DISCRIMINATION AND CLASSIFICATION

Tuesday, April 9, 2019 10:34 PM

INTRODUCTION

DISCRIMINATION: GENERATE A RULE/DISCRIMINANT/CLASSIFIER TO SEPARATE DISTINCT SETS OF OBSERVATIONS.

CLASSIFICATION: TO ALLOCATE NEW OBSERVATIONS TO PREVIOUSLY DEFINED SETS.

SOMETIMES A FUNCTION CAN SEPARATE OBSERVATIONS AND ALLOCATE NEW OBSERVATIONS AT THE SAME TIME, SO THE DISTINCTION BETWEEN SEPARATION AND ALLOCATION BECOMES BLURRED. SO, IN THE REST CONTENT WE SHALL USE "CLASSIFICATION" TO INDICATE ALL THE WORK.

CLASSIFICATION WITH TWO GROUPS

NOW SUPPOSE WE WANT TO SEPARATE TWO GROUPS π_1 AND π_2 WITHOUT KNOWING THEIR EXACT DISTRIBUTIONS, HERE WE STATE THE INGREDIENTS WE NEED FOR CONSTRUCTING THE CLASSIFICATION MODEL:

1. ABASTRACT DISTRIBUTION OF TWO GROUPS, DENOTED AS $f_1(\underline{x})$ AND $f_2(\underline{x})$, WHERE $\underline{x} \in \mathbb{R}^n$.
2. PRIOR PROBABILITY, THAT IS, THE PRIOR POSSIBILITY OF BELONGING TO EACH GROUP FOR OBSERVATIONS, DENOTED AS p_1 AND p_2, $p_1+p_2=1$.
3. COST OF MISCLASSIFICATION:

	COST	CLASSIFIED AS π_1	CLASSIFIED AS π_2
TRUE POPULATION:	π_1	0	$c(2\|1)$
	π_2	$c(1\|2)$	0

WHERE $c(1|2)$ IS WHEN AN OBSERVATION FROM π_2 IS INCORRECTLY CLASSIFIED AS π_1; AND $c(2|1)$ IS SIMILARILY DEFINED.

AFTER HAVING LISTED ALL THE INGREDIENTS, NOW WE CONSTRUCT THE DISCRIMINANT/CLASSIFIER STEP BY STEP.

PERSONAL HABIT, $\mathbb{P}$ REPRESENTS THE PROBABILITY MEASURE

PROBABILITY OF MISCLASSIFICATION

LET $P(i|j) := \mathbb{P}($OBSERVATION COMES FROM GROUP j, AND IS CLASSIFIED AS IN GROUP $i)$, $(i,j=1,2)$

AND $R_i := \{\underline{x} : \underline{x}$ BE CLASSIFIED AS IN GROUP $i\}$, THEN OBVIOUSLY,

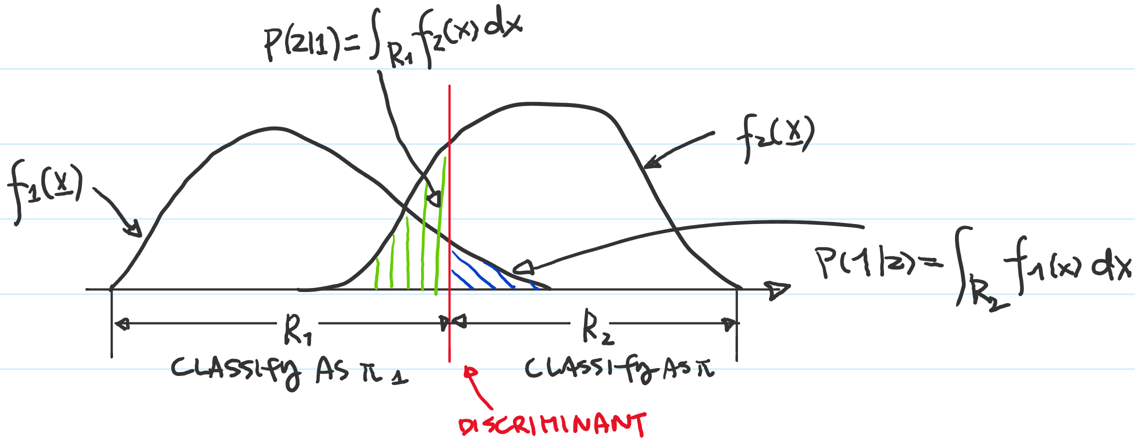
$\mathbb{P}($OBSERVATION IS CORRECTLY CLASSIFIED AS $\pi_1) = P(1|1)\cdot p_1$

