General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

NASA

Technical Memorandum 80303

Decision Rules For Unbiased Inventory Estimates

P. Argentiero
D. Koch

(NASA-TM-80303) DECISION RULES FOR UNBIASED INVENTORY ESTIMATES (NASA) 27 p HC A03/MF A01 CSCL 05B
N79-30610
Unclas
G3/43 36068

JULY 1979

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

SEP 1979
RECEIVED
NASA STI FACILITY
ACCESS DEPT.

TM 80303

DECISION RULES FOR UNBIASED INVENTORY ESTIMATES

P. Argentiero
D. Koch

July 1979

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland 20771

CONTENTS

	Page
ABSTRACT	v
1. INTRODUCTION	1
2. INVENTORY ESTIMATION	3
3. A NUMERICAL SIMULATION	9
4. SUMMARY	14
ACKNOWLEDGMENTS	15
REFERENCES	16
APPENDIX: OPTIMAL ONE DIMENSIONAL FEATURE EXTRACTION	17

LIST OF TABLES

Table		Page
1	Statistics of LANDSAT-2 MSS Signatures Acquired May 1975 Over Finney County, Kansas	10
2	Confusion Matrix for Four Dimensional Bayes Classifier	11
3	Confusion Matrix for One Dimensional Feature Extraction and Bayes Classifier	12

DECISION RULES FOR UNBIASED INVENTORY ESTIMATES

P. Argentiero
D. Koch

ABSTRACT

This paper presents an efficient and accurate procedure for estimating inventories from remote sensing scenes. In place of the conventional and expensive full dimensional Bayes decision rule, a one-dimensional feature extraction and classification technique is employed. It is shown that this efficient decision rule can be used to develop unbiased inventory estimates and that for large sample sizes typical of satellite derived remote sensing scenes, resulting accuracies are comparable or superior to more expensive alternative procedures. Mathematical details of the procedure are provided in the body of the report and in the Appendix.

Results of a numerical simulation of the technique using statistics obtained from an observed LANDSAT scene are included. The simulation demonstrates the effectiveness of the technique in computing accurate inventory estimates.

PRECEDING PAGE BLANK NOT FILMED

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

DECISION RULES FOR UNBIASED INVENTORY ESTIMATES

1. INTRODUCTION

A satellite based remote sensing scene is usually composed of a large number of multidimensional samples each of which consists of radiation intensities obtained from a specific ground area. For the purposes of interpretation and information extraction, it is frequently useful to assume that the portion of the Earth's surface under observation in a scene includes a small number of identifiable classes of ground cover. To interpret the scene a decision rule is required which allocates the samples among the classes. For many applications, it is important to know how each class of ground cover is distributed throughout a scene. When this type of information is conveyed by means of a color coded display it is called a *thematic map*. When thematic mapping is needed it is essential for a decision rule to correctly classify as many samples as possible. It is known that under a reasonable set of assumptions a Bayes or maximum likelihood classifier minimizes the global error rate. Hence, this continues to be the most commonly used classification procedure for applications which require thematic mapping.

Some applications call for only an estimate of the proportion of a scene accounted for by each class of ground cover. This is referred to as *inventory estimation*. The information needed for inventory estimation is different in character from what is needed for thematic mapping. For instance, when only the proportions of a scene allocated to each class are estimated, a certain amount of cancellation of errors of commission and omission is possible. In the case of thematic mapping no such cancellation can occur. Hence, it should not surprise that the minimal error rate criterion might not be the most appropriate basis for choosing a decision rule for inventory estimation.

This paper presents a procedure for inventory estimation which is both efficient and accurate. In place of the conventional and expensive Bayes decision rule, a one dimensional feature extractic and classification technique is employed. It is shown that this computationally inexpensive decisio rule can be used to develop unbiased inventory estimates and that for the large sample sizes typical of satellite derived remote sensing scenes, resulting accuracies are comparable to those of any alternative procedure. Section 2 and the Appendix provide the mathematical details of this approach to inventory estimation. Section 3 shows the results of a numerical simulation of the technique using class statistics obtained from an observed LANDSAT scene. Section 4 provides a summary of the results.

2. INVENTORY ESTIMATION

Inventory estimates can be obtained from a remote sensing scene by applying a decision rule to each sample and then determining by a simple counting procedure the proportions of the total sample set allocated to each class. The quality of the resulting estimates will depend on the choice of a decision rule. In order to choose intelligently it is useful to analyze the simple relationship which exists between the statistical properties of a decision rule and the expected values of errors in inventory estimates.

