

DATA PROCESSING I: ADVANCEMENTS IN MACHINE

ANALYSIS OF MULTISPECTRAL DATA

by

Philip H. Swain and Staff
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana

INTRODUCTION

Research in multispectral data processing at LARS/Purdue is directed at supporting a substantial level of applications research as well as advancing the technology of remote sensing data processing. During the past year significant progress has been made in both respects. Almost the entire multispectral data analysis process, from data editing to results evaluation, has been impacted, and the new level of technology has been vigorously tested by the data analysis operations associated with the 1971 Corn Blight Watch Experiment.¹

The following discussion of these advancements is organized to follow generally the steps utilized in the multispectral data analysis procedure. In terms of Figure 1, we begin with the data display process used to accomplish data editing and proceed clockwise through clustering, statistics computation, etc. In the interest of brevity, each result will be treated here in a general way and references given to available sources where a more detailed treatment may be found.

DATA EDITING FACILITY

The special-purpose digital display system delivered to LARS/Purdue late in 1970 [1] represents a tremendous potential for facilitating the man/data interface. During 1971 the first software for utilizing this system became operational and was made available to LARSYS users [2]. With this software, the user can display a television-quality image of digitized multispectral data and, by means of a light pen and keyboard, accurately specify areas in the data to receive special attention

¹The 1971 Corn Blight Watch Experiment is described elsewhere in these proceedings.

(Figure 2). Two advantages of this mode of man/data interface over the familiar gray-scale line-printer output (Figure 3) are the higher quality of the image available to the researcher and the ease and accuracy with which features in the data can be located and designated to the computer by means of the light pen. These features greatly improve both the speed and accuracy with which the data analysis can be executed.

Data editing represents only one of many potential uses of the digital display hardware. Examples of other applications to be studied include on-line display and evaluation of analysis results and implementation of a highly interactive data analysis capability.

One feels compelled to note at this point, however, that line-printer output still represents a proven and acceptable means for displaying both data and analysis results. But as technological advances bring down the cost of video-type displays and step up the speed of digital data transmission, digital display systems suitable for image data -- now available only on a limited basis as research tools -- will become increasingly attractive as a standard means of interfacing man with such data.

CLUSTER ANALYSIS

Multispectral cluster analysis (sometimes referred to in the literature as unsupervised classification) has been under study for some years as a means for data compression and similarity analysis. A clustering technique has been developed at LARS/Purdue, for use in conjunction with supervised classification, as an aid in class definition and training sample selection. A computer program [3] prints point-by-point maps of the clustering results (Figure 4), indicating the relative homogeneity of the analyzed areas; this information assists in the process of selecting training samples for characterizing the different spectral classes in the data. Also provided is a quantitative analysis of the separability of the clusters in the multivariate measurement ("feature") space.

The clustering technique described above processes data points in the measurement space. Another promising approach, currently under investigation and discussed further in a later section of this paper, is the clustering of sample statistics in parameter space.

FEATURE SELECTION

A feature selection criterion has been developed [4] which eliminates the considerable level of human interaction with the

computational processing heretofore required for the selection of data channels preferred for classification. The basic problem faced in connection with feature selection is finding a means for estimating error probabilities (or probabilities of correct classification) accurately since for multivariate problems it is generally not feasible to calculate these probabilities directly even in the relatively simple case in which Gaussian distribution of the data within classes is assumed. The problem of finding an estimator of probability of correct classification in the multiclass and multivariate case is unsolved. What is commonly done in practice is to estimate the probabilities associated with all pairs of classes and take an average or weighted average of the pairwise probabilities as an estimate of the overall probability of correct classification [5]. To do this effectively, however, requires availability of a function, based on the statistical separability of pairs of classes, which behaves like the probability of correctly discriminating between the classes.

