

DATA CLASSIFICATION

Algorithms and Applications

Edited by

Charu C. Aggarwal

IBM T. J. Watson Research Center
Yorktown Heights, New York, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

Contents

Editor Biography	xxiii
Contributors	xxv
Preface	xxvii
1 An Introduction to Data Classification	1
<i>Charu C. Aggarwal</i>	
1.1 Introduction	2
1.2 Common Techniques in Data Classification	4
1.2.1 Feature Selection Methods	4
1.2.2 Probabilistic Methods	6
1.2.3 Decision Trees	7
1.2.4 Rule-Based Methods	9
1.2.5 Instance-Based Learning	11
1.2.6 SVM Classifiers	11
1.2.7 Neural Networks	14
1.3 Handling Different Data Types	16
1.3.1 Large Scale Data: Big Data and Data Streams	16
1.3.1.1 Data Streams	16
1.3.1.2 The Big Data Framework	17
1.3.2 Text Classification	18
1.3.3 Multimedia Classification	20
1.3.4 Time Series and Sequence Data Classification	20
1.3.5 Network Data Classification	21
1.3.6 Uncertain Data Classification	21
1.4 Variations on Data Classification	22
1.4.1 Rare Class Learning	22
1.4.2 Distance Function Learning	22
1.4.3 Ensemble Learning for Data Classification	23
1.4.4 Enhancing Classification Methods with Additional Data	24
1.4.4.1 Semi-Supervised Learning	24
1.4.4.2 Transfer Learning	26
1.4.5 Incorporating Human Feedback	27
1.4.5.1 Active Learning	28
1.4.5.2 Visual Learning	29
1.4.6 Evaluating Classification Algorithms	30
1.5 Discussion and Conclusions	31

2	Feature Selection for Classification: A Review	37
	<i>Jiliang Tang, Salem Alelyani, and Huan Liu</i>	
2.1	Introduction	38
2.1.1	Data Classification	39
2.1.2	Feature Selection	40
2.1.3	Feature Selection for Classification	42
2.2	Algorithms for Flat Features	43
2.2.1	Filter Models	44
2.2.2	Wrapper Models	46
2.2.3	Embedded Models	47
2.3	Algorithms for Structured Features	49
2.3.1	Features with Group Structure	50
2.3.2	Features with Tree Structure	51
2.3.3	Features with Graph Structure	53
2.4	Algorithms for Streaming Features	55
2.4.1	The Grafting Algorithm	56
2.4.2	The Alpha-Investing Algorithm	56
2.4.3	The Online Streaming Feature Selection Algorithm	57
2.5	Discussions and Challenges	57
2.5.1	Scalability	57
2.5.2	Stability	58
2.5.3	Linked Data	58
3	Probabilistic Models for Classification	65
	<i>Hongbo Deng, Yizhou Sun, Yi Chang, and Jiawei Han</i>	
3.1	Introduction	66
3.2	Naive Bayes Classification	67
3.2.1	Bayes' Theorem and Preliminary	67
3.2.2	Naive Bayes Classifier	69
3.2.3	Maximum-Likelihood Estimates for Naive Bayes Models	70
3.2.4	Applications	71
3.3	Logistic Regression Classification	72
3.3.1	Logistic Regression	73
3.3.2	Parameters Estimation for Logistic Regression	74
3.3.3	Regularization in Logistic Regression	75
3.3.4	Applications	76
3.4	Probabilistic Graphical Models for Classification	76
3.4.1	Bayesian Networks	76
3.4.1.1	Bayesian Network Construction	77
3.4.1.2	Inference in a Bayesian Network	78
3.4.1.3	Learning Bayesian Networks	78
3.4.2	Hidden Markov Models	78
3.4.2.1	The Inference and Learning Algorithms	79
3.4.3	Markov Random Fields	81
3.4.3.1	Conditional Independence	81
3.4.3.2	Clique Factorization	81
3.4.3.3	The Inference and Learning Algorithms	82
3.4.4	Conditional Random Fields	82
3.4.4.1	The Learning Algorithms	83
3.5	Summary	83

