N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE

AgRISTARS

A Joint Program for
Agriculture and
Resources Inventory
Surveys Through
Aerospace
Remote Sensing

Supporting Research

March 1980

ESTIMATION OF PROBABILITIES OF LABEL IMPERFECTIONS
AND CORRECTION OF MISLABELS

C. B. Chittineni

1. Report No. JSC-16342; SR-L0-00427	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Estimation of Probabilities of Label Imperfections and Correction of Mislabels		5. Report Date March 1980
		6. Performing Organization Code
7. Author(s) C. B. Chittineni		8. Performing Organization Report No. LEMSCO-14356
		10. Work Unit No.
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058		11. Contract or Grant No. NAS 9-15800
		13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Tech. Monitor: Dr. R. Heydorn		Technical Report
		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Schemes that use both the imperfectly labeled and unlabeled pattern sets for the estimation of probabilities of label imperfections and correction of mislabels are presented in this paper. Using relationships between the class conditional densities, a priori probabilities with and without imperfections in the labels, the problem of estimating probabilities of label imperfections is formulated as that of minimizing the Bayes probability of error. Experimental results are presented from the processing of remotely sensed multispectral scanner imagery data. A thresholding scheme is proposed for the correction of pattern mislabels. For a symmetric mislabeling case, a relationship is developed between the probability that such a scheme gives a bad label to a pattern and the probability that the scheme accepts the original label of the pattern. This relationship could be used for computing the threshold from unlabeled samples for a specified probability of bad labeling. An example is presented to illustrate the behavior of the scheme. Furthermore, bounds are presented between the Bayes errors with and without imperfections in the labels and are shown to become equalities when the imperfections in the labels become symmetric.

17. Key Words (Suggested by Author(s)) Correction of mislabels, pattern recognition, probability of bad labeling, probability of error, probability of label imperfections, remote sensing, symmetric labeling errors, thresholding scheme.	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 44	22. Price*

TECHNICAL REPORT

ESTIMATION OF PROBABILITIES OF LABEL IMPERFECTIONS
AND CORRECTION OF MISLABELS

Job Order 73-302

This report describes Classification activities of the
Supporting Research project of the AgRISTARS program.

PREPARED BY

C. B. Chittineni

APPROVED BY

T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

John C. Wainwright

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division

Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

March 1980

PREFACE

The research which is the subject of this paper was accomplished in support of the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program within the Earth Observations Division, Space and Life Sciences Directorate, of the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center. The author, a principal scientist for Lockheed Engineering and Management Services Company, Inc., performed this work under Contract NAS 9-15800.

CONTENTS

TABLES

FIGURES

1. INTRODUCTION

In the practical applications of pattern recognition such as in the classification of remotely sensed multispectral scanner (MSS) imagery data, it is usually difficult to obtain labels for the training patterns. The labels for the training patterns are provided by an analyst-interpreter (AI), who examines the imagery film and uses ancillary information (such as historical information, cropping practices, and crop calendar models for agricultural imagery). Very often these labels are imperfect, and acquiring labels for the training patterns is costly.

In the literature, several researchers investigated the problem of pattern recognition with imperfectly labeled patterns. Kashyap (ref. 1) proposed an iterative training procedure for a two-class case. Shanmugam and Breiphol (ref. 2) developed an error-correcting procedure for disjoint densities using Parzen density estimators (refs. 3-6). Chittineni (refs. 7-9) investigated the problem of learning with imperfectly labeled patterns and studied the applicability of probabilistic distance measures for feature selection with imperfectly labeled patterns. Most of these proposed schemes require the knowledge of probabilities of label imperfections, which usually are not available.

Several scientists considered the problem of estimation of recognition system performance (refs. 10-15). Highleyman (ref. 12) investigated the problem of estimating the probability of error of a given classifier both for known and unknown a priori probabilities. Fukunaga and Kessell (ref. 13) examined the problem of estimating the probability of error using unlabeled samples. Chow (ref. 14) established a relationship between error and reject rates, which is useful in estimating the probability of error using unlabeled samples. Chittineni (ref. 15) investigated the problem of estimating recognition system performance and probabilities of label imperfections as maximum likelihood estimates from the classifier decisions of labeled and unlabeled patterns.

1-7

/

It is the purpose of this paper to present schemes for estimating the probabilities of label imperfections and correcting the labels of mislabeled patterns with the specified probability that the label correction scheme gives a bad label to a pattern. It is assumed that a set of imperfectly labeled patterns and a set of unlabeled patterns are given. The proposed schemes use Parzen density estimators and both imperfectly labeled and unlabeled pattern sets.

The paper is organized in the following manner. Section 2 presents a model for label imperfections and develops relationships between the densities, with and without imperfections in the labels. Section 3 develops a scheme for estimating the probability of label imperfections using Parzen density estimators and presents experimental results in the processing of remotely sensed MSS imagery data. Section 4-1 presents a thresholding scheme for the correction of pattern mislabels; in section 4-2, a relationship between the probability that the label correction scheme gives a bad label to a pattern and the probability that it accepts the original label of a pattern is developed for a symmetric mislabeling case. In section 4-3, an example is presented for normal distributions having equal a priori probabilities and equal covariance matrices to illustrate the behavior of the mislabel correction scheme. Conclusions are presented in section 5. In appendix A, a relationship is developed between the Bayes probability of errors with and without imperfections in the labels for symmetric probabilities of label imperfections. For a two-class case, bounds are presented on the probability of error without imperfections in the labels in terms of label imperfection probabilities and probability of error with imperfections in the labels. These bounds are shown to become identities when the imperfections in the labels become symmetric. In appendix B, a thresholding scheme is proposed for the correction of mislabels when $\beta < b$.

2. A MODEL FOR LABEL IMPERFECTIONS

Let ω and ω' be the perfect and imperfect labels, respectively, each of which takes the values $1,2,\cdots,M$, where M is the number of classes. Let $P(\omega = i)$ and $p(X|\omega = i)$ be the a priori probabilities and the class conditional densities, respectively, of the patterns in classes $\omega = i$. The imperfections in the labels are described by the probabilities

$$\beta_{ji} = P(\omega' = i|\omega = j) \quad ; \quad i,j = 1,2,\cdots,M \tag{1}$$

where i and j indicate class. We have the constraint

$$\sum_{i=1}^{M} \beta_{ji} = 1 \tag{2}$$

It is assumed that

$$p(X|\omega = j) = p(X|\omega' = i, \omega = j) \tag{3}$$

That is, the density of a pattern, given its true label, does not depend on its imperfect label. To obtain a relationship between $p(X|\omega = i)$ and $p(X|\omega' = i)$, consider

$$p(X|\omega' = i) = \frac{1}{P(\omega' = i)} \sum_{j=1}^{M} p(X,\omega' = i,\omega = j)$$

$$= \frac{1}{P(\omega' = i)} \sum_{j=1}^{M} p(X|\omega' = i,\omega = j)P(\omega' = i|\omega = j)P(\omega = j)$$

$$= \frac{1}{P(\omega' = i)} \sum_{j=1}^{M} \beta_{ji}P(\omega = j)p(X|\omega = j) \tag{4}$$

Cross-multiplying and dividing equation (4) by $p(X)$ establishes the relationship between a posteriori probabilities:

$$p(\omega' = i|X) = \sum_{j=1}^{M} \beta_{ji}p(\omega = j|X) \tag{5}$$

Similarly, the a priori probabilities are related as follows.

2-1

3

$$P(\omega' = i) = \sum_{j=1}^{M} \beta_{ji} P(\omega = j) \tag{6}$$

Let
$$\Delta = \beta_{11}\beta_{22} - \beta_{12}\beta_{21} \tag{7}$$

Inverting equation (4) yields the following result for a two-class case.

$$P(\omega = i)p(X|\omega = i) = \frac{1}{\Delta}[\beta_{jj}P(\omega' = i)p(X|\omega' = i) - \beta_{ji}P(\omega' = j)p(X|\omega' = j)]$$
$$i,j = 1,2 \atop i \neq j \tag{8}$$

Similarly, for the a priori and a posteriori probabilities,

$$P(\omega = i) = \frac{1}{\Delta}[\beta_{jj}P(\omega' = i) - \beta_{ji}P(\omega' = j)] \quad ; \quad i,j = 1,2 \atop i \neq j \tag{9}$$

$$p(\omega = i|X) = \frac{1}{\Delta}[\beta_{jj}p(\omega' = i|X) - \beta_{ji}p(\omega' = j|X)] \quad ; \quad i,j = 1,2 \atop i \neq j \tag{10}$$

For a symmetric case, when

$$\beta_{11} = \beta_{22} = \beta \text{ and } \beta_{12} = \beta_{21} = 1 - \beta \tag{11}$$

then
$$\Delta = (2\beta - 1) \tag{12}$$

From equations (7), (11), and (12),

$$[P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2)] = \frac{1}{(2\beta - 1)}[P(\omega' = 1)p(X|\omega' = 1)$$
$$- P(\omega' = 2)p(X|\omega' = 2)] \tag{13}$$

2-2

4

3. ESTIMATION OF LABEL IMPERFECTION PROBABILITIES

In this section, the problem of estimating probabilities of label imperfections β_{ji} is considered. It is assumed that a set of patterns $X_i(j)$, $j = 1,2,\cdots,N_i$ is given with imperfect labels $\omega' = i$, $i = 1,2,\cdots,M$, and a set of unlabeled patterns X_j, $j = 1,2,\cdots,N$. It is also assumed that the a priori probabilities $P(\omega' = i)$ of the imperfectly labeled classes are available.