Divergence is a monotonic function of statistical separability of two classes which has been used in this manner. However, this separability measure has the disadvantage that it increases without bound as separability increases, whereas probability of correct classification saturates at 100 percent (see Figure 5). This difficulty has been circumvented by writing the feature selection program to allow the user to specify a limiting value (MAX) which artificially saturates the separability measure. To do this properly, however, the user must learn to judge for a given type of problem what constitutes an appropriate saturation value.

In an effort to remove this latter shortcoming, alternative separability measures have been investigated. In particular, a separability measure referred to here as Bhattacharyya distance, or B-distance, has been found to have the sort of behavior sought and indeed to provide a much more reliable feature selection criterion than divergence [4]. This further suggested a transformation of divergence which closely approximates the feature selection properties of the B-distance but requires far less computation. The transformed divergence has been implemented at LARS/Purdue as the standard feature selection criterion.

POINT CLASSIFICATION

The next step in the procedure for multispectral data analysis, the multivariate classification method, has not been altered, but some newly completed research has reconfirmed the wisdom (from a practical viewpoint) of selecting the Gaussian maximum-likelihood approach for analysis of real-world multispectral data. This approach [6] assumes that the class-conditional distributions of the data in all classes to be recognized can be adequately represented by multivariate Gaussian distributions, or, in any case, by the union of a small number of such distributions. Although

pattern classifiers based on this approach have been applied successfully at various remote sensing facilities involved with machine analysis, some important questions regarding this choice of approach have remained open: How much improvement in classification accuracy could be obtained by using a nonparametric classification method which requires no a priori assumptions regarding the data distributions? How much would classification accuracy degrade if the classifier were of the computationally faster and simpler linear variety?

An experimental investigation yielding a considerable volume of results [3] has demonstrated that, for agricultural remote sensing data, very general nonparametric models can be expected to produce only marginally better results than the Gaussian classifier. In general the improvement is not sufficient to warrant the substantial increase in computational resources required (time, machine memory). On the other hand, another study [7] suggests that the extra cost of the Gaussian classifier by comparison with linear classifiers is generally well justified. The linear classifiers investigated have shown markedly poorer ability to generalize from training fields to data not used for training the classifier.

SAMPLE CLUSTERING AND SAMPLE CLASSIFICATION

The term "perfield classification" has been used in the literature to refer to the classification of an entire agricultural field based on all data drawn from that field. This approach takes advantage of the spatial context of the data, the fact that local regions tend to be composed of members of the same class (the same "population," in statistical terminology), by using the combined information in a number of observations to infer the classification of the aggregate. To divorce this concept from the agricultural frame of reference, "sample classification" is defined as the classification of any aggregate of data points assumed to be from the same population. It is often the case that decisions concerning the aggregate can be made faster and more reliably than decisions concerning the data points taken individually.¹

As intensive study of this approach [3] has been completed in which both sample clustering and sample classification were investigated. The results of this study are too extensive, both in number and in scope, to receive adequate treatment here. Following are some highlights.

¹The greatest benefits in this respect generally accrue when the aggregation is performed before the decision process is applied (eg. by finding a parametric characterization of the aggregate) rather than after (eg., poll-taking after classification).

For agricultural remote sensing data, the accuracy of sample classification is relatively insensitive to whether parametric or nonparametric methods are used to estimate probability distributions. As noted earlier in this paper the potential improvement in accuracy obtainable using nonparametric methods is too small to justify the considerable increase in computation time and complexity.

Although many measures of statistical separability are available for use in sample classification, the experimental results using agricultural data were relatively insensitive to the choice of separability measure used. However, a separability measure known as the Jeffries-Matusita distance does have some theoretical as well as practical advantages worth exploiting:

1. Its behavior as a function of dimensionality resembles that of probability of correct classification (in the parametric case).
2. It is a metric over a large space of distribution functions.
3. It is among the simplest separability functions to compute.

Sample clustering, achieved by first computing a parametric characterization of the samples and then applying cluster analysis to the statistical parameters (Figure 6), appears to offer several advantages over the more conventional point-by-point clustering. In experiments with agricultural remote sensing data, sample clustering has exhibited a distinct tendency to produce more appropriate class/subclass structures leading to better classification accuracy for both point and sample classification. In addition, a dramatic time saving is achieved for cluster processing because of the considerable degree of data reduction accomplished by representing a large number of data points by relatively few statistical parameters.

