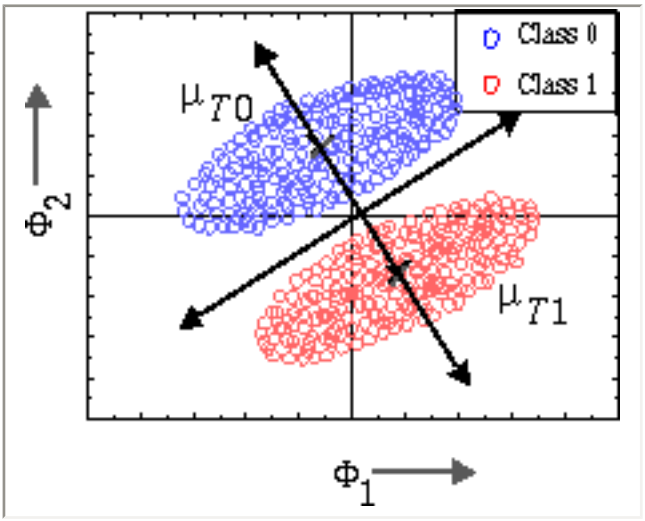
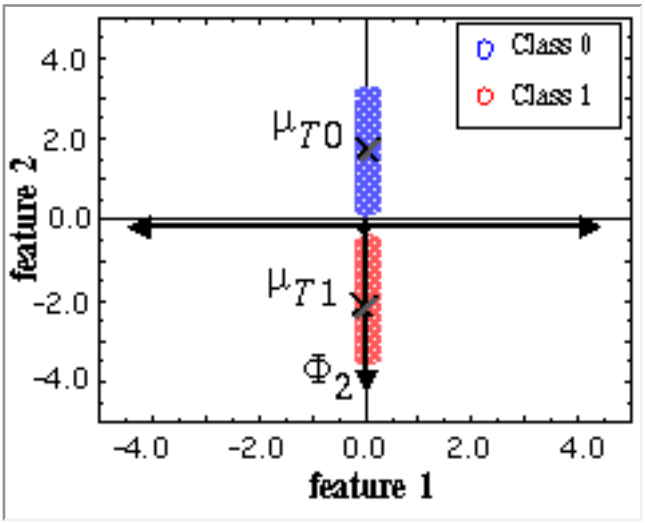
$$criterion = inv(S_w) \times S_b \qquad criterion = inv(S_m) \times S_b$$

- Transformation matrix is given by:

$$T = \Phi$$

where Φ are the eigenvectors corresponding to non-zero eigenvalues.

A COMPARISON OF PCA AND LDA

Original Data	
Class-Independent PCA	
Class-Independent LDA	

LINEAR DISCRIMINANT ANALYSIS - A BRIEF TUTORIAL

S. Balakrishnama, A. Ganapathiraju

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
Box 9571, 216 Simrall, Hardy Rd.
Mississippi State, Mississippi 39762
Tel: 601-325-8335, Fax: 601-325-3149
Email: {balakris, ganapath}@isip.msstate.edu

1. INTRODUCTION

There are many possible techniques for classification of data. Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are two commonly used techniques for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. The use of Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition.We decided to implement an algorithm for LDA in hopes of providing better classification compared to Principle Components Analysis. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between the given classes.This method also helps to better understand the distribution of the feature data. Figure 1 will be used as an example to explain and illustrate the theory of LDA.

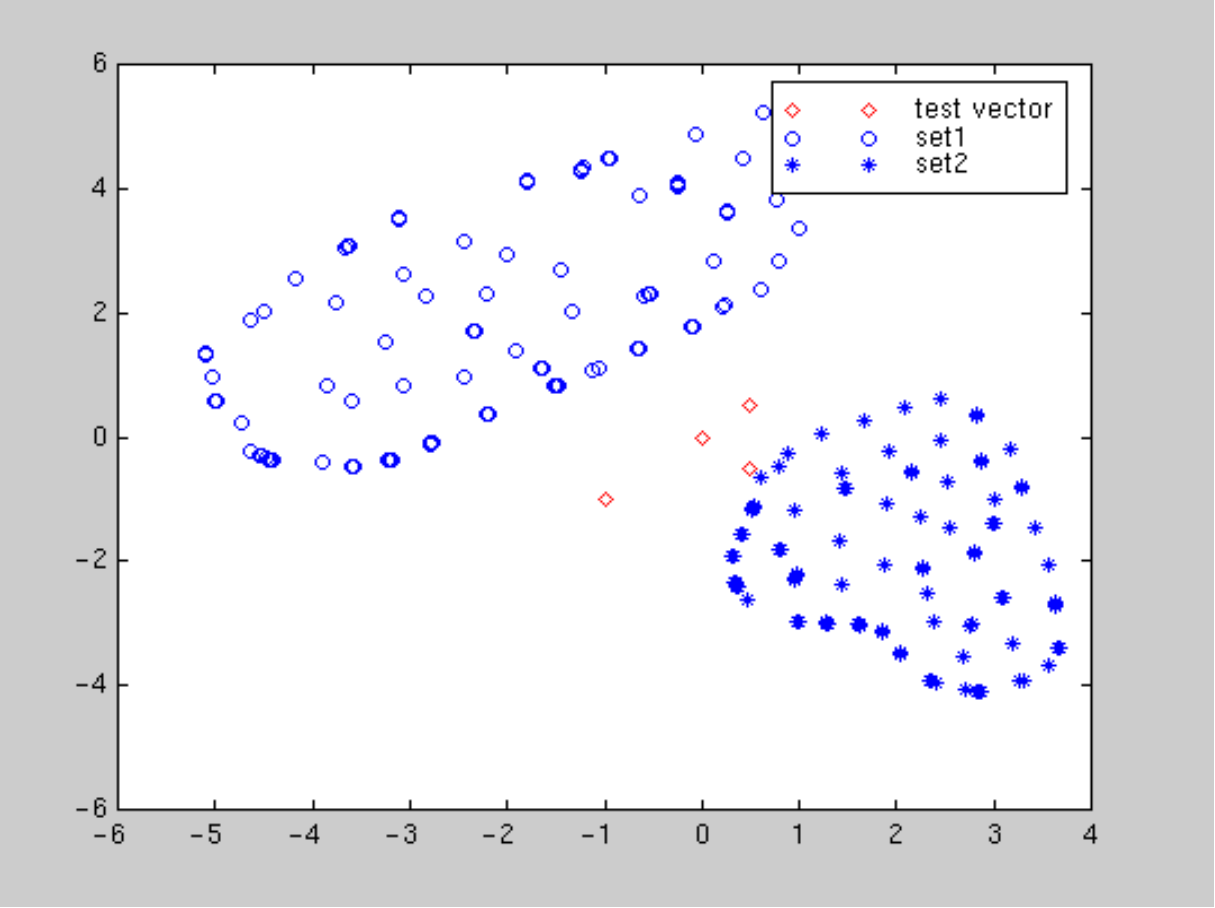


Figure 1. Figure showing data sets and test vectors in original

2. DIFFERENT APPROACHES TO LDA

Data sets can be transformed and test vectors can be classified in the transformed space by two different approaches.

Class-dependent transformation: This type of approach involves maximizing the ratio of between class variance to within class variance. The main objective is to maximize this ratio so that adequate class separability is obtained. The class-specific type approach involves using two optimizing criteria for transforming the data sets independently.

Class-independent transformation: This approach involves maximizing the ratio of overall variance to within class variance. This approach uses only one optimizing criterion to transform the data sets and hence all data points irrespective of their class identity are transformed using this transform. In this type of LDA, each class is considered as a separate class against all other classes.

3. MATHEMATICAL OPERATIONS

In this section, the mathematical operations involved in using LDA will be analyzed the aid of sample set in Figure 1. For ease of understanding, this concept is applied to a two-class problem. Each data set has 100 2-D data points. Note that the mathematical formulation of this classification strategy parallels the Matlab implementation associated with this work.