$\mathbb{P}($OBSERVATION IS MISCLASSIFIED AS $\pi_1) = P(1|2)\cdot p_2$

$\mathbb{P}($OBSERVATION IS CORRECTLY CLASSIFIED AS $\pi_2) = P(2|2)\cdot p_2$

$\mathbb{P}($OBSERVATION IS MISCLASSIFIED AS $\pi_2) = P(2|1)\, p_1$

IF $\mathbb{R}^n = \mathbb{R}^1$, THEN WE CAN HAVE THE FOLLOWING FIGURE ILLUSTRATION:



EXPECTED COST OF MISCLASSIFICATION (ECM)

INTUITIVELY, WE DEFINE EXPECTED COST OF MISCLASSIFICATION AS:

$$ECM = c(2|1)\,P(2|1)\,p_1 + c(1|2)\,P(1|2)\,p_2$$

A REASONABLE CLASSIFICATION RULE SHOULD HAVE AN ECM AS SMALL, OR NEARLY AS SMALL, AS POSSIBLE. THEREFORE, WE HAVE THE FOLLOWING RESULT:

RESULT 1 THE REGIONS R_1 AND R_2 THAT MINIMIZE THE ECM ARE DEFINED BY THE VALUES $\underline{x}$ FOR WHICH THE FOLLOWING INEQUALITIES HOLD:

DENSITY RATIO COST RATIO PRIOR PROBABILITY RATIO

$$R_1: \quad \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1|2)}{c(2|1)} \cdot \frac{p_2}{p_1}$$

$$R_2: \quad \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{c(1|2)}{c(2|1)} \cdot \frac{p_2}{p_1}$$

PROOF

$$\begin{aligned} ECM &= c(2|1)\,P(2|1)\,p_1 + c(1|2)\,P(1|2)\,p_2 \\ &= c(2|1) \cdot \int_{R_2} f_1(\underline{x})\,d\underline{x} \cdot p_1 + c(1|2) \cdot \int_{R_1} f_2(\underline{x})\,d\underline{x} \cdot p_2 \\ &= c(2|1) \cdot \left[1 - \int_{R_1} f_1(\underline{x})\,d\underline{x}\right] \cdot p_1 + c(1|2) \cdot \int_{R_1} f_2(\underline{x})\,d\underline{x} \cdot p_2 \\ &= \int_{R_1} \left[f_2(\underline{x}) \cdot c(1|2) \cdot p_2 - f_1(\underline{x}) \cdot c(2|1) \cdot p_1\right] d\underline{x} + c(2|1) \cdot p_1 \end{aligned}$$

IF WE WANT TO MINIMIZE THE ECM WITH RESPECT TO R_1, JUST LET R_1 INCLUDE THOSE VALUES $\underline{x}$ FOR WHICH THE INTEGRAND

$$f_2(\underline{x})\,c(1|2)\,p_2 - f_1(\underline{x})\,c(2|1)\,p_1 \leq 0 \qquad (*)$$

AND EXCLUDE THOSE $\underline{x}$ FOR WHICH THIS QUANTITY IS POSITIVE.

FROM $(*)$ WE OBTAIN:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1|2)}{c(2|1)} \cdot \frac{p_2}{p_1} \qquad \blacksquare$$

COROLLARY 2 [SPECIAL CASES OF MINIMUM EXPECTED REGION]

1) IF WE HAVE EQUAL PRIOR PROBABILITY, THEN:

$$R_1 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1|2)}{c(2|1)} \right\}, \quad R_2 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{c(1|2)}{c(2|1)} \right\}$$

2) IF WE HAVE EQUAL MISCLASSIFICATION COST, THEN:

$$R_1 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{p_2}{p_1} \right\}, \quad R_2 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_2}{p_1} \right\}$$

3) IF WE HAVE BOTH EQUAL PRIOR PROBABILITY AND EQUAL MISCLASSIFICATION COST, THEN:

$$R_1 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq 1 \right\}, \quad R_2 = \left\{ \underline{x} : \frac{f_1(\underline{x})}{f_2(\underline{x})} < 1 \right\}$$