STATISTICAL DESIGN AND ANALYSIS

Finally, the effective utilization of large quantities of remote sensing data demands the development of statistical models which can be used for specifying data collection and data analysis schemes and for evaluating the results produced by such schemes. The 1971 Corn Blight Watch Experiment and forward-looking considerations related to the ERTS and SKYLAB satellites have particularly highlighted this need. Conventional models developed for ground data collection alone are simply not adequate.

A recent study [8] has formulated a three-stage sampling model for remote sensing and used the model to evaluate the precision of crop acreage estimates and to determine the effects of the number of flightlines, number of segments within flightlines, and the subsampling density within segments on the precision of these estimates. While this work has

has perhaps raised as many important questions as it has answered, it represents the initiation of a significant effort to determine systematically the cost-benefit relationships associated with the remote sensing technology and to utilize these relationships both in guiding and evaluating its application.

ACKNOWLEDGEMENT

The research summarized herein was supported by NASA under Grant NGL 15-005-112. Grateful appreciation is expressed to NASA for this support. The various individual studies were carried out by LARS staff members G. Santner, K. Hunt, Dr. L. Eisgruber; graduate students T. Robertson, A. Wacker, P. Scheffer; and the author.

REFERENCES

1. IBM 4507 Digital Image Display System, IBM Federal Systems Division, Gaithersburg, Maryland, October 1970.
2. FIELDSEL: LARS Data Editing Program, LARS Program Abstract No. 255, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana. 47906.
3. Wacker, A.G., "The Minimum Distance Approach to Classification," Technical Report Number TR-EE71-37, School of Electrical Engineering, Purdue University, Lafayette, Indiana 47906, October 1971. Also LARS Information Note 100771, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47906, October 1971.
4. Swain, P.H., T.V. Robertson, A.G. Wacker, "Comparison of the Divergence and B-Distance in Feature Selection," LARS Information Note 020871, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47906, February 1971.
5. Fu, K.S., P.J. Min, "On Feature Selection in Multiclass Pattern Recognition," Technical Report Number TR-EE68-17, School of Electrical Engineering, Purdue University, Lafayette, Indiana 47906, July 1968.
6. Landgrebe, D.A., P.J. Min, P.H. Swain, K.S. Fu, "The Application of Pattern Recognition to a Remote Sensing Problem," Proceedings of the Seventh Symposium on Adaptive Processes, UCLA, Los Angeles, California, December 1968.
7. Robertson, T.V., "Some Experimental Results for Linear Classifiers Applied to Agricultural Remote Sensing Data," LARS Information Note 020972, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47906 (in preparation).
8. Eisgruber, L.M., "The Effect of Subsampling Ratios on Precision of Estimates from Remote Sensing," LARS Information Note 021072, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47906 (in preparation).