1. Formulate the data sets and the test sets, which are to be classified in the original space. The given data sets and the test vectors are formulated, a graphical plot of the data sets and test vectors for the example considered in original space is shown in Figure 1. For ease of understanding let us represent the data sets as a matrix consisting of features in the form given below:

$$set1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \dots & \dots \\ \dots & \dots \\ a_{m1} & a_{m2} \end{bmatrix} \qquad set2 = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \dots & \dots \\ \dots & \dots \\ b_{m1} & b_{m2} \end{bmatrix} \tag{1}$$

2. Compute the mean of each data set and mean of entire data set. Let μ_1 and μ_2 be the mean of set 1 and set 2 respectively and μ_3 be mean of entire data, which is obtained by merging set 1 and set 2, is given by Equation 1.

$$\mu_3 = p_1 \times \mu_1 + p_2 \times \mu_2 \tag{2}$$

where p_1 and p_2 are the apriori probabilities of the classes. In the case of this simple two class problem, the probability factor is assumed to be 0.5.

3. In LDA, within-class and between-class scatter are used to formulate criteria for class separability. Within-class scatter is the expected covariance of each of the classes. The scatter measures are computed using Equations 3 and 4.

$$S_w = \sum_j p_j \times (cov_j) \tag{3}$$

Therefore, for the two-class problem,

$$S_w = 0.5 \times cov_1 + 0.5 \times cov_2 \tag{4}$$

All the covariance matrices are symmetric. Let $cov1$ and $cov2$ be the covariance of set 1 and set 2 respectively. Covariance matrix is computed using the following equation.

$$cov_j = (\boldsymbol{x}_j - \mu_j)(\boldsymbol{x}_j - \mu_j)^T \tag{5}$$

The between-class scatter is computes using the following equation.

$$S_b = \sum_j (\mu_j - \mu_3) \times (\mu_j - \mu_3)^T \tag{6}$$

Note that S_b can be thought of as the covariance of data set whose members are the mean vectors of each class. As defined earlier, the optimizing criterion in LDA is the ratio of between-class scatter to the within-class scatter. The solution obtained by maximizing this criterion defines the axes of the transformed space. However for the class-dependent transform the optimizing criterion is computed using equations and (5). It should be noted that if the LDA is a class dependent type, for *L*-class L separate optimizing criterion are required for each class. The optimizing factors in case of class dependent type are computed as

$$criterion_j = inv(cov_j) \times S_b \tag{7}$$

For the class independent transform, the optimizing criterion is computed as

$$criterion = inv(S_w) \times S_b \tag{8}$$

4. By definition, an eigen vector of a transformation represents a 1-D invariant subspace of the vector space in which the transformation is applied. A set of these eigen vectors whose corresponding eigen values are non-zero are all linearly independent and are invariant under the transformation. Thus any vector space can be represented in terms of linear combinations of the eigen vectors. A linear dependency between features is indicated by a

zero eigen value. To obtain a non-redundant set of features all eigen vectors corresponding to non-zero eigen values only are considered and the ones corresponding to zero eigen values are neglected. In the case of LDA, the transformations are found as the eigen vector matrix of the different criteria defined in Equations 7 and 8.

5. For any *L*-class problem we would always have *L-1* non-zero eigen values. This is attributed to the constraints on the mean vectors of the classes in Equation 2. The eigen vectors corresponding to non-zero eigen values for the definition of the transformation.

   For our 2-class example, Figures 2 and 3 show the direction of the significant eigen vector along which there is maximum discrimination information. Having obtained the transformation matrices, we transform the data sets using the single LDA transform or the class specific transforms which ever the case may be. From the figures it can be observed that, transforming the entire data set to one axis provides definite boundaries to classify the data. The decision region in the transformed space is a solid line separating the transformed data sets thus

   For the class dependent LDA,

$$transformed\_set\_j = transform\_j^{T} \times set\_j \tag{9}$$

   For the class independent LDA,

$$transformed\_set = transform\_spec^{T} \times data\_set^{T} \tag{10}$$

   Similarly the test vectors are transformed and are classified using the euclidean distance of the test vectors from each class mean.

   The two Figures 4 and 5 clearly illustrate the theory of Linear Discriminant Analysis applied to a 2-class problem. The original data sets are shown and the same data sets after transformation are also illustrated. It is quite clear from these figures that transformation provides a boundary for proper classification. In this example the classes were properly defined but cases where there is overlap between classes, obtaining a decision region in original space will be very difficult and in such cases transformation proves to be very essential. Transformation along largest eigen vector axis is the best transformation.

   Figures 6 and 7, are interesting in that they show how the linear transformation process can be viewed as projecting data points onto the maximally discriminating axes represented by the eigen vectors.

6. Once the transformations are completed using the LDA transforms, Euclidean distance or RMS distance is used to classify data points. Euclidean distance is computed using Equation 11 where μ_{ntrans} is the mean of the transformed data set, n is the class index and $\boldsymbol{x}$ is the test vector. Thus for n classes, n euclidean distances are obtained for each test point.