GAUSSIANITY MEANS OF NORMAL DISTRIBUTION

CLASSIFICATION WITH TWO GAUSSIAN GROUPS

CLASSIFICATION PROCEDURES BASED ON NORMAL POPULATIONS PREDOMINATE IN STATISTICAL PRACTICE BECAUSE OF THEIR SIMPLICITY AND REASONABLY HIGH EFFICIENCY ACROSS A WIDE VARIETY OF POPULATION MODELS. WE KNOW ASSUME THAT $f_1(\underline{x})$ AND $f_2(\underline{x})$ ARE MULTIVARIATE NORMAL DENSITIES, THE FIRST WITH MEAN VECTOR μ_1 AND COVARIANCE MATRIX Σ_1 AND THE SECOND ONE WITH μ_2 AND Σ_2.

WE SHALL DISCUSS THIS PROBLEM UNDER TWO DIFFERENT CASES:

ALSO CALLED LINEAR DISCRIMINANT ANALYSIS (LDA)

- CLASSIFICATION OF NORMAL GROUPS WHEN $\Sigma_1 = \Sigma_2 = \Sigma$

FIRST WE RECALL THE DENSITY FUNCTION OF MULTIVARIATE GAUSSIAN DISTRIBUTION:

$$f(x) = \frac{1}{\ldots} \exp\left\{ -\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu) \right\} \qquad x \in \mathbb{R}^n$$

DISTRIBUTION:

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})\right\}, \quad \underline{x} \in \mathbb{R}^n$$

SUPPOSE ALSO THAT THE POPULATION PARAMETERS $\underline{\mu}_1$, $\underline{\mu}_2$ AND Σ ARE KNOWN. THEN ACCORDING TO RESULT 1, WE OBTAIN:

$$R_1: \quad \frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1)\right\}}{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right\}}$$

$$= \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right\}$$

$$\geq \frac{C(1|2)}{C(2|1)} \cdot \frac{p_2}{p_1}$$

DIRECTLY,

$$R_2: \quad \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right\} < \frac{C(1|2)}{C(2|1)} \cdot \frac{p_2}{p_1}$$

GIVEN THESE REGIONS R_1 AND R_2, WE CAN CONSTRUCT THE CLASSIFICATION RULE AS FOLLOWING:

RESULT 3 UNDER ABOVE ASSUMPTIONS, THEN THE ALLOCATION RULE THAT MINIMIZES THE ECM IS AS FOLLOWS:

ALLOCATE $\underline{x}_0$ TO π_1 IF

$$(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x}_0 - \frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2) \geq \ln\left[\frac{C(1|2)}{C(2|1)} \cdot \frac{p_2}{p_1}\right]$$

ALLOCATE $\underline{x}_0$ TO π_2 OTHERWISE.

PROOF WE ALREADY KNOW THAT

ALWAYS POSITIVE

$$R_1: \quad \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)\right\} \geq \underline{\frac{C(1|2)}{C(2|1)} \cdot \frac{p_2}{p_1}}$$

$$\Rightarrow \quad -\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2) \geq \ln\left[\frac{C(1|2)}{C(2|1)} \cdot \frac{p_2}{p_1}\right]$$

MOREOVER,

Σ IS SYMMETRIC

MOREOVER,

$$-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma^{-1}(\underline{x}-\underline{\mu}_1)+\frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma^{-1}(\underline{x}-\underline{\mu}_2)$$

CANCEL $\quad$ Σ IS SYMMETRIC COMBINE

$$=\frac{1}{2}\left[-\underline{x}'\Sigma^{-1}\underline{x}+\underline{x}'\Sigma^{-1}\underline{\mu}_1+\underline{\mu}_1'\Sigma^{-1}\underline{x}-\underline{\mu}_1'\Sigma^{-1}\underline{\mu}_1+\underline{x}'\Sigma^{-1}\underline{x}-\underline{x}'\Sigma^{-1}\underline{\mu}_2-\underline{\mu}_2'\Sigma^{-1}\underline{x}+\underline{\mu}_2'\Sigma^{-1}\underline{\mu}_2\right]$$

$$=\frac{1}{2}\left[2(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x}-(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)\right]$$

$$=(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x}-\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)$$