4	Decision Trees: Theory and Algorithms	87
	<i>Victor E. Lee, Lin Liu, and Ruoming Jin</i>	
4.1	Introduction	87
4.2	Top-Down Decision Tree Induction	91
	4.2.1 Node Splitting	92
	4.2.2 Tree Pruning	97
4.3	Case Studies with C4.5 and CART	99
	4.3.1 Splitting Criteria	100
	4.3.2 Stopping Conditions	100
	4.3.3 Pruning Strategy	101
	4.3.4 Handling Unknown Values: Induction and Prediction	101
	4.3.5 Other Issues: Windowing and Multivariate Criteria	102
4.4	Scalable Decision Tree Construction	103
	4.4.1 RainForest-Based Approach	104
	4.4.2 SPIES Approach	105
	4.4.3 Parallel Decision Tree Construction	107
4.5	Incremental Decision Tree Induction	108
	4.5.1 ID3 Family	108
	4.5.2 VFDT Family	110
	4.5.3 Ensemble Method for Streaming Data	113
4.6	Summary	114
5	Rule-Based Classification	121
	<i>Xiao-Li Li and Bing Liu</i>	
5.1	Introduction	121
5.2	Rule Induction	123
	5.2.1 Two Algorithms for Rule Induction	123
	5.2.1.1 CN2 Induction Algorithm (Ordered Rules)	124
	5.2.1.2 RIPPER Algorithm and Its Variations (Ordered Classes)	125
	5.2.2 Learn One Rule in Rule Learning	126
5.3	Classification Based on Association Rule Mining	129
	5.3.1 Association Rule Mining	130
	5.3.1.1 Definitions of Association Rules, Support, and Confidence	131
	5.3.1.2 The Introduction of Apriori Algorithm	133
	5.3.2 Mining Class Association Rules	136
	5.3.3 Classification Based on Associations	139
	5.3.3.1 Additional Discussion for CARs Mining	139
	5.3.3.2 Building a Classifier Using CARs	140
	5.3.4 Other Techniques for Association Rule-Based Classification	142
5.4	Applications	144
	5.4.1 Text Categorization	144
	5.4.2 Intrusion Detection	147
	5.4.3 Using Class Association Rules for Diagnostic Data Mining	148
	5.4.4 Gene Expression Data Analysis	149
5.5	Discussion and Conclusion	150

6	Instance-Based Learning: A Survey	157
	<i>Charu C. Aggarwal</i>	
6.1	Introduction	157
6.2	Instance-Based Learning Framework	159
6.3	The Nearest Neighbor Classifier	160
6.3.1	Handling Symbolic Attributes	163
6.3.2	Distance-Weighted Nearest Neighbor Methods	163
6.3.3	Local Distance Scaling	164
6.3.4	Attribute-Weighted Nearest Neighbor Methods	164
6.3.5	Locally Adaptive Nearest Neighbor Classifier	167
6.3.6	Combining with Ensemble Methods	169
6.3.7	Multi-Label Learning	169
6.4	Lazy SVM Classification	171
6.5	Locally Weighted Regression	172
6.6	Lazy Naive Bayes	173
6.7	Lazy Decision Trees	173
6.8	Rule-Based Classification	174
6.9	Radial Basis Function Networks: Leveraging Neural Networks for Instance-Based Learning	175
6.10	Lazy Methods for Diagnostic and Visual Classification	176
6.11	Conclusions and Summary	180
7	Support Vector Machines	187
	<i>Po-Wei Wang and Chih-Jen Lin</i>	
7.1	Introduction	187
7.2	The Maximum Margin Perspective	188
7.3	The Regularization Perspective	190
7.4	The Support Vector Perspective	191
7.5	Kernel Tricks	194
7.6	Solvers and Algorithms	196
7.7	Multiclass Strategies	198
7.8	Conclusion	201
8	Neural Networks: A Review	205
	<i>Alain Biem</i>	
8.1	Introduction	206
8.2	Fundamental Concepts	208
8.2.1	Mathematical Model of a Neuron	208
8.2.2	Types of Units	209
8.2.2.1	McCullough Pitts Binary Threshold Unit	209
8.2.2.2	Linear Unit	210
8.2.2.3	Linear Threshold Unit	211
8.2.2.4	Sigmoidal Unit	211
8.2.2.5	Distance Unit	211
8.2.2.6	Radial Basis Unit	211
8.2.2.7	Polynomial Unit	212
8.2.3	Network Topology	212
8.2.3.1	Layered Network	212
8.2.3.2	Networks with Feedback	212
8.2.3.3	Modular Networks	213
8.2.4	Computation and Knowledge Representation	213