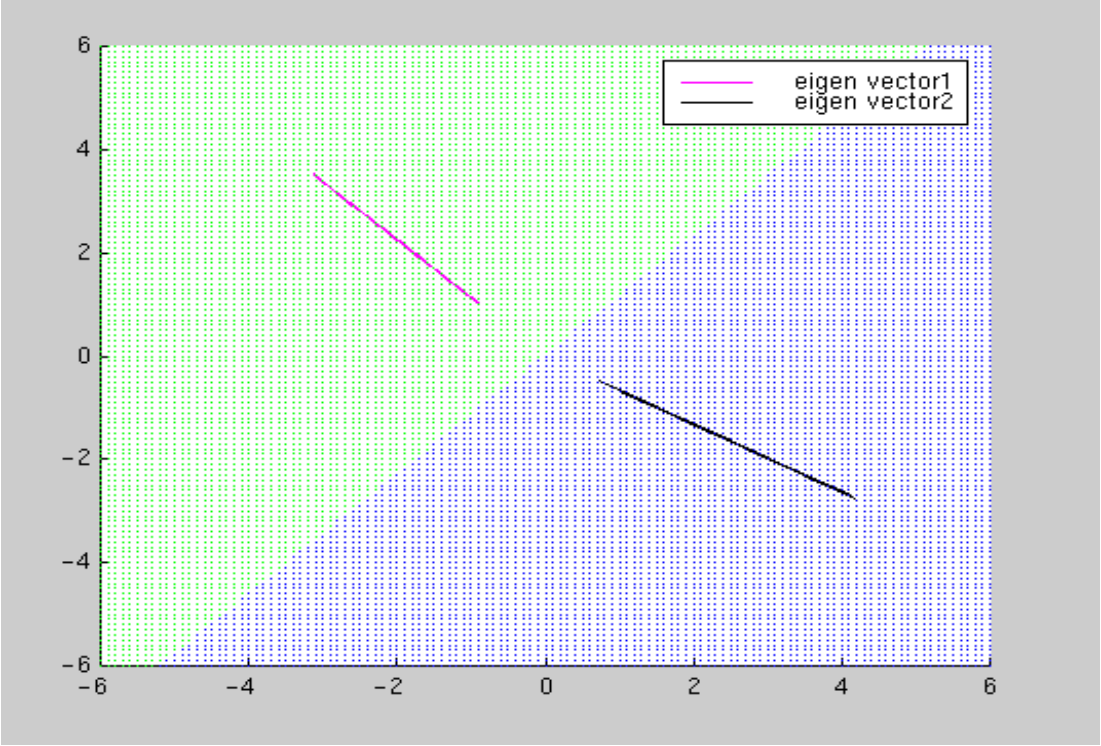


Figure 2. Figure for eigen vector direction in class dependent type

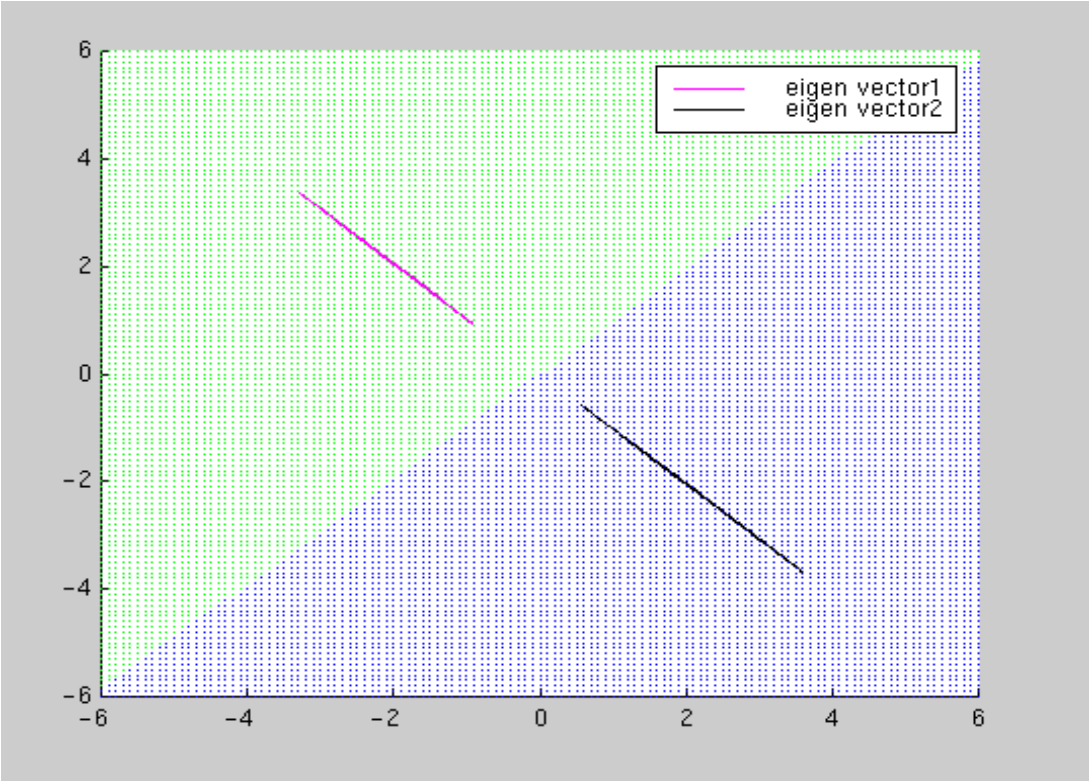


Figure 3. Figure for eigen vector direction in class independent type

$$dist\_n = (transform\_n\_spec)^T \times \boldsymbol{x} - \mu_{ntrans} \quad (11)$$

7. The smallest Euclidean distance among the n distances classifies the test vector as belonging to class n.

4. CONCLUSIONS

We have presented the theory and implementation of LDA as a classification technique. Throughout the tutorial we have used a 2-class problem as an exemplar. Two approaches to LDA, namely, class independent and class dependent, have been explained. The choice of the type of LDA depends on the data set and the goals of the classification problem. If generalization is of importance, the class independent tranformation is preferred. However, if good discrimination is what is aimed for, the class dependent type should be the first choice. As part of our future work, we plan to work on a Java-based demonstration which could be used to visualize LDA based transformations on user defined data sets and also help the user apperiaciate the difference between the various classification techniques.

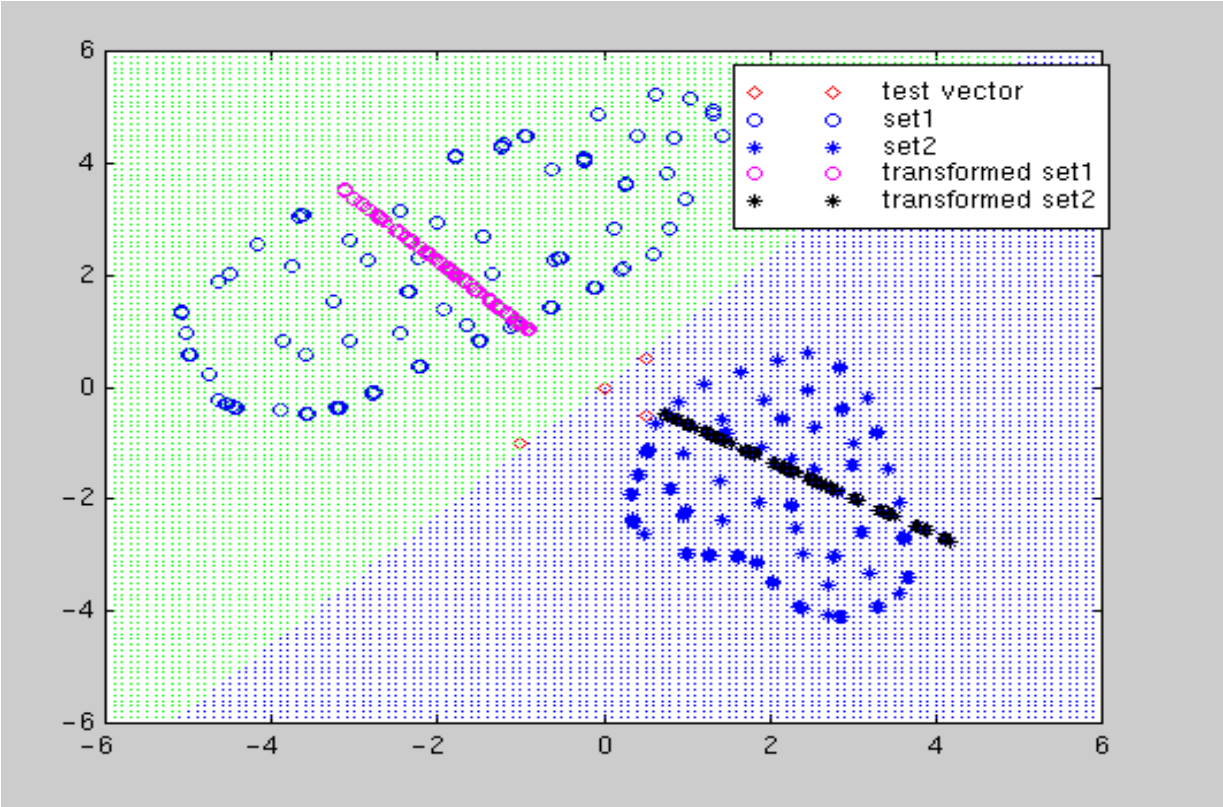


Figure 4. Data sets in original space and transformed space along with the tranformation axis for class dependent LDA of a 2-class problem