Assume that a total of N samples are to be allocated among K classes. Assume further that a decision rule has been chosen and associate with this decision rule a *confusion matrix* C defined as

$C(i,j) \rightarrow$ conditional probability that a sample from class j is assigned to class i.

In addition, define the following symbols:

$N_i \rightarrow$ number of samples from class i included in the total sample set.

$\hat{N}_i \rightarrow$ number of samples estimated to be in class i by applying the decision rule and counting the number of samples allocated to class i.

$n_{i,j} \rightarrow$ actual number of samples from class j that are assigned by the decision rule to class i

Our first relationship is

$$\hat{N}_i = \sum_{j=1}^{K} n_{i,j} \tag{1}$$

The symbol $n_{i,j}$ represents a binomial random variable with mean and variance given by

$$E(n_{i,j}) = N_j C(i,j) \tag{2a}$$

$$E([E(n_{i,j}) - n_{i,j}]^2) = N_j C(i,j)[1 - C(i,j)] \tag{2b}$$

Hence

$$E(\hat{N}_i) \cdot \sum_{j=1}^{K} N_j C(i,j) \tag{3}$$

Define

$$p_i \rightarrow \frac{N_i}{N}$$

$$\hat{p}_i \rightarrow \frac{\hat{N}_i}{N}$$

Equation 3 yields

$$E(\hat{p}_i) = \sum_{j=1}^{K} p_j C(i,j) \tag{(4}$$

At this point it is useful to introduce matrix notation. Define K dimensional column vectors P and $\hat{P}$ as

$$P \rightarrow \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix} \quad , \quad \hat{P} \rightarrow \begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \vdots \\ \hat{p}_K \end{bmatrix}$$

The symbol P represents a vector of correct proportions of the scene accounted for by each of the K classes. The symbol $\hat{P}$ represents the proportions estimated by applying the decision rule and counting the number of samples allocated to each of the K classes. The information included in equation 4 can be expressed in matrix form as

$$CP = E(\hat{P}) \tag{5}$$

Equation 5 implies that inventory estimates obtained from the application of a gi rule are unbiased if and only if the associated confusion matrix is the identity matrix.

Assume that the confusion matrix C is invertible and define a vector $\hat{P}_-$ of modifie estimates as

$$\hat{P}_- = C^{-1} \hat{P}$$

It follows that $\hat{P}_-$ is an unbiased estimator of P. Let D represent the covariance matrix

$$E([\hat{P}_- - P][\hat{P}_- - P]^T) = C^{-1} D (C^{-1})^T$$

From the definition of $\hat{P}$ and equations 1 and 2b, one can show that the ith diagor of D is

$$D(i,i) = \frac{1}{N^2} \sum_{j=1}^{K-1} N_j C(i,j)(1 - C(i,j))$$

Since

$$C(i,j)(1 - C(i,j)) \leqslant 1/4$$

it follows that an upper bound on a diagonal element of D is

$$D(i,i) \leqslant \frac{1}{4N}$$

We have shown that a sufficient condition for the derivation of unbiased inventory estin use of a decision rule whose confusion matrix is computable and invertible. The varianc ing unbiased inventory estimates are bounded by terms which are proportional to 1/4N wl total sample size. Hence for large sample sizes typical of satellite derived remote sensing inventory estimates given by the right side of equation 6 should be very accurate.

Based on the results of this section, we can specify the following criteria for choosing a decision rule for inventory estimation

A) computational efficiency

B) the computability of the associated confusion matrix

C) the invertibility of the associated confusion matrix.

With regard to criterion B it is worth mentioning that there is no available practical algorithm for computing the confusion matrix of a multidimensional Bayes classifier. For this reason as well as for reasons of computational efficiency, the conventional Bayes or maximum likelihood decision rule is not an optimal choice for use in inventory estimation.

In order to satisfy criterion A, our search for an appropriate decision rule was limited to those which employ dimensionality reduction. A simple example of dimensionality reduction is the selection of certain subcomponents of the sample vectors for use in classification. This is referred to as *feature selection* with the components as features. On a more sophisticated level, one can introduce a priori information or specific problem knowledge to determine relationships among sample vector components. These relationships implicitly define a dimensionality reducing transformation. This procedure is called *feature extraction*. But since useful problem knowledge is not always available, the feature extraction approach preferred for our application depends entirely on training sample statistics. With this approach one attempts to find the dimensionality reducing transformation conditioned on training sample statistics which optimizes a performance measure. The advantages of this type of feature extraction when applied to remote sensing data are its ability to be implemented in an autonomous or non-interactive fashion, and its interdisciplinary character.