3.1 ESTIMATION OF BAYES PROBABILITY OF ERROR

The risk incurred by the Bayes classifier is the minimum risk that can be achieved. The labeled and unlabeled samples can be used to estimate the Bayes probability of error as follows: Let $p(X)$ be the mixture density function. That is,

$$p(X) = \bar{P}(\omega = 1)p(X|\omega = 1) + \bar{P}(\omega = 2)p(X|\omega = 2) + \cdots + P(\omega = M)p(X|\omega = M)$$

$$(14)$$

where values for $P(\omega = i)$ are the a priori probabilities and those for $p(X|\omega = i)$ are the class conditional densities. The Bayes classifier classifies a pattern X into a class, the a posteriori probability of which is largest. When X is classified according to the Bayes decision rule, the conditional probability of error is

$$r(X) = 1 - \max_i[p(\omega = i|X)] \tag{15}$$

The Bayes probability of error is then given by

$$P_e = E[r(X)] = \int r(X)p(X)dx \tag{16}$$

Thus, if we know $r(X)$ as a function of X, the Bayes probability of error P_e can be estimated by the sample mean $r(X_i)$ of N test patterns as

$$\hat{P}_e = \frac{1}{N}\sum_{i=1}^{N} r(X_i) \tag{17}$$

where X_i is drawn from the mixture density and the labels of X_i are not needed. The estimate of equation (17) is unbiased; that is,

$$P_e = E(\hat{P}_e) \tag{18}$$

since

$$0 \leq r(X) \leq \frac{M-1}{M} \tag{19}$$

The variance of $\hat{P}_e$ is given by

$$Var\ (\hat{P}_e) = \frac{E[r^2(X)] - P_e^2}{N}$$

$$\leq \frac{P_e(1 - P_e)}{N} - \frac{P_e}{MN} \tag{20}$$

The variance of $\hat{P}_e$ is at least $\frac{P_e}{MN}$ less than the variance of the error estimate, based on counting misclassified labeled test patterns $\frac{P_e(1 - P_e)}{N}$. This is because the error count estimate gives a binary quantization of the error on the test pattern while $r(X)$ assigns a real value. To use equation (17) in estimating the Bayes probability of error, knowledge of the risk function is required. The risk function $r(X)$ can be obtained using density estimators for class conditional densities.

3.2 PARZEN ESTIMATE OF $r(X)$

Given a sequence of independent, identically distributed, random n-dimensional vectors $X_1, X_2, \cdots, X_N$ from a distribution with probability density function $p(X)$, the Parzen estimate of $p(X)$ is given (refs. 3-6) by

$$p_N(X) = \frac{1}{N[h(N)]^n} \sum_{i=1}^{N} K\left(\frac{X - X_i}{h(N)}\right) \tag{21}$$

With the proper choice of the weighting function $h(N)$ and kernel $K(\cdot)$, $p_N(X)$ tends uniformly in probability to $p(X)$. Choosing a normal kernel

$$h(N)^{-n} K\left(\frac{X - X_i}{h(N)}\right) = \frac{(2\pi)^{-n/2} N^{1/2}}{\sqrt{|\Sigma|}} \exp\left[-\frac{N^{1/n}}{2}(X - X_i)^T \Sigma^{-1}(X - X_i)\right] \tag{22}$$

where Σ is the sample covariance matrix of the data. The Parzen estimate of the conditional error for any X is given by

$$r_p(X) = 1 - \max_i \left[\frac{P(\omega = i)p_{N_i}p(X|\omega_i)}{P(\omega = 1)p_{N_1}(X|\omega = 1) + \cdots + P(\omega = M)p_{N_M}(X|\omega = M)} \right] \tag{23}$$

3.3 ESTIMATION OF β_{ij} IN THE MULTICLASS CASE

Let

$$
B = \begin{bmatrix}
\beta_{11} & \beta_{21} & \cdots & \beta_{M1} \\
\beta_{12} & \beta_{22} & \cdots & \beta_{M2} \\
& \vdots & & \\
\beta_{1M} & \beta_{2M} & \cdots & \beta_{MM}
\end{bmatrix}
$$

$$P_{\omega|X} = [p(\omega = 1|X), p(\omega = 2|X), \cdots, p(\omega = M|X)]^T$$

$$P_{\omega'|X} = [p(\omega' = 1|X), p(\omega' = 2|X), \cdots, p(\omega' = M|X)]^T$$

$$\delta = B^{-1}$$

$$\tag{24}$$

From equations (5) and (24), we obtain

$$P_{\omega|X} = \delta P_{\omega'|X} \tag{25}$$

Now the problem of estimating probabilities of label imperfections β_{ij} is formulated as follows.

Find: β_{ij}, $i,j = 1,2,\cdots,M$ such that P_e is minimized, where

$$P_e = 1 - \left\{ \frac{1}{N} \sum_{\ell=1}^{N} \max_i [p(\omega = i|X_\ell)] \right\} \tag{26}$$

3-3

7

subject to the constraints

$$\sum_{i=1}^{M} \beta_{ji} = 1 \quad ; \quad j = 1,2,\cdots,M$$

$$0 \leq \beta_{ji} \quad ; \quad i,j = 1,2,\cdots,M$$

(27)

and

From the given set of imperfectly labeled patterns and unlabeled patterns, various quantities in equation (26) are estimated as follows.

$$p(X|\omega' = i) = \frac{(2\pi)^{-n/2}}{N_i^{1/2}\sqrt{|\Sigma_i|}} \sum_{\ell=1}^{N_i} \exp\left\{ -\frac{N_i^{1/n}}{2} [X - X_i(\ell)]^T \Sigma_i^{-1}[X - X_i(\ell)]\right\}$$

(28)

where Σ_i is the sample covariance matrix of the patterns in the class $\omega' = i$.

$$p(\omega' = i|X) = \frac{P(\omega' = i)p(X|\omega' = i)}{P(\omega' = 1)p(X|\omega' = 1) + P(\omega' = 2)p(X|\omega' = 2) + \cdots + P(\omega' = M)p(X|\omega' = M)}$$

(29)

and $p(\omega = i|X)$, $i = 1,2,\cdots,M$, is obtained from equations (25), (28), and (29). The estimates of β_{ij} that minimize P_e subject to the constraints of equation (27) can easily be obtained using optimization techniques such as the Davidon-Fletcher-Powell procedure (refs. 16-18).

3.4 ESTIMATION OF β_{ij} IN THE TWO-CLASS CASE

From equation (10), in a two-class case the Bayes risk $r(X)$ becomes

$$r(X) = \min[p(\omega = 1|X), p(\omega = 2|X)]$$

$$= \frac{1}{2} - \frac{1}{2}\left|p(\omega = 1|X) - p(\omega = 2|X)\right|$$

$$= \frac{1}{2} - \frac{1}{2|\beta_{11}\beta_{22} - \beta_{12}\beta_{21}|}\left|[(\beta_{22} + \beta_{12})p(\omega' = 1|X) - (\beta_{11} + \beta_{21})p(\omega' = 2|X)]\right|$$

$$= \frac{1}{2} - \frac{1}{|\beta_{11} + \beta_{22} - 1|}\left|2p(\omega' = 1|X) + \beta_{22} - \beta_{11} - 1\right|$$

(30)

For a two-class case, the problem of estimating β_{ij} may now be formulated as follows.

3-4

8

Find: β_{11}, β_{22}

such that P_e is minimized, where

$$P_e = \frac{1}{2} - \frac{1}{2|\beta_{11} + \beta_{22} - 1|} \frac{1}{N} \sum_{j=1}^{N} \left| [2p(\omega' = 1|X_j) + \beta_{22} - \beta_{11} - 1] \right| \quad (31)$$

subject to the constraints

$$\left. \begin{array}{c} 0 \le \beta_{11} \le 1 \\ 0 \le \beta_{22} \le 1 \end{array} \right\} \quad (32)$$

The a posteriori probability $p(\omega' = 1|X)$, can be estimated using equations (28) and (29). The probabilities of label imperfections β_{ii} ($i = 1,2$) that minimize the P_e of equation (31), subject to the constraints of the inequalities in equation (32), can be easily obtained using an optimization technique such as that of Davidon-Fletcher-Powell. Experimental results in processing remotely sensed MSS data are presented in the next section.

