## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>xix</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xxiii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Class Discovery</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Dimensional Reduction</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Class Prediction</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Classification Rules of Thumb</td>
<td>5</td>
</tr>
<tr>
<td>1.5 DNA Microarray Datasets Used</td>
<td>9</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
</tr>
<tr>
<td>PART I CLASS DISCOVERY</td>
<td>13</td>
</tr>
<tr>
<td>2 Crisp K-Means Cluster Analysis</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Algorithm</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Implementation</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Distance Metrics</td>
<td>20</td>
</tr>
<tr>
<td>2.5 Cluster Validity</td>
<td>24</td>
</tr>
<tr>
<td>2.5.1 Davies–Bouldin Index</td>
<td>25</td>
</tr>
<tr>
<td>2.5.2 Dunn’s Index</td>
<td>25</td>
</tr>
<tr>
<td>2.5.3 Intracluster Distance</td>
<td>26</td>
</tr>
<tr>
<td>2.5.4 Intercluster Distance</td>
<td>27</td>
</tr>
<tr>
<td>2.5.5 Silhouette Index</td>
<td>30</td>
</tr>
<tr>
<td>2.5.6 Hubert’s Γ Statistic</td>
<td>31</td>
</tr>
<tr>
<td>2.5.7 Randomization Tests for Optimal Value of K</td>
<td>31</td>
</tr>
<tr>
<td>2.6 V-Fold Cross-Validation</td>
<td>35</td>
</tr>
<tr>
<td>2.7 Cluster Initialization</td>
<td>37</td>
</tr>
</tbody>
</table>
CONTENTS

2.7.1 K Randomly Selected Microarrays 37
2.7.2 K Random Partitions 40
2.7.3 Prototype Splitting 41
2.8 Cluster Outliers 44
2.9 Summary 44
References 45

3 Fuzzy K-Means Cluster Analysis 47
3.1 Introduction 47
3.2 Fuzzy K-Means Algorithm 47
3.3 Implementation 49
3.4 Summary 54
References 54

4 Self-Organizing Maps 57
4.1 Introduction 57
4.2 Algorithm 57
4.2.1 Feature Transformation and Reference Vector Initialization 59
4.2.2 Learning 60
4.2.3 Conscience 61
4.3 Implementation 63
4.3.1 Feature Transformation and Reference Vector Initialization 63
4.3.2 Reference Vector Weight Learning 66
4.4 Cluster Visualization 67
4.4.1 Crisp K-Means Cluster Analysis 67
4.4.2 Adjacency Matrix Method 68
4.4.3 Cluster Connectivity Method 69
4.4.4 Hue–Saturation–Value (HSV) Color Normalization 69
4.5 Unified Distance Matrix (U Matrix) 71
4.6 Component Map 71
4.7 Map Quality 73
4.8 Nonlinear Dimension Reduction 75
References 79

5 Unsupervised Neural Gas 81
5.1 Introduction 81
5.2 Algorithm 82
5.3 Implementation 82
CONTENTS

5.3.1 Feature Transformation and Prototype Initialization 82
5.3.2 Prototype Learning 83
5.4 Nonlinear Dimension Reduction 85
5.5 Summary 87
References 88

6 Hierarchical Cluster Analysis 91
6.1 Introduction 91
6.2 Methods 91
6.2.1 General Programming Methods 91
6.2.2 Step 1: Cluster-Analyzing Arrays as Objects with Genes as Attributes 92
6.2.3 Step 2: Cluster-Analyzing Genes as Objects with Arrays as Attributes 94
6.3 Algorithm 96
6.4 Implementation 96
6.4.1 Heatmap Color Control 96
6.4.2 User Choices for Clustering Arrays and Genes 97
6.4.3 Distance Matrices and Agglomeration Sequences 98
6.4.4 Drawing Dendograms and Heatmaps 104
References 105

7 Model-Based Clustering 107
7.1 Introduction 107
7.2 Algorithm 110
7.3 Implementation 111
7.4 Summary 116
References 117

8 Text Mining: Document Clustering 119
8.1 Introduction 119
8.2 Duo-Mining 119
8.3 Streams and Documents 120
8.4 Lexical Analysis 120
8.4.1 Automatic Indexing 120
8.4.2 Removing Stopwords 121
8.5 Stemming 121
8.6 Term Weighting 121
8.7 Concept Vectors 124
CONTENTS

8.8 Main Terms Representing Concept Vectors 124
8.9 Algorithm 125
8.10 Preprocessing 127
8.11 Summary 137
    References 137

9 Text Mining: N-Gram Analysis 139
9.1 Introduction 139
9.2 Algorithm 140
9.3 Implementation 141
9.4 Summary 154
    References 156