In the case of a one dimensional Bayes decision rule it is possible to compute the associated confusion matrix. Details of the computation are provided in the Appendix. Hence, in order to

satisfy criterion B attention can be restricted to linear one dimensional feature extraction and Bayes classification procedures. This approach to inventory estimation can be outlined as follows:

A) linearly map all samples onto one dimension.

B) obtain one dimensional a priori class statistics from training sample statistics by applying the usual laws for the behavior of first and second order moments under linear transformations.

C) employ a normality assumption and allocate the one dimensional samples among classes using the conventional Bayes or maximum likelihood decision rule.

D) construct a vector of inventory estimates by determining the proportion of the total sample set allocated to each class

E) use the algorithm described in the Appendix to compute the confusion matrix of the one dimensional Bayes decision rule.

F) premultiply the vector of inventory estimates by the inverse of the confusion matrix to obtain unbiased estimates.

If the procedure outlined in steps A through F were to begin with a one dimensional linear transformation chosen more or less at random the probable result would be failure at step F because the confusion matrix in question is apt to be singular. Hence, to satisfy criterion C a computationally efficient procedure is required for choosing the coefficients of a one dimensional linear transformation such that the resulting confusion matrix used in step F is invertible. Since the elements of a confusion matrix are positive and the columns must sum to one, a logical candidate is the feature extraction whose associated confusion matrix has maximum trace. Notice that this choice also has the property that among all possible linear one dimensional feature extractions, it provides the minimal error rate when a priori class probabilities are equal. In this sense, it is the optimal one dimensional feature extraction. The coefficients of the optimal one dimensional feature extraction can be

obtained by means of a conventional multidimensional optimization procedure which is outlined in the Appendix. This feature extraction is recommended for use in the inventory estimation process described in steps A through F and is used in the numerical example of Section 3.

3. A NUMERICAL SIMULATION

The feature extraction and classification scheme described in Section 2 and the Appendix was incorporated into a computer program on the HP9825A programmable desk calculator. The program is in the HP basic language and occupies about 4000 bytes of the available memory. The input to the program consists of four dimensional mean vector and covariance matrices, and convergence control parameters for the Newton-Raphson optimization procedure. The output of the program consists of a four dimensional vector V which defines the one dimensional feature extraction, the means and variances of the resulting one dimensional classes, and the confusion matrix.

For this numerical simulation class statistics were obtained from a LANDSAT 2 scene obtained over Finney County, Kansas during May of 1975 [1]. The five classes consisted of two types of winter wheat and three confuser crops. The class statistics were obtained from well known sites in Finney County. The four channels are those of the Multispectral Scanner on board the LANDSAT 2. The sizes of the training sample sets range from about one hundred to about three hundred. The class statistics are shown in Table 1. Among the several factors which limit the accuracy of a Bayes decision rule in classifying remote sensing data are

A) Significant deviations of class populations from normality

B) Errors introduced by small sample sizes

C) Training samples are not randomly selected from the populations in question and, thus, are not representative.

Because this simulation ignores such factors, misclassification probabilities will be somewhat optimistic compared to what might be obtained in an actual application. To measure the quality of our feature extraction and classification procedure it was decided to compare its performance to that of a Bayes decision rule operating on all four channels of the Multispectral Scanner data. This decision rule is optimal in the sense that it minimizes global misclassification probability. The performance

Table 1
Statistics of LANDSAT-2 MSS Signatures Acquired May 1975 Over
Finney County, Kansas