3.5 EXPERIMENTAL RESULTS

In this section, some results are obtained by applying the theory presented in the previous sections for estimating the probabilities of label imperfections in processing remotely sensed Landsat MSS imagery data. The images are of a 5- by 6-nautical-mile area called a segment. The image is divided into a rectangular array of pixels, 117 rows by 196 columns. The images are overlaid with a rectangular grid. Two classes are considered: class 1 is wheat, and class 2 is "other." The pixels at the grid intersections are labeled by an AI using film products of the images and ancillary data such as historic information and crop growth stage models. These labels are imperfect labels. Also, ground truth (GT) labels, which are the true labels for these pixels, are acquired. Twelve features and 836 unlabeled patterns are used. The numbers of imperfectly labeled patterns in each class are listed in table 3-1, and the a priori probabilities $P(\omega' = i)$ are estimated from the number of imperfectly labeled patterns in each class. The Davidon-Fletcher-Powell optimization method is used to estimate β_{ij} by minimizing P_e of equation (31) subject to

3-5

9

TABLE 3-1.- COMPARISON OF ESTIMATED LABELING ACCURACIES WITH THE ONES COMPUTED FROM GT LABELS

| Segment | Description | No. of AI labeled patterns | | Probabilities of label imperfections $P(\omega'=i\,|\,\omega=j)$ | |
		Wheat	"Other"	Estimated using proposed method	Computed from comparison of AI, and GT labels
1231	Jackson County, Okla.	71	25	0.9586 0.0414 0.1330 0.8670	0.9315 0.0685 0.1304 0.8696
1520	Bigstone County, Mont.	20	71	0.8629 0.1371 0.0363 0.9637	0.7917 0.2083 0.0150 0.9850

the constraints of inequalities set out in equation (31). Table 3-1 summarizes the estimated labeling accuracies using the method proposed in the paper and computed labeling accuracies using AI and GT labels.

From table 3-1, it is seen that the labeling accuracies estimated using the proposed method are in reasonable agreement with the ones computed from the GT labels. Also, it is to be noted that, although the GT labels of remote sensing data are fairly accurate, they are not perfect.

4. MISLABEL CORRECTION WITH SPECIFIED PROBABILITY OF BAD LABELING

In this section, the problem of identification and correction of mislabels of patterns using unlabeled patterns is considered. In particular, thresholding schemes are proposed for the identification and correction of mislabels. A relationship is developed between the probability that such a scheme gives a bad label to a pattern and the probability that the scheme accepts the original label of the pattern. This relationship could be used in computing the threshold with a specified probability of bad labeling. It is assumed that the probabilities of label imperfections are symmetric. That is,

$$\beta = \beta_{ii} \quad ; \quad i = 1,2,\cdots,M$$

and

$$b = \beta_{ij} \quad ; \quad \begin{matrix} i,j = 1,2,\cdots,M \\ i \neq j \end{matrix}$$

(33)

It is also assumed that β_{ij} is estimated using a technique such as the one in section 3. The following thresholding scheme is proposed when $\beta > b$. The case when $\beta < b$ is treated in appendix B.

4.1 A THRESHOLDING SCHEME FOR MISLABEL CORRECTION WHEN $\beta > b$

For the identification and correction of mislabels of the patterns when $\beta > b$, the following scheme is proposed.

Change the label of X to $\omega = i$ whenever

$$\max_{i}[p(\omega' = i|X)] > 1 - t \tag{34}$$

where t is some threshold; otherwise, do not change the label of X. That is, the label of X is changed whenever there is enough confidence in the thresholding scheme to change the label. (Section 4.2 discusses the computation of t.) Define a random variable U(X),

$$U(X) = \max_{i}[p(\omega' = i|X)] \tag{35}$$

Let $V_{DCL}(t)$ be the region of the feature space over which, for a particular threshold t, the original label of pattern X is accepted. That is,

$$V_{DCL}(t) = [X|U(X) \leq (1 - t)] \qquad (36)$$

Let $P_{DCL}(t)$ be the probability that the thresholding scheme will not change the label of a pattern at threshold t and the probability that a pattern lies in the region $V_{DCL}(t)$. Using the above label correction scheme, whenever the label of a particular pattern X is changed, let $P_{BL}(t)$ be the probability that a bad label will be given to a pattern. The threshold t can be determined by specifying the P_{BL}. A relationship between $P_{DCL}(t)$ and $P_{BL}(t)$ which can be used to compute the threshold t using unlabeled patterns is developed in the next section.

4.2 A RELATIONSHIP BETWEEN $P_{BL}(t)$ AND $P_{DCL}(t)$

In this section, a relationship is developed between $P_{BL}(t)$ and $P_{DCL}(t)$ for symmetric probabilities of label imperfections when $\beta > b$. Suppose that the threshold is decreased from t to t - Δt. Then let the region $V_{DCL}(t)$ be expanded from $V_{DCL}(t)$ to $V_{DCL}(t - \Delta t)$. At threshold t, the labels of the patterns in the incremental region $\Delta V_{DCL}(t)$ are changed; but, at threshold t - Δt, they are not changed. The patterns in the region $\Delta V_{DCL}(t)$ satisfy the relation

$$(1 - t)p(X) \leq \max_i [P(\omega' = i)p(X|\omega' = i)] \leq (1 - t + \Delta t)p(X) \qquad (37)$$

Let $\Delta P_{DCL}(t)$ be the increment in the probability $P_{DCL}(t)$. It is also the probability that a pattern lies in the region $\Delta V_{DCL}(t)$. That is,

$$\Delta P_{DCL}(t) = \int_{\Delta V_{DCL}(t)} p(X)dx \qquad (38)$$

Let $\Delta P_{CL}(t)$ and $\Delta P_{BL}(t)$ be the increments of the probability of correct labeling and of the probability of bad labeling, respectively, when the threshold is decreased from t to t - Δt. Because in the increase of region $V_{DCL}(t)$, there will be a decrease in the probabilities $P_{CL}(t)$ and $P_{BL}(t)$. When $\beta > b$, $\Delta P_{CL}(t)$ and $\Delta P_{BL}(t)$ can be written as

$$-\Delta P_{CL}(t) = \int_{\Delta V_{DCL}(t)} \max_i [P(\omega = i)p(X|\omega = i)]dx \qquad (39)$$

4-2

/3

and
$$-\Delta P_{BL}(t) = \int_{\Delta V_{DCL}(t)} [1 - \max_i p(\omega = i|X)]p(X)dx \tag{40}$$

In the region $\Delta V_{DCL}(t)$, the probability $\Delta P_{DCL}(t)$ can be split into two parts: (1) the decrease in the probability of correct labeling of a pattern $\Delta P_{CL}(t)$ and (2) the decrease in the probability of bad labeling of a pattern $\Delta P_{BL}(t)$. That is, from equations (38), (39), and (40), we obtain

$$\Delta P_{DCL}(t) = -\Delta P_{CL}(t) - \Delta P_{BL}(t) \tag{41}$$

Consider

$$p(\omega' = i|X) = \sum_{j=1}^{M} \beta_{ji} p(\omega = j|X)$$

$$= \beta p(\omega = i|X) + b \sum_{\substack{j=1 \\ j \neq i}}^{M} p(\omega = j|X)$$

$$= (\beta - b)p(\omega = i|X) + b \tag{42}$$

From equation (42), we obtain

$$\max_i [p(\omega = i|X)] = \frac{1}{(\beta - b)} \max_i [p(\omega' = i|X)] - \frac{b}{(\beta - b)} \tag{43}$$

Using equations (39) and (43) in equation (40) yields

$$\int_{\Delta V_{DCL}(t)} \max_i [p(\omega' = i|X)]p(X)dx = (\beta - b)\Delta P_{BL}(t) + \beta \Delta P_{DCL}(t) \tag{44}$$

Therefore, from equations (37) and (44), in the incremental region $\Delta V_{DCL}(t)$, we have

$$\frac{(1 - \beta - t)}{(\beta - b)} \Delta P_{DCL}(t) \leq \Delta P_{BL}(t) \leq \frac{(1 - t - \beta)}{(\beta - b)} \Delta P_{DCL}(t) + \frac{\Delta t \, \Delta P_{DCL}(t)}{(\beta - b)} \tag{45}$$

Summing equation (45), with t steadily decreasing from t to 0, yields the following.

$$\Sigma \frac{(1 - \beta - t)}{(\beta - b)} \Delta P_{DCL}(t) \leq \Sigma \Delta P_{BL}(t)$$

$$\leq \Sigma \frac{(1 - t - \beta)}{(\beta - b)} \Delta P_{DCL}(t) + \Sigma \frac{\Delta t \, \Delta P_{DCL}(t)}{(\beta - b)} \tag{46}$$

If we let Δt tend to zero, the last sum of the above equation vanishes, resulting in

$$P_{BL}(t) = \int_0^t \frac{(1 - \beta - \xi)}{(\beta - b)} \, dP_{DCL}(\xi) \tag{47}$$

Equation (46) shows a relationship between $P_{DCL}(t)$ and $P_{BL}(t)$. Once the densities are estimated from the imperfectly labeled patterns, $P_{DCL}(t)$ can be computed from the unlabeled samples. For a specified P_{BL}, equation (47) can be used to compute the threshold t.

