Return to Main

Objectives

Distance Measures:
Distance
Weighted Distance
Prewhitening
Relationship
Noise Reduction
Computation

Maximum Likelihood:
Classification
Maximization
Mahalanobis

On-Line Resources:
PCA
ICA
Factor Analysis
Statistical Normalization
Pattern Recognition Applet

LECTURE 19: PRINCIPLE COMPONENTS ANALYSIS

- Objectives:
  - Introduce the concept of a distance measure
  - Introduce statistically-weighted distance measures
  - Review maximum likelihood classification
  - Explore the relationship between weighted distance measures and maximum likelihood classification
  - Introduce the Mahalanobis distance measure

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.

[Return to Main](#)

Introduction:

Speech Signals:

Signal Processing:

Parameterization:

Statistical Modeling:

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 19: PRINCIPLE COMPONENTS ANALYSIS

- Objectives:
  - Introduce the concept of a distance measure
  - Introduce statistically-weighted distance measures
  - Review maximum likelihood classification
  - Explore the relationship between weighted distance measures and maximum likelihood classification
  - Introduce the Mahalanobis distance measure

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.

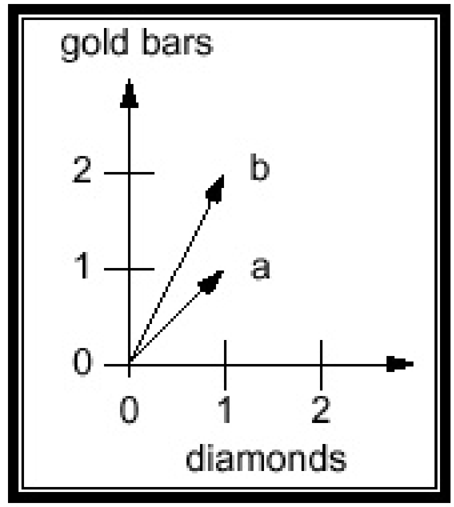
and use in our pattern recognition course. The material in this lecture follows this textbook closely:

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.

DISTANCE MEASURES



What is the distance between pt. a and pt. b?

The N-dimensional real Cartesian space, denoted $\Re^N$ is the collection of all N-dimensional vectors with real elements. A metric, or distance measure, is a real-valued function with three properties:

$\forall \bar{x}, \bar{y}, \bar{z} \in \Re^N$:

1. $d(\bar{x}, \bar{y}) \geq 0$.
2. $d(\bar{x}, \bar{y}) = 0$ if and only if $\bar{x} = \bar{y}$
3. $d(\bar{x}, \bar{y}) \leq d(\bar{x}, \bar{z}) + d(\bar{z}, \bar{y})$

The Minkowski metric of order s, or the l_s metric, between $\bar{x}$ and $\bar{y}$ is:

$$d_s(\bar{x}, \bar{y}) \equiv \sqrt[s]{\sum_{k=1}^{N} |x_k - y_k|^s} = \|\bar{x} - \bar{y}\|_s$$

(the norm of the difference vector).

Important cases are:

1. l_1 or city block metric (sum of absolute values),

$$d_1(\bar{x}, \bar{y}) = \sum_{k=1}^{N} |x_k - y_k|$$

2. l_2, or Euclidean metric (mean-squared error),

$$d_2(\bar{x}, \bar{y}) = \sqrt{\sum_{k=1}^{N} |x_k - y_k|^2}$$

3. l_∞ or Chebyshev metric,

$$d_\infty(\bar{x}, \bar{y}) = \max_k |x_k - y_k|$$

WEIGHTED EUCLIDEAN DISTANCES

We can similarly define a weighted Euclidean distance metric:

$$d_{2w}(\bar{x}, \bar{y}) = \sqrt{|\bar{x} - \bar{y}|^T \underline{W} |\bar{x} - \bar{y}|}$$

where:

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_k \end{bmatrix}, \bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_k \end{bmatrix}, \text{ and } \underline{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ w_{k1} & w_{k2} & \cdots & w_{kk} \end{bmatrix}.$$

Why are Euclidean distances so popular?