Channel	Mean	Std. Dev.	Covariance Matrix 1	2	3	4
(1) 184 Pixels of Non-Wheat						
1	27.7	3.6	12.7		SYMMETRIC	
2	24.5	8.0	25.0	63.4		
3	75.1	20.4	-51.4	-140.7	415.5	
4	37.4	12.0	-30.8	-84.2	242.1	143.4
(2) 333 Pixels of Non-Wheat						
1	34.7	3.6	12.7			
2	40.4	5.5	17.2	30.0		
3	47.0	5.2	8.8	9.9	27.3	
4	19.7	2.5	0.6	-1.2	10.4	6.0
(3) 324 Pixels of Non-Wheat						
1	33.3	1.6	2.6			
2	38.5	2.7	2.6	7.2		
3	44.1	6.4	4.3	2.5	41.2	
4	18.7	3.3	1.9	0.3	19.9	11.1
(4) 106 Pixels of Winter Wheat						
1	28.5	2.4	5.8			
2	27.5	4.0	7.4	16.2		
3	51.2	5.2	-6.0	-14.4	26.7	
4	24.0	3.0	-4.3	-8.9	14.1	9.0
(5) 127 Pixels of Winter Wheat						
1	21.5	2.7	7.3			
2	16.7	4.2	10.3	18.0		
3	54.9	5.1	4.1	4.9	26.0	
4	29.1	2.8	-1.0	-2.8	11.4	8.1

of the four dimensional Bayes classifier was determined by means of a monte carlo program written for the HP9825A programmable desk calculator. The assumptions of the monte carlo simulation were

A. The five class populations are normally distributed

B. The class statistics are those of Table 1

C. *A priori* class probabilities are equal

D. Monte carlo sample sizes are sufficiently large that sampling error is insignificant.

One thousand samples were obtained from each class and classified into one of the five classes according to a four dimensional Bayes decision rule as described in the Appendix. The element in th ith row and the ith column of the confusion matrix was estimated as the proportion of samples chosen from the jth class which were assigned to the ith class. The estimated confusion matrix is shown in Table 2. The table also shows the expected values of inventory estimates as obtained from the four dimensional Bayes decision rule. In this case expected values of inventory estimates can be obtained by averaging the rows of the confusion matrix. From Table 2 it is also seen that the combined misclassification probability for classes 4 and 5 which are associated with winter wheat is 0.1.

Table 2
Confusion Matrix for Four Dimensional Bayes Classifier

Class	1	2	3	4	5
1	0.81	0.01	0.01	0.01	0.02
2	0.02	0.58	0.12	0.03	0
3	0.08	0.34	0.84	0.03	0
4	0.07	0.07	0.03	0.88	0.06
5	0.02	0	0	0.05	0.92
Expected Values of Inventory Estimate	0.17	0.15	0.26	0.22	0.20

The statistics shown in Table 1 were used as input to the one dimensional feature extraction and classification program. For this simulation *a priori* class probabilities were assumed to be equal. Hence the correct value for each class inventory estimate is 0.2. The resulting confusion matrix along with expected values of inventory estimates are shown in Table 3. The confusion matrix was computed analytically according to the methods described in the Appendix. The combined misclassification probability for classes 4 and 5 is 0.14. However, expected values of inventory estimates for winter wheat classes 4 and 5 are seen to be considerably worse than corresponding estimates for the optimal four dimensional Bayes classifier. The primary reason is errors of commission introduced by misclassification of samples from class 1 into classes 4 and 5. But section 2 describes how the confusion matrix can be used to correct inventory estimates to yield unbiased estimates. As in Section 2, let $\hat{P}$ be a five dimensional vector representing inventory estimates for the five classes. Repeating Equation 6, an unbiased inventory estimate $\hat{P}_{-}$ for these classes can be obtained as

$$\hat{P}_{-} = C^{-1} \hat{P} \tag{11}$$

where C is the confusion matrix given in Table 3.

Table 3
Confusion Matrix for One Dimensional Feature Extraction and Bayes Classifier

Class	1	2	3	4	5
1	0.22	0	0	0	0.04
2	0.07	0.39	0.20	0.03	0
3	0.04	0.54	0.78	0.04	0
4	0.34	0.07	0.02	0.84	0.08
5	0.34	0	0	0.09	0.88
Expected Value of Inventory Estimate	0.05	0.14	0.28	0.27	0.26

An explicit form for C^{-1} is

$$C^{-1} = \begin{bmatrix} 4.85 & -0.01 & 0 & 0.02 & -0.22 \\ -1.02 & 4.00 & -1.02 & -0.10 & 0.06 \\ 0.54 & -2.75 & 1.99 & 0.01 & -0.03 \\ -1.73 & -0.27 & 0.04 & 1.20 & -0.03 \\ -1.70 & -0.03 & 0 & -1.13 & 1.23 \end{bmatrix} \tag{12}$$

From Table 3 we have

$$E(\hat{P}) = \begin{bmatrix} 0.05 \\ 0.14 \\ 0.28 \\ 0.27 \\ 0.26 \end{bmatrix} \tag{13}$$

Also

$$E(\hat{P}_{_}) = C^{-1} E(\hat{P}) \tag{14}$$

Equations 11, 12, 13 and 14 yield

$$E(\hat{P}_{_}) = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} \tag{15}$$

Equation 15 is a logical consequence of the unbiased property of $\hat{P}_{_}$ as an estimator. But the results of this simulation serve as a useful numerical check on the validity of our development.