8.2.5	Learning	213
8.2.5.1	Hebbian Rule	213
8.2.5.2	The Delta Rule	214
8.3	Single-Layer Neural Network	214
8.3.1	The Single-Layer Perceptron	214
8.3.1.1	Perceptron Criterion	214
8.3.1.2	Multi-Class Perceptrons	216
8.3.1.3	Perceptron Enhancements	216
8.3.2	Adaline	217
8.3.2.1	Two-Class Adaline	217
8.3.2.2	Multi-Class Adaline	218
8.3.3	Learning Vector Quantization (LVQ)	219
8.3.3.1	LVQ1 Training	219
8.3.3.2	LVQ2 Training	219
8.3.3.3	Application and Limitations	220
8.4	Kernel Neural Network	220
8.4.1	Radial Basis Function Network	220
8.4.2	RBFN Training	222
8.4.2.1	Using Training Samples as Centers	222
8.4.2.2	Random Selection of Centers	222
8.4.2.3	Unsupervised Selection of Centers	222
8.4.2.4	Supervised Estimation of Centers	223
8.4.2.5	Linear Optimization of Weights	223
8.4.2.6	Gradient Descent and Enhancements	223
8.4.3	RBF Applications	223
8.5	Multi-Layer Feedforward Network	224
8.5.1	MLP Architecture for Classification	224
8.5.1.1	Two-Class Problems	225
8.5.1.2	Multi-Class Problems	225
8.5.1.3	Forward Propagation	226
8.5.2	Error Metrics	227
8.5.2.1	Mean Square Error (MSE)	227
8.5.2.2	Cross-Entropy (CE)	227
8.5.2.3	Minimum Classification Error (MCE)	228
8.5.3	Learning by Backpropagation	228
8.5.4	Enhancing Backpropagation	229
8.5.4.1	Backpropagation with Momentum	230
8.5.4.2	Delta-Bar-Delta	231
8.5.4.3	Rprop Algorithm	231
8.5.4.4	Quick-Prop	231
8.5.5	Generalization Issues	232
8.5.6	Model Selection	232
8.6	Deep Neural Networks	232
8.6.1	Use of Prior Knowledge	233
8.6.2	Layer-Wise Greedy Training	234
8.6.2.1	Deep Belief Networks (DBNs)	234
8.6.2.2	Stack Auto-Encoder	235
8.6.3	Limits and Applications	235
8.7	Summary	235

9	A Survey of Stream Classification Algorithms	245
	<i>Charu C. Aggarwal</i>	
9.1	Introduction	245
9.2	Generic Stream Classification Algorithms	247
9.2.1	Decision Trees for Data Streams	247
9.2.2	Rule-Based Methods for Data Streams	249
9.2.3	Nearest Neighbor Methods for Data Streams	250
9.2.4	SVM Methods for Data Streams	251
9.2.5	Neural Network Classifiers for Data Streams	252
9.2.6	Ensemble Methods for Data Streams	253
9.3	Rare Class Stream Classification	254
9.3.1	Detecting Rare Classes	255
9.3.2	Detecting Novel Classes	255
9.3.3	Detecting Infrequently Recurring Classes	256
9.4	Discrete Attributes: The Massive Domain Scenario	256
9.5	Other Data Domains	262
9.5.1	Text Streams	262
9.5.2	Graph Streams	264
9.5.3	Uncertain Data Streams	267
9.6	Conclusions and Summary	267
10	Big Data Classification	275
	<i>Hanghang Tong</i>	
10.1	Introduction	275
10.2	Scale-Up on a Single Machine	276
10.2.1	Background	276
10.2.2	SVMPerf	276
10.2.3	Pegasos	277
10.2.4	Bundle Methods	279
10.3	Scale-Up by Parallelism	280
10.3.1	Parallel Decision Trees	280
10.3.2	Parallel SVMs	281
10.3.3	MRM-ML	281
10.3.4	SystemML	282
10.4	Conclusion	283
11	Text Classification	287
	<i>Charu C. Aggarwal and ChengXiang Zhai</i>	
11.1	Introduction	288
11.2	Feature Selection for Text Classification	290
11.2.1	Gini Index	291
11.2.2	Information Gain	292
11.2.3	Mutual Information	292
11.2.4	χ^2 -Statistic	292
11.2.5	Feature Transformation Methods: Unsupervised and Supervised LSI	293
11.2.6	Supervised Clustering for Dimensionality Reduction	294
11.2.7	Linear Discriminant Analysis	294
11.2.8	Generalized Singular Value Decomposition	295
11.2.9	Interaction of Feature Selection with Classification	296
11.3	Decision Tree Classifiers	296
11.4	Rule-Based Classifiers	298