One reason is efficient computation. Suppose we are given a set of M reference vectors, $\bar{x}_m$, a measurement, $\bar{y}$, and we want to find the nearest neighbor:

$$NN = \min_m d_2(\bar{x}_m, \bar{y})$$

This can be simplified as follows:

We note the minimum of a square root is the same as the minimum of a square (both are monotonically increasing functions):

$$\begin{aligned} d_2(\bar{x}_m, \bar{y})^2 &= \sum_{j=1}^{k} (x_{m_j} - y_j)^2 = \sum_{j=1}^{k} x_{m_j}{}^2 - 2x_{m_j} y_j + y_j^2 \\ &= \|\bar{x}_m\|^2 - 2\bar{x}_m \bullet \bar{y} + \|\bar{y}\|^2 \\ &= C_m + C_y - 2\bar{x}_m \bullet \bar{y} \end{aligned}$$

Therefore,

$$NN = \min_m d_2(\bar{x}_m, \bar{y}) = C_m - 2\bar{x}_m \bullet \bar{y}$$

Thus, a Euclidean distance is virtually equivalent to a dot product (which can be computed very quickly on a vector processor). In fact, if all reference vectors have the same magnitude, C_m can be ignored (normalized codebook).

PREWHITENING OF FEATURES

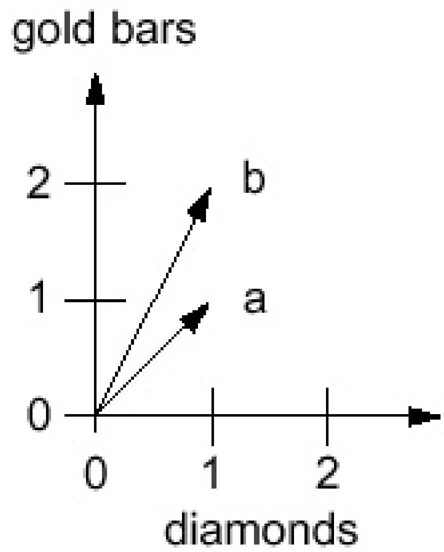
Consider the problem of comparing features of different scales:

Suppose we represent these points in space in two coordinate systems using the transformation:

$$\bar{z} = \underline{V}\bar{x}$$

System 1:

$$\beta_1 = 1\hat{i} + 0\hat{j} \text{ and } \beta_2 = 0\hat{i} + 1\hat{j}$$

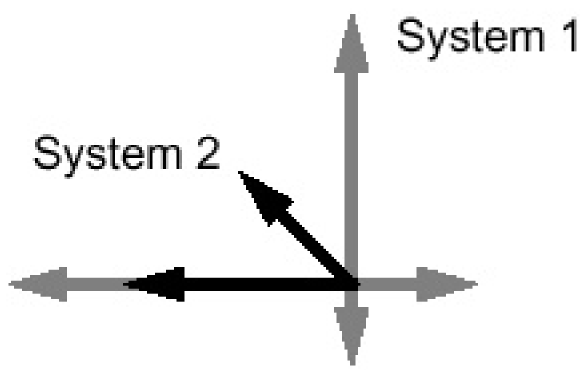


$$\bar{a} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \bar{b} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$d_2(\bar{a}, \bar{b}) = \sqrt{0^2 + 1^2} = 1$$

System 2:

$$\gamma_1 = -2\hat{i} + 0\hat{j} \text{ and } \gamma_2 = -1\hat{i} + 1\hat{j}$$



$$\bar{a} = \begin{bmatrix} -2 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \bar{b} = \begin{bmatrix} -2 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{3}{2} \\ 2 \end{bmatrix}$$

$$d_2(\bar{a}, \bar{b}) = \sqrt{\left(-1 - \left(-\frac{3}{2}\right)\right)^2 + (1-2)^2} = \sqrt{\frac{5}{4}}$$

The magnitude of the distance has changed. Though the rank-ordering of distances under such linear transformations won't change, the cumulative effects of such changes in distances can be damaging in pattern recognition. Why?

We can simplify the distance calculation in the transformed space:

$$\begin{aligned} d_2(\underline{V}\bar{x}, \underline{V}\bar{y}) &= \sqrt{[\underline{V}\bar{x} - \underline{V}\bar{y}]^T [\underline{V}\bar{x} - \underline{V}\bar{y}]} \\ &= \sqrt{[\bar{x} - \bar{y}]^T \underline{V}^T \underline{V}[\bar{x} - \bar{y}]} \\ &= d_{2W}(\bar{x}, \bar{y}) \end{aligned}$$

This is just a weighted Euclidean distance.

