## **CONTENTS**

PREFA	CE	xi
CONTR	CONTRIBUTORS	
n.o.	AD THE STRUCTURE A DISPLACEMENT	
1 <i>RO</i> /	AD TO STATISTICAL BIOINFORMATICS	1
Challeng	ge 1: Multiple-Comparisons Issue	1
Challen	Challenge 2: High-Dimensional Biological Data	
Challenge 3: Small-n and Large-p Problem		3
Challenge 4: Noisy High-Throughput Biological Data		3
Challeng	Challenge 5: Integration of Multiple, Heterogeneous Biological Data Information	
Reference	ces	5
2 PR(	OBABILITY CONCEPTS AND DISTRIBUTIONS FOR ANALYZING	
	RGE BIOLOGICAL DATA	7
	troduction	7
	asic Concepts	8
	onditional Probability and Independence	10
	andom Variables	13
	spected Value and Variance	15
	stributions of Random Variables	19
	int and Marginal Distribution	39
	ultivariate Distribution	42
	empling Distribution	46
2.10 Su	ımmary	54
3 QU	ALITY CONTROL OF HIGH-THROUGHPUT BIOLOGICAL DATA	57
3.1 Sc	ources of Error in High-Throughput Biological Experiments	57
3.2 St	atistical Techniques for Quality Control	59
3.3 Iss	sues Specific to Microarray Gene Expression Experiments	66
3.4 Co	onclusion	69
Reference	ces	69
4 STA	TISTICAL TESTING AND SIGNIFICANCE FOR LARGE BIOLOGICAL	
	TA ANALYSIS	71
41 7		71
	troduction	71
	atistical Testing	72
4.3 Er	ror Controlling	78

## viii CONTENTS

Acknowledgments		
Acknowledgments         8           References         8           6         CLUSTERING: UNSUPERVISED LEARNING IN LARGE           BIOLOGICAL DATA         8           5.1         Measures of Similarity         9           5.2         Clustering         9           5.3         Assessment of Cluster Quality         11:           5.4         Conclusion         12:           References         12:           6.         CLASSIFICATION: SUPERVISED LEARNING WITH         HIGH-DIMENSIONAL BIOLOGICAL DATA         12:           6.1         Introduction         12:           6.2         Classification and Prediction Methods         13:           6.3         Feature Selection and Ranking         14:           6.4         Cross-Validation         14:           6.6         Comparison of Classification Methods Using High-Dimensional Data         14:           6.7         Software Examples for Classification Methods         15:           7         MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE         15:           8.7         Junction         15:           7.1         Introduction         15:           7.2         Classical Multidimensional Visualization Techniques         15: <t< td=""><td>4.4 Real Data Analysis</td><td>81</td></t<>	4.4 Real Data Analysis	81
References         8           5 CLUSTERING: UNSUPERVISED LEARNING IN LARGE BIOLOGICAL DATA         8           5.1 Measures of Similarity         9           5.2 Clustering         9           5.3 Assessment of Cluster Quality         11           5.4 Conclusion         12           References         12           6 CLASSIFICATION: SUPERVISED LEARNING WITH HIGH-DIMENSIONAL BIOLOGICAL DATA         12           6.1 Introduction         12           6.2 Classification and Prediction Methods         13           6.3 Feature Selection and Ranking         14           6.4 Cross-Validation         14           6.5 Enhancement of Class Prediction by Ensemble Voting Methods         14           6.6 Comparison of Classification Methods Using High-Dimensional Data         14           6.7 Software Examples for Classification Methods         15           7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA         15           7.1 Introduction         15           7.2 Classical Multidimensional Visualization Techniques         15           7.3 Two-Dimensional Projections         16           7.4 Issues and Challenges         16           7.5 Systematic Exploration of Low-Dimensional Projections         16           7.6 One-Dimensional Histogram Ordering	· · · · · · · · · · · · · · · · · · ·	87
Scalabor	· ·	87
Statistical Data   Statistical Probabilistic Models   Statistical Probabilistic Mode	References	88