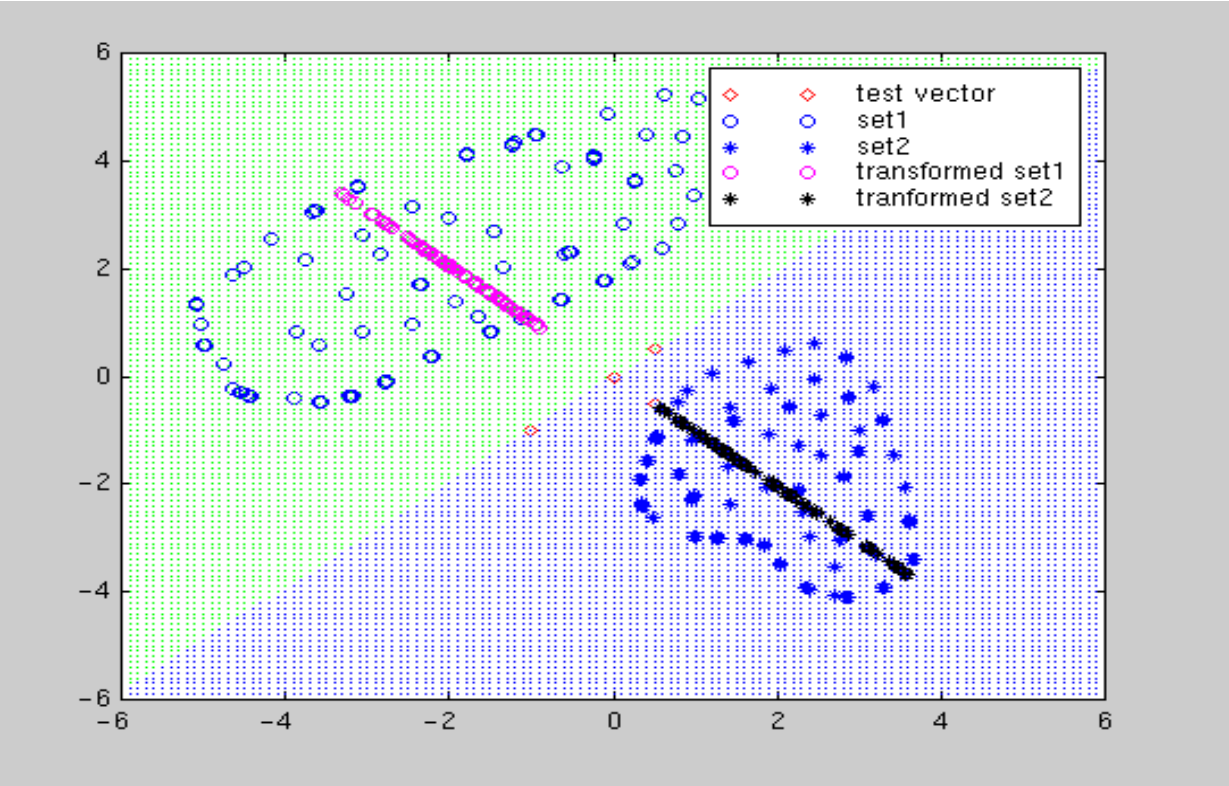


Figure 5. Data sets in original space and transformed space for class independent type of LDA of a 2-class problem

5. SOFTWARE

All Matlab code written for this project is available for public from our website at *www.isip.msstate.edu*

6. REFERENCES

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

[2] S. Axler, *Linear Algebra Done Right*, Springer-Verlag New York Inc., New York, New York, 1995.

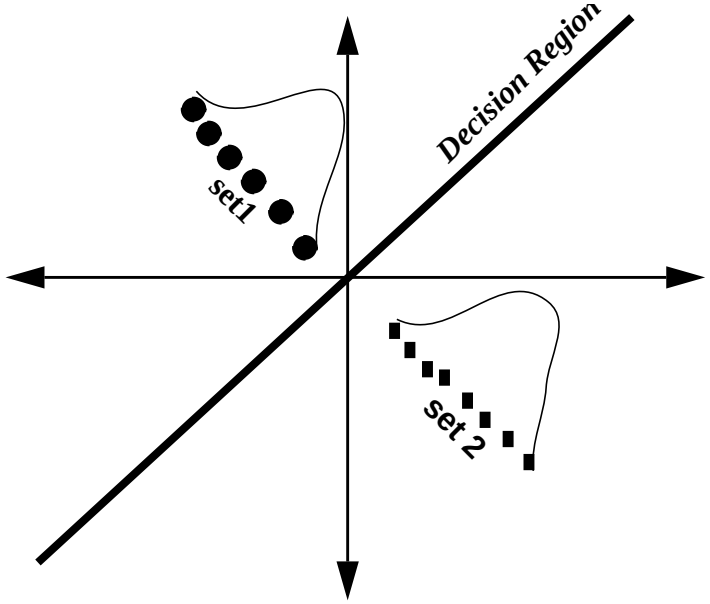


Figure 6. Figure showing histogram plot of transformed data with decision region in class independent type and the amount of class separability obtained in transformed space

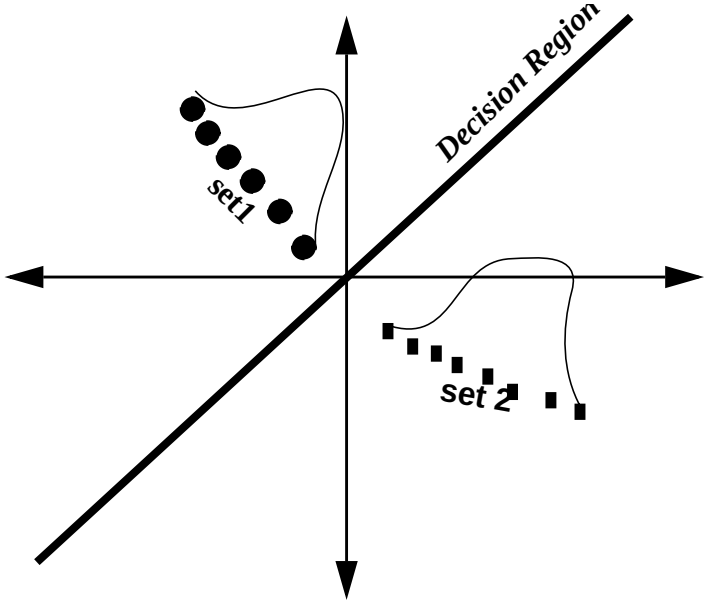


Figure 7. Histogram plot of transformed data with decision region in class dependent type and the amount of class separability obtained in transformed space

Paris' Independent Component Analysis & Blind Source Separation page

Independent Component Analysis (ICA) and Blind Source Separation (BSS) have received quite a lot of attention lately so that I decided to compile a list of online resources for whoever is interested. By no means is this page complete and if you have any additions do send me mail. In the papers section I do not list all of the papers of every author (that's why you should check their homepages) but the really good ones are here. Also not all ICA & BSS people have home pages so if you discover any or if I missed yours tell me about it and I'll add them.

People working on ICA & BSS

- Pierre Comon, one of the first people to look at ICA, at the Laboratoire I3S of Universit? de Nice - Sophia Antipolis
- Tony Bell at the Salk Institute's Computational Neurobiology Lab
- Shun-ichi Amari at RIKEN's Lab for Information Synthesis
- Andrzej Cichocki at RIKEN's Laboratory for Open Information Systems
- Wlodzimierz Kasprzak at RIKEN's Laboratory for Open Information Systems
- Kari Torkkola at Motorola
- Erkki Oja at Helsinki University of Technology Laboratory of Computer and Information Science
- Juha Karhunen at Helsinki University of Technology Laboratory of Computer and Information Science
- Aapo Hyvarinen at Helsinki University of Technology Laboratory of Computer and Information Science
- Jean-François Cardoso at the Ecole Nationale Supérieure des Télécommunications Signal Department
- Barak Pearlmutter at the University of New Mexico CS Department
- Mark Girolami at the University of Paisley Department of Computing and Information Systems (Looking for students)
- Te-Won Lee at the Salk Institute's Computational Neurobiology Lab
- Adel Belouchrani at Villanova University's Department of Electrical and Computer Engineering
- Simon Godsill and Dominic Chan at the Signal Processing and Communications Group at Cambridge University
- Juergen Schmidhuber at IDSIA
- Henrik Sahlin at Chalmers University of Technology.
- Seungjin Choi at Pohang University of Science and Technology.
- Gustavo Deco at Siemens.
- Hans van Hateren at the University of Groningen.
- Daniël Schobben formerly at the Eindhoven University of Technology now at Phillips
- Vicente Zarzoso at the University of Strathclyde
- Kevin Knuth at the Albert Einstein College of Medicine
- Russ Lambert doing very cool things with sTrAnGe matrices!
- Alex Westner a brave man who scoffs at the complexity of real-world mixtures
- Daniel Rowe at Caltech, doing Bayesian BSS
- Michael Zibulevsky at the University of New Mexico
- Na Kyungmin at the Seoul National University
- Lieven De Lathauwer at the Katholieke Universiteit Leuven
- Scott Rickard at Princeton
- Lucas Parra at the Sarnoff Corp.
- Simone Fiori at the Perugia University
- Carlos G. Puntonet at the University of Granada
- Andreas Ziehe at GMD FIRST
- Fathi Salam at the Circuits, Systems and Artificial Neural Networks Laboratory, Michigan State University

Other ICA Pages

- Te-Won Lee's ICA-CNL Page.
- Allan Barros' ICA Page.
- Tony Bell's ICA Page.
- Daniel Rowe's Bayesian BSS Page.