11.5	Probabilistic and Naive Bayes Classifiers	300
11.5.1	Bernoulli Multivariate Model	301
11.5.2	Multinomial Distribution	304
11.5.3	Mixture Modeling for Text Classification	305
11.6	Linear Classifiers	308
11.6.1	SVM Classifiers	308
11.6.2	Regression-Based Classifiers	311
11.6.3	Neural Network Classifiers	312
11.6.4	Some Observations about Linear Classifiers	315
11.7	Proximity-Based Classifiers	315
11.8	Classification of Linked and Web Data	317
11.9	Meta-Algorithms for Text Classification	321
11.9.1	Classifier Ensemble Learning	321
11.9.2	Data Centered Methods: Boosting and Bagging	322
11.9.3	Optimizing Specific Measures of Accuracy	322
11.10	Leveraging Additional Training Data	323
11.10.1	Semi-Supervised Learning	324
11.10.2	Transfer Learning	326
11.10.3	Active Learning	327
11.11	Conclusions and Summary	327

12 Multimedia Classification **337**

Shiyu Chang, Wei Han, Xianming Liu, Ning Xu, Pooya Khorrami, and Thomas S. Huang

12.1	Introduction	338
12.1.1	Overview	338
12.2	Feature Extraction and Data Pre-Processing	339
12.2.1	Text Features	340
12.2.2	Image Features	341
12.2.3	Audio Features	344
12.2.4	Video Features	345
12.3	Audio Visual Fusion	345
12.3.1	Fusion Methods	346
12.3.2	Audio Visual Speech Recognition	346
12.3.2.1	Visual Front End	347
12.3.2.2	Decision Fusion on HMM	348
12.3.3	Other Applications	349
12.4	Ontology-Based Classification and Inference	349
12.4.1	Popular Applied Ontology	350
12.4.2	Ontological Relations	350
12.4.2.1	Definition	351
12.4.2.2	Subclass Relation	351
12.4.2.3	Co-Occurrence Relation	352
12.4.2.4	Combination of the Two Relations	352
12.4.2.5	Inherently Used Ontology	353
12.5	Geographical Classification with Multimedia Data	353
12.5.1	Data Modalities	353
12.5.2	Challenges in Geographical Classification	354
12.5.3	Geo-Classification for Images	355
12.5.3.1	Classifiers	356
12.5.4	Geo-Classification for Web Videos	356

12.6	Conclusion	356
13	Time Series Data Classification	365
	<i>Dimitrios Kotsakos and Dimitrios Gunopoulos</i>	
13.1	Introduction	365
13.2	Time Series Representation	367
13.3	Distance Measures	367
13.3.1	L_p -Norms	367
13.3.2	Dynamic Time Warping (DTW)	367
13.3.3	Edit Distance	368
13.3.4	Longest Common Subsequence (LCSS)	369
13.4	k -NN	369
13.4.1	Speeding up the k -NN	370
13.5	Support Vector Machines (SVMs)	371
13.6	Classification Trees	372
13.7	Model-Based Classification	374
13.8	Distributed Time Series Classification	375
13.9	Conclusion	375
14	Discrete Sequence Classification	379
	<i>Mohammad Al Hasan</i>	
14.1	Introduction	379
14.2	Background	380
14.2.1	Sequence	380
14.2.2	Sequence Classification	381
14.2.3	Frequent Sequential Patterns	381
14.2.4	n -Grams	382
14.3	Sequence Classification Methods	382
14.4	Feature-Based Classification	382
14.4.1	Filtering Method for Sequential Feature Selection	383
14.4.2	Pattern Mining Framework for Mining Sequential Features	385
14.4.3	A Wrapper-Based Method for Mining Sequential Features	386
14.5	Distance-Based Methods	386
14.5.0.1	Alignment-Based Distance	387
14.5.0.2	Keyword-Based Distance	388
14.5.0.3	Kernel-Based Similarity	388
14.5.0.4	Model-Based Similarity	388
14.5.0.5	Time Series Distance Metrics	388
14.6	Model-Based Method	389
14.7	Hybrid Methods	390
14.8	Non-Traditional Sequence Classification	391
14.8.1	Semi-Supervised Sequence Classification	391
14.8.2	Classification of Label Sequences	392
14.8.3	Classification of Sequence of Vector Data	392
14.9	Conclusions	393
15	Collective Classification of Network Data	399
	<i>Ben London and Lise Getoor</i>	
15.1	Introduction	399
15.2	Collective Classification Problem Definition	400
15.2.1	Inductive vs. Transductive Learning	401

