

GEOFFREY J. McLACHLAN

**DISCRIMINANT ANALYSIS
AND STATISTICAL PATTERN
RECOGNITION**

**WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS**

Contents

Preface

xiii

1. General Introduction

1

- 1.1. Introduction, 1
- 1.2. Basic Notation, 4
- 1.3. Allocation Rules, 6
- 1.4. Decision-Theoretic Approach, 7
- 1.5. Unavailability of Group-Prior Probabilities, 9
- 1.6. Training Data, 11
- 1.7. Sample-Based Allocation Rules, 12
- 1.8. Parametric Allocation Rules, 13
- 1.9. Assessment of Model Fit, 16
- 1.10. Error Rates of Allocation Rules, 17
- 1.11. Posterior Probabilities of Group Membership, 21
- 1.12. Distances Between Groups, 22

2. Likelihood-Based Approaches to Discrimination

27

- 2.1. Maximum Likelihood Estimation of Group Parameters, 27
- 2.2. A Bayesian Approach, 29
- 2.3. Estimation of Group Proportions, 31
- 2.4. Estimating Disease Prevalence, 33
- 2.5. Misclassified Training Data, 35
- 2.6. Partially Classified Training Data, 37
- 2.7. Maximum Likelihood Estimation for Partial Classification, 39

vii

- 2.8. Maximum Likelihood Estimation for Partial Nonrandom Classification, 43
- 2.9. Classification Likelihood Approach, 45
- 2.10. Absence of Classified Data, 46
- 2.11. Group-Conditional Mixture Densities, 50
- 3. Discrimination via Normal Models** 52
 - 3.1. Introduction, 52
 - 3.2. Heteroscedastic Normal Model, 52
 - 3.3. Homoscedastic Normal Model, 59
 - 3.4. Some Other Normal-Theory Based Rules, 65
 - 3.5. Predictive Discrimination, 67
 - 3.6. Covariance-Adjusted Discrimination, 74
 - 3.7. Discrimination with Repeated Measurements, 78
 - 3.8. Partially Classified Data, 86
 - 3.9. Linear Projections of Homoscedastic Feature Data, 87
 - 3.10. Linear Projections of Heteroscedastic Feature Data, 96
- 4. Distributional Results for Discrimination via Normal Models** 101
 - 4.1. Introduction, 101
 - 4.2. Distribution of Sample NLDF (W -Statistic), 101
 - 4.3. Moments of Conditional Error Rates of Sample NLDR, 107
 - 4.4. Distributions of Conditional Error Rates of Sample NLDR, 112
 - 4.5. Constrained Allocation with the Sample NLDR, 118
 - 4.6. Distributional Results for Quadratic Discrimination, 122
- 5. Some Practical Aspects and Variants of Normal Theory-Based Discriminant Rules** 129
 - 5.1. Introduction, 129
 - 5.2. Regularization in Quadratic Discrimination, 130
 - 5.3. Linear Versus Quadratic Normal-Based Discriminant Analysis, 132
 - 5.4. Some Models for Variants of the Sample NQDR, 137
 - 5.5. Regularized Discriminant Analysis (RDA), 144
 - 5.6. Robustness of NLDR and NQDR, 152
 - 5.7. Robust Estimation of Group Parameters, 161

6. Data Analytic Considerations with Normal Theory-Based Discriminant Analysis	168
6.1. Introduction, 168	
6.2. Assessment of Normality and Homoscedasticity, 169	
6.3. Data-Based Transformations of Feature Data, 178	
6.4. Typicality of a Feature Vector, 181	
6.5. Sample Canonical Variates, 185	
6.6. Some Other Methods of Dimension Reduction to Reveal Group Structure, 196	
6.7. Example: Detection of Hemophilia A Carriers, 201	
6.8. Example: Statistical Diagnosis of Diabetes, 206	
6.9. Example: Testing for Existence of Subspecies in Fisher's <i>Iris</i> Data, 211	
7. Parametric Discrimination via Nonnormal Models	216
7.1. Introduction, 216	
7.2. Discrete Feature Data, 216	
7.3. Parametric Formulation for Discrete Feature Data, 218	
7.4. Location Model for Mixed Features, 220	
7.5. Error Rates of Location Model-Based Rules, 229	
7.6. Adjustments to Sample NLDR for Mixed Feature Data, 232	
7.7. Some Nonnormal Models for Continuous Feature Data, 238	
7.8. Case Study of Renal Venous Renin in Hypertension, 243	
7.9. Example: Discrimination Between Depositional Environments, 249	
8. Logistic Discrimination	255
8.1. Introduction, 255	
8.2. Maximum Likelihood Estimation of Logistic Regression Coefficients, 259	
8.3. Bias Correction of MLE for $g = 2$ Groups, 266	
8.4. Assessing the Fit and Performance of Logistic Model, 270	
8.5. Logistic Versus Normal-Based Linear Discriminant Analysis, 276	
8.6. Example: Differential Diagnosis of Some Liver Diseases, 279	

- | | |
|---|------------|
| 9. Nonparametric Discrimination | 283 |
| 9.1. Introduction, 283 | |
| 9.2. Multinomial-Based Discrimination, 284 | |
| 9.3. Nonparametric Estimation of Group-Conditional Densities, 291 | |
| 9.4. Selection of Smoothing Parameters in Kernel Estimates of Group-Conditional Densities, 300 | |
| 9.5. Alternatives to Fixed Kernel Density Estimates, 308 | |
| 9.6. Comparative Performance of Kernel-Based Discriminant Rules, 312 | |
| 9.7. Nearest Neighbor Rules, 319 | |
| 9.8. Tree-Structured Allocation Rules, 323 | |
| 9.9. Some Other Nonparametric Discriminant Procedures, 332 | |
| 10. Estimation of Error Rates | 337 |
| 10.1. Introduction, 337 | |
| 10.2. Some Nonparametric Error-Rate Estimators, 339 | |
| 10.3. The Bootstrap, 346 | |
| 10.4. Variants of the Bootstrap, 353 | |
| 10.5. Smoothing of the Apparent Error Rate, 360 | |
| 10.6. Parametric Error-Rate Estimators, 366 | |
| 10.7. Confidence Intervals, 370 | |
| 10.8. Some Other Topics in Error-Rate Estimation, 373 | |
| 11. Assessing the Reliability of the Estimated Posterior Probabilities of Group Membership | 378 |
| 11.1. Introduction, 378 | |
| 11.2. Distribution of Sample Posterior Probabilities, 379 | |
| 11.3. Further Approaches to Interval Estimation of Posterior Probabilities of Group Membership, 384 | |
| 12. Selection of Feature Variables in Discriminant Analysis | 389 |
| 12.1. Introduction, 389 | |
| 12.2. Test for No Additional Information, 392 | |
| 12.3. Some Selection Procedures, 396 | |
| 12.4. Error-Rate-Based Procedures, 400 | |
| 12.5. The <i>F</i> -Test and Error-Rate-Based Variable Selections, 406 | |

12.6. Assessment of the Allocatory Capacity of the Selected Feature Variables, 410	
13. Statistical Image Analysis	413
13.1. Introduction, 413	
13.2. Markov Random Fields, 417	
13.3. Noncontextual Methods of Segmentation, 421	
13.4. Smoothing Methods, 422	
13.5. Individual Contextual Allocation of Pixels, 425	
13.6. ICM Algorithm, 428	
13.7. Global Maximization of the Posterior Distribution of the Image, 435	
13.8. Incomplete-Data Formulation of Image Segmentation, 438	
13.9. Correlated Training Data, 443	
References	447
Author Index	507
Subject Index	519