Suppose all dimensions of the vector are not equal in importance. For example, suppose one dimension has virtually no variation, while another is very reliable. Suppose two dimensions are statistically correlated. What is a statistically optimal transformation?

Consider a decomposition of the covariance matrix (which is symmetric):

$$\underline{C} = \Phi\Delta\Phi^T$$

where Φ denotes a matrix of eigenvectors of $\underline{C}$ and Δ denotes a diagonal matrix whose elements are the eigenvalues of $\underline{C}$. Consider:

$$\bar{z} = \underline{\Lambda}^{-1/2}\underline{\Phi}\bar{x}$$

The covariance of $\bar{z}$, $\underline{C}_{\bar{z}}$ is easily shown to be an identity matrix (prove this!) We can also show that:

$$d_2(\bar{z}_1, \bar{z}_2) = \sqrt{[\bar{x}_1 - \bar{x}_2]^T \underline{C}_{\bar{x}}^{-1} [\bar{x}_1 - \bar{x}_2]}$$

Again, just a weighted Euclidean distance.

- If the covariance matrix of the transformed vector is a diagonal matrix, the transformation is said to be an orthogonal transform.

- If the covariance matrix is an identity matrix, the transform is said to be an orthonormal transform.

- A common approximation to this procedure is to assume the dimensions of $\bar{x}$ are uncorrelated but of unequal variances, and to approximate $\underline{C}$ by a diagonal matrix, $\underline{\Lambda}$. Why? This is known as variance-weighting.

The prewhitening transform, $\bar{z} = \underline{\Lambda}^{-1/2}\underline{\Phi}\bar{x}$, is normally created as a $k \times k$ matrix in which the eigenvalues are ordered from largest to smallest:

$$\begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_k \end{bmatrix} = \begin{bmatrix} \lambda_1^{-1/2} & ? & ? & ? \\ ? & \lambda_2^{-1/2} & ? & ? \\ ? & ? & \dots & ? \\ ? & ? & ? & \lambda_k^{-1/2} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{13} \\ v_{21} & v_{22} & \cdots & v_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ v_{k1} & v_{k2} & \cdots & v_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{bmatrix}$$

where

$$\lambda_1 > \lambda_2 > \dots > \lambda_k .$$

In this case, a new feature vector can be formed by truncating the transformation matrix to $l < k$ rows. This is essentially discarding the least important features.

A measure of the amount of discriminatory power contained in a feature, or a set of features, can be defined as follows:

$$\%\ \mathrm{var} = \frac{\sum_{j=1}^{l} \lambda_j}{\sum_{j=1}^{k} \lambda_j}$$

This is the percent of the variance accounted for by the first l features.

Similarly, the coefficients of the eigenvectors tell us which dimensions of the input feature vector contribute most heavily to a dimension of the output feature vector. This is useful in determining the “meaning” of a particular feature (for example, the first decorrelated feature often is correlated with the overall spectral slope in a speech recognition system — this is sometimes an indication of the type of microphone).

Computing a "noise-free" covariance matrix is often difficult. One might attempt to do something simple, such as:

$$c_{ij} = \sum_{n=0}^{N-1} (x_i - \mu_i)(x_j - \mu_j) \text{ and } \mu_i = \sum_{n=0}^{N-1} x_i$$

On paper, this appears reasonable. However, often, the complete set of feature vectors contains valid data (speech signals) and noise (nonspeech signals). Hence, we will often compute the covariance matrix across a subset of the data, such as the particular acoustic event (a phoneme or word) we are interested in.

Second, the covariance matrix is often ill-conditioned. Stabilization procedures are used in which the elements of the covariance matrix are limited by some minimum value (a noise-floor or minimum SNR) so that the covariance matrix is better conditioned.

But how do we compute eigenvalues and eigenvectors on a computer?
One of the hardest things to do numerically! Why?

Suggestion: use a canned routine (see Numerical Recipes in C).

The definitive source is EISPACK (originally implemented in Fortran, now available in C). A simple method for symmetric matrices is known as the Jacobi transformation. In this method, a sequence of transformations are applied that set one off-diagonal element to zero at a time. The product of the subsequent transformations is the eigenvector matrix.