4.3 AN EXAMPLE

The mislabel correction scheme presented in section 4.1 does not change the label of pattern X whenever $\max_i p(\omega' = i|X) \leq (1 - t)$. For a two-class case, the region of $V_{DCL}(t)$ also can be described as those X-values for which the following relation is satisfied:

$$\frac{t}{1 - t} \leq \frac{P(\omega' = 1)p(X|\omega' = 1)}{P(\omega' = 2)p(X|\omega' = 2)} \leq \frac{1 - t}{t} \tag{48}$$

Using equation (4) for a symmetric mislabeling case, equation (48) can be written as

$$\frac{t - (1 - \beta)}{\beta - t} \leq \frac{P(\omega = 1)p(X|\omega = 1)}{P(\omega = 2)p(X|\omega = 2)} \leq \frac{\beta - t}{t - (1 - \beta)} \tag{49}$$

For this example, it is assumed that the a priori probabilities are equal and the class conditional densities are Gaussian with equal covariance matrices. That is,

$$\left. \begin{array}{c} P(\omega = i) = 0.5 \\[2mm] p(X|\omega = i) \sim N(M_i, \Sigma) \\[2mm] i = 1,2 \end{array} \right\} \tag{50}$$

4-4

15

Let
$$v(X) = \log\left[\frac{P(\omega = 1)p(X|\omega = 1)}{P(\omega = 2)p(X|\omega = 2)}\right]$$

$$= X^T \Sigma^{-1}(M_1 - M_2) - \frac{1}{2}\left(M_1^T \Sigma^{-1} M_1 - M_2^T \Sigma^{-1} M_2\right) \tag{51}$$

Let
$$s^2 = (M_1 - M_2)^T \Sigma^{-1}(M_1 - M_2) \tag{52}$$

where s is the Mahalanobis distance between the pattern classes.

Since $v(X)$ is a linear combination of Gaussian random variables, it is also normally distributed. The class conditional densities of $v(X)$ can be written as

$$p[v(X)|\omega = 1] \sim N(\frac{1}{2}s^2, s^2)$$

$$P[v(X)|\omega = 2] \sim N(-\frac{1}{2}s^2, s^2) \tag{53}$$

Then the probabilities $P_{DCL}(t)$ and $P_{BL}(t)$ can be computed as follows:

Let

$$\ell_1 = \log\left[\frac{t - (1 - \beta)}{\beta - t}\right]$$

$$\ell_2 = \log\left[\frac{\beta - t}{t - (1 - \beta)}\right]$$

$$\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\varepsilon^2) \tag{54}$$

$$a = -\frac{s}{2} - \frac{1}{s}\log\left[\frac{\beta - t}{t - (1 - \beta)}\right]$$

and

$$b = -\frac{s}{2} + \frac{1}{s}\log\left[\frac{\beta - t}{t - (1 - \beta)}\right]$$

Consider the following.

$$P_{DCL}(t) = \left\{ P(\omega = 1) \int_{\ell_1}^{\ell_2} P[v(X)|\omega = 1]d[v(X)] + P(\omega = 2) \int_{\ell_1}^{\ell_2} p[v(X)|\omega = 2]d[v(X)] \right\}$$

$$= 0.5 \left[\int_{\frac{\ell_1 - \frac{1}{2}s^2}{s}}^{\frac{\ell_2 - \frac{1}{2}s^2}{s}} \phi(\xi)d\xi + \int_{\frac{\ell_1 + \frac{1}{2}s^2}{s}}^{\frac{\ell_2 + \frac{1}{2}s^2}{s}} \phi(\xi)d\xi \right]$$

$$= 0.5 \left[\int_a^b \phi(\xi)d\xi + \int_{-b}^{-a} \phi(\xi)d\xi \right]$$

$$= \Phi(b) - \Phi(a)$$

$$(55)$$

where $\Phi(a) = \int_{-\infty}^a \phi(\xi)d\xi$.

Similarly, for a two-class case, the probability that the algorithm gives a bad label can be written as

$$P_{BL}(t) = P(\omega = 1)P\left[0 \leq \frac{P(\omega = 1)p(X|\omega = 1)}{P(\omega = 2)p(X|\omega = 2)} \leq \frac{t - (1 - \beta)}{\beta - t} \middle| \omega = 1 \right]$$

$$+ P(\omega = 2)P\left[\frac{\beta - t}{t - (1 - \beta)} \leq \frac{P(\omega = 1)p(X|\omega = 1)}{P(\omega = 2)p(X|\omega = 2)} \leq \infty \middle| \omega = 2 \right]$$

$$= 0.5 \left[\int_{-\infty}^{\frac{\ell_1 - \frac{1}{2}s^2}{s}} \Phi(\xi)d\xi + \int_{\frac{\ell_2 + \frac{1}{2}s^2}{s}}^{\infty} \Phi(\xi)d\xi \right]$$

$$= 0.5 \left[\int_{-\infty}^a \Phi(\xi)d\xi + \int_{-a}^{\infty} \Phi(\xi)d\xi \right]$$

$$= \Phi(a)$$

$$(56)$$

4-6

17

Figures 4-1 through 4-4 show the plots of P_{BL} versus t, P_{DCL} versus t, P_{CL} versus t, and P_{BL} versus P_{DCL}, respectively, for values of β = 0.95, 0.91, 0.85, and 0.81 and for values of s = 1, 2, 3, and 4. The tip of the arrow points to the direction of increase in the value of β. From figure 4-1, it is seen that for a specified P_{BL} the threshold t increases with the increase in the probability of imperfections in the labels or with the decrease in the value of β.
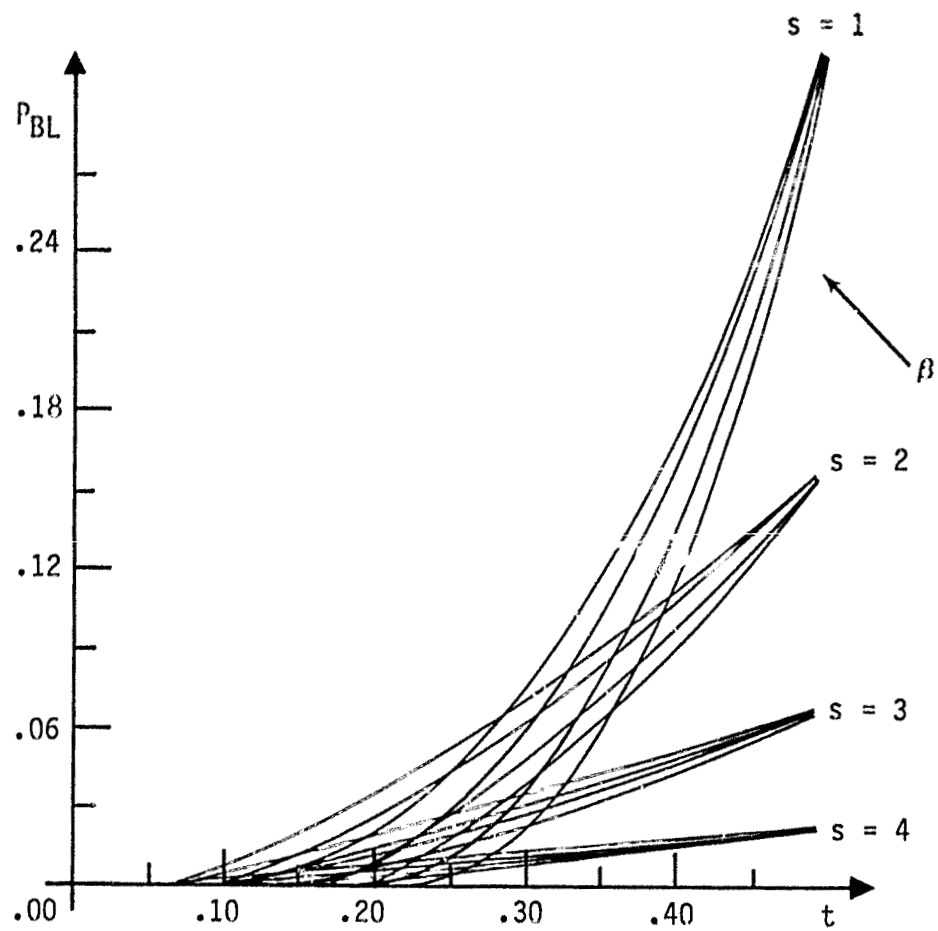
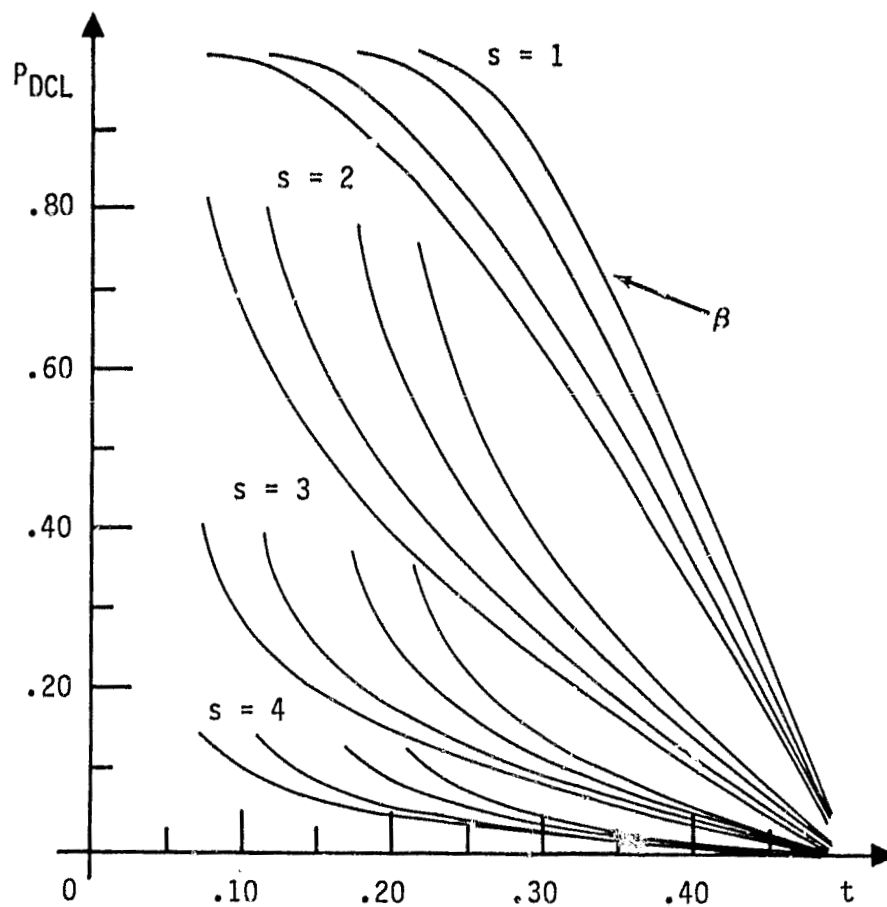Figure 4-1.- Plot of probability of bad label, P_{BL}, versus threshold t.