Benchmarks

- Daniël Schobben, Kari Torkkola and me maintain these:
  - Real world benchmarks
  - Synthetic benchmarks

Code and Software ...

- Code from Pierre Comon.
- The FastICA package by Aapo Hyvärinen (very cool, get it!)
- OOLABSS by Cichocki and Orsier.
- By Tony Bell, Here
- By Jean-François Cardoso:
  - The EASI Algorithm.
  - The JADE Algorithm (and its calling program).
- By Dominic Chan, Here
- RICA by Cichocki and Barros.
- Genie by Na Kyungmin.
- By me:
  - Mostly old instantaneous ICA code (in MATLAB)

Online Demos of BSS

- Barak Pearlmutter's Demo on Contextual ICA
- Te-Won Lee's Demo
- Dominic Chan's Demo
- My little frequency domain algorithm: (This is actually a static mix, I just put it up cause it sounds cool, but the algorithm can deal with convolved mixtures too)
  - The input sound.
  - The speech output (and in slow motion)
  - The noise output (and in slow motion)
- Henrik Sahlin's demo.
- Hans van Hateren's demo on images.
- Daniël Schobben's demo page
- Shiro Ikeda's demos
- Alex Westner's demos
- Scott Rickard's demo page
- Christian Jutten's JAVA demo

Online Papers on ICA & BSS

(since I don't really sit around all day playing with this page, there some links that are extinct by now. Rather than giving up, check out the home page of the corresponding author in the top of the page. You are most likely to find their papers there. You are also most likely to find their newer papers there too).

- **Yes, I actually finished my dissertation!**
  - Smaragdis, P. 2001. Redundancy reduction for computational audition, a unifying approach. *Ph.D. dissertation*, MAS department, Massachusetts Institute of Technology.
  - Smaragdis, P. 1997. Information Theoretic Approaches to Source Separation, *Masters Thesis*, MAS Department, Massachusetts Institute of Technology.
    I realize that the name of the thesis is not terribly enlightening. What happens in there, apart from the obligatory background stuff, is the development of a new frequency domain separation algorithm to deal with convolved mixtures. The idea is to move to a space where the separation parameters are orthogonal, to assist convergence, and to be able to implement at the same time faster convolution schemes. In addition to this the algorithm is on-line and real-time so that you can actually use it. Results are nice too!
  - Smaragdis, P. 1997. Efficient Blind Separation of Convolved Sound Mixtures,IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz NY, October 1997.
    Pretty much the same material, geared towards DSP-heads. Written before my thesis so it is a little outdated.
  - Smaragdis, P. 1998. Blind Separation of Convolved Mixtures in the Frequency Domain. International Workshop on Independence & Artificial Neural Networks University of La Laguna, Tenerife, Spain, February 9 - 10, 1998.
    Condenced version of my thesis. Most up to date compared to my other offerings.
- **Tony Bell has some neat papers on Blind Source Separation:**
  - Bell A.J. & Sejnowski T.J. 1995. An information-maximization approach to blind separation and blind deconvolution, Neural Computation, 7, 1129-1159.
  - Bell A.J. & Sejnowski T.J. 1995. Fast blind separation based on information theory, in *Proc. Intern. Symp. on Nonlinear Theory and Applications*, vol. 1, 43-47, Las Vegas, Dec. 1995

  And a couple of papers on ICA alone:
  - Bell A.J. & Sejnowski T.J. 1996. Learning the higher-order structure of a natural sound, *Network: Computation in Neural Systems*, to appear
  - Bell A.J. & Sejnowski T.J. 1996. The `Independent Components' of natural scenes are edge filters, *Vision Research*, under review [Please note that this is a draft].
- **Kari Torkkola has some practical papers on simultaneous Blind Source Separation and Deconvolution:**
  - Torkkola, K.:Blind Separation of Delayed Sources Based on Information Maximization. *Proceedings of the IEEE Conference on Acoustics, Speech and Signal Processing*, May 7-10 1996, Atlanta, GA, USA.
  - Torkkola, K.:Blind Separation of Convolved Sources Based on Information Maximization. *IEEE Workshop on Neural Networks for Signal Processing*, Sept 4-6 1996, Kyoto, Japan.
  - Torkkola, K.:IIR Filters for Blind Deconvolution Using Information Maximizationrm. *NIPS96 Workshop: Blind Signal Processing and Their Applications*, Snowmaas (Aspen), Colorado.
- **Barak Pearlmutter has a paper on context sensitive ICA:**
  - Barak A. Pearlmutter and Lucas C. Parra. A context-sensitive generalization of ICA. 1996 International Conference on Neural Information Processing. September 1996, Hong Kong.
- **Erkki Oja has papers on PCA, nonlinear PCA and ICA:**
  - Oja, E., Karhunen, J., Wang, L., and Vigario, R.:Principal and independent components in neural networks - recent developments. *Proc. VII Italian Workshop on Neural Nets WIRN'95*, May 18 - 20, 1995, Vietri sul Mare, Italy (1995).
  - Oja, E.:The nonlinear PCA learning rule and signal separation - mathematical analysis. Helsinki University of Technology, Laboratory of Computer and Information Science, *Report A26* (1995).
  - Oja, E. and Taipale, O.:Applications of learning and intelligent systems - the Finnish technology programme. *Proc. Int. Conf. on Artificial Neural Networks ICANN-95, Industrial Conference*, Oct. 9 - 13, 1995, Paris, France (1995).
  - Oja, E.: PCA, ICA, and nonlinear Hebbian learning. *Proc. Int. Conf. on Artificial Neural Networks ICANN-95*, Oct. 9 - 13, 1995, Paris, France, pp. 89 - 94 (1995).
  - Oja, E. and Karhunen, J.:Signal separation by nonlinear Hebbian learning. In M. Palaniswami, Y. Attikiouzel, R. Marks II, D. Fogel, and T. Fukuda (Eds.), *Computational Intelligence - a Dynamic System Perspective*. New York: IEEE Press, pp. 83 - 97 (1995).
- **Juha Karhunen has written ICA & BSS papers with Oja (right above) and some on his own:**
  - Karhunen, J.:Neural Approaches to Independent Component Analysis and Source Separation. *To appear in Proc. 4th European Symposium on Artificial Neural Networks (ESANN'96)*, April 24 - 26, 1996, Bruges, Belgium (invited paper).
  - Karhunen, J., Wang, L., and Vigario, R.,Nonlinear PCA Type Approaches for Source Separation and Independent Component Analysis*Proc. of the 1995 IEEE Int. Conf. on Neural Networks (ICNN'95)*, Perth, Australia, November 27 - December 1, 1995, pp. 995-1000.
  - Karhunen, J., Wang, L., and Joutsensalo, J.,Neural Estimation of Basis Vectors in Independent Component Analysis *Proc. of the Int. Conf. on Artificial Neural Networks (ICANN'95)*, Paris, France, October 9-13, 1995, pp. 317-322.
- **Andrzej Cichocki organized a special invited session on BSS in Nolta '95 and has a nice list of papers on the subject:**
  (another apparently defunct set of links ...)
  - Shun-ichi Amari, Andrzej Cichocki and Howard Hua Yang, *"Recurrent Neural Networks for Blind Separation of Sources",* , pp.37-42.
  - Anthony J. Bell and Terrence J. Sejnowski, *"Fast Blind Separation based on Information Theory",* pp. 43-47.
  - Adel Belouchrani and Jean-Francois Cardoso, *"Maximum Likelihood Source Separation by the Expectation-Maximization Technique: Deterministic and Stochastic Implementation",* pp.49-53.
  - Jean-Francois Cardoso, *"The Invariant Approach to Source Separation",* pp. 55-60.
  - Andrzej Cichocki, Wlodzimierz Kasprzak and Shun-ichi Amari, *"Multi-Layer Neural Networks with Local Adaptive Learning Rules for Blind Separation of Source Signals",* , pp.61-65.
  - Yannick Deville and Laurence Andry, *"Application of Blind Source Separation Techniques to Multi-Tag Contactless Identification Systems",* , pp. 73-78.
  - Jie Huang , Noboru Ohnishi and Naboru Sugie *"Sound SeparatioN Based on Perceptual Grouping of Sound Segments",* , pp.67-72.
  - Christian Jutten and Jean-Francois Cardoso, *"Separation of Sources: Really Blind ?"* , pp. 79-84.
  - Kiyotoshi Matsuoka and Mitsuru Kawamoto, *"Blind Signal Separation Based on a Mutual Information Criterion",* pp. 85-91.
  - Lieven De Lathauwer, Pierre Comon, Bart De Moor and Joos Vandewalle, *"Higher-Order Power Method - Application in Independent Component Analysis"* , pp. 91-96.
  - Jie Zhu, Xi-Ren Cao, and Ruey-Wen Liu, *"Blind Source Separation Based on Output Independence - Theory and Implementation"* , pp. 97-102.