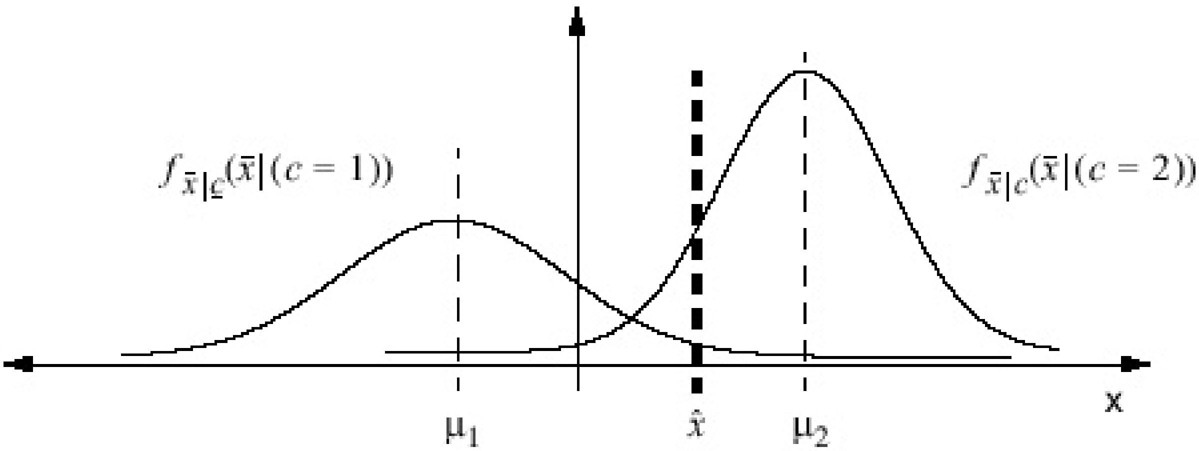
Another method, known as the QR decomposition, factors the covariance matrix into a series of transformations:

$$\underline{C} = \underline{Q}\underline{R}$$

where $\underline{Q}$ is orthogonal and $\underline{R}$ is upper diagonal. This is based on a transformation known as the Householder transform that reduces columns of a matrix below the diagonal to zero.

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:

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

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

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

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

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

from which we have

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

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

$$c^* = \underset{c}{\operatorname{argmax}}\ [P((\bar{x} = \hat{\bar{x}}) \mid (c = \hat{c}))]$$

$$= \underset{c}{\operatorname{argmx}}[P((\bar{x} = \hat{\bar{x}}) \mid (c = \hat{c}))P(c = \hat{c})]$$

if the class probabilities are equal,

$$c^* = \underset{c}{\operatorname{argmx}}[P((\bar{x} = \hat{\bar{x}}) \mid (c = \hat{c}))]$$

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:

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

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

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:

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

(Note that the maximization became a minimization.)

We can define a distance measure based on this as:

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

THE MAHALANOBIS DISTANCE

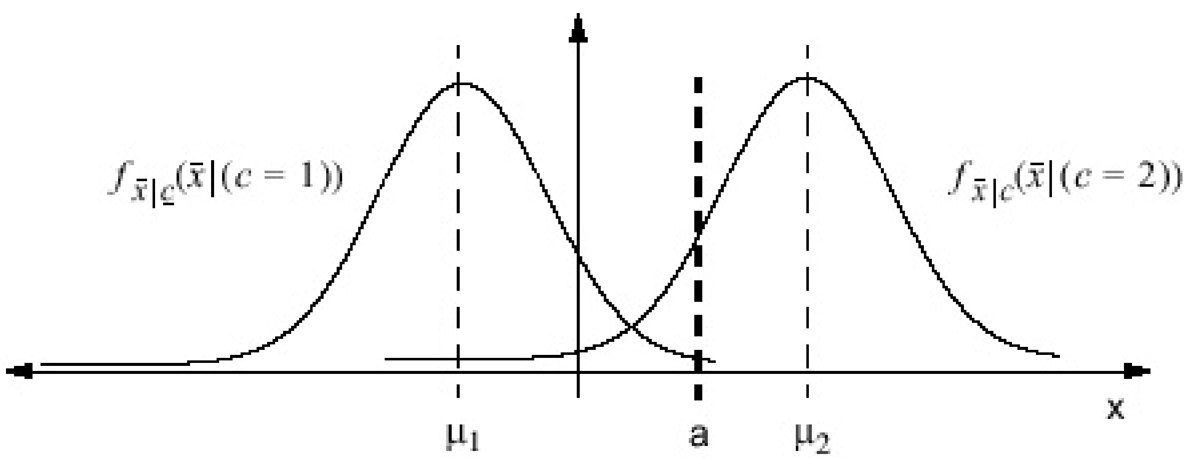
Note that the distance is conditioned on each class mean and covariance. This is why "generic" distance comparisons are a joke.