4. SUMMARY

The results of this paper suggest that a decision rule used for inventory estimation should satisfy the following criteria

A) computational efficiency

B) the computability of the associated confusion matrix

C) the invertibility of the associated confusion matrix.

When criteria B and C are satisfied, it is shown that unbiased inventory estimates can be constructed. The variances of these estimates are bounded by terms which are proportional to 1/N where N is the total sample size. Hence, for the large sample sizes typical of satellite based remote sensing scenes very accurate inventory estimates should be achievable.

The computational procedure recommended in this paper and which satisfies the above criteria is outlined in steps A through F as given in Section 2, with the one dimensional feature extraction chosen so as to maximize the trace of the associated confusion matrix. The mathematical details of the confusion matrix computation and the required optimization process are included in the Appendix.

The computational algorithms needed for this approach to inventory estimation were programmed for a small desk top computing machine and a simulated inventory estimation was performed using class statistics obtained from an observed LANDSAT scene. The simulation demonstrated the unbiased nature of resulting inventory estimates.

ACKNOWLEDGMENTS

We thank Dr. Paul Beaudet of Business and Technological Systems, Inc. and Dr. John Barker of the Goddard Space Flight Center for their aid during the course of this research.

REFERENCES

1. Σ^2 Study, Capability Demonstration for LANDSAT-D System, Final Report, by General Electric, Contract NAS 5-23412, Mod 30.

2. Patrick, E., "Fundamentals of Pattern Recognition", Prentice-Hall, Inc., 1972.

3. Davenport, W., and Root, W., "Random Signals and Noise", McGraw-Hill, Inc., 1958.

4. Conte, S. and DeBoor, C., "Elementary Numerical Analysis", McGraw-Hill, Inc., 1972.

5. Fukunaga, K., "Introduction to Statistical Pattern Recognition", Academic Press, 1972.

APPENDIX

OPTIMAL ONE DIMENSIONAL FEATURE EXTRACTION

Assume K normal random variables $\eta(\eta_i, \Sigma_i)$, $i = 1, 2, \ldots K$ each of dimension N. Samples are chosen from these populations according to a probability law defined by *a priori* class probabilities P_i, $i = 1, 2, \ldots K$. The problem is to classify each sample by assigning it to the class from which it was chosen. In this case the Bayes classifier can be obtained by defining K functions, each with an Ndimensional argument

$$S_i(X) = (X - \eta_i)^T \Sigma_i^{-1} (X - \eta_i) + \ln |\Sigma_i| - 2\ln P_i, \; i = 1, 2, \ldots K \tag{1}$$

Assign a sample X to the class indexed by i if and only if

$$S_i(X) < S_j(X) \text{ for all } j \text{ not equal to } i. \tag{2}$$

The decision rule defined by equation 2 can be shown to minimize risk under a zero-one loss matrix [2]. It also maximizes likelihood. Hence, by the Neyman-Pearson Lemma [3], it minimizes the global probability of misclassification.

If the number of samples to be classified is large and if the integers N and K are relatively large, the above defined decision rule can be expensive to implement. An alternative is to map the N dimensional samples to a one dimensional space and perform the Bayes classification in one dimension. Let V be the N dimensional vector which defines the transformation. Each N dimensional sample X to be classified is mapped onto a scalar χ by

$$\chi = V^T X \tag{3}$$

The classification problem can be redefined in the following fashion. Assume K one dimensional normal random variables $\eta(\mu_i, \sigma_i)$, $i = 1, 2, \ldots K$. The required means and variances are defined as

$$\mu_i = V^T \eta_i, i = 1, 2, \ldots K \tag{4a}$$

$$\sigma_i^2 = V^T \Sigma_i V, i = 1, 2, \ldots K \tag{4b}$$

The scalar X defined by equation 3 can be considered as having been chosen from one of the above defined populations according to a probability law defined by *a priori* class probabilities P_i, i = 1, 2, . . . K. Define the K functions

$$f_i(X) = \frac{(X - \mu_i)^2}{\sigma_i^2} + \ln \sigma_i^2 - 2\ln P_i, i = 1, 2, \ldots K \tag{5}$$