FINALLY, WE GET THE RESULT

$$R_1:\quad (\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}\underline{x}-\frac{1}{2}(\underline{\mu}_1-\underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1+\underline{\mu}_2)\geq \ln\left[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right]$$

NOTICE THAT IT IS A LINEAR FUNCTION WITH RESPECT TO $\underline{x}$, THIS IS WHY THIS METHOD IS CALLED LINEAR DISCRIMINANT ANALYSIS. $\blacksquare$

IN MOST PRACTICE CASE, THE POPULATION QUANTITIES $\underline{\mu}_1$, $\underline{\mu}_2$, AND Σ ARE UNKNOWN, SO THE AFOREMENTIONED RULE MUST BE MODIFIED, THAT IS, WE USE THE INFORMATION PROVIDED BY OBSERVATIONS TO ESTIMATE THESE QUANTITIES.

SUPPOSE THAT WE HAVE n_1 OBSERVATIONS OF THE MULTIVARIATE RANDOM VARIABLE $X'=[X_1, X_2, \dots, X_p]$ FROM π_1, AND n_2 OBSERVATIONS OF THIS QUANTITY FROM π_2, WITH $n_1+n_2-2\geq p$. THEN THE RESPECTIVE DATA MATRICES ARE

$$\underline{X}_1=\begin{bmatrix}\underline{x}_{11}'\\ \underline{x}_{12}'\\ \vdots\\ \underline{x}_{1n_1}'\end{bmatrix}\qquad \underline{X}_2=\begin{bmatrix}\underline{x}_{21}'\\ \underline{x}_{22}'\\ \vdots\\ \underline{x}_{2n_2}'\end{bmatrix}$$

FROM THESE DATA MATRICES, THE SAMPLE MEAN VECTORS AND CO-

VARIANCE MATRICES ARE DETERMINATED BY

$$\bar{\underline{x}}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} \underline{x}_{1j}, \quad S_1 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(\underline{x}_{1j}-\bar{\underline{x}}_1)(\underline{x}_{1j}-\bar{\underline{x}}_1)'$$

$$\bar{\underline{x}}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} \underline{x}_{2j}, \quad S_2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(\underline{x}_{2j}-\bar{\underline{x}}_2)(\underline{x}_{2j}-\bar{\underline{x}}_2)'$$

SINCE THE TWO GROUPS ARE ASSUMED TO HAVE THE SAME CO-VARIANCE, NOW WE INTRODUCE A NEW TERMINOLOGY: POOLED COVARIANCE TO ESTIMATE THE TRUE COVARIANCE.

$$S_{pooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{(n_1-1)+(n_2-1)}$$

ONE CAN VERIFY THAT S_{pooled} IS AN UNBIASED ESTIMATOR OF Σ.

THEREFORE WE CAN UPDATE OUR CLASSIFICATION RULE AS:
ALLOCATE $\underline{x}_0$ TO π_1 IF

$$(\bar{\underline{x}}_1-\bar{\underline{x}}_2)' S_{pooled}^{-1}\underline{x}_0 - \frac{1}{2}(\bar{\underline{x}}_1-\bar{\underline{x}}_2)' S_{pooled}^{-1}(\bar{\underline{x}}_1+\bar{\underline{x}}_2) \geq \ln\left[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right]$$

IT'S ALSO CALLED QUADRATIC DETERMINANT ANALYSIS (Q.D.A.)

- CLASSFICATION OF NORMAL GROUPS WHEN $\Sigma_1 \neq \Sigma_2$

STILL WE USE THE CONCLUTION OBTAINED IN RESULT 1. BUT WHEN $\Sigma_1 \neq \Sigma_2$, WE GET

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}}\cdot \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma_1^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma_2^{-1}(\underline{x}-\underline{\mu}_2)\right\}$$

$$\geq \frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}$$

ALSO, WE TAKE THE LOGARITHM AT BOTH SIDES, THEN

$$\ln\frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} + \left[-\frac{1}{2}(\underline{x}-\underline{\mu}_1)'\Sigma_1^{-1}(\underline{x}-\underline{\mu}_1) + \frac{1}{2}(\underline{x}-\underline{\mu}_2)'\Sigma_2^{-1}(\underline{x}-\underline{\mu}_2)\right] \geq \ln\left[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right]$$

