

1. Record Nr.	UNINA9910830222903321
Autore	Braga-Neto Ulisses de Mendon AAcA
Titolo	Error estimation for pattern recognition // Ulisses M. Braga Neto, Edward R. Dougherty
Pubbl/distr/stampa	Chichester, West Sussex : , : Wiley Blackwell, , 2015 [Piscataway, New Jersey] : , : IEEE Xplore, , [2015]
ISBN	1-119-07937-3 1-119-07933-0
Descrizione fisica	1 online resource (336 p.)
Collana	IEEE press series on biomedical engineering
Disciplina	610.28
Soggetti	Biomedical engineering -- Congresses Image processing -- Digital techniques Optical pattern recognition Pattern recognition systems - Mathematics Pattern perception Error analysis (Mathematics) Electrical Engineering Electrical & Computer Engineering Engineering & Applied Sciences
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references (pages 291-300) and indexes.
Nota di contenuto	PREFACE XIII -- ACKNOWLEDGMENTS XIX -- LIST OF SYMBOLS XXI -- 1 CLASSIFICATION 1 -- 1.1 Classifiers 1 -- 1.2 Population-Based Discriminants 3 -- 1.3 Classification Rules 8 -- 1.4 Sample-Based Discriminants 13 -- 1.4.1 Quadratic Discriminants 14 -- 1.4.2 Linear Discriminants 15 -- 1.4.3 Kernel Discriminants 16 -- 1.5 Histogram Rule 16 -- 1.6 Other Classification Rules 20 -- 1.6.1 k-Nearest-Neighbor Rules 20 -- 1.6.2 Support Vector Machines 21 -- 1.6.3 Neural Networks 22 -- 1.6.4 Classification Trees 23 -- 1.6.5 Rank-Based Rules 24 -- 1.7 Feature Selection 25 -- Exercises 28 -- 2 ERROR ESTIMATION 35 -- 2.1 Error Estimation Rules 35 -- 2.2 Performance Metrics 38 -- 2.2.1 Deviation Distribution 39 -- 2.2.2 Consistency 41 -- 2.2.3 Conditional Expectation 41 -- 2.2.4 Linear Regression 42 --