5.1 Measures of Similarity       99         5.2 Clustering       99         5.2 Clustering       99         5.3 Assessment of Cluster Quality       11:         5.4 Conclusion       12:         References       12:         6 CLASSIFICATION: SUPERVISED LEARNING WITH HIGH-DIMENSIONAL BIOLOGICAL DATA       12:         6.1 Introduction       12:         6.2 Classification and Prediction Methods       13:         6.3 Feature Selection and Ranking       14:         6.4 Cross-Validation       14:         6.5 Enhancement of Class Prediction by Ensemble Voting Methods       14:         6.6 Comparison of Classification Methods Using High-Dimensional Data       14:         6.7 Software Examples for Classification Methods       15:         References       15:         7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA       15:         7.1 Introduction       15:         7.2 Classical Multidimensional Visualization Techniques       15:         7.3 Two-Dimensional Projections       16:         7.4 Systematic Exploration of Low-Dimensional Projections       16:         7.5 Systematic Exploration of Low-Dimensional Projections       16:         7.6 One-Dimensional Histogram Ordering       17:         7.7 Two-Dimensional Sc	5 CLUSTERING: UNSUPERVISED LEARNING IN LARGE	
5.2 Clustering       99         5.3 Assessment of Cluster Quality       11:         5.4 Conclusion       12:         References       12:         6 CLASSIFICATION: SUPERVISED LEARNING WITH HIGH-DIMENSIONAL BIOLOGICAL DATA       12:         6.1 Introduction       12:         6.2 Classification and Prediction Methods       13:         6.3 Feature Selection and Ranking       14:         6.4 Cross-Validation       14:         6.5 Enhancement of Class Prediction by Ensemble Voting Methods       14:         6.6 Comparison of Classification Methods Using High-Dimensional Data       14:         6.7 Software Examples for Classification Methods       15:         References       15:         7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA       15         7.1 Introduction       15:         7.2 Classical Multidimensional Visualization Techniques       15:         7.3 Two-Dimensional Projections       16:         7.4 Issues and Challenges       16:         7.5 Systematic Exploration of Low-Dimensional Projections       16:         7.6 One-Dimensional Scatterplot Ordering       17:         7.7 Two-Dimensional Scatterplot Ordering       17:         7.8 Conclusion       18:         8.1 Introduction       18: </td <td>BIOLOGICAL DATA</td> <td>89</td>	BIOLOGICAL DATA	89
5.2 Clustering       99         5.3 Assessment of Cluster Quality       112         5.4 Conclusion       122         References       12         6 CLASSIFICATION: SUPERVISED LEARNING WITH HIGH-DIMENSIONAL BIOLOGICAL DATA       12         6.1 Introduction       12         6.2 Classification and Prediction Methods       13         6.3 Feature Selection and Ranking       14         6.4 Cross-Validation       14         6.5 Enhancement of Class Prediction by Ensemble Voting Methods       14         6.6 Comparison of Classification Methods Using High-Dimensional Data       14         6.7 Software Examples for Classification Methods       15         7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA       15         7.1 Introduction       15         7.2 Classical Multidimensional Visualization Techniques       15         7.3 Two-Dimensional Projections       16         7.5 Systematic Exploration of Low-Dimensional Projections       16         7.5 Systematic Exploration of Low-Dimensional Projections       16         7.6 One-Dimensional Histogram Ordering       17         7.7 Two-Dimensional Scatterplot Ordering       17         7.8 Conclusion       18         8.1 Introduction       18         8.2 Statistical/Pro	5.1 Measures of Similarity	90
12	5.2 Clustering	99
12	5.3 Assessment of Cluster Quality	115