COMBINE $\quad$ Σ_1, Σ_2 SYMMETRIC $\quad$ IRRELEVANT TO $\underline{x}$

$$\ln \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} + \Big[-\tfrac{1}{2}(\underline{x}-\underline{\mu_1})'\Sigma_1^{-1}(\underline{x}-\underline{\mu_1}) + \tfrac{1}{2}(\underline{x}-\underline{\mu_2})'\Sigma_2^{-1}(\underline{x}-\underline{\mu_2})\Big] \geq \ln\Big[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\Big]$$

COMBINE $\qquad \Sigma_1, \Sigma_2$ SYMMETRIC $\qquad$ IRRELEVANT TO $\underline{x}$

$$-\frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} + \frac{1}{2}\Big[\underline{x}'\Sigma_2^{-1}\underline{x} - \underline{x}'\Sigma_2^{-1}\underline{\mu_2} - \underline{\mu_2}'\Sigma_2^{-1}\underline{x} + \underline{\mu_2}'\Sigma_2^{-1}\underline{\mu_2} - \dots$$

$$- \underline{x}'\Sigma_1^{-1}\underline{x} + \underline{x}'\Sigma_1^{-1}\underline{\mu_1} + \underline{\mu_1}'\Sigma_1^{-1}\underline{x} - \underline{\mu_1}'\Sigma_1^{-1}\underline{\mu_1}\Big] \geq \ln\Big[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\Big]$$

$$R_1: -\frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2}\underline{x}'(\Sigma_1^{-1}-\Sigma_2^{-1})\underline{x} + (\underline{\mu_1}'\Sigma_1^{-1} - \underline{\mu_2}'\Sigma_2^{-1})\underline{x} - k \geq \ln\Big[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\Big] \quad (*)$$

WHERE $k := \frac{1}{2}\ln\frac{|\Sigma_1|}{|\Sigma_2|} + \frac{1}{2}(\underline{\mu_1}'\Sigma_1^{-1}\underline{\mu_1} - \underline{\mu_2}'\Sigma_2^{-1}\underline{\mu_2})$.

NOTICE IT'S A QUADRATIC FUNCTION WITH RESPECT TO $\underline{x}$, THIS IS WHY THIS METHOD IS CALLED QUADRATIC DISCRIMINANT ANALYSIS. SO WE CONCLUDE:

RESULT 4 WHEN $\underline{\mu_1}, \underline{\mu_2}, \Sigma_1$ AND Σ_2 ARE KNOWN, MOREOVER $\Sigma_1 \neq \Sigma_2$, THEN THE CLASSIFICATION RULE IS GIVEN BY $(*)$.

IN PRACTICE, THE CLASSIFICATION RULE $(*)$ IS IMPLEMENTED BY SUBSTITUDING THE SAMPLE QUANTITIES $\overline{\underline{x}}_1, \overline{\underline{x}}_2$, S_1 AND S_2 FOR $\underline{\mu_1}, \underline{\mu_2}$, Σ_1 AND Σ_2, RESPECTIVELY.

SIMILARLY, WE CAN UPDATE OUR CLASSIFICATION RULE AS:

ALLOCATE $\underline{x}_0$ TO π_1 IF:

$$-\frac{1}{2}\underline{x}_0'(S_1^{-1}-S_2^{-1})\underline{x}_0 + (\overline{\underline{x}}_1'S_1^{-1} - \overline{\underline{x}}_2'S_2^{-1})\underline{x}_0 - k \geq \ln\Big[\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\Big]$$

WHERE $k = \frac{1}{2}\ln\frac{|S_1|}{|S_2|} + \frac{1}{2}(\overline{\underline{x}}_1'\Sigma_1^{-1}\overline{\underline{x}}_1 - \overline{\underline{x}}_2'\Sigma_2^{-1}\overline{\underline{x}}_2)$,

ALLOCATE $\underline{x}_0$ TO π_2 OTHERWISE.

EVALUATION OF CLASSIFICATION FUNCTION

ONE IMPORTANT WAY OF JUDGING THE PERFORMANCE OF ANY CLASSI-

EVALUATION OF CLASSIFICATION FUNCTIONS

ONE IMPORTANT WAY OF JUDGING THE PERFORMANCE OF ANY CLASSIFICATION PROCEDURE IS TO CALCULATE ITS "ERROR RATES", OR MISCLASSIFICATION PROBABILITY.