The scalar X, and hence the vector X, is assigned to the class indexed by i if and only if

$$f_i(X) < f_j(X) \text{ for all } j \text{ not equal to } i. \tag{6}$$

For every vector V, equations 3 through 6 provide a well defined decision rule with which one can associate a global misclassification probability P. Our object is to obtain a V which minimizes P. Represent the relationship between V and P as

$$P = F(V) \tag{7}$$

We require an explicit and executable algorithm for the function symbolized in equation 7. Given that, conventional multidimensional optimization techniques can be used to complete our task.

For any vector V, associate a so-called confusion matrix D_V defined as follows:

$D_V[i, j] \rightarrow$ conditional probability that a sample chosen from class i is assigned to class j under the decision rule defined by equations 3 through 6 (8)

It follows that

$$P = 1 - \sum_{i=1}^{k} P_i D_V[i, i] \tag{9}$$

For equation 9 to be of use, a computable expression for the elements of D_V must be provided. For simplicity of exposition we assume that if $i \neq j$, $\sigma_i \neq \sigma_j$. It is not difficult to modify the results to account for duplications.

For each j, define the set T_j as follows

$$T_j = \left\{\text{all } x \text{ such that } f_j(x) \leq f_i(x), i = 1, 2, \ldots K\right\} \tag{10}$$

Then

$$D_V[i, j] = \int_{T_j} \eta(\mu_i, \sigma_i)\, dX \tag{11}$$

It remains to provide an explicit description of the sets T_j, $j = 1, 2, \ldots K$.

For each $i \leq K$ and $j \leq K$ define the sets $\beta_{i,j}$ as

$$\beta_{i,j} = \left\{\text{all } x \text{ such that } f_j(x) \leq f_i(x)\right\} \tag{12}$$

Then

$$T_j = \bigcap_i \beta_{i,j} \tag{13}$$

From equation 5 and 6 it follows that the boundary points of $\beta_{i,j}$ are the solution of the quadratic equation

$$Ax^2 + Bx + C = 0 \tag{14}$$

with

$$A = \frac{1}{\sigma_j^2} - \frac{1}{\sigma_i^2} \tag{15a}$$

$$B = \frac{2\mu_i}{\sigma_i^2} - \frac{2\mu_j}{\sigma_j^2} \tag{15b}$$

$$C = \frac{\mu_j^2}{\sigma_j^2} - \frac{\mu_i^2}{\sigma_i^2} + \ln\left(\frac{\sigma_j^2}{\sigma_i^2}\right) - 2\ln\left(\frac{P_j}{P_i}\right) \tag{15c}$$

Define the boundary points of $\beta_{i,j}$ as

$$T_{1,i,j} = \min \left[\frac{-B + \sqrt{B^2 - 4AC}}{2A} , \frac{-B - \sqrt{B^2 - 4AC}}{2A} \right] \tag{16a}$$

$$T_{2,i,j} = \max \left[\frac{-B + \sqrt{B^2 - 4AC}}{2A} , \frac{-B - \sqrt{B^2 - 4AC}}{2A} \right] \tag{16b}$$

Hence,

$$\beta_{i,j} = \begin{cases} [T_{1,i,j}, T_{2,i,j}] \text{ if } \sigma_j < \sigma_i \\ [-\infty, T_{1,i,j}] \cup [T_{2,i,j}, \infty] \text{ if } \sigma_j > \sigma_i \end{cases} \tag{17}$$

It will be convenient to represent the set of boundary points of an arbitrary set S by the symbolism b (s). From equation 17

$$b(\beta_{i,j}) = T_{1,i,j} \cup T_{2,i,j} \tag{18}$$

From equations 13 and 17 it follows that the set T_j can be represented as the disjoint union of a finite set of intervals and that

$$b(T_j) \subset \bigcup_i b(\beta_{i,j}) \tag{19}$$

When the elements of $b(T_j)$ are linearly ordered according to size, left sided and right sided boundary points must alternate. Hence, to reconstruct the set T_j it is sufficient to know which elements of $\bigcup_i b(\beta_{i,j})$ are elements of $b(T_j)$ and the classification of the smallest element in $b(T_j)$ as either a left sided or a right sided boundary point. For each element $z \in \bigcup_i b(\beta_{i,j})$, z is placed in $b(T_j)$ if and only if for each $i \leqslant K$

$$\begin{aligned} &T_{1,i,j} \leqslant z \leqslant T_{2,i,j}, \text{ if } \sigma_j < \sigma_i \\ &z \leqslant T_{1,i,j} \text{ or } z \geqslant T_{2,i,j}, \text{ if } \sigma_j > \sigma_i \end{aligned} \tag{20}$$