6. CLASSIFICATION: SUPERVISED LEARNING WITH HIGH-DIMENSIONAL BIOLOGICAL DATA  6.1 Introduction 6.2 Classification and Prediction Methods 6.3 Feature Selection and Ranking 6.4 Cross-Validation 6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods 6.8 References 6.9 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 6.1 Introduction 6.2 Classical Multidimensional Visualization Techniques 6.3 Two-Dimensional Projections 6.4 Issues and Challenges 6.6 One-Dimensional Histogram Ordering 6.7 Two-Dimensional Scatterplot Ordering 6.8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 6.1 Introduction 6.2 Statistical/Probabilistic Models 6.3 Estimation Methods 6.4 Numerical Algorithms 6.5 Examples 6.6 Conclusion 6.7 Conclusion 6.8 Conclusion 6.8 Conclusion 6.9 Supermatic Exploration of Low-Dimensional Projections 6.9 Statistical/Probabilistic Models 6.9 Statistical/Probabilistic Models 6.1 Statistical/Probabilistic Models 6.2 Conclusion 6.3 Conclusion 6.4 Numerical Algorithms 6.5 Examples 6.6 Conclusion 6.7 Conclusion 6.8 Statistical/Probabilistic Models 6.8 Conclusion 6.9 Statistical/Probabilistic Models 6.0 Statistical/Probabilistic Models 6.1 Statistical/Probabilistic Models 6.2 Statistical/Probabilistic Models 6.3 Statistical/Probabilistic Models 6.4 Supermatical Algorithms 6.5 Examples 6.6 Conclusion 6.7 Software Examples 6.7 Software Examples 6.8 Conclusion 6.9 Statistical/Probabilistic Models 6.9 Statistica	5.4 Conclusion	123
HIGH-DIMENSIONAL BIOLOGICAL DATA   12:	References	123
HIGH-DIMENSIONAL BIOLOGICAL DATA   12:	6 CLASSIFICATION: SUPERVISED LEARNING WITH	
6.1 Introduction 6.2 Classification and Prediction Methods 6.3 Feature Selection and Ranking 6.4 Cross-Validation 6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods 6.8 References 6.9 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE 6.0 BIOMEDICAL DATA 6.1 Introduction 6.2 Classical Multidimensional Visualization Techniques 6.3 Two-Dimensional Projections 6.4 Issues and Challenges 6.5 Systematic Exploration of Low-Dimensional Projections 6.6 Conclusion 6.7 Two-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Seferences 7.9 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 7.1 Introduction 7.2 Statistical/Probabilistic Models 7.3 Statistical/Probabilistic Models 7.4 Introduction 7.5 Statistical/Probabilistic Models 7.6 Conclusion 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Statistical/Probabilistic Models 7.9 Stat		129
6.2 Classification and Prediction Methods 6.3 Feature Selection and Ranking 6.4 Cross-Validation 6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods 6.8 References 6.9 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE 6.0 BIOMEDICAL DATA 7.1 Introduction 7.2 Classical Multidimensional Visualization Techniques 7.3 Two-Dimensional Projections 7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Conclusion 7.9 Conclusion 7.0 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 7.1 Introduction 7.2 Statistical/Probabilistic Models 7.3 Statistical/Probabilistic Models 7.4 Introduction 7.5 Statistical/Probabilistic Models 7.6 Statistical/Probabilistic Models 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Statistical/Probabilistic Models 7.0 Statistical/Probabilistic Models 7.1 Numerical Algorithms 7.2 Examples 7.3 Conclusion 7.4 Statistical/Probabilistic Models 7.5 Statistical/Probabilistic Models 7.6 Numerical Algorithms 7.7 Statistical/Probabilistic Models 7.8 Conclusion 7.9 Conclusion 7.9 Statistical/Probabilistic Models		
6.3 Feature Selection and Ranking 6.4 Cross-Validation 6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods 6.8 References 6.7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE 6.8 BIOMEDICAL DATA 7.1 Introduction 7.2 Classical Multidimensional Visualization Techniques 7.3 Two-Dimensional Projections 7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Conclusion 7.9 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 7.1 Introduction 7.2 Statistical/Probabilistic Models 7.3 Statistical/Probabilistic Models 7.4 Introduction 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Scatterplot Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Statistical/Probabilistic Models 7.0 Statistical/Probabilistic Models 7.1 Introduction 7.2 Classical Methods 7.3 Introduction 7.4 Introduction 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 7.9 Statistical/Probabilistic Models 7.0 Statistical/Probabilistic M		129