PART II DIMENSION REDUCTION 159

10 Principal Components Analysis 161
10.1 Introduction 161
10.2 Multivariate Statistical Theory 161
    10.2.1 Matrix Definitions 162
    10.2.2 Principal Component Solution of $R$ 163
    10.2.3 Extraction of Principal Components 164
    10.2.4 Varimax Orthogonal Rotation of Components 166
    10.2.5 Principal Component Score Coefficients 168
    10.2.6 Principal Component Scores 169
10.3 Algorithm 170
10.4 When to Use Loadings and PC Scores 170
10.5 Implementation 171
    10.5.1 Correlation Matrix $R$ 171
    10.5.2 Eigenanalysis of Correlation Matrix $R$ 172
    10.5.3 Determination of Loadings and Varimax Rotation 174
    10.5.4 Calculating Principal Component (PC) Scores 176
10.6 Rules of Thumb For PCA 182
10.7 Summary 186
    References 187

11 Nonlinear Manifold Learning 189
11.1 Introduction 189
11.2 Correlation-Based PCA 190
11.3 Kernel PCA 191
11.4 Diffusion Maps 192
## CONTENTS

11.5 Laplacian Eigenmaps 192  
11.6 Local Linear Embedding 193  
11.7 Locality Preserving Projections 194  
11.8 Sammon Mapping 195  
11.9 NLML Prior to Classification Analysis 195  
11.10 Classification Results 197  
11.11 Summary 200  
References 203

### PART III  CLASS PREDICTION  205

12 Feature Selection  207  
12.1 Introduction  207  
12.2 Filtering versus Wrapping  208  
12.3 Data  209  
  12.3.1 Numbers  209  
  12.3.2 Responses  209  
  12.3.3 Measurement Scales  210  
  12.3.4 Variables  211  
12.4 Data Arrangement  211  
12.5 Filtering  213  
  12.5.1 Continuous Features  213  
  12.5.2 Best Rank Filters  219  
  12.5.3 Randomization Tests  236  
  12.5.4 Multitestinig Problem  237  
  12.5.5 Filtering Qualitative Features  242  
  12.5.6 Multiclass Gini Diversity Index  246  
  12.5.7 Class Comparison Techniques  247  
  12.5.8 Generation of Nonredundant Gene List  250  
12.6 Selection Methods  254  
  12.6.1 Greedy Plus Takeaway (Greedy PTA)  254  
  12.6.2 Best Ranked Genes  258  
12.7 Multicollinearity  259  
12.8 Summary  270  
References  270

13 Classifier Performance  273  
13.1 Introduction  273  
13.2 Input–Output, Speed, and Efficiency  273  
13.3 Training, Testing, and Validation  277
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.4</td>
<td>Ensemble Classifier Fusion</td>
<td>280</td>
</tr>
<tr>
<td>13.5</td>
<td>Sensitivity and Specificity</td>
<td>283</td>
</tr>
<tr>
<td>13.6</td>
<td>Bias</td>
<td>284</td>
</tr>
<tr>
<td>13.7</td>
<td>Variance</td>
<td>285</td>
</tr>
<tr>
<td>13.8</td>
<td>Receiver–Operator Characteristic (ROC) Curves</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>295</td>
</tr>
<tr>
<td>14</td>
<td>Linear Regression</td>
<td>297</td>
</tr>
<tr>
<td>14.1</td>
<td>Introduction</td>
<td>297</td>
</tr>
<tr>
<td>14.2</td>
<td>Algorithm</td>
<td>299</td>
</tr>
<tr>
<td>14.3</td>
<td>Implementation</td>
<td>299</td>
</tr>
<tr>
<td>14.4</td>
<td>Cross-Validation Results</td>
<td>300</td>
</tr>
<tr>
<td>14.5</td>
<td>Bootstrap Bias</td>
<td>303</td>
</tr>
<tr>
<td>14.6</td>
<td>Multiclass ROC Curves</td>
<td>306</td>
</tr>
<tr>
<td>14.7</td>
<td>Decision Boundaries</td>
<td>308</td>
</tr>
<tr>
<td>14.8</td>
<td>Summary</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>310</td>
</tr>
<tr>
<td>15</td>
<td>Decision Tree Classification</td>
<td>311</td>
</tr>
<tr>
<td>15.1</td>
<td>Introduction</td>
<td>311</td>
</tr>
<tr>
<td>15.2</td>
<td>Features Used</td>
<td>314</td>
</tr>
<tr>
<td>15.3</td>
<td>Terminal Nodes and Stopping Criteria</td>
<td>315</td>
</tr>
<tr>
<td>15.4</td>
<td>Algorithm</td>
<td>315</td>
</tr>
<tr>
<td>15.5</td>
<td>Implementation</td>
<td>315</td>
</tr>
<tr>
<td>15.6</td>
<td>Cross-Validation Results</td>
<td>318</td>
</tr>
<tr>
<td>15.7</td>
<td>Decision Boundaries</td>
<td>326</td>
</tr>
<tr>
<td>15.8</td>
<td>Summary</td>
<td>327</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>329</td>
</tr>
<tr>
<td>16</td>
<td>Random Forests</td>
<td>331</td>
</tr>
<tr>
<td>16.1</td>
<td>Introduction</td>
<td>331</td>
</tr>
<tr>
<td>16.2</td>
<td>Algorithm</td>
<td>333</td>
</tr>
<tr>
<td>16.3</td>
<td>Importance Scores</td>
<td>334</td>
</tr>
<tr>
<td>16.4</td>
<td>Strength and Correlation</td>
<td>338</td>
</tr>
<tr>
<td>16.5</td>
<td>Proximity and Supervised Clustering</td>
<td>342</td>
</tr>
<tr>
<td>16.6</td>
<td>Unsupervised Clustering</td>
<td>345</td>
</tr>
<tr>
<td>16.7</td>
<td>Class Outlier Detection</td>
<td>348</td>
</tr>
<tr>
<td>16.8</td>
<td>Implementation</td>
<td>350</td>
</tr>
<tr>
<td>16.9</td>
<td>Parameter Effects</td>
<td>350</td>
</tr>
<tr>
<td>16.10</td>
<td>Summary</td>
<td>357</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>358</td>
</tr>
</tbody>
</table>
17 K Nearest Neighbor 361
  17.1 Introduction 361
  17.2 Algorithm 362
  17.3 Implementation 363
  17.4 Cross-Validation Results 364
  17.5 Bootstrap Bias 369
  17.6 Multiclass ROC Curves 373
  17.7 Decision Boundaries 374
  17.8 Summary 377
  References 378