4-8

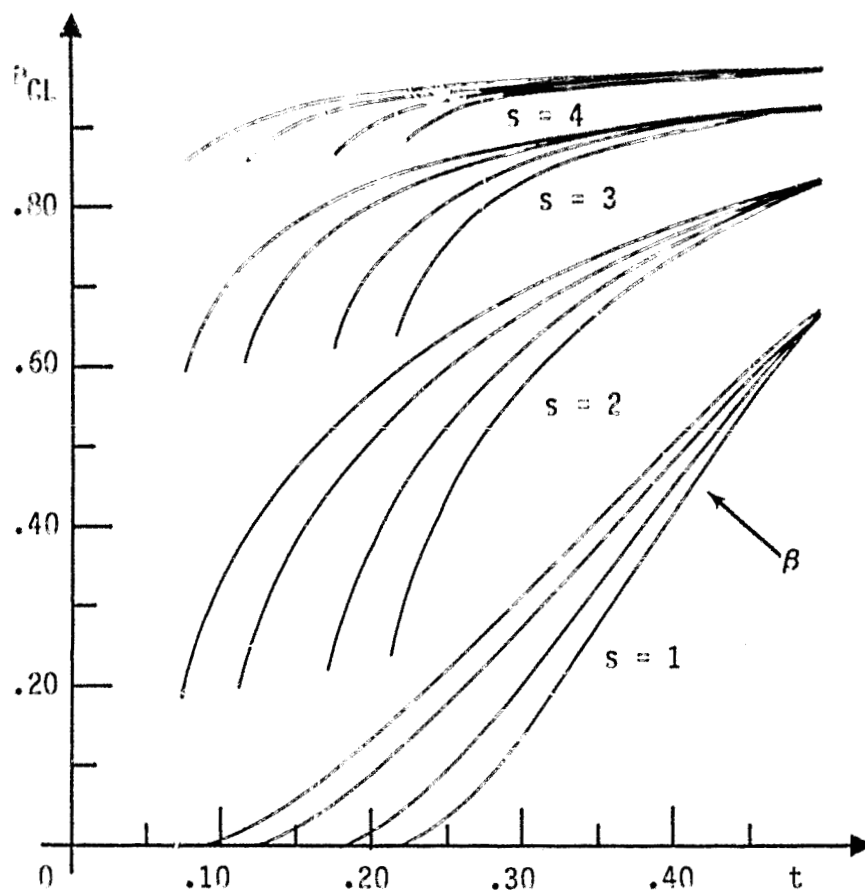Figure 4-2.- Plot of probability of not changing the label, P_{DCL}, versus threshold t.

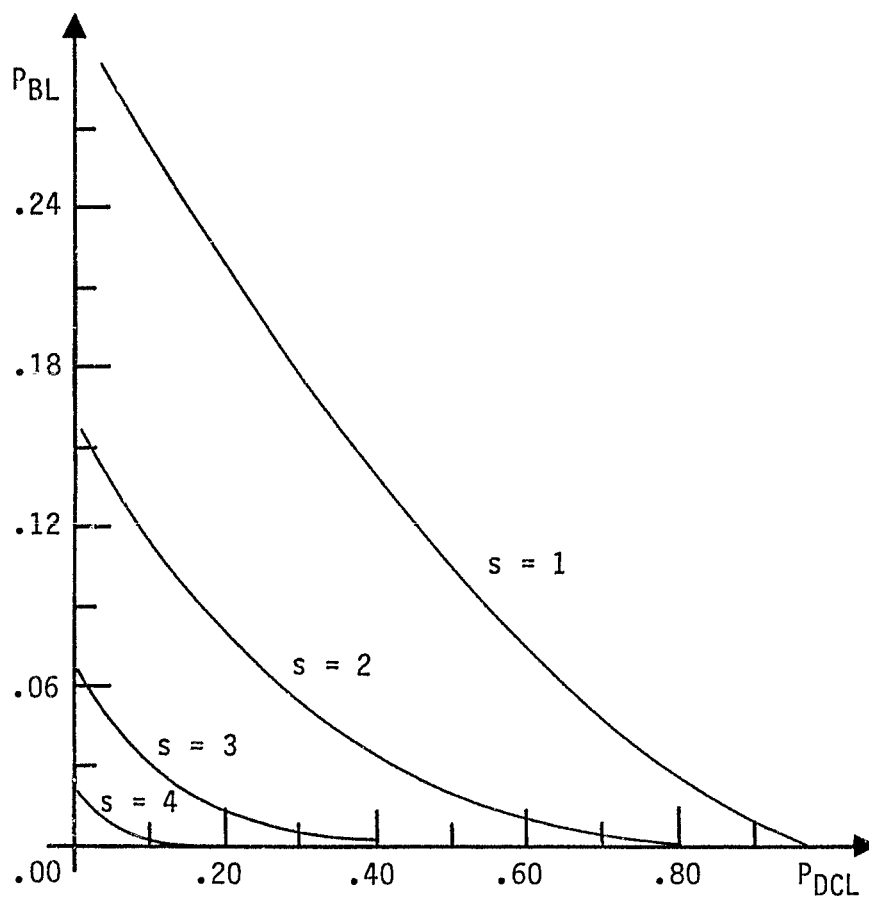Figure 4-3.- Plot of probability of giving a correct label, P_{CL}, versus threshold t.

4-10

Figure 4-4.- Trade-off between P_{BL} and P_{DCL}.

5. CONCLUSIONS

In the practical applications of pattern recognition, obtaining labels for the training patterns is expensive, and very often these labels are imperfect. Schemes are presented in this paper for the estimation of probabilities of label imperfections and correction of mislabels.

The risk incurred by the Bayes classifier is the minimum risk that can be achieved. The conditional risk $r(X)$ can be obtained as a function of X, using estimated densities from the labeled patterns. The probability of error can be estimated as an average value of $r(X)$ over the unlabeled patterns. The resulting estimated probability of error has less variance when compared to the variance of the error estimate based on counting the misclassified labeled test set. Using the relationships between the probability densities with and without imperfections in the labels, the problem of estimating the probabilities of label imperfections is formulated for the two-class and multiclass cases as that of minimizing the Bayes probability of error with probability constraints. Optimization techniques, such as the Davidon-Fletcher-Powell procedure, can be used to estimate the probabilities of label imperfections. Experimental results from processing remotely sensed MSS imagery data are presented. The estimated probabilities of label imperfections using the proposed method and the probabilities of label imperfections computed using the imperfect (AI) and the GT labels are in good agreement.

Thresholding schemes are proposed for correcting mislabels of the patterns. Whenever there is enough confidence in the scheme (as determined by the threshold), the correct label of the pattern is determined. A relationship between the probability that such a scheme will give a bad label to a pattern and the probability that the scheme will accept the original label of the pattern is developed for a symmetric mislabeling case. This relationship could be used to compute the threshold from the relatively inexpensive unlabeled patterns, for a specified probability of bad labeling.

An example is presented for Gaussian distributions with equal covariance matrices and equal a priori probabilities. This illustrates the behavior of the probability that the scheme gives a bad label, the probability that the scheme gives a correct label, and the probability that the scheme accepts the original label. All are functions of the threshold, of various probabilities of label imperfections, and of different Mahalanobis distances between the classes. The trade of curves between P_{BL} and P_{DCL} are presented for this example.