6.4 Cross-Validation 6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods 8.156 References 8.157  **MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 8.158  7.1 Introduction 7.2 Classical Multidimensional Visualization Techniques 8.159  7.3 Two-Dimensional Projections 9.169  7.4 Issues and Challenges 9.160  7.5 Systematic Exploration of Low-Dimensional Projections 9.160  7.6 One-Dimensional Histogram Ordering 9.177  7.7 Two-Dimensional Scatterplot Ordering 9.178  7.8 Conclusion 9.188  8.8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 9.189  8.1 Introduction 9.189  8.2 Statistical/Probabilistic Models 9.3 Estimation Methods 9.3 Estimation Methods 9.3 Numerical Algorithms 9.3 Examples 9.3 Conclusion 9.3 Conclusion 9.3 Conclusion 9.3 Conclusion 9.3 Conclusion 9.3 Examples 9.4 Conclusion 9.5 Examples 9.5 Conclusion 9.7 Conclusion 9.8 Conclusion 9.8 Conclusion 9.9 Co		132
6.5 Enhancement of Class Prediction by Ensemble Voting Methods 6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods  References 156 7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 157 7.1 Introduction 157 7.2 Classical Multidimensional Visualization Techniques 158 7.3 Two-Dimensional Projections 168 7.4 Issues and Challenges 169 7.5 Systematic Exploration of Low-Dimensional Projections 160 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 180 8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 181 8.1 Introduction 182 8.2 Statistical/Probabilistic Models 183 8.3 Estimation Methods 184 8.4 Numerical Algorithms 195 8.5 Examples 196 8.6 Conclusion 197	6.3 Feature Selection and Ranking	140
6.6 Comparison of Classification Methods Using High-Dimensional Data 6.7 Software Examples for Classification Methods References 156 7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 157 7.1 Introduction 157 7.2 Classical Multidimensional Visualization Techniques 158 7.3 Two-Dimensional Projections 169 7.4 Issues and Challenges 160 7.5 Systematic Exploration of Low-Dimensional Projections 160 7.6 One-Dimensional Histogram Ordering 170 7.7 Two-Dimensional Scatterplot Ordering 170 7.8 Conclusion 180 References 181 8. STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 181 8.1 Introduction 182 8.2 Statistical/Probabilistic Models 183 8.3 Estimation Methods 184 8.4 Numerical Algorithms 195 8.5 Examples 196 8.6 Conclusion 197		144
6.7 Software Examples for Classification Methods References  156 References 156 References 157 RULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 157 R.1 Introduction 157 R.2 Classical Multidimensional Visualization Techniques 158 R.3 Two-Dimensional Projections 169 R.4 Issues and Challenges 160 R.5 Systematic Exploration of Low-Dimensional Projections 160 R.6 One-Dimensional Histogram Ordering 170 R.7 Two-Dimensional Scatterplot Ordering 171 References 180 References 181 References 182 References 183 References 184 References 185 References 185 References 186 References 187 References 187 References 188 References 189 References 180 References 180 References 180 References 181 References 181 References 182 References 182 References 183 References 184 References 185 References 185 References 186 References 187 References 187 References 188 References 188 References 189 References 180 Refere	- J G	145
References 15-  7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA 15-  7.1 Introduction 15-  7.2 Classical Multidimensional Visualization Techniques 15-  7.3 Two-Dimensional Projections 16-  7.4 Issues and Challenges 16-  7.5 Systematic Exploration of Low-Dimensional Projections 16-  7.6 One-Dimensional Histogram Ordering 17-  7.7 Two-Dimensional Scatterplot Ordering 17-  7.8 Conclusion 18-  References 18-  8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18-  8.1 Introduction 18-  8.2 Statistical/Probabilistic Models 18-  8.3 Estimation Methods 18-  8.4 Numerical Algorithms 19-  8.5 Examples 19-  8.6 Conclusion 19-  15-  16-  17-  18-  18-  18-  18-  18-  18-  18	-	147
