

Contents

Foreword v

Preface xxiii

Updated and revised content xxvii

Acknowledgments xxix

Part I Machine learning tools and techniques 1

1 What's it all about? 3

- 1.1 **Data mining and machine learning** 4
 - Describing structural patterns* 6
 - Machine learning* 7
 - Data mining* 9
- 1.2 **Simple examples: The weather problem and others** 9
 - The weather problem* 10
 - Contact lenses: An idealized problem* 13
 - Irises: A classic numeric dataset* 15
 - CPU performance: Introducing numeric prediction* 16
 - Labor negotiations: A more realistic example* 17
 - Soybean classification: A classic machine learning success* 18
- 1.3 **Fielded applications** 22
 - Decisions involving judgment* 22
 - Screening images* 23
 - Load forecasting* 24
 - Diagnosis* 25
 - Marketing and sales* 26
 - Other applications* 28

- 1.4 Machine learning and statistics 29
- 1.5 Generalization as search 30
 - Enumerating the concept space* 31
 - Bias* 32
- 1.6 Data mining and ethics 35
- 1.7 Further reading 37

- 2 Input: Concepts, instances, and attributes 41**
 - 2.1 What's a concept? 42
 - 2.2 What's in an example? 45
 - 2.3 What's in an attribute? 49
 - 2.4 Preparing the input 52
 - Gathering the data together* 52
 - ARFF format* 53
 - Sparse data* 55
 - Attribute types* 56
 - Missing values* 58
 - Inaccurate values* 59
 - Getting to know your data* 60
 - 2.5 Further reading 60

- 3 Output: Knowledge representation 61**
 - 3.1 Decision tables 62
 - 3.2 Decision trees 62
 - 3.3 Classification rules 65
 - 3.4 Association rules 69
 - 3.5 Rules with exceptions 70
 - 3.6 Rules involving relations 73
 - 3.7 Trees for numeric prediction 76
 - 3.8 Instance-based representation 76
 - 3.9 Clusters 81
 - 3.10 Further reading 82

- 4 Algorithms: The basic methods 83**
 - 4.1 Inferring rudimentary rules 84**
 - Missing values and numeric attributes 86*
 - Discussion 88*
 - 4.2 Statistical modeling 88**
 - Missing values and numeric attributes 92*
 - Bayesian models for document classification 94*
 - Discussion 96*
 - 4.3 Divide-and-conquer: Constructing decision trees 97**
 - Calculating information 100*
 - Highly branching attributes 102*
 - Discussion 105*
 - 4.4 Covering algorithms: Constructing rules 105**
 - Rules versus trees 107*
 - A simple covering algorithm 107*
 - Rules versus decision lists 111*
 - 4.5 Mining association rules 112**
 - Item sets 113*
 - Association rules 113*
 - Generating rules efficiently 117*
 - Discussion 118*
 - 4.6 Linear models 119**
 - Numeric prediction: Linear regression 119*
 - Linear classification: Logistic regression 121*
 - Linear classification using the perceptron 124*
 - Linear classification using Winnow 126*
 - 4.7 Instance-based learning 128**
 - The distance function 128*
 - Finding nearest neighbors efficiently 129*
 - Discussion 135*
 - 4.8 Clustering 136**
 - Iterative distance-based clustering 137*
 - Faster distance calculations 138*
 - Discussion 139*
 - 4.9 Further reading 139**

5	Credibility: Evaluating what's been learned	143
5.1	Training and testing	144
5.2	Predicting performance	146
5.3	Cross-validation	149
5.4	Other estimates	151
	<i>Leave-one-out</i>	151
	<i>The bootstrap</i>	152
5.5	Comparing data mining methods	153
5.6	Predicting probabilities	157
	<i>Quadratic loss function</i>	158
	<i>Informational loss function</i>	159
	<i>Discussion</i>	160
5.7	Counting the cost	161
	<i>Cost-sensitive classification</i>	164
	<i>Cost-sensitive learning</i>	165
	<i>Lift charts</i>	166
	<i>ROC curves</i>	168
	<i>Recall-precision curves</i>	171
	<i>Discussion</i>	172
	<i>Cost curves</i>	173
5.8	Evaluating numeric prediction	176
5.9	The minimum description length principle	179
5.10	Applying the MDL principle to clustering	183
5.11	Further reading	184
6	Implementations: Real machine learning schemes	187
6.1	Decision trees	189
	<i>Numeric attributes</i>	189
	<i>Missing values</i>	191
	<i>Pruning</i>	192
	<i>Estimating error rates</i>	193
	<i>Complexity of decision tree induction</i>	196
	<i>From trees to rules</i>	198
	<i>C4.5: Choices and options</i>	198
	<i>Discussion</i>	199
6.2	Classification rules	200
	<i>Criteria for choosing tests</i>	200
	<i>Missing values, numeric attributes</i>	201