For a two-class case, bounds are presented between the Bayes probabilities of error with and without imperfections in the labels. Furthermore, it is shown that these bounds become identical when the imperfections in the labels become symmetric.

6. REFERENCES

1. Kashyap, R. L.: Algorithms for Pattern Classification. In Adaptive Learning, and Pattern Recognition Systems: Theory and Applications, J. M. Mendel and K. S. Fu, Eds., Academic Press (N.Y.), 1970, pp. 81-113.

2. Shanmugam, K.; and Breiphol, A. M.: An Error Correcting Procedure for Learning With an Imperfect Teacher. IEEE Transactions: Systems, Man, and Cybernetics, vol. SMC-1, pp. 223-229, July 1971.

3. Murty, V. K.: Estimation of Probability Density. Ann. Math Stat., vol. 36, pp. 1027-1031, June 1965.

4. Murty, V. K.: Nonparametric Estimation of Multivariate Densities With Applications. In Multivariate Analysis, P. R. Krishnaiah, Ed., Academic press (N.Y.), 1966, pp. 43-56.

5. Parzen, E.: On Estimation of a Probability Density Function and Mode. Ann. Math. Stat., vol. 33, pp. 1065-1076, Sept. 1962.

6. Cacoullos, T.: Estimation of a Multivariate Density. Ann. Inst. Stat. Math., vol. 18, pp. 179-189, 1966.

7. Chittineni, C. B.: Learning With Imperfectly Labeled Patterns. JSC-14867, LEC-13068, National Aeronautics and Space Administration, Lyndon B. Johnson Space Center (NASA/JSC, Houston), Apr. 1979. Also in Proc. 1979 IEEE Conf. on Pattern Recognition and Image Processing (79 CH 1628-2C), pp. 52-62.

8. Chittineni, C. B.: On the Selection of Effective Features From the Imperfectly Labeled Patterns. Int. J. Computer and Information Sciences, vol. 2, no. 2, pp. 103-114, 1973.

9. Chittineni, C. B.: On Feature Extraction From Imperfectly Labeled Patterns. IEEE Transactions: Systems, Man, and Cybernatics, vol. SMC-3, pp. 290-292, May 1973.

10. Chittineni, C. B.: On the Estimation of Probability of Error. Pattern Recognition, vol. 9, no. 6, pp. 191-196, 1977.

11. Chittineni, C. B.: Fisher Classifier and Its Probability of Error Estimation. JSC-14865, LEC-13301, NASA/JSC (Houston), May 1979.

12. Highleyman, W. H.: The Design and Analysis of Pattern Recognition Experiments. Bell System Tech. J., vol. 41, pp. 723-744, Mar. 1962.

13. Fukunaga, K.; and Kessell, D. L.: Nonparametric Bayes Error Estimation Using Unclassified Samples. IEEE Transactions on Information Theory, vol. IT-19, pp. 434-440, July 1973.

14. Chow, C. K.: On Optimum Recognition Error and Reject Tradeoff. IEEE Transactions on Information Theory, vol. IT-26, pp. 41-46, Jan. 1970.

15. Chittineni, C. B.: On the Maximum Likelihood Estimation of Label Imperfections and Its Use in the Identification of Mislabeled Patterns, JSC-16067, LEC-13678, NASA/JSC (Houston), Sept. 1979.

16. Fletcher, R.; and Powell, M. J. D.: A Rapidly Convergent Descent Method for Minimization. Computer J., vol. 6, pp. 163-168, Apr. 1963.

17. Johnson, I. L.: The Davidon-Fletcher-Powell Penalty Function Method: A Generalized Iterative Technique for Solving Parameter Optimization Problems. Technical note TN D-8251, NASA/JSC (Houston), May 1976.

18. Cooper, L.; and Steinberg, D.: Introduction to Methods of Optimization. W. B. Saunders Co. (Philadelphia), 1970.

APPENDIX A

BAYES ERROR PROBABILITIES WITH AND WITHOUT
IMPERFECTIONS IN THE LABELS

APPENDIX A

BAYES ERROR PROBABILITIES WITH AND WITHOUT
IMPERFECTIONS IN THE LABELS

The Bayes risk in classifying a pattern X can be written as

$$r(X) = 1 - \max_i [p(\omega = i|X)] \tag{A-1}$$

and

$$r'(X) = 1 - \max_i [p(\omega' = i|X)] \tag{A-2}$$

where $r(X)$ is the conditional error with the densities without label imperfections and $r'(X)$ is the conditional risk with imperfections in the labels. The probability of errors can be written as

$$P_e = E_{p(X)} [r(X)] \tag{A-3}$$

and

$$P'_e = E_{p(X)} [r'(X)] \tag{A-4}$$

where E is the expectation operator. For symmetric probabilities of label imperfections of equation (33), theorem A-1 gives the relationship between the error probabilities P_e and P'_e.

Theorem A-1: If the probabilities of imperfections in the labels are symmetric, as given in equation (33) and $\beta > b$, then the Bayes probability of error with and without imperfections in the labels are related as

$$P'_e = (\beta - b)P_e + (1 - \beta) \tag{A-5}$$

Proof: From equations (5), (33), (A-1), and (A-2), we obtain

$$r'(x) = 1 - \max_i \left[\sum_{j=1}^{M} \beta_{ji} p(\omega = j|X) \right] = 1 - \{(\beta - b)\max_i [p(\omega = i|X)] + b\}$$

$$= 1 - \{(\beta - b)[1 - r(X)] + b\}$$

$$= (\beta - b)r(X) + (1 - \beta) \tag{A-6}$$

A-1

28

Taking exceptions on both sides of equation (A-6) yields equation (A-5).

For two-class symmetric probabilities of label imperfections, the error probabilities are related as

$$P_e' = (2\beta - 1)P_e + (1 - \beta) \tag{A-7}$$

If the label imperfection probabilities are not symmetric, the Bayes errors depend on the particular probability density functions of the patterns. However, for a two-class case, the following bounds are obtained between P_e and P_e' and are shown to be an identity of equation (A-7) when the imperfections in the labels become symmetric.

A.1 LOWER BOUND ON P_e

Case (a): $\beta_{11} > \beta_{22}$

Consider the case when $\beta_{11} > \beta_{22}$. From equation (8), we obtain

$$[P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2)]$$

$$= \frac{1}{\Delta}\,[(\beta_{22} + \beta_{12})P(\omega' = 1)p(X|\omega' = 1) - (\beta_{11} + \beta_{21})P(\omega' = 2)p(X|\omega' = 2)]$$

$$= \alpha_1[P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)] - \alpha_2 P(\omega' = 2)p(X|\omega' = 2)$$

$$\tag{A-8}$$

where
$$\alpha_1 = \frac{-\beta_{11} + \beta_{22} + 1}{\beta_{11} + \beta_{22} - 1} > 0$$

and
$$\alpha_2 = \frac{2(\beta_{11} - \beta_{22})}{\beta_{11} + \beta_{22} - 1} > 0 \tag{A-9}$$

Define the regions Ω_1 and Ω_2 as

$$\Omega_1 = \{X | P(\omega' = 1)p(X|\omega' = 1) > P(\omega' = 2)p(X|\omega' = 2)\} \tag{A-10}$$

and
$$\Omega_2 = \{X | P(\omega' = 1)p(X|\omega' = 1) < P(\omega' = 2)p(X|\omega' = 2)\} \tag{A-11}$$

A-2

29

Let Ω_{11} and Ω_{12} be subsets of region Ω_1, where

$$\Omega_{11} = \{X|\alpha_1[P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)]$$

$$> \alpha_2 P(\omega' = 2)p(X|\omega' = 2)\} \tag{A-12}$$

and $\quad \Omega_{12} = \{X|\alpha_1[P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)]$

$$> \alpha_2 P(\omega' = 2)p(x|\omega' = 2)\} \tag{A-13}$$

Let $\qquad \alpha_3 = \dfrac{\alpha_1}{\alpha_1 + \alpha_2} = \dfrac{1 + \beta_{22} - \beta_{11}}{1 + \beta_{11} - \beta_{22}} < 1 \tag{A-14}$

In terms of α_3, the regions Ω_{11} and Ω_{12} become

$$\Omega_{11} = \{X|\alpha_3 P(\omega' = 1)p(X|\omega' = 1) > P(\omega' = 2)p(X|\omega' = 2)\} \tag{A-15}$$

and $\quad \Omega_{12} = \{X|\alpha_3 P(\omega' = 1)p(X|\omega' = 1) < P(\omega' = 2)p(X|\omega' = 2)\} \tag{A-16}$

From equations (A-8) through (A-16), equation (A-17) is obtained.