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?

Principal Components and Factor Analysis

Principal components analysis (PCA) and factor analysis (FA) are statistical techniques applied to a single set of variables to discover which sets of variables in the set form coherent subsets that are relatively independent of one another. Variables that are correlated with one another which are also largely independent of other subsets of variables are combined into factors. Factors which are generated are thought to be representative of the underlying processes that have created the correlations among variables.

PCA and FA can be exploratory in nature, FA is used as a tool in attempts to reduce a large set of variables to a more meaningful, smaller set of variables. As both FA and PCA are sensitive to the magnitude of correlations robust comparisons must be made to ensure the quality of the analysis. Accordingly, PCA and FA are sensitive to outliers, missing data, and poor correlations between variables due to poorly distributed variables (See normality link for more information on distributions.) As a result data transformations have a large impact upon FA and PCA.

Correlation coefficients tend to be less reliable when estimated from small sample sizes. In general it is a minimum to have at least five cases for each observed variable. Missing data need be dealt with to provide for the best possible relationships between variables. Fitting missing data through regression techniques are likely to over fit the data and result in correlations to be unrealistically high and may as a result manufacture factors. Normality provides for an enhanced solution, but some inference may still be derived from nonnormal data. Multivariate normality also implies that the relationships between variables are linear. Linearity is required to ensure that correlation coefficients are generated form appropriate data, meeting the assumptions necessary for the use of the general linear model. Univariate and multivariate outliers need to be screened out due to a heavy influence upon the calculation of correlation coefficients, which in turn has a strong influence on the calculation of factors. In PCA multicollinearity is not a problem as matrix inversion is not required, yet for most forms of FA singularity and multicollinearity is a problem. If the determinant of R and eigenvalues associated with some factors approach zero, multicollinearity or singularity may be present. Deletion of singular or multicollinear variables is required.

Uses of Principle Components Analysis and Factor Analysis

Direct Uses: - identification of groups of inter-related variables,
- reduction of number of variables,

Indirect Uses: - a method of transforming data. Transformation of data through rewriting the data with properties the original data did not have. The data may be efficiently simplified prior to a classification while also removing artifacts such as multicollinearity.

Theory to Common Factor Analysis and Factor Analysis

The key underlying base to Common Factor Analysis (PCA and FA) is that the chosen variables can be transformed into linear combinations of an underlying set of hypothesized or unobserved components (factors). Factors may either be associated with 2 or more of the original variables (common factors) or associated with an individual variable (unique factors). Loadings relate the specific association between factors and original variables. Therefore, it is necessary to find the loadings, then solve for the factors, which will approximate the relationship between the original variables and underlying factors. The loadings are derived from the magnitude of eigenvalues associated to individual variables.

The difference between PCA and FA is that is that for the purposes of matrix computations PCA assumes that all variance is common, with all unique factors set equal to zero; while FA assumes that there is some unique variance. The level of unique variance is dictated by the FA model which is chosen. Accordingly, PCA is a model of a closed system, while FA is a model of an open system.

Rotation attempts to put the factors in a simpler position with respect to the original variables, which aids in the interpretation of factors. Rotation places the factors into positions that only the variables which are distinctly related to a factor will be associated. Varimax, quartimax, and equimax are all orthogonal rotations, while oblique rotations are non-orthogonal. The varimax rotation maximizes the variance of the loadings, and is also the most commonly used.

To run a PCA or FA

To analyze data with either PCA or FA 3 key decisions must be made:
- the factor extraction method,
- the number of factors to extract, and
- the transformation method to be used.

Interpretation of a Factor Analysis

Determination of number of factors to extract

- significance test, difficult to meet assumptions required to significance tests, therefore the following heuristics are used.