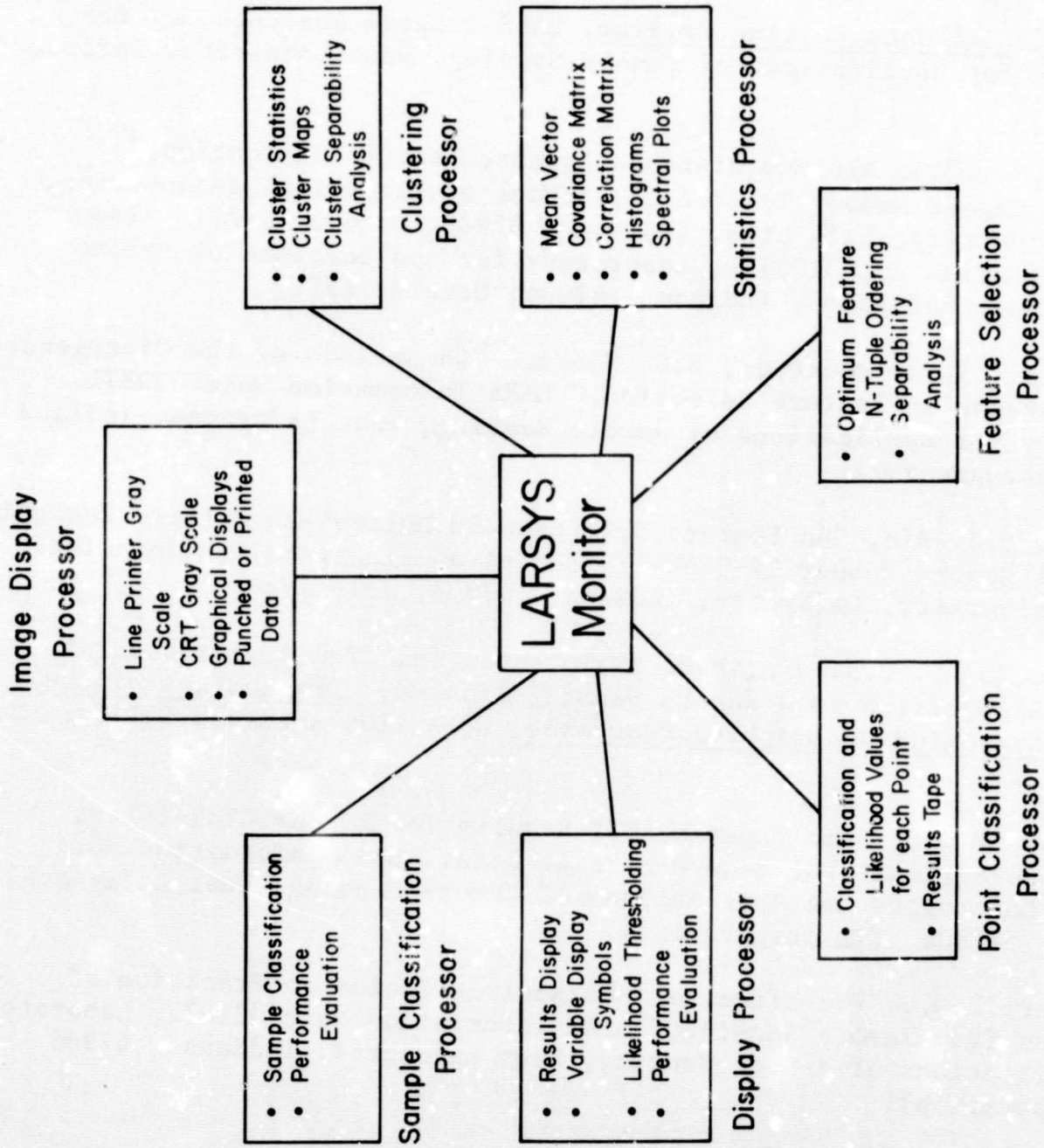


Figure 1. LARSYS: a software system for the analysis of multispectral remote sensing data.