$\int |P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2)|dx$

$\quad = \alpha_1 \int\limits_{\Omega_2} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)|dx + \alpha_2 \int\limits_{\Omega_2} P(\omega' = 2)p(X|\omega' = 2)dx$

$\quad + \alpha_1 \int\limits_{\Omega_{11}} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)|dx - \alpha_2 \int\limits_{\Omega_{11}} P(\omega' = 2)p(X|\omega' = 2)dx$

$\quad - \alpha_1 \int\limits_{\Omega_{12}} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)|dx + \alpha_2 \int\limits_{\Omega_{12}} P(\omega' = 2)p(X|\omega' = 2)dx$

$\quad = \alpha_1 \int |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)|dx + \alpha_2 P(\omega' = 2)$

$\quad - 2\alpha_1 \int\limits_{\Omega_{12}} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)|dx - 2\alpha_2 P(\omega' = 2) \int\limits_{\Omega_{11}} p(X|\omega' = 1)dx$

$$\tag{A-17}$$

The regions Ω_{11} and Ω_{12} are illustrated in figure A-1. For a two-class case, from equations (A-1) through (A-4), the following relationships are developed.

$$P_e = \frac{1}{2} - \frac{1}{2} \int |P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2| dx \qquad (A\text{-}18)$$

and

$$P'_e = \frac{1}{2} - \frac{1}{2} \int |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)| dx \qquad (A\text{-}19)$$

Using equations (A-18) and (A-19) in equation (A-17) yields the following.

$$P_e = \frac{1}{2} - \frac{1}{2} \alpha_1 - \frac{\alpha_2}{2} P(\omega' = 2) + \alpha_1 P'_e + \alpha_2 P(\omega' = 2) \int_{\Omega_{11}} p(X|\omega' = 2) dx$$

$$+ \alpha_1 \int_{\Omega_{12}} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)| dx$$

$$\geq \frac{1}{2} - \frac{1}{2} \frac{(-\beta_{11} + \beta_{22} + 1)}{(\beta_{11} + \beta_{22} - 1)} - \frac{(\beta_{11} - \beta_{22})}{(\beta_{11} + \beta_{22} - 1)} P(\omega' = 1) + \frac{(-\beta_{11} + \beta_{22} + 1)}{(\beta_{11} + \beta_{22} - 1)} P'_e$$

$$(A\text{-}21)$$

When the imperfections in the labels become symmetric, it is easily seen that $\alpha_2 = 0$ and the region Ω_{12} becomes the null set. The inequality of equation (A-21) then becomes equal and is identical to equation (A-7).

Case (b): $\beta_{11} < \beta_{22}$

Consider the case when $\beta_{11} < \beta_{22}$. Let

$$\left. \begin{array}{l} \alpha_1 = \dfrac{\beta_{11} - \beta_{22} + 1}{\beta_{11} + \beta_{22} - 1} \\[2em] \alpha_2 = \dfrac{2(\beta_{22} - \beta_{11})}{\beta_{11} + \beta_{22} - 1} \end{array} \right\} \qquad (A\text{-}22)$$

and

A-4

Proceeding as before, we obtain

$$P_e = \frac{1}{2} - \frac{\alpha_1}{2} - \frac{\alpha_2}{2} P(\omega' = 1) + \alpha_1 P'_e$$

$$+ \alpha_1 \int_{\Omega_{22}} |P(\omega' = 1)p(X|\omega' = 1) - P(\omega' = 2)p(X|\omega' = 2)| dx$$

$$+ \alpha_2 P(\omega' = 1) \int_{\Omega_{21}} p(X|\omega' = 1) dx$$

$$\geq \frac{1}{2} - \frac{1}{2} \frac{(\beta_{11} - \beta_{22} + 1)}{(\beta_{11} + \beta_{22} - 1)} - \frac{(\beta_{22} - \beta_{11})}{(\beta_{11} + \beta_{22} - 1)} P(\omega' = 1) + \frac{(\beta_{11} - \beta_{22} + 1)}{(\beta_{11} + \beta_{22} - 1)} P'_e$$

$$(A-23)$$

The regions Ω_{21} and Ω_{22} are illustrated in figure A-2.

When the imperfections in the labels are symmetric, it is easily seen that $\alpha_2 = 0$, and the region Ω_{22} becomes the null set. The inequality in equation (A-23) then becomes equal and is identical to equation (A-8).

A.2 UPPERBOUND ON P_e

Case (a): $\beta_{11} > \beta_{22}$

Let
$$\alpha_1 = (2\beta_{22} - 1)$$

$$(A-24)$$

and
$$\alpha_2 = 2(\beta_{11} - \beta_{22})$$

Proceeding in a manner similar to case (a) of section A.1, the probability of errors P_e an P'_e are related as follows.
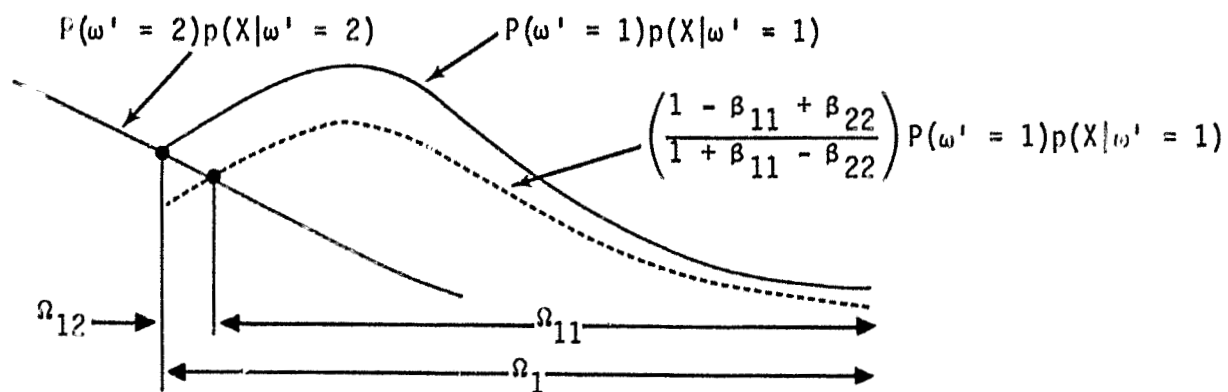
Figure A-1.- Illustration of regions Ω_{11} and Ω_{12} for the lower bound on P_e when $\beta_{11} > \beta_{22}$.
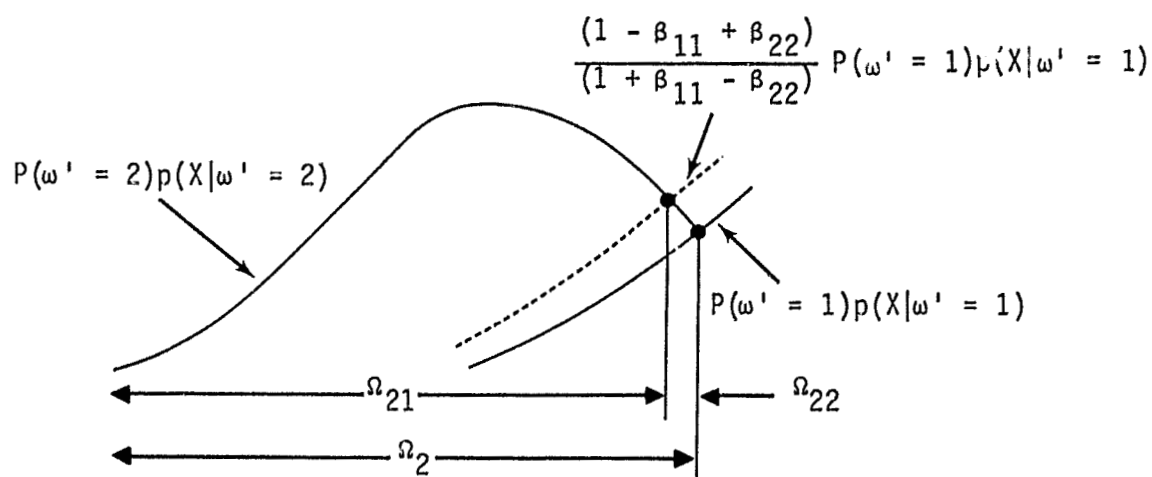

Figure A-2.- Illustration of regions Ω_{21} and Ω_{22} for the lower bound on P_e when $\beta_{11} < \beta_{22}$.

A-6

33

$$P_e = \frac{1}{2} - \frac{1}{2\alpha_1} + \frac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right) P(\omega = 1) + \frac{P'_e}{\alpha_1} - \frac{\alpha_2}{\alpha_1} P(\omega = 1) \int_{\Omega_{21}} p(X|\omega = 1)dx$$

$$- \int_{\Omega_{22}} |P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2)|dx$$

$$\leq \frac{1}{2} - \frac{1}{2(2\beta_{22} - 1)} + \frac{(\beta_{11} - \beta_{22})}{(2\beta_{22} - 1)} P(\omega = 1) + \frac{P'_e}{(2\beta_{22} - 1)} \tag{A-25}$$

The regions Ω_{21} and Ω_{22} are illustrated in figure A-3.

When the imperfections in the labels are symmetric, it is easily seen that $\alpha_2 = 0$, and the region Ω_{22} becomes null. The inequality of equation (A-25) becomes equal and is identical to equation (A-8).