- magnitude of eigenvalues,
Assess the amount of original variance accounted for. Retain factors whose eigenvalues are greater than 1. (Ignore those with eigen values less than one as the factor is accounting for less variance than an original variable).

Eigenvalues and Proportion of Original Variance

	Magnitude	Variance Prop.
Value 1	11.181	.532
Value 2	3.141	.15
Value 3	1.712	.082
Value 4	1.255	.06
Value 5	1.074	.051
Value 6	.78	.037
Value 7	.478	.023
Value 8	.394	.019
Value 9	.363	.017
Value 10	.217	.01
Value 11	.183	.009

> the figure relates that five factors are significant.

- substantive importance,
an absolute test of eigen values in a proportional sense. Retain any eigenvalue that accounts for at least 5% of the variance,
- skree test,
plot magnitude of eigen values (Y axis) versus components (X axis), retain factors which are above the inflection point of the slope.
- interpretability,
a battery of tests where the above heuristics may all be applied, assess magnitude of eigen values, substantive importance, and a skree test.

Which variables are best summarized by the model?

- interpret communalities (final estimates of communalities),
high = most important,
low = least important.

Communality Summary

	SMC	Final Estimate
POP86	1	.995
POP91	1	.993
POPDEN	.923	.902
CDAREA	.498	.775
FRENCH	.943	.727
NONNAT	.991	.881
NONMOV1	1	.993
NONMOV5	.999	.989
LOWED	.996	.966
HIGHED	.993	.974
MUNEMP	.895	.94
FUNEMP	.884	.929
LABOUR	1	.991
AGRIC	.671	.746
MINES	.614	.609
GOV	.938	.716

	SMC	Final Estimate
HOUSEV	.806	.866
CROWD	.723	.813
PPFAM	.635	.784
AVGINC	.725	.798
LOWINC	.997	.978

> the communalities relate the overall effect of the factors.

Naming of Factors

- look at individual factor scores, see which variables have the highest factor scores. Also look at the factor scores to see if the initial interpritations are confimed by the factor scores (Factor scores are normally distributed only when the input variables are normally distributed. Therefore, when interpreting factors the greatest concern is with the tail values. The normal distribution of factor scores also acts as a data transformation and prepares the data for other multivariate analyses.)

What is meant by an Illconditioned Correlation Matrix

- an illconditioned correlation matrix is a manifestation of multicollinearity. FA is sensitive to an illconditioned matrix while PCA is not. To solve for the factor scores matrix inversion is required, which is not possible with a singular matrix. To solve this problem click on the multicollinearity link.

To assess the value of input variables to the model

- assess the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO), which provides results in the range from 0.5 to 0.9.

Measures of Variable Sampling Adequacy

Total matrix sampling adequacy : .833

POP86	.865
POP91	.824
POPDEN	.911
CDAREA	.629
FRENCH	.693
NONNAT	.81
NONMOV1	.82
NONMOV5	.852
LOWED	.848
HIGHED	.901
MUNEMP	.645
FUNEMP	.674
LABOUR	.911
AGRIC	.758
MINES	.534
GOV	.85

HOUSEV	.884
CROWD	.647
PPFAM	.628
AVGINC	.821
LOWINC	.88

Bartlett Test of Sphericity - DF : 230 Chi Square : 16626.24 P : ●

> For example: MINES is a less valid variable than POPDEN in this model.

A value of 1 relates a complete relationship, totally related, which is bad. The range which is provided as a heuristic is:
0.9 - marvelous,
0.8 - meritorious,
0.7 - middling,
0.6 - mediocre, or
0.5 - miserable (perfectly uncorrelated).

Barlett test of sphericity, variable projected upon an n-dimensional spheroid, the significance of the relationship is then evaluated. (see figure above, where the p value is significant, and does not fit in the allocated space)

Multivariate Statistics: A Practical Guide

Back to the Practical Guide to Multivariate Statistics Index Page
Back to the Mike Wulder Personal Web Pages

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

Multivariate Statistics: Factor Analysis

Factor Analysis can be seen as the granddaddy of all the multivariate techniques we are looking at here. Of the three, it is the most-frequently used, and has the largest amount of literature devoted to it. See references for some places to start.)