15.2.2	Active Collective Classification	402
15.3	Iterative Methods	402
15.3.1	Label Propagation	402
15.3.2	Iterative Classification Algorithms	404
15.4	Graph-Based Regularization	405
15.5	Probabilistic Graphical Models	406
15.5.1	Directed Models	406
15.5.2	Undirected Models	408
15.5.3	Approximate Inference in Graphical Models	409
15.5.3.1	Gibbs Sampling	409
15.5.3.2	Loopy Belief Propagation (LBP)	410
15.6	Feature Construction	410
15.6.1	Data Graph	411
15.6.2	Relational Features	412
15.7	Applications of Collective Classification	412
15.8	Conclusion	413
16	Uncertain Data Classification	417
	<i>Reynold Cheng, Yixiang Fang, and Matthias Renz</i>	
16.1	Introduction	417
16.2	Preliminaries	419
16.2.1	Data Uncertainty Models	419
16.2.2	Classification Framework	419
16.3	Classification Algorithms	420
16.3.1	Decision Trees	420
16.3.2	Rule-Based Classification	424
16.3.3	Associative Classification	426
16.3.4	Density-Based Classification	429
16.3.5	Nearest Neighbor-Based Classification	432
16.3.6	Support Vector Classification	436
16.3.7	Naive Bayes Classification	438
16.4	Conclusions	441
17	Rare Class Learning	445
	<i>Charu C. Aggarwal</i>	
17.1	Introduction	445
17.2	Rare Class Detection	448
17.2.1	Cost Sensitive Learning	449
17.2.1.1	MetaCost: A Relabeling Approach	449
17.2.1.2	Weighting Methods	450
17.2.1.3	Bayes Classifiers	450
17.2.1.4	Proximity-Based Classifiers	451
17.2.1.5	Rule-Based Classifiers	451
17.2.1.6	Decision Trees	451
17.2.1.7	SVM Classifier	452
17.2.2	Adaptive Re-Sampling	452
17.2.2.1	Relation between Weighting and Sampling	453
17.2.2.2	Synthetic Over-Sampling: SMOTE	453
17.2.2.3	One Class Learning with Positive Class	453
17.2.2.4	Ensemble Techniques	454
17.2.3	Boosting Methods	454

17.3	The Semi-Supervised Scenario: Positive and Unlabeled Data	455
17.3.1	Difficult Cases and One-Class Learning	456
17.4	The Semi-Supervised Scenario: Novel Class Detection	456
17.4.1	One Class Novelty Detection	457
17.4.2	Combining Novel Class Detection with Rare Class Detection	458
17.4.3	Online Novelty Detection	458
17.5	Human Supervision	459
17.6	Other Work	461
17.7	Conclusions and Summary	462
18	Distance Metric Learning for Data Classification	469
	<i>Fei Wang</i>	
18.1	Introduction	469
18.2	The Definition of Distance Metric Learning	470
18.3	Supervised Distance Metric Learning Algorithms	471
18.3.1	Linear Discriminant Analysis (LDA)	472
18.3.2	Margin Maximizing Discriminant Analysis (MMDA)	473
18.3.3	Learning with Side Information (LSI)	474
18.3.4	Relevant Component Analysis (RCA)	474
18.3.5	Information Theoretic Metric Learning (ITML)	475
18.3.6	Neighborhood Component Analysis (NCA)	475
18.3.7	Average Neighborhood Margin Maximization (ANMM)	476
18.3.8	Large Margin Nearest Neighbor Classifier (LMNN)	476
18.4	Advanced Topics	477
18.4.1	Semi-Supervised Metric Learning	477
18.4.1.1	Laplacian Regularized Metric Learning (LRML)	477
18.4.1.2	Constraint Margin Maximization (CMM)	478
18.4.2	Online Learning	478
18.4.2.1	Pseudo-Metric Online Learning Algorithm (POLA)	479
18.4.2.2	Online Information Theoretic Metric Learning (OITML)	480
18.5	Conclusions and Discussions	480
19	Ensemble Learning	483
	<i>Yaliang Li, Jing Gao, Qi Li, and Wei Fan</i>	
19.1	Introduction	484
19.2	Bayesian Methods	487
19.2.1	Bayes Optimal Classifier	487
19.2.2	Bayesian Model Averaging	488
19.2.3	Bayesian Model Combination	490
19.3	Bagging	491
19.3.1	General Idea	491
19.3.2	Random Forest	493
19.4	Boosting	495
19.4.1	General Boosting Procedure	495
19.4.2	AdaBoost	496
19.5	Stacking	498
19.5.1	General Stacking Procedure	498
19.5.2	Stacking and Cross-Validation	500
19.5.3	Discussions	501
19.6	Recent Advances in Ensemble Learning	502
19.7	Conclusions	503