· WHEN THE POPULATION DENSITY FUNCTIONS ARE KNOWN

IF THE POPULATION DENSITY FUNCTIONS ARE KNOWN, EVERYTHING IS CLEAR AND ACCURATE FROM THE VIEW OF PROBABILITY. WE SHALL DEFINE THE TOTAL PROBABILITY OF MISCLASSIFICATION (TPM) AS:

$$TPM = p_1 \int_{R_2} f_1(\underline{x})\, d\underline{x} + p_2 \int_{R_1} f_2(\underline{x})\, d\underline{x}$$

· WHEN THE POPULATION DENSITY FUNCTIONS ARE UNKNOWN

HERE WE CAN INTRODUCE A MEASURE OF PERFORMANCE THAT DOES NOT DEPEND ON THE DENSITY FUNCTIONS, IT IS CALLED APPARENT ERROR RATE (APER), WHICH COULD BE EASILY CALCULATED FROM THE CONFUSION MATRIX:

		PREDICTED MEMBERSHIP π_1	PREDICTED MEMBERSHIP π_2	
ACTUAL MEMBERSHIP	π_1	n_{1C} CORRECTLY	$n_{1M} = n_1 - n_{1C}$ MISCLASSIFICATION	n_1
	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

TOTAL NUMBER OF OBSERVATIONS ALLOCATED AS IN GROUP 1

THE APPARENT ERROR IS DEFINED AS:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2}$$

CLASSIFICATION WITH SEVERAL POPULATIONS

IN THEORY, THE GENERALIZATION OF CLASSIFICATIONS FROM 2 TO $g \geq 2$ GROUPS IS STRAIGHTFORWARD. HOWEVER, NOT MUCH IS KNOWN ABOUT THE PROPERTIES OF THE CORRESPONDING SAMPLE CLASSIFICATION FUNCTIONS, AND IN PARTICULAR, THEIR ERROR RATES HAVE NOT BEEN FULLY INVESTIGATED.

LET $f_i(\underline{x})$ BE THE DENSITY FUNCTION ASSOCIATED WITH POPULATION π_i;
p_i BE THE PRIOR PROBABILITY OF POPULATION π_i;
$c(k|i)$ COST OF ALLOCATING AN ITEM TO π_k WHEN IN FACT IT IS FROM π_i;
$P(k|i) := \mathbb{P}(\text{CLASSIFY ITEM AS } \pi_k | \pi_i) = \int_{R_k} f_i(\underline{x})\, d\underline{x}$,
WHERE $i = 1, \dots, g$. THEN, WE CAN DERIVE THE CONDITIONAL EXPECTED COST OF MISCLASSIFYING $\underline{x}$ FROM π_1 INTO OTHER GROUPS IS

$$ECM(1) = \sum_{k=1}^{g} P(k|1)\, c(k|1) = \sum_{k=2}^{g} P(k|1)\, c(k|1)$$

$$\Rightarrow \quad ECM = \sum_{i=1}^{g} p_i\, ECM(i)$$
$$= \sum_{i=1}^{g} p_i \cdot \left[\sum_{k \neq i}^{g} P(k|i)\, c(k|i) \right] \qquad (*)$$

DETERMINING AN OPTIMAL CLASSIFICATION PROCEDURE ACCOUNTS TO CHOOSING THE MUTUALLY EXCLUSIVE AND EXHAUSTIVE CLASSIFICATION REGIONS $R_1, R_2, \dots, R_g$ SUCH THAT (*) IS MINIMAL. THEN HERE WE GIVE THE RESULT WITHOUT PROOF.

RESULT: THE CLASSIFICATION REGIONS THAT MINIMIZE THE (*)

RESULT 5 THE CLASSIFICATION REGIONS THAT MINIMIZE THE (*) ARE DEFINED BY ALLOCATING $\underline{x}$ TO THAT POPULATION π_k, $k = 1, 2, \ldots, g$, FOR WHICH

$$\sum_{i \neq k}^{g} p_i f_i(\underline{x}) \, c(k \mid i)$$

IS SMALLEST.

SIMILARLY, IF WE ASSUME ALL p_i OR ALL $c(k \mid i)$ ARE IDENTICAL AMONG $i = 1, 2, \ldots, g$, THEN THE ABOVE RESULT COULD BE SIMPLIFIED.