2.2.5 Confidence Intervals	42
2.3 Test-Set Error Estimation	43
2.4 Resubstitution	46
2.5 Cross-Validation	48
2.6 Bootstrap	55
2.7 Convex Error Estimation	57
2.8 Smoothed Error Estimation	61
2.9 Bolstered Error Estimation	63
2.9.1 Gaussian-Bolstered Error Estimation	67
2.9.2 Choosing the Amount of Bolstering	68
2.9.3 Calibrating the Amount of Bolstering	71
Exercises	73
3 PERFORMANCE ANALYSIS	77
3.1 Empirical Deviation Distribution	77
3.2 Regression	79
3.3 Impact on Feature Selection	82
3.4 Multiple-Data-Set Reporting Bias	84
3.5 Multiple-Rule Bias	86
3.6 Performance Reproducibility	92
Exercises	94
4 ERROR ESTIMATION FOR DISCRETE CLASSIFICATION	97
4.1 Error Estimators	98
4.1.1 Resubstitution Error	98
4.1.2 Leave-One-Out Error	98
4.1.3 Cross-Validation Error	99
4.1.4 Bootstrap Error	99
4.2 Small-Sample Performance	101
4.2.1 Bias	101
4.2.2 Variance	103
4.2.3 Deviation Variance, RMS, and Correlation	105
4.2.4 Numerical Example	106
4.2.5 Complete Enumeration Approach	108
4.3 Large-Sample Performance	110
Exercises	114
5 DISTRIBUTION THEORY	115
5.1 Mixture Sampling Versus Separate Sampling	115
5.2 Sample-Based Discriminants Revisited	119
5.3 True Error	120
5.4 Error Estimators	121
5.4.1 Resubstitution Error	121
5.4.2 Leave-One-Out Error	122
5.4.3 Cross-Validation Error	122
5.4.4 Bootstrap Error	124
5.5 Expected Error Rates	125
5.5.1 True Error	125
5.5.2 Resubstitution Error	128
5.5.3 Leave-One-Out Error	130
5.5.4 Cross-Validation Error	132
5.5.5 Bootstrap Error	133
5.6 Higher-Order Moments of Error Rates	136
5.6.1 True Error	136
5.6.2 Resubstitution Error	137
5.6.3 Leave-One-Out Error	139
5.7 Sampling Distribution of Error Rates	140
5.7.1 Resubstitution Error	140
5.7.2 Leave-One-Out Error	141
Exercises	142
6 GAUSSIAN DISTRIBUTION THEORY: UNIVARIATE CASE	145
6.1 Historical Remarks	146
6.2 Univariate Discriminant	147
6.3 Expected Error Rates	148
6.3.1 True Error	148
6.3.2 Resubstitution Error	151
6.3.3 Leave-One-Out Error	152
6.3.4 Bootstrap Error	152
6.4 Higher-Order Moments of Error Rates	154
6.4.1 True Error	154
6.4.2 Resubstitution Error	157
6.4.3 Leave-One-Out Error	160
6.4.4 Numerical Example	165
6.5 Sampling Distributions of Error Rates	166
6.5.1 Marginal Distribution of Resubstitution Error	166
6.5.2 Marginal Distribution of Leave-One-Out Error	169
6.5.3 Joint Distribution of Estimated and True Errors	174
Exercises	176
7 GAUSSIAN DISTRIBUTION THEORY: MULTIVARIATE CASE	179
7.1 Multivariate Discriminants	179
7.2 Small-Sample Methods	180
7.2.1 Statistical Representations	181
7.2.2 Computational Methods	194
7.3 Large-Sample Methods	199
7.3.1 Expected Error Rates	200
7.3.2 Second-Order Moments of Error Rates	207
Exercises	218
8 BAYESIAN MMSE ERROR ESTIMATION	221
8.1 The Bayesian MMSE Error Estimator	222
8.2 Sample-Conditioned MSE	226
8.3 Discrete Classification	227
8.4 Linear Classification of Gaussian Distributions	238
8.5 Consistency	246
8.6 Calibration	253
8.7 Concluding Remarks	255
Exercises	257
A BASIC PROBABILITY REVIEW	259
A.1 Sample Spaces and Events	259
A.2 Definition of Probability	260
A.3 Borel-Cantelli Lemmas	261
A.4 Conditional Probability	262
A.5 Random Variables	263
A.6 Discrete Random Variables	265
A.7 Expectation	266
A.8 Conditional Expectation	268
A.9 Variance	269
A.10 Vector Random Variables	270
A.11 The Multivariate Gaussian	271
A.12 Convergence of Random Sequences	273
A.13 Limiting Theorems	275
B VAPNIK-CHEVONENKIS THEORY	277
B.1 Shatter Coefficients	277
B.2 The VC Dimension	278
B.3 VC Theory of Classification	279
B.3.1	

Linear Classification Rules 279 -- B.3.2 kNN Classification Rule 280 --
B.3.3 Classification Trees 280 -- B.3.4 Nonlinear SVMs 281 -- B.3.5
Neural Networks 281 -- B.3.6 Histogram Rules 281 -- B.4 Vapnik-
Chervonenkis Theorem 282 -- C DOUBLE ASYMPTOTICS 285 --
BIBLIOGRAPHY 291 -- AUTHOR INDEX 301 -- SUBJECT INDEX 305.

Sommario/riassunto

This book is the first of its kind to discuss error estimation with a model-based approach. From the basics of classifiers and error estimators to more specialized classifiers, it covers important topics and essential issues pertaining to the scientific validity of pattern classification. Error Estimation for Pattern Recognition focuses on error estimation, which is a broad and poorly understood topic that reaches all research areas using pattern classification. It includes model-based approaches and discussions of newer error estimators such as bolstered and Bayesian estimators. This book was motivated by the application of pattern recognition to high-throughput data with limited replicates, which is a basic problem now appearing in many areas. The first two chapters cover basic issues in classification error estimation, such as definitions, test-set error estimation, and training-set error estimation. The remaining chapters in this book cover results on the performance and representation of training-set error estimators for various pattern classifiers. Additional features of the book include: . The latest results on the accuracy of error estimation. Performance analysis of resubstitution, cross-validation, and bootstrap error estimators using analytical and simulation approaches. Highly interactive computer-based exercises and end-of-chapter Problems This is the first book exclusively about error estimation for pattern recognition.