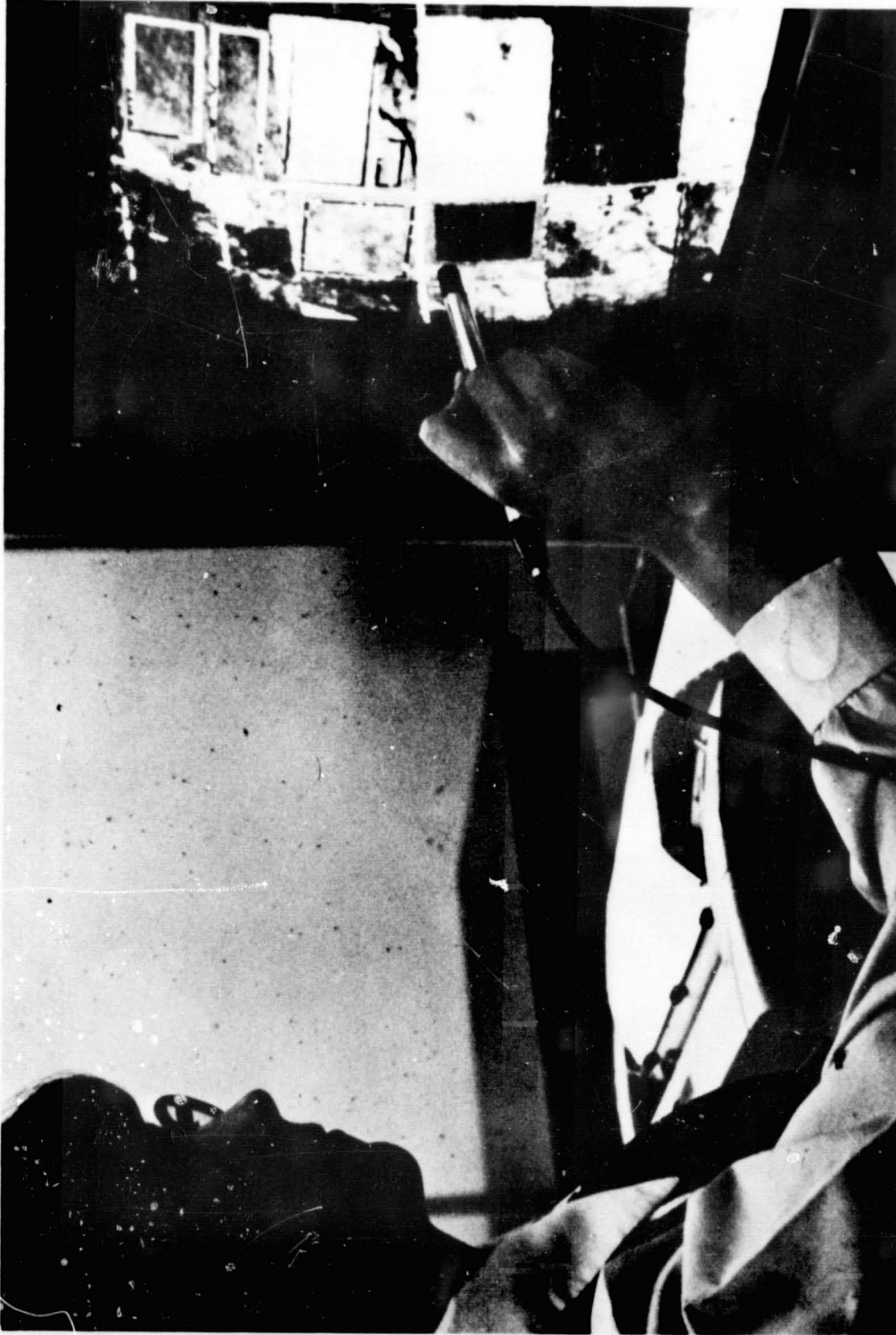
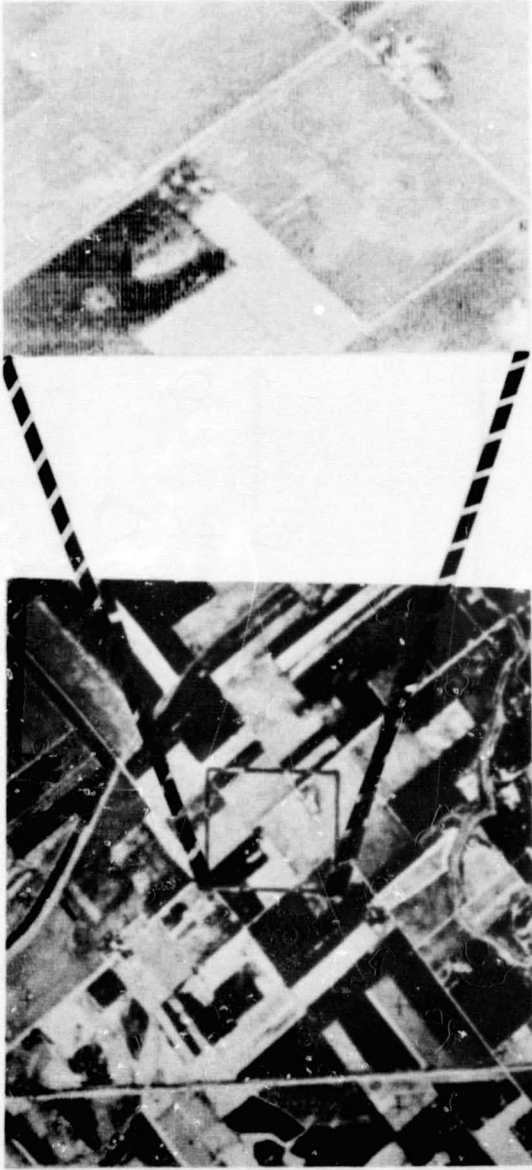