20	Semi-Supervised Learning	511
	<i>Kaushik Sinha</i>	
20.1	Introduction	511
20.1.1	Transductive vs. Inductive Semi-Supervised Learning	514
20.1.2	Semi-Supervised Learning Framework and Assumptions	514
20.2	Generative Models	515
20.2.1	Algorithms	516
20.2.2	Description of a Representative Algorithm	516
20.2.3	Theoretical Justification and Relevant Results	517
20.3	Co-Training	519
20.3.1	Algorithms	520
20.3.2	Description of a Representative Algorithm	520
20.3.3	Theoretical Justification and Relevant Results	520
20.4	Graph-Based Methods	522
20.4.1	Algorithms	522
20.4.1.1	Graph Cut	522
20.4.1.2	Graph Transduction	523
20.4.1.3	Manifold Regularization	524
20.4.1.4	Random Walk	525
20.4.1.5	Large Scale Learning	526
20.4.2	Description of a Representative Algorithm	526
20.4.3	Theoretical Justification and Relevant Results	527
20.5	Semi-Supervised Learning Methods Based on Cluster Assumption	528
20.5.1	Algorithms	528
20.5.2	Description of a Representative Algorithm	529
20.5.3	Theoretical Justification and Relevant Results	529
20.6	Related Areas	531
20.7	Concluding Remarks	531
21	Transfer Learning	537
	<i>Sinno Jialin Pan</i>	
21.1	Introduction	538
21.2	Transfer Learning Overview	541
21.2.1	Background	541
21.2.2	Notations and Definitions	541
21.3	Homogenous Transfer Learning	542
21.3.1	Instance-Based Approach	542
21.3.1.1	Case I: No Target Labeled Data	543
21.3.1.2	Case II: A Few Target Labeled Data	544
21.3.2	Feature-Representation-Based Approach	545
21.3.2.1	Encoding Specific Knowledge for Feature Learning	545
21.3.2.2	Learning Features by Minimizing Distance between Distributions	548
21.3.2.3	Learning Features Inspired by Multi-Task Learning	549
21.3.2.4	Learning Features Inspired by Self-Taught Learning	550
21.3.2.5	Other Feature Learning Approaches	550
21.3.3	Model-Parameter-Based Approach	550
21.3.4	Relational-Information-Based Approaches	552
21.4	Heterogeneous Transfer Learning	553
21.4.1	Heterogeneous Feature Spaces	553
21.4.2	Different Label Spaces	554

21.5	Transfer Bounds and Negative Transfer	554
21.6	Other Research Issues	555
21.6.1	Binary Classification vs. Multi-Class Classification	556
21.6.2	Knowledge Transfer from Multiple Source Domains	556
21.6.3	Transfer Learning Meets Active Learning	556
21.7	Applications of Transfer Learning	557
21.7.1	NLP Applications	557
21.7.2	Web-Based Applications	557
21.7.3	Sensor-Based Applications	557
21.7.4	Applications to Computer Vision	557
21.7.5	Applications to Bioinformatics	557
21.7.6	Other Applications	558
21.8	Concluding Remarks	558
22	Active Learning: A Survey	571
	<i>Charu C. Aggarwal, Xiangnan Kong, Quanquan Gu, Jiawei Han, and Philip S. Yu</i>	
22.1	Introduction	572
22.2	Motivation and Comparisons to Other Strategies	574
22.2.1	Comparison with Other Forms of Human Feedback	575
22.2.2	Comparisons with Semi-Supervised and Transfer Learning	576
22.3	Querying Strategies	576
22.3.1	Heterogeneity-Based Models	577
22.3.1.1	Uncertainty Sampling	577
22.3.1.2	Query-by-Committee	578
22.3.1.3	Expected Model Change	578
22.3.2	Performance-Based Models	579
22.3.2.1	Expected Error Reduction	579
22.3.2.2	Expected Variance Reduction	580
22.3.3	Representativeness-Based Models	580
22.3.4	Hybrid Models	580
22.4	Active Learning with Theoretical Guarantees	581
22.4.1	A Simple Example	581
22.4.2	Existing Works	582
22.4.3	Preliminaries	582
22.4.4	Importance Weighted Active Learning	582
22.4.4.1	Algorithm	583
22.4.4.2	Consistency	583
22.4.4.3	Label Complexity	584
22.5	Dependency-Oriented Data Types for Active Learning	585
22.5.1	Active Learning in Sequences	585
22.5.2	Active Learning in Graphs	585
22.5.2.1	Classification of Many Small Graphs	586
22.5.2.2	Node Classification in a Single Large Graph	587
22.6	Advanced Methods	589
22.6.1	Active Learning of Features	589
22.6.2	Active Learning of Kernels	590
22.6.3	Active Learning of Classes	591
22.6.4	Streaming Active Learning	591
22.6.5	Multi-Instance Active Learning	592
22.6.6	Multi-Label Active Learning	593
22.6.7	Multi-Task Active Learning	593