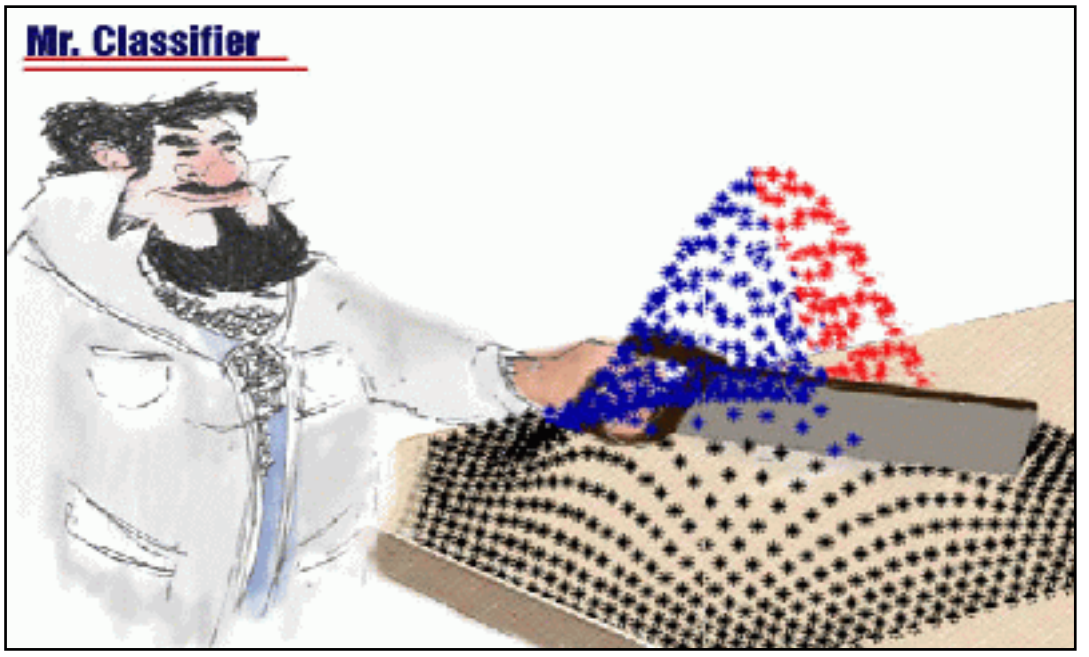
  Papers are included in *Proceedings 1995 International Symposium on Nonlinear Theory and Applications NOLTA'95, Vol.1,* NTA Research Society of IEICE, Tokyo, Japan, 1995.
- **Shun-ichi Amari wrote some excelent papers with the RIKEN people on BSS and the math behind it:**
  - S. Amari, A. Cichocki and H. H. Yang, A New Learning Algorithm for Blind Signal Separation (128K), In: Advances in Neural Information Processing Systems 8, Editors D. Touretzky, M. Mozer, and M. Hasselmo, pp.?-?(to appear), MIT Press, Cambridge MA, 1996.
  - Shun-ichi Amari, Neural Learning in Structured Parameter Spaces , NIPS'96
  - Shun-ichi Amari, Information Geometry of Neural Networks - New Bayesian Duality Theory - , ICONIP'96
  - Shun-ichi Amari, Gradient Learning in Structured Parameter Spaces: Adaptive Blind Separation of Signal Sources , WCNN'96
  - Shun-ichi Amari and Jean-Francois Cardoso, Blind Source Separation - Semiparametric Statistical Approach, sumitted to IEEE Tr. on Signal Processing.
  - Shun-ichi Amari, Natural Gradient Works Efficiently in Learning, sumitted to Neural Computation.
  - Howard Hua Yang and Shun-ichi Amari, Adaptive On-Line Learning Algorithms for Blind Separation - Maximum Entropy and Minimum Mutual Information , accepted for Neural Computation.
  - Shun-ichi Amari, Tian-Ping CHEN, Andrzej CICHOCKI, Stability Analysis of Adaptive Blind Source Separation , accepted for Neural Neworks.
  - Shun-ichi Amari, Superefficiency in Blind Source Separation , sumitted to IEEE Tr. on Signal Processing.
  - Shun-ichi Amari and Noboru Murata, Statistical Analysis of Regularization Constant - From Bayes, MDL and NIC Points of View, International Work-Donf. on Artificial and Natural Neural Networks 97.
  - Shun-ichi Amari, Geometry of Semiparametric Models and Applications , ISI'97
- **Jean-François Cardoso has lots (and lots, and lots, ...) of papers on BSS!**
  - Jean-François Cardoso and Beate Laheld. Equivariant adaptive source separation. To appear in IEEE Trans. on S.P.
  - Jean-François Cardoso. Performance and implementation of invariant source separation algorithms In *Proc. ISCAS'96*, 1996.
  - Jean-François Cardoso, Sandip Bose, and Benjamin Friedlander. On optimal source separation based on second and fourth order cumulants In *Proc. IEEE Workshop on SSAP*, Corfou, Greece, 1996.
  - Jean-François Cardoso. The equivariant approach to source separation In *Proc. NOLTA*, pages 55-60, 1995.
  - Jean-François Cardoso. Séparation de sources dans l'espace signal In *Proc. GRETSI, Juan les Pins,France*, 1995.
  - Jean-François Cardoso. A tetradic decomposition of 4th-order tensors: application to the source separation problem In M. Moonen and B. de Moor, editors, *Algorithms, architectures and applications*, volume III of *SVD and signal processing*, pages 375-382. Elsevier, 1995.
  - Jean-François Cardoso, Sandip Bose, and Benjamin Friedlander. Output cumulant matching for source separation In *Proc. IEEE SP Workshop on Higher-Order Stat., Aiguablava, Spain*, pages 44-48, 1995.
  - Adel Belouchrani and Jean-François Cardoso. Maximum likelihood source separation for discrete sources In *Proc. EUSIPCO*, pages 768-771, Edinburgh, September 1994.
  - Jean-François Cardoso. On the performance of source separation algorithms In *Proc. EUSIPCO*, pages 776-779, Edinburgh, September 1994.
  - Jean-François Cardoso, Adel Belouchrani, and Beate Laheld. A new composite criterion for adaptive and iterative blind source separation In *Proc. ICASSP*, volume 4, pages 273-276, April 1994.
  - Beate Laheld and Jean-François Cardoso. Adaptive source separation with uniform performance In *Proc. EUSIPCO*, pages 183-186, Edinburgh, September 1994.
  - Jean-François Cardoso and Antoine Souloumiac. An efficient technique for blind separation of complex sources In *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, pages 275-279, 1993.
  - Jean-François Cardoso. Iterative techniques for blind source separation using only fourth order cumulants In *Proc. EUSIPCO*, pages 739-742, 1992.
  - Jean-François Cardoso and Beate Laheld. Adaptive blind source separation for channel spatial equalization In *Proc. of COST 229 workshop on adaptive signal processing*, pages 19-26, 1992.
  - Jean-François Cardoso. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem In *Proc. ICASSP*, pages 2655-2658, 1990.
  - Jean-François Cardoso. Source separation using higher order moments In *Proc. ICASSP*, pages 2109-2112, 1989.
  - Jean-François Cardoso and Pierre Comon. Independent component analysis, a survey of some algebraic methods In *Proc. ISCAS'96*, vol.2, pp. 93-96, 1996.
  - Jean-François Cardoso, Infomax and maximum likelihood for source separation,. To appear in IEEE Letters on Signal Processing, April, 1997.
- **Cichocki & Kasprzak have a nice collection of papers:**
  - Cichocki A., Kasprzak W.: *Nonlinear Learning Algorithms for Blind Separation of Natural Images* , **Neural Network World**, vol.6, 1996, No.4, IDG Co., Prague, 515-523.
  - Cichocki A., Kasprzak W., Amari S.-I.: *Neural Network Approach to Blind Separation And Enhancement of Images* , **EUSIPCO'96** , (Trieste, Italy, September 1996).
  - Kasprzak W., Cichocki A.: *Hidden Image Separation From Incomplete Image Mixtures by Independent Component Analysis* , **ICPR'96** , Vienna, August 1996.
  - Cichocki A., Amari S., Adachi M., Kasprzak W.: *Self-Adaptive Neural Networks for Blind Separation of Sources* , **1996 IEEE International Symposium on Circuits and Systems, ISCAS'96, Vol. 2**, IEEE, Piscataway, NJ, 1996, 157-160.
- **Mark Girolami at University of Paisley has some papers too:**
  - Girolami, M and Fyfe, C. Blind Separation of Sources Using Exploratory Projection Pursuit Networks. *Speech and Signal Processing, International Conference on the Engineering Applications of Neural Networks,* ISBN 952-90-7517-0, London, pp249-252, 1996.
  - Girolami, M and Fyfe, C. Higher Order Cumulant Maximisation Using Nonlinear Hebbian and Anti-Hebbian Learning for Adaptive Blind Separation of Source Signals, *Proc IWSIP-96, IEEE/IEE International Workshop on Signal and Image Processing, Advances in Computational Intelligence,* Elsevier Science, pp141 - 144, Manchester, 4-7 November 1996.
  - Girolami, M and Fyfe, C. Multivariate Density Factorisation for Independent Component Analysis : An Unsupervised Artificial Neural Network Approach, *AISTATS-97, 3'rd International Workshop on Artificial Intelligence and Statistics,* Fort Lauderdale, Florida, Jan 1997.
  - Girolami, M and Fyfe, C. Negentropy and Kurtosis as Projection Pursuit Indices Provide Generalised ICA Algorithms, *NIPS-96 Blind Signal Separation Workshop, (Org A. Cichocki & A.Back)*, Aspen, Colorado, 7 Dec, 1996.
  - Girolami, M and Fyfe, C. A Temporal Model of Linear Anti-Hebbian Learning, *Neural Processing Letters*, In Press Vol 4, Issue 3, Jan 1997.
- **Te-Won Lee at the Salk Institute has some interesting papers:**
  - Blind separation of delayed and convolved sources, *T.W. Lee and A.J. Bell and R. Lambert,* accepted for publication in "Advances in Neural Information Processing Systems", MIT Press, Cambridge MA, 1996
  - Blind Source Separation of Real World Signals, *T.W. Lee, A.J. Bell and R. Orglmeister,* to appear in "IEEE International Conference Neural Networks ", Houston, 1997
- **Aapo Hyvarinen has a nice bunch of papers:**
  - A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks* 10(3):626-634, 1999.
  - A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation,* 9(7):1483-1492, 1997.
  - A. Hyvärinen. New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit. In *Advances in Neural Information Processing Systems 10* (NIPS*97), pp. 273-279, MIT Press, 1998.
- **Juergen Schmidhuber at IDSIA has related papers:**
  - S. Hochreiter and J. Schmidhuber. LOCOCODE. TR FKI-222-97, June 1997.
  - J. Schmidhuber and M. Eldracher and B. Foltin. Semilinear predictability minimzation produces well-known feature detectors. Neural Computation, 8(4):773-786, 1996.
  - J. Schmidhuber and D. Prelinger. Discovering predictable classifications. Neural Computation, 5(4):625-635, 1993.
  - J. Schmidhuber. Learning factorial codes by predictability minimization. Neural Computation, 4(6):863-879, 1992.
- **Henrik Sahlin's papers on separation and second order statistics:**
  - H. Broman, U. Lindgren, H. Sahlin and P. Stoica ``Source Separation: A TITO system identification approach", Tech. rep. CTH-TE-33, Department of Applied Electronics, Chalmers University of Technology, Sept. 1995, Submitted to IEEE Trans. SP.
  - H. Sahlin and U. Lindgren "The Asymptotic Cramer-Rao Lower Bound for Blind Signal Separation". "The Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing", Corfu, Greece, 1996
  - H. Sahlin "Asymptotic parameter variance analysis for Blind Signal Separation" "The proceedings of RVK96", Lulea, Sweden, 1996
  - H. Sahlin and H. Broman "Blind Separation of Images" "The Proceedings of Asilomar Conference on Signals, Systems, and Computers", Pacific Grove, CA, USA, 1996.
  - U. Lindgren, H. Sahlin and H. Broman "Source separation using second order statistics" "The Proceedings of EUSIPCO-96", Trieste, Italy, 1996.
  - H. Sahlin, U. Lindgren and H. Broman "Multi Input Multi Output Blind Signal Separation Using Second Order Statistics" Tech. rep. CTH-TE-54, Department of Applied Electronics, Chalmers University of Technology, Dec., 1996.
  - H. Sahlin "On Signal Separation by Second Order Statistics" Licentiate Thesis, Technical Report No. 251L, Department of Applied Electronics, Chalmers University of Technology, 1997.
  - H. Sahlin and H. Broman "Signal Separation Applied to Real World Signals" Proceedings of International Workshop on Accoustic Echo and Noise Control, London, UK, September,1997.
- **Gustavo Deco and Dragan Obradovic have an excellent book on ICA and related matters:**
  - An Information-Theoretic Approach to Neural Computing
  - Second link
- **Kevin Knuth on ICA:**
  - Bayesian source separation and localization. To be published in: SPIE'98 Proceedings: Bayesian Inference for Inverse Problems, San Diego, July 1998.