Definition and an example

Factor analysis is:

a statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors). The statistical approach involving finding a way of condensing the information contained in a number of original variables into a smaller set of dimensions (factors) with a minimum loss of information (Hair et al., 1992).

Factor analysis could be used to verify your conceptualization of a construct of interest. For example, in many studies, the construct of "leadership" has been observed to be composed of "task skills" and "people skills." Let's say that, for some reason, you are developing a new questionnaire about leadership and you create 20 items. You think 10 will reflect "task" elements and 10 "people" elements, but since your items are new, you want to test your conceptualization.

Before you use the questionnaire on your sample, you decide to pretest it (always wise!) on a group of people who are like those who will be completing your survey. When you analyze your data, you do a factor analysis to see if there are really two factors, and if those factors represent the dimensions of task and people skills. If they do, you will be able to create two separate scales, by summing the items on each dimension. If they don't, well it's back to the drawing board.

What you need in order to do a factor analysis

Remember, factor analysis requires that you have data in the form of correlations, so all of the assumptions that apply to correlations, are relevent here.

Types of factor analysis: Two main types:

- **Principal component analysis** -- this method provides a *unique solution*, so that the original data can be reconstructed from the results. It looks at the total variance among the variables, so the solution generated will include as many factors as there are variables, although it is unlikely that they will all meet the criteria for retention. There is only one method for completing a principal components analysis; this is not true of any of the other multidimensional methods described here.
- **Common factor analysis --** this is what people generally mean when they say "factor analysis." This family of techniques uses an estimate of common variance among the original variables to generate the factor solution. Because of this, the number of factors will always be less than the number of original variables. So, choosing the number of factors to keep for further analysis is more problematic using common factor analysis than in principle components.

Steps in conducting a factor analysis

There are four basic factor analysis steps:

- data collection and generation of the correlation matrix
- extraction of initial factor solution
- rotation and interpretation
- construction of scales or factor scores to use in further analyses

Extraction of an initial solution

The output of a factor analysis will give you several things. The table below shows how output helps to determine the number of components/factors to be retained for futher analysis. One good rule of thumb for determining the number of factors, is the "eigenvalue greater than 1" criteria. For the moment, let's not worry about the meaning of eigenvalues, however this criteria allows us to be fairly sure that any factors we keep will account for at least the variance of one of the variables used in the analysis. However, when applying this rule, keep in mind that when the number of variables is small, the analysis may result in fewer factors than "really" exist in the data, while a large number of variables may produce more factors meeting the criteria than are meaningful. There *are* other criteria for selecting the number of factors to keep, but this is the easiest to apply, since it is the default of most statistical computer programs.

Note that the factors will all be ***orthogonal*** to one another, meaning that they will be uncorrelated.

Remember that in our hypothetical leadership example, we expected to find two factors, representing task and people skills. The first output is the results of the extraction of components/factors, which will look something like this:

Table #1: Sample extraction of components/factors

Factors	Eigenvalue	% of variance	Cumulative % of variance
1	2.6379	44.5	37.6
2	1.9890	39.3	83.8
3	0.8065	8.4	92.2
4	0.6783	7.8	100.0

Interpreting your results

Since the first two factors were the only ones that had eigenvalues > 1, the final factor solution will only represent 83.8% of the variance in the data. The ***loadings*** listed under the "Factor" headings represent a correlation between that item and the overall factor. Like Pearson correlations, they range from -1 to 1. The next panel of factor analysis output might look something like this:

Table #2: Unrotated Factor Matrix

Variables	Factor 1	Factor 2	Communality
Ability to define problems	**.81**	-.45	.87
Ability to supervise others	**.84**	-.31	.79
Ability to make decisions	**.80**	-.29	.90
Ability to build consensus	**.89**	.37	.88
Ability to facilitate decision-making	**.79**	.51	.67
Ability to work on a team	**.45**	.43	.72

This table shows the difficulty of interpreting an unrotated factor solution. All of the most significant loadings (highlighted) are on Factor #1. This is a common pattern. One way to obtain more intepretable results is to ***rotate*** your solution. Most computer packages use ***varimax*** rotation, although there are other techniques.

Below is an example of what the factors might look like if we rotated them. Notice that the loadings are distributed between the factors, and that the results are easier to interpret.