7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE BIOMEDICAL DATA  15.  7.1 Introduction 7.2 Classical Multidimensional Visualization Techniques 15.  7.3 Two-Dimensional Projections 16.  7.4 Issues and Challenges 16.  7.5 Systematic Exploration of Low-Dimensional Projections 16.  7.6 One-Dimensional Histogram Ordering 17.  7.7 Two-Dimensional Scatterplot Ordering 18.  8 Conclusion 18  8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS  8.1 Introduction 8.2 Statistical/Probabilistic Models 8.3 Estimation Methods 18  8.4 Numerical Algorithms 19  8.5 Examples 19  8.6 Conclusion 19		150
BIOMEDICAL DATA   15   7.1   Introduction   157   7.2   Classical Multidimensional Visualization Techniques   158   7.3   Two-Dimensional Projections   168   7.4   Issues and Challenges   169   7.5   Systematic Exploration of Low-Dimensional Projections   169   7.6   One-Dimensional Histogram Ordering   179   7.7   Two-Dimensional Scatterplot Ordering   179   7.8   Conclusion   188   References   189   8   STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS   189   8.1   Introduction   189   8.2   Statistical/Probabilistic Models   189   8.3   Estimation Methods   189   8.4   Numerical Algorithms   199   8.5   Examples   199   8.6   Conclusion   199   8.6   Conclusion   199   8.7   Classical Multidimensional Visualization Techniques   199   150	References	154
7.1 Introduction 7.2 Classical Multidimensional Visualization Techniques 7.3 Two-Dimensional Projections 16 7.4 Issues and Challenges 16 7.5 Systematic Exploration of Low-Dimensional Projections 16 7.6 One-Dimensional Histogram Ordering 17 7.7 Two-Dimensional Scatterplot Ordering 17 7.8 Conclusion 18 References 18 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18 8.1 Introduction 18 8.2 Statistical/Probabilistic Models 18 8.3 Estimation Methods 18 8.4 Numerical Algorithms 19 8.5 Examples 19 8.6 Conclusion 19	7 MULTIDIMENSIONAL ANALYSIS AND VISUALIZATION ON LARGE	
7.2 Classical Multidimensional Visualization Techniques 7.3 Two-Dimensional Projections 7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 8 Conclusion 18 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 8.1 Introduction 8.2 Statistical/Probabilistic Models 8.3 Estimation Methods 8.4 Numerical Algorithms 9 S.5 Examples 8.6 Conclusion 196	BIOMEDICAL DATA	157
7.2 Classical Multidimensional Visualization Techniques 7.3 Two-Dimensional Projections 7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 8 Conclusion 18 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 8.1 Introduction 8.2 Statistical/Probabilistic Models 8.3 Estimation Methods 8.4 Numerical Algorithms 9 S.5 Examples 8.6 Conclusion 196	7.1 Introduction	157
7.3 Two-Dimensional Projections 7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 8 Conclusion 18 8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18 8.1 Introduction 18 8.2 Statistical/Probabilistic Models 18 8.3 Estimation Methods 18 8.4 Numerical Algorithms 19 8.5 Examples 19 8.6 Conclusion 19		
7.4 Issues and Challenges 7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 8 Conclusion 8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 8 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 8.1 Introduction 8.2 Statistical/Probabilistic Models 8.3 Estimation Methods 8.4 Numerical Algorithms 9 S.5 Examples 8.6 Conclusion 9 Section 19 Sec		
7.5 Systematic Exploration of Low-Dimensional Projections 7.6 One-Dimensional Histogram Ordering 7.7 Two-Dimensional Scatterplot Ordering 7.8 Conclusion 18 References 18:  3 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18.  3.1 Introduction 18.2 Statistical/Probabilistic Models 18.3 Estimation Methods 18.4 Numerical Algorithms 19  3.5 Examples 19  3.6 Conclusion 19	<b>J</b>	
7.6 One-Dimensional Histogram Ordering       176         7.7 Two-Dimensional Scatterplot Ordering       177         7.8 Conclusion       18         References       18         3 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS       18         8.1 Introduction       18         8.2 Statistical/Probabilistic Models       18         8.3 Estimation Methods       18         8.4 Numerical Algorithms       19         8.5 Examples       19         8.6 Conclusion       19		