18 Naive Bayes Classifier 379
  18.1 Introduction 379
  18.2 Algorithm 380
  18.3 Cross-Validation Results 380
  18.4 Bootstrap Bias 384
  18.5 Multiclass ROC Curves 386
  18.6 Decision Boundaries 386
  18.7 Summary 389
  References 391

19 Linear Discriminant Analysis 393
  19.1 Introduction 393
  19.2 Multivariate Matrix Definitions 394
  19.3 Linear Discriminant Analysis 396
    19.3.1 Algorithm 397
    19.3.2 Cross-Validation Results 397
    19.3.3 Bootstrap Bias 401
    19.3.4 Multiclass ROC Curves 402
    19.3.5 Decision Boundaries 403
  19.4 Quadratic Discriminant Analysis 403
  19.5 Fisher’s Discriminant Analysis 406
  19.6 Summary 411
  References 412

20 Learning Vector Quantization 415
  20.1 Introduction 415
  20.2 Cross-Validation Results 417
  20.3 Bootstrap Bias 417
  20.4 Multiclass ROC Curves 426
20.5 Decision Boundaries 428
20.6 Summary 428
References 430

21 Logistic Regression 433
21.1 Introduction 433
21.2 Binary Logistic Regression 434
21.3 Polytomous Logistic Regression 439
21.4 Cross-Validation Results 443
21.5 Decision Boundaries 444
21.6 Summary 444
References 447

22 Support Vector Machines 449
22.1 Introduction 449
22.2 Hard-Margin SVM for Linearly Separable Classes 449
22.3 Kernel Mapping into Nonlinear Feature Space 452
22.4 Soft-Margin SVM for Nonlinearly Separable Classes 452
22.5 Gradient Ascent Soft-Margin SVM 454
22.5.1 Cross-Validation Results 455
22.5.2 Bootstrap Bias 457
22.5.3 Multiclass ROC Curves 465
22.5.4 Decision Boundaries 465
22.6 Least-Squares Soft-Margin SVM 465
22.6.1 Cross-Validation Results 470
22.6.2 Bootstrap Bias 477
22.6.3 Multiclass ROC Curves 477
22.6.4 Decision Boundaries 477
22.7 Summary 481
References 483

23 Artificial Neural Networks 487
23.1 Introduction 487
23.2 ANN Architecture 488
23.3 Basics of ANN Training 488
23.3.1 Backpropagation Learning 493
23.3.2 Resilient Backpropagation (RPROP) Learning 496
23.3.3 Cycles and Epochs 496
23.4 ANN Training Methods 497
23.4.1 Method 1: Gene Dimensional Reduction and Recursive Feature Elimination for Large Gene Lists 497
23.4.2 Method 2: Gene Filtering and Selection 502
## CONTENTS

### 27 Mixture of Experts
- 27.1 Introduction 591
- 27.2 Algorithm 595
- 27.3 Cross-Validation Results 596
- 27.4 Decision Boundaries 597
- 27.5 Summary 597

References 599

### 28 Covariance Matrix Filtering
- 28.1 Introduction 601
- 28.2 Covariance and Correlation Matrices 601
- 28.3 Random Matrices 602
- 28.4 Component Subtraction 608
- 28.5 Covariance Matrix Shrinkage 610
- 28.6 Covariance Matrix Filtering 613
- 28.7 Summary 621

References 622

### APPENDIXES

### A Probability Primer
- A.1 Choices 627
- A.2 Permutations 628
- A.3 Combinations 630
- A.4 Probability 632
  - A.4.1 Addition Rule 633
  - A.4.2 Multiplication Rule and Conditional Probabilities 634
  - A.4.3 Multiplication Rule for Independent Events 635
  - A.4.4 Elimination Rule (Disease Prevalence) 636
  - A.4.5 Bayes’ Rule (Pathway Probabilities) 637

### B Matrix Algebra
- B.1 Vectors 639
- B.2 Matrices 642
- B.3 Sample Mean, Covariance, and Correlation 647
- B.4 Diagonal Matrices 648
- B.5 Identity Matrices 649
- B.6 Trace of a Matrix 650