    [pdf475K], [doc500K].
  - Knuth K.H. 1999. A Bayesian approach to source separation. In: J.-F. Cardoso, C. Jutten and P. Loubaton (eds.), Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99, Aussios, France, Jan. 1999, pp. 283-288.
  - Knuth K.H. and Vaughan H.G., Jr. 1998. Convergent Bayesian formulations of blind source separation and electromagnetic source estimation. Presented at the MaxEnt98 workshop in Munich, July 1998.
- **Alex Westner has worked on real world experiments:**
  - Westner, Alex. (1999) Object-Based Audio Capture: Separating Acoustically-Mixed Sounds. M.S. Thesis, Massachusetts Institute of Technology
  - Westner, Alex and V. Michael Bove, Jr. (1999) cBlind separation of real world audio signals using overdetermined mixtures. To appear in Proceedings of ICA'99, Jan. 11-15, Aussois, France
- **Roberto Manduchi on ICA and textures:**
  - R. Manduchi, J. Portilla, "Independent Component Analysis of Textures", accepted for presentation at IEEE International Conference on Computer Vision, Kerkyra, Greece, September 1999
- **Michael Zibulevsky:**
  - Zibulevsky, M. and Pearlmutter, B.A. "Blind Source Separation by Sparse Decomposition", Technical Report CS99-1, Computer Science Department, University of New Mexico.
- **Lucas Parra:**
  - Lucas Parra, Clay Spence, "On-line convolutive source separation of non-stationary signals", Journal of VLSI Signal Processing, Special issue on the 1998 IEEE Neural Networks and Signal Processing Workshop, to appear in summer 2000, (.ps.gz 249K, .pdf 482K)
  - Lucas Parra, Klaus-Robert Mueller, Clay Spence, Andreas Ziehe, Paul Sajda, "Unmixing Hyperspectral Data", Advances in Neural Information Processing Systems 12, MIT Press, to appear 1999. (.ps.gz 76K)
- **Simone Fiori has some papers on ICA by the adaptive activation function networks and learning on Stiefel-Grassman manifold:**
  - S. Fiori, "Entropy Optimization by the PFANN Network: Application to Independent Component Analysis", Network: Computation in Neural Systems, Vol. 10, No. 2, pp. 171 - 186, May 1999
  - S. Fiori, "'Mechanical' Neural Learning for Blind Source Separation", Electronics Letters, Vol. 35, No. 22, pp. 1963 - 1964, Oct. 1999
  - S. Fiori, "Blind Separation of Circularly Distributed Source Signals by the Neural Extended APEX Algorithm", Neurocomputing, Vol. 34, No. 1-4, pp. 239 - 252, August 2000
  - S. Fiori, "Blind Signal Processing by the Adaptive Activation Function Neurons", Neural Networks, Vol. 13, No. 6, pp. 597 - 611, August 2000
  - S. Fiori, "A Theory for Learning by Weight Flow on Stiefel-Grassman Manifold", Neural Computation. Accepted for publication: To appear.