Case (b): $\beta_{11} < \beta_{22}$

Let
$$\alpha_1 = 2\beta_{11} - 1$$

and
$$\alpha_2 = 2(\beta_{22} - \beta_{11}) \tag{A-26}$$

Proceeding in a manner similar to case (a) of section A.1, the probability of errors P_e and P'_e are related as

$$P_e = \frac{1}{2} - \frac{1}{2\alpha_1} + \frac{1}{2}\left(\frac{\alpha_2}{\alpha_1}\right) P(\omega = 2) + \frac{P'_e}{\alpha_1} - \frac{\alpha_2}{\alpha_1} P(\omega = 2) \int_{\Omega_{11}} p(X|\omega = 2)dx$$

$$- \int_{\Omega_{12}} |P(\omega = 1)p(X|\omega = 1) - P(\omega = 2)p(X|\omega = 2)|dx$$

$$\leq \frac{1}{2} - \frac{1}{2(2\beta_{11} - 1)} + \frac{(\beta_{22} - \beta_{11})}{(2\beta_{11} - 1)} P(\omega = 2) + \frac{P'_e}{(2\beta_{11} - 1)} \tag{A-27}$$

The regions Ω_{11} and Ω_{12} are illustratated in figure A-4.

When the imperfections in the labels become symmetric, it is easily seen that $\alpha_2 = 0$, and the region Ω_{12} becomes null. The inequality of equation (A-27) becomes equal and is identical to equation (A-8).
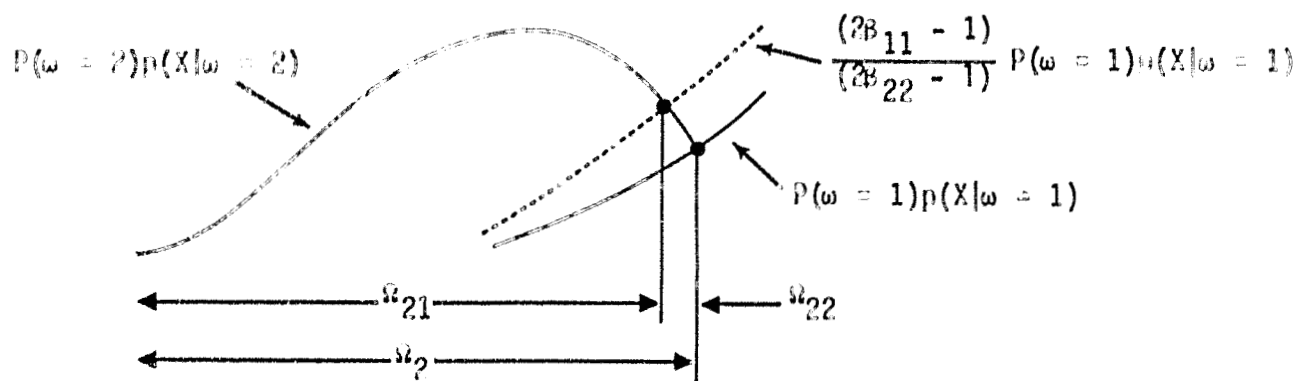
Figure A-3.- Illustration of regions Ω_{21} and Ω_{22} for the upper bound on P_e, when $\beta_{11} > \beta_{22}$.
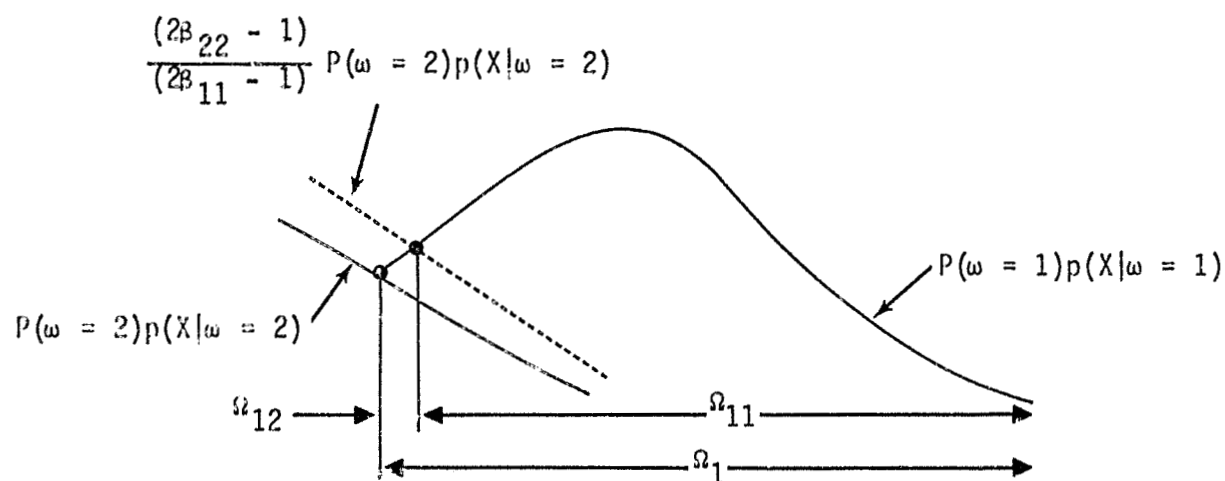

Figure A-4.- Illustration of regions ω_{12} and ω_{11} for the upper bound on P_e when $\beta_{11} < \beta_{22}$.

APPENDIX B

A THRESHOLDING SCHEME FOR THE CORRECTION
OF MISLABELS WHEN $\beta < b$

APPENDIX B

A THRESHOLDING SCHEME FOR THE CORRECTION
OF MISLABELS WHEN $\beta < b$

When $\beta < b$, the following scheme is proposed for identifying mislabeled patterns with symmetric probabilities of label imperfections, as given in equation (33).

Change the label of X to $\omega = i$ whenever

$$\min_i[p(\omega' = i|X)] < 1 - t \tag{B-1}$$

where t is some threshold; otherwise, do not change the label of X. For this scheme, a relationship between $P_{BL}(t)$ and $P_{DCL}(t)$ is obtained in the following and is shown to be equivalent to equation (47).

Let
$$U(X) = \min_i[p(\omega' = i|X)] \tag{B-2}$$

and
$$V_{DCL}(t) = [X|U(X) \geq 1 - t] \tag{B-3}$$

Suppose that the threshold t is decreased from t to $t - \Delta t$. Let $\Delta V_{DCL}(t)$ be the decremental region of $V_{DCL}(t) - V_{DCL}(t - \Delta t)$. For patterns in the region $\Delta V_{DCL}(t)$, we have

$$(1 - t)p(X) \leq \min_i[P(\omega' = i)p(X|\omega' = i)] \leq (1 - t + \Delta t)p(X) \tag{B-4}$$

Let $\Delta P_{DCL}(t)$, $\Delta P_{CL}(t)$, and $\Delta P_{BL}(t)$ be the increments in the probabilities $P_{DCL}(t)$, $P_{CL}(t)$, and $P_{BL}(t)$, respectively, because of the decrease in the threshold from t to $t - \Delta t$. Then, we have

$$-\Delta P_{DCL}(t) = \int_{\Delta V_{DCL}(t)} p(X)dx \tag{B-5}$$

$$\Delta P_{CL}(t) = \int_{\Delta V_{DCL}(t)} \max_i[P(\omega = i)p(X|\omega = i)]dx \tag{B-6}$$

and
$$\Delta P_{BL}(t) = \int_{\Delta V_{DCL}(t)} \{1 - \max_i[p(\omega = i|X)]\}p(X)dx \tag{B-7}$$

B-1

From equations (B-5) through (B-7), we get

$$-\Delta P_{DCL}(t) = \Delta P_{CL}(t) + \Delta P_{BL}(t) \qquad (B-8)$$

When $\beta < b$, from equation (42), we obtain

$$-\max_i [p(\omega = i|X)] = \frac{1}{(b - \beta)} \min_i [p(\omega' = i|X)] - \frac{b}{(b - \beta)} \qquad (B-9)$$

Using equations (B-5) and (B-9) in equation (B-7) yields

$$(b - \beta)\Delta P_{BL}(t) = \beta \, \Delta P_{DCL}(t) + \int_{\Delta V_{DCL}(t)} \min_i [P(\omega' = i)p(X|\omega' = i)]dx$$

$$(B-10)$$

From equations (B-4) and (B-10), in the decremental region $\Delta V_{DCL}(t)$, we have

$$\frac{(t - 1 + \beta)}{(b - \beta)} \Delta P_{DCL}(t) \leq \Delta P_{BL}(t) \leq \frac{(t - 1 + \beta)}{(b - \beta)} \Delta P_{DCL}(t) - \frac{\Delta t \, \Delta P_{DCL}(t)}{(b - \beta)}$$

$$(B-11)$$

Summing equation (B-11), with t steadily decreasing from t to zero, and letting Δt tend to zero results in

$$P_{BL}(t) = \int_0^t \frac{(\xi - 1 + \beta)}{(b - \beta)} dP_{DCL}(\xi) \qquad (B-12)$$

It is seen that equation (B-12) is identical to equation (47).

B-2