7.7 Two-Dimensional Scatterplot Ordering       17-         7.8 Conclusion       18         References       18:         3 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS       18.         8.1 Introduction       18:         8.2 Statistical/Probabilistic Models       18:         8.3 Estimation Methods       18:         8.4 Numerical Algorithms       19:         8.5 Examples       19:         8.6 Conclusion       19:		
7.8 Conclusion 18 References 18:  3 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18.  3.1 Introduction 18: 3.2 Statistical/Probabilistic Models 18: 3.3 Estimation Methods 18: 3.4 Numerical Algorithms 19 3.5 Examples 19: 3.6 Conclusion 19:		
References 18:  3 STATISTICAL MODELS, INFERENCE, AND ALGORITHMS FOR LARGE BIOLOGICAL DATA ANALYSIS 18.  3.1 Introduction 18: 3.2 Statistical/Probabilistic Models 18: 3.3 Estimation Methods 18: 3.4 Numerical Algorithms 19 3.5 Examples 19: 3.6 Conclusion 19:		_
LARGE BIOLOGICAL DATA ANALYSIS       18.         8.1 Introduction       18.         8.2 Statistical/Probabilistic Models       18.         8.3 Estimation Methods       18.         8.4 Numerical Algorithms       19.         8.5 Examples       19.         8.6 Conclusion       19.	References	182
LARGE BIOLOGICAL DATA ANALYSIS       18.         8.1 Introduction       18.         8.2 Statistical/Probabilistic Models       18.         8.3 Estimation Methods       18.         8.4 Numerical Algorithms       19.         8.5 Examples       19.         8.6 Conclusion       19.	8 STATISTICAL MODELS INFERENCE AND ALCOPITUMS FOR	
8.2       Statistical/Probabilistic Models       18'         8.3       Estimation Methods       18'         8.4       Numerical Algorithms       19         8.5       Examples       19'         8.6       Conclusion       19'		185
8.2       Statistical/Probabilistic Models       18'         8.3       Estimation Methods       18'         8.4       Numerical Algorithms       19         8.5       Examples       19'         8.6       Conclusion       19'		
8.3Estimation Methods1898.4Numerical Algorithms198.5Examples198.6Conclusion19		
8.4       Numerical Algorithms       19         8.5       Examples       19         8.6       Conclusion       19		
3.5 Examples       192         3.6 Conclusion       193		
3.6 Conclusion 198		
	8.6 Conclusion References	198 199

ix

9	EXPERIMENTAL DESIGNS ON HIGH-THROUGHPUT BIOLOGICAL	
	EXPERIMENTS	201
9.1	Randomization	201
9.2	Replication	202
9.3	Pooling	209
9.4	Blocking	210
9.5	Design for Classifications	214
9.6	Design for Time Course Experiments	215
9.7	Design for eQTL Studies	215
	erences	216
10	STATISTICAL RESAMPLING TECHNIQUES FOR LARGE BIOLOGICAL	
	DATA ANALYSIS	219
10.	1 Introduction	219
10.2	Resampling Methods for Prediction Error Assessment and Model Selection	221
10.3	Feature Selection	225
،10	4 Resampling-Based Classification Algorithms	226
10.3	5 Practical Example: Lymphoma	226
10.0	6 Resampling Methods	227
10.	7 Bootstrap Methods	232
10.3	Sample Size Issues	233
10.9	O Loss Functions	235
10.	10 Bootstrap Resampling for Quantifying Uncertainty	236
	11 Markov Chain Monte Carlo Methods	238
10.	12 Conclusions	240
Ref	erences	247
11	STATISTICAL NETWORK ANALYSIS FOR BIOLOGICAL SYSTEMS	
	AND PATHWAYS	249
11.	1 Introduction	249
11.3	2 Boolean Network Modeling	250
11		259
11.4	·	273
Ref	erences	279
12	TRENDS AND STATISTICAL CHALLENGES IN GENOMEWIDE	
	ASSOCIATION STUDIES	283
12.	I Introduction	283
12.2	2 Alleles, Linkage Disequilibrium, and Haplotype	283
12.		285
12.4		286
12.:	*** *	287
12.0		290
12.		296
12.8	1 31 3	298
12.9	76 7 11 6	298
	10 Gene and Pathway-Based Analysis	299
	,	

## **X** CONTENTS

12.11	2.11 Disease Risk Estimates			
12.12	12.12 Meta-Analysis 12.13 Rare Variants and Sequence-Based Analysis			
12.13				
12.14	12.14 Conclusions			
Ackn	Acknowledgments			
Refer	References			
13	R AND BIOCONDUCTOR PACKAGES IN BIOINFORMATICS: TOWARDS			
	SYSTEMS BIOLOGY	309		
13.1	Introduction	309		
13.2	Brief overview of the Bioconductor Project	310		
13.3	Experimental Data	311		
13.4	Annotation	318		
13.5	Models of Biological Systems	328		
13.6	Conclusion	335		
13.7	Acknowledgments	336		
References		336		
INDI	EX	339		