Paris Smaragdis <paris@media.mit.edu>

MIT Media Lab
20 Ames Street, Rm E15-401
Cambridge, MA 02139

STATISTICAL NORMALIZATION FUNCTIONS FOR SIGNAL PROCESSING PROBLEMS



Suresh Balakrishnama
Candidate for Master of Science in Electrical Engineering
Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University, Mississippi State, MS 39762
Email: balakris@isip.msstate.edu
http://www.isip.msstate.edu/publications/seminars/masters_oral/1999/lda/index.html

PATTERN RECOGNITION

"All applets on our web site now require the java plugin to run. This is necessary so we can bring state-of-the-art features to you which are not currently supported by browser vendors. You can find the plugin at http://java.sun.com/products/plugin. For additional information or suggestions please contact help@isip.msstate.edu"
No JDK 1.2 support for APPLET!!

- Source Code: Download the source code for this applet.
- Tutorial: Learn how to use this applet.

All applets on our web site now require a Java plug-in. This is necessary so we can bring state-of-the-art Java features to you which are not currently supported by browser vendors such as Netscape. You can find the appropriate plug-in at http://java.sun.com/products/plugin. We have generated a list of steps necessary for installing the plug-in in a Unix environment. For additional information or help with your installation please contact help@isip.msstate.edu.

Lincoln Laboratory
Massachusetts Institute of Technology
244 Wood Street
Lexington, MA 02420-9108
Phone (781) 981-5500

webmaster@ll.mit.edu

Information Systems Technology Group

LNKnet Pattern Classification Software

LNKnet is a software package developed at MIT Lincoln Laboratory which integrates more than 20 neural network, statistical, and machine learning classification, clustering, and feature selection algorithms into a modular software package.

Documentation includes a short quick-start guide, a long 200-page user's guide, and UNIX man pages for the many LNKnet commands.

Downloading LNKnet

LNKnet is public domain software available at no cost with a lenient open-distribution license.

LNKnet software was originally developed under Sun Microsystem's Solaris 2.5.1 (SunOS 5.5.1) UNIX operating system and then ported to Solaris 2.6 (SunOS 5.6) under SUN Open Windows. It was recently ported to Red Hat Linux, and a test version is available under Linux. Source code is available both for individual programs (e.g. mlp, gauss, knn) and for the graphical user interface.

The following three gzipped tar files contain executables, man pages, documentation (the Quick-Start Guide and the LNKnet User's Guide), and sample data sets for Solaris 2.5.1 (SunOS 5.5.1), Solaris 2.6 (SunOS 5.6), and Red Hat Linux:

LNKnet executables for Solaris 2.5.1 with Sample Data and Man Pages - gzipped tar file (36 Mbytes)

LNKnet executables for Solaris 2.6 with Sample Data and Man Pages - gzipped tar file (37 Mbytes)

LNKnet executables for Red Hat Linux with Sample Data and Man Pages - gzipped tar file (39 Mbytes)

LNKnet now uses the GNU Autoconf tool to simplify compiling LNKnet on different host platforms. The following gzipped tar file uses Autoconf and contains source code for all LNKnet programs, man pages, documentation, and sample data sets:

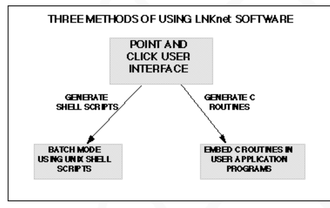
Source Code - gzipped tar file (8 Mbytes)

The following two pdf files contain the Quick-Start Guide and the LNKnet User's Guide:

Quick-Start Guide - pdf file (0.1 Mbytes)

Users-Guide - pdf file (5.2 Mbytes)

Methods Used to Access LNKnet Classifiers



The above image shows three different methods are used to access LNKnet pattern classification software. The first and most popular approach is to take advantage of the high-level graphical user interface (GUI) to run experiments. This GUI allows a non-programmer who knows at least a little about pattern classification to run experiments by pressing buttons and making selections using a mouse. A second approach, used to create automatic batch jobs, is to run LNKnet classification driver programs from the UNIX command line or in shell scripts. A third popular approach, used by C-programmers to embed LNKnet classifiers in application programs, is to use the LNKnet GUI to automatically produce C source code which implements a trained classifier. This C-code can be copied into an application program and used with little knowledge concerning details of the classifier being used. These three interfaces have been used successfully for speech recognition, talker identification, medical risk prediction, chemical analysis, and other problems at MIT Lincoln Laboratory and many other government labs and Universities. The GUI has substantially reduced the time required to run classification experiments and explore new data sets.

Algorithms Included in LNKnet

	Supervised Training	Combine Unsupervised-Supervised Training	Unsupervised Training (Clustering)
Neural Network Algorithms	Back-Propagation (MLP) Hypersphere Classifier Perceptron Committee	Radial Basis Function (RBF) Incremental RBF Learning Vector Quantizer (LVQ) Nearest Cluster Classifier Fuzzy ARTMap Classifier	
Statistical and Machine Learning Algorithms	Gaussian Linear Discriminant Gaussian Quadratic K-Nearest Neighbor Binary Decision Treee Parzen Window Histogram	Gaussian Mixture Classifier - Diagonal/Full Covariance - Tied/Per-Class Centers	K-Means Clustering EM Clustering Leader Clustering Random Clustering
Feature Selection Algorithms	Linear Discriminant (LDA) Forward/Backward Search		Principal Components (PCA)

The above table shows the most comonly used LNKnet neural network, statistical, and machine learning classification and clustering algorithms and feature selection algorithms. Classification algorithms include: multi-layer sigmoid nets (multi-layer perceptrons) with squared-error, cross-entropy, maximum likelihood, and classification-figure-of-merit cost functions; hypersphere classifiers (sometimes called RCE classifiers); unimodal Gaussian classifier with full/diagonal and grand/per-class covariance matrices (linear discriminant and quadratic discriminant classifiers); k-nearest neighbor classifier, condensed k-nearest neighbor classifier; binary decision tree (similar to CART); radial basis function classifier, incremental adaptive center and variance radial basis function classifier with four cost functions; four versions of learning vector quantizer classifiers; nearest cluster classifier, perceptron convergence procedure; Parzen widow classifier, histogram classifier, and Gaussian mixture classifier with full/ diagonal and tied per-class covariance matrices. Clustering algorithms include: leader clustering (similar to ART); k-means clustering with binary splitting initialization; and Gaussian mixture clustering with binary splitting initialization. Feature reduction algorithms include principal components analysis and canonical linear discriminant analysis. Feature selection algorithms available with all classifiers include forward and backward search feature selection. In addition, n-fold cross validation is available with all classifiers and C source code is produced which replicates any trained classifier.

LNKnet Graphics and Plots

LNKnet produces a wide variety of plots and tables. The above plot is a decision region plot for a vowel classification problem. Plots produced include 2-D decision region plots, output profile and histogram plots, structure plots describing the computing architecture, and plots illustrating how the error rate and other cost functions vary with time. In addition, after an experiment is run, outputs include confusion matrices, tables of the error rate for each class, and overall error rates.

Questions and Comments

email: KUKOLICH@SST.LL.MIT.EDU