Table #3: Rotated Factor Matrix

Variables	Factor 1	Factor 2	Communality
Ability to define problems	**.68**	.17	.87
Ability to supervise others	**.87**	.24	.79
Ability to make decisions	**.65**	.07	.90
Ability to build consensus	.16	**.76**	.88
Ability to facilitate decision-making	.30	**.83**	.67
Ability to work on a team	.19	**.69**	.72

Naming the factors

Now we have a highly interpretable solution, which represents almost 90% of the data. The next step is to name the factors. There are a few rules suggested by methodologists:

Factor names should

- be brief, one or two words
- communicate the nature of the underlying construct

Look for patterns of similarity between items that load on a factor. If you are seeking to validate a theoretical structure, you may want to use the factor names that already exist in the literature. Otherwise, use names that will communicate your conceptual structure to others. In addition, you can try looking at what items **do not** load on a factor, to determine what that factor isn't. Also, try reversing loadings to get a better interpretation.

Using the factor scores

It is possible to do several things with factor analysis results, but the most common are to use factor scores, or to make summated scales based on the factor structure.

Because the results of a factor analysis can be strongly influenced by the presence of error in the original data, Hait, et al. recommend using factor scores if the scales used to collect the original data are "well-constructed, valid, and reliable" instruments. Otherwise, they suggest that if the scales are "untested and exploratory, with little or no evidence of reliability or validity," summated scores should be constructed. An added benefit of summated scores is that if they are to be used in further analysis, they preserve the variation in the data.

Other links

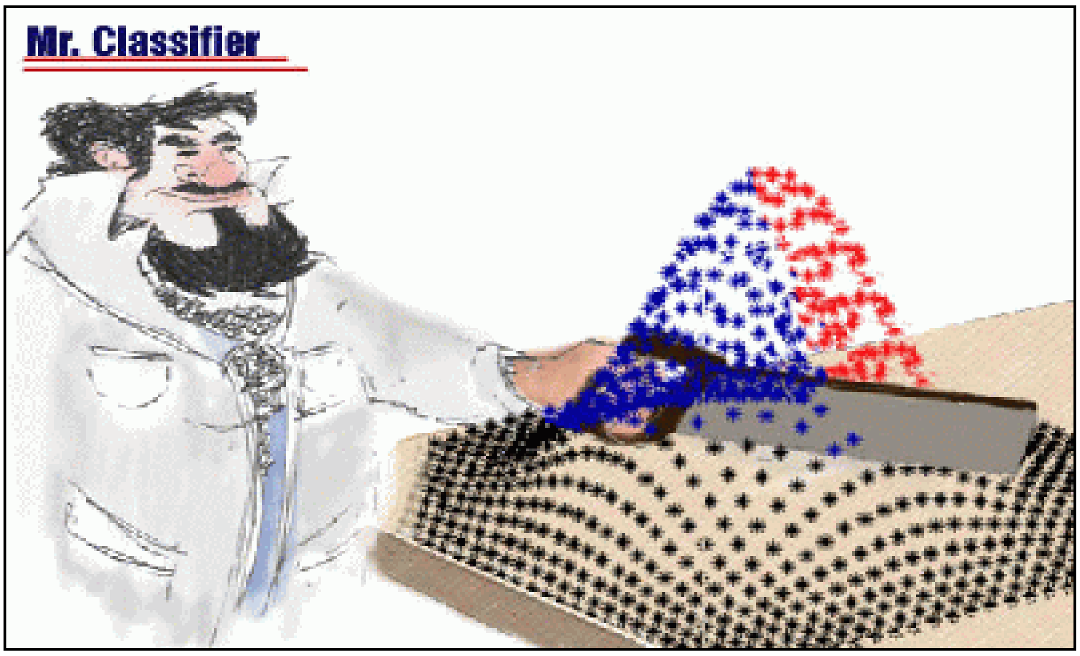
Phillip Ingram, of the School of Earth Sciences, Macquarie University, Sydney, Australia, has a Statistics Page, which includes separate pages for Multivariable Statistics, including principal components and factor analysis. The material is more advanced than that presented here, but very useful for those who will be employing these analyses techniques.

Back to the Multivariate Statistics home page

Forward to the Multidimensional Scaling (MDS) page

Forward to the Cluster Analysis page

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.