Order the elements of $b(T_j)$ by increasing size to create the indexed set $\{s_{j,m}\}$. There exists an $\ell \leqslant z$ and an $i \leqslant K$ such that

$$s_{j,1} = T_{\ell,i,j} \tag{21}$$

Classify $s_{j,1}$ as either a left sided or right sided boundary point according to the following rule

$$\left.\begin{array}{l} \ell = 1 \text{ and } \sigma_j < \sigma_i \\ \ell = 2 \text{ and } \sigma_j > \sigma_i \end{array}\right\} \rightarrow \text{left sided}$$
$$\left.\begin{array}{l} \ell = 1 \text{ and } \sigma_i < \sigma_j \\ \ell = 2 \text{ and } \sigma_i > \sigma_j \end{array}\right\} \rightarrow \text{right sided} \quad . \tag{22}$$

In order to conveniently compute the integral on the right side of equation 11, transform the indexed set $\{s_{j,m}\}$ into the indexed set $\{r_{j,m}\}$ by the mapping

$$r_{j,m} = (s_{j,m} - \mu_i)/\sigma_i \tag{23}$$

Let Q be the index of the largest element in $\{r_{j,m}\}$ and let $d(x)$ represent the distribution function of the standard normal random variable. Then $D_V[i,j]$ can be computed as follows when $s_{j,1}$ is a right sided boundary point

$$D_V[i,j] = 1 - d(r_{j,Q}) + d(r_{j,1}) + \sum_{\ell=2}^{Q-2} d(r_{j,\ell+1}) - d(r_{j,\ell}) \tag{24}$$

and when $s_{j,1}$ is a left sided boundary point

$$D_V[i,j] = \sum_{\ell=1}^{Q-1} d(r_{j,\ell+1}) - d(r_{j,\ell}) \tag{25}$$

Equations 8 through 25 give explicit expression to the functional relationship symbolized by equation 7. The vector V which minimizes P can be estimated by a conventional multidimensional Newton-Raphson iteration procedure [4]. This procedure requires all first and second order derivatives of P with respect to the elements of V. These derivatives can be obtained analytically by implicit differentiation and the application of the chain rule on equations 9, 15, 16, 23, 24, and 25. A sufficiently accurate first guess is also required. An adequate first guess can be obtained as follows. The Fisher criterion [5] for class separability between classes indexed by i and j is given by

$$f_{i,j} = (\mu_i - \mu_j)^2/(\sigma_i^2 + \sigma_j^2) \tag{26}$$

The feature extraction procedure which maximizes $f_{i,j}$ is defined by the vector $V_{i,j}$ which is

$$V_{i,j} = (\tfrac{1}{2}\,\Sigma_i + \tfrac{1}{2}\,\Sigma_j)^{-1}(\eta_i - \eta_j) \tag{27}$$

The first guess V_0 is obtained as the average $V_{i,j}$ over all pairs (i, j), $i \leq K$, $j < i$. Hence

$$V_0 = \frac{2\,(K-2)!}{K!} \sum_{i=2}^{K} \sum_{j=1}^{i-1} V_{i,j} \tag{28}$$

For some applications, the correct classification of samples of certain types may be more important than correct classification of samples of other types. Hence, it is worthwhile to introduce a simple generalization to the feature extraction procedure defined by equations 3 through 28. Let G be a K dimensional vector consisting of zeroes and ones. Instead of choosing a V to minimize the global probability of misclassification as defined by equation 9, choose a V which minimizes

$$P_G = \sum_{i=1}^{K} G(i)\, P_i\,(1 - D_V[i, j]) \tag{29}$$

The symbol P_G represents the probability of misclassification of samples chosen from all classes indexed by an integer i such that $G(i) = 1$. The generalization is obtained by substituting equation 29 for equation 9. The rest of the development given by equations 10 through 28 is left undisturbed.