Figure 2. Specifying field boundaries on the digital image-editing display system.



Figure 3. Outlining field boundaries on gray-scale printouts.



a) Digital display image

b) Computer printout

Figure 4. Map-like display of cluster analysis results.

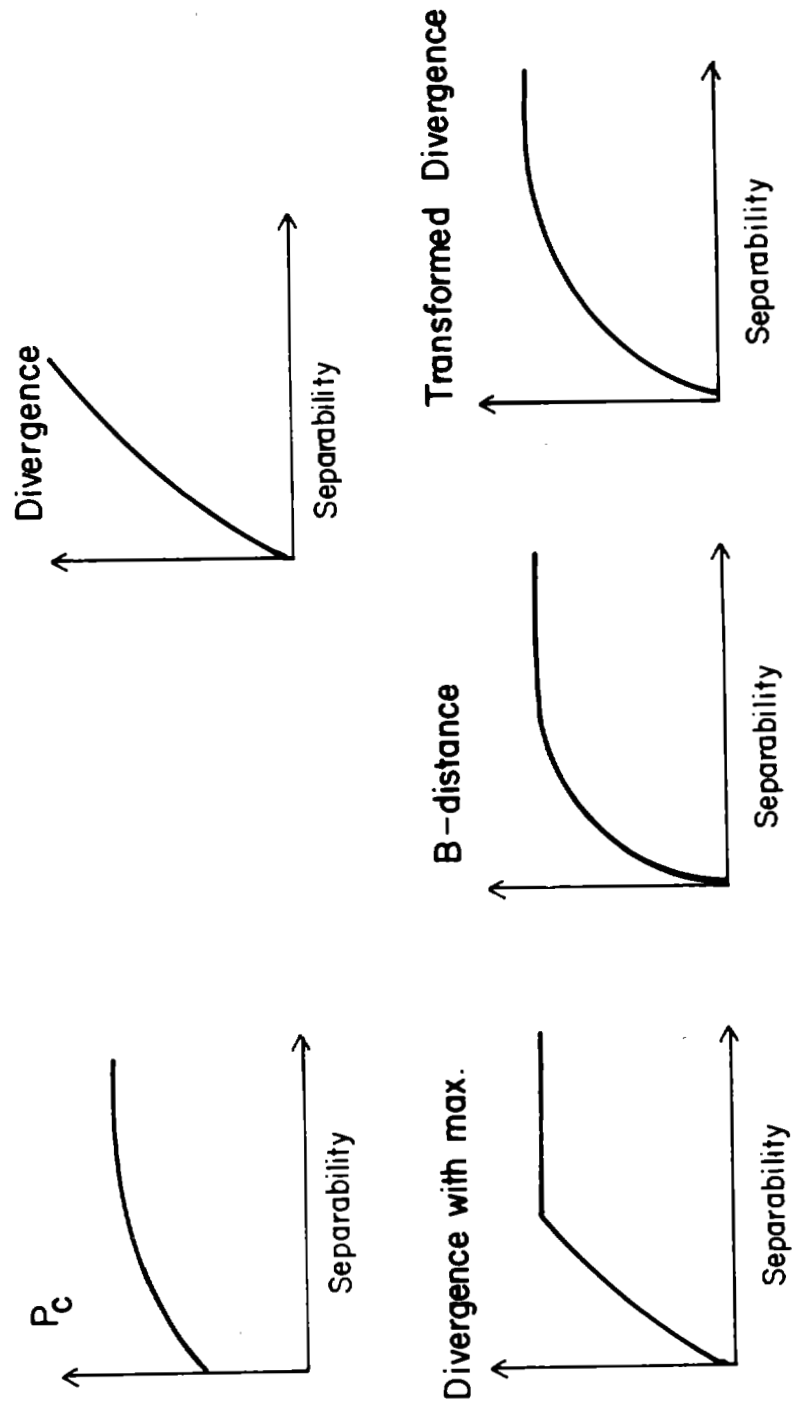


Figure 5. Behavior of probability of correct classification and various measures of statistical separability.

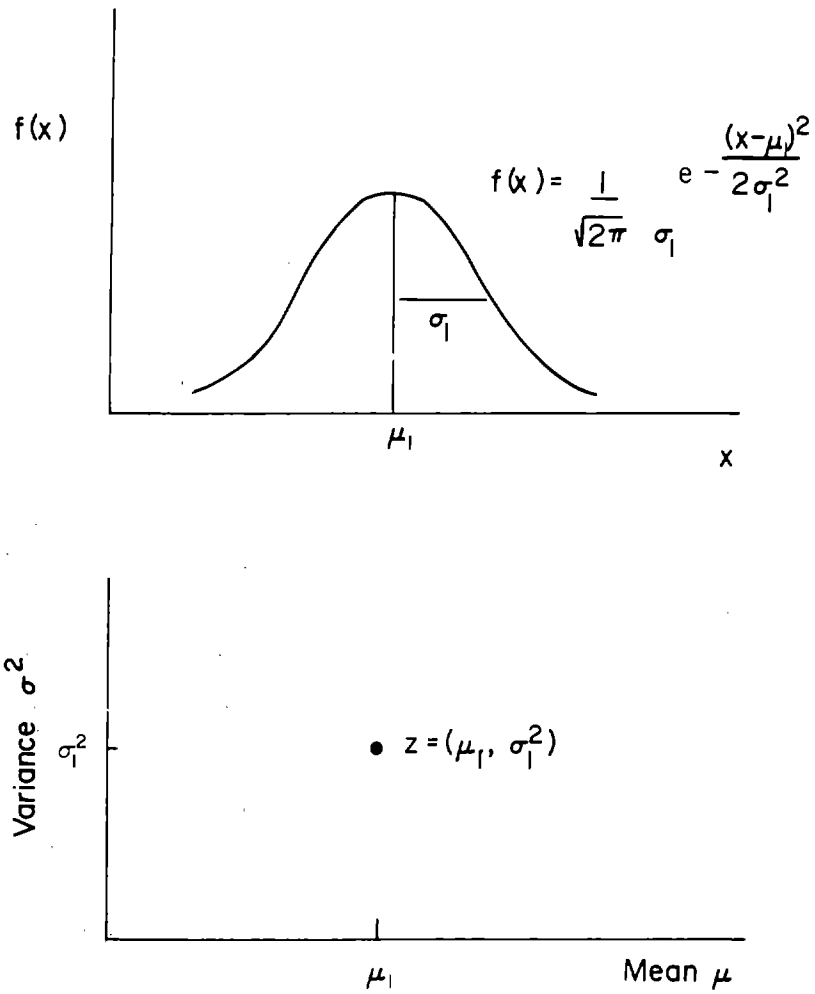


Figure 6. Parameter space representation of a sample (one-dimensional Gaussian case).