22.6.8	Multi-View Active Learning	594
22.6.9	Multi-Oracle Active Learning	594
22.6.10	Multi-Objective Active Learning	595
22.6.11	Variable Labeling Costs	596
22.6.12	Active Transfer Learning	596
22.6.13	Active Reinforcement Learning	597
22.7	Conclusions	597
23	Visual Classification	607
	<i>Giorgio Maria Di Nunzio</i>	
23.1	Introduction	608
23.1.1	Requirements for Visual Classification	609
23.1.2	Visualization Metaphors	610
23.1.2.1	2D and 3D Spaces	610
23.1.2.2	More Complex Metaphors	610
23.1.3	Challenges in Visual Classification	611
23.1.4	Related Works	611
23.2	Approaches	612
23.2.1	Nomograms	612
23.2.1.1	Naïve Bayes Nomogram	613
23.2.2	Parallel Coordinates	613
23.2.2.1	Edge Cluttering	614
23.2.3	Radial Visualizations	614
23.2.3.1	Star Coordinates	615
23.2.4	Scatter Plots	616
23.2.4.1	Clustering	617
23.2.4.2	Naïve Bayes Classification	617
23.2.5	Topological Maps	619
23.2.5.1	Self-Organizing Maps	619
23.2.5.2	Generative Topographic Mapping	619
23.2.6	Trees	620
23.2.6.1	Decision Trees	621
23.2.6.2	Treemap	622
23.2.6.3	Hyperbolic Tree	623
23.2.6.4	Phylogenetic Trees	623
23.3	Systems	623
23.3.1	EnsembleMatrix and ManiMatrix	623
23.3.2	Systematic Mapping	624
23.3.3	iVisClassifier	624
23.3.4	ParallelTopics	625
23.3.5	VisBricks	625
23.3.6	WHITE	625
23.3.7	Text Document Retrieval	625
23.4	Summary and Conclusions	626
24	Evaluation of Classification Methods	633
	<i>Nele Verbiest, Karel Vermeulen, and Ankur Teredesai</i>	
24.1	Introduction	633
24.2	Validation Schemes	634
24.3	Evaluation Measures	636
24.3.1	Accuracy Related Measures	636

24.3.1.1	Discrete Classifiers	636
24.3.1.2	Probabilistic Classifiers	638
24.3.2	Additional Measures	642
24.4	Comparing Classifiers	643
24.4.1	Parametric Statistical Comparisons	644
24.4.1.1	Pairwise Comparisons	644
24.4.1.2	Multiple Comparisons	644
24.4.2	Non-Parametric Statistical Comparisons	646
24.4.2.1	Pairwise Comparisons	646
24.4.2.2	Multiple Comparisons	647
24.4.2.3	Permutation Tests	651
24.5	Concluding Remarks	652
25	Educational and Software Resources for Data Classification	657
	<i>Charu C. Aggarwal</i>	
25.1	Introduction	657
25.2	Educational Resources	658
25.2.1	Books on Data Classification	658
25.2.2	Popular Survey Papers on Data Classification	658
25.3	Software for Data Classification	659
25.3.1	Data Benchmarks for Software and Research	660
25.4	Summary	661
	Index	667