	<i>Generating good rules</i>	202
	<i>Using global optimization</i>	205
	<i>Obtaining rules from partial decision trees</i>	207
	<i>Rules with exceptions</i>	210
	<i>Discussion</i>	213
6.3	Extending linear models	214
	<i>The maximum margin hyperplane</i>	215
	<i>Nonlinear class boundaries</i>	217
	<i>Support vector regression</i>	219
	<i>The kernel perceptron</i>	222
	<i>Multilayer perceptrons</i>	223
	<i>Discussion</i>	235
6.4	Instance-based learning	235
	<i>Reducing the number of exemplars</i>	236
	<i>Pruning noisy exemplars</i>	236
	<i>Weighting attributes</i>	237
	<i>Generalizing exemplars</i>	238
	<i>Distance functions for generalized exemplars</i>	239
	<i>Generalized distance functions</i>	241
	<i>Discussion</i>	242
6.5	Numeric prediction	243
	<i>Model trees</i>	244
	<i>Building the tree</i>	245
	<i>Pruning the tree</i>	245
	<i>Nominal attributes</i>	246
	<i>Missing values</i>	246
	<i>Pseudocode for model tree induction</i>	247
	<i>Rules from model trees</i>	250
	<i>Locally weighted linear regression</i>	251
	<i>Discussion</i>	253
6.6	Clustering	254
	<i>Choosing the number of clusters</i>	254
	<i>Incremental clustering</i>	255
	<i>Category utility</i>	260
	<i>Probability-based clustering</i>	262
	<i>The EM algorithm</i>	265
	<i>Extending the mixture model</i>	266
	<i>Bayesian clustering</i>	268
	<i>Discussion</i>	270
6.7	Bayesian networks	271
	<i>Making predictions</i>	272
	<i>Learning Bayesian networks</i>	276

<i>Specific algorithms</i>	278
<i>Data structures for fast learning</i>	280
<i>Discussion</i>	283

7 Transformations: Engineering the input and output 285

7.1 Attribute selection	288
<i>Scheme-independent selection</i>	290
<i>Searching the attribute space</i>	292
<i>Scheme-specific selection</i>	294
7.2 Discretizing numeric attributes	296
<i>Unsupervised discretization</i>	297
<i>Entropy-based discretization</i>	298
<i>Other discretization methods</i>	302
<i>Entropy-based versus error-based discretization</i>	302
<i>Converting discrete to numeric attributes</i>	304
7.3 Some useful transformations	305
<i>Principal components analysis</i>	306
<i>Random projections</i>	309
<i>Text to attribute vectors</i>	309
<i>Time series</i>	311
7.4 Automatic data cleansing	312
<i>Improving decision trees</i>	312
<i>Robust regression</i>	313
<i>Detecting anomalies</i>	314
7.5 Combining multiple models	315
<i>Bagging</i>	316
<i>Bagging with costs</i>	319
<i>Randomization</i>	320
<i>Boosting</i>	321
<i>Additive regression</i>	325
<i>Additive logistic regression</i>	327
<i>Option trees</i>	328
<i>Logistic model trees</i>	331
<i>Stacking</i>	332
<i>Error-correcting output codes</i>	334
7.6 Using unlabeled data	337
<i>Clustering for classification</i>	337
<i>Co-training</i>	339
<i>EM and co-training</i>	340
7.7 Further reading	341

8 Moving on: Extensions and applications 345

- 8.1 Learning from massive datasets 346
- 8.2 Incorporating domain knowledge 349
- 8.3 Text and Web mining 351
- 8.4 Adversarial situations 356
- 8.5 Ubiquitous data mining 358
- 8.6 Further reading 361

Part II The Weka machine learning workbench 363

9 Introduction to Weka 365

- 9.1 What's in Weka? 366
- 9.2 How do you use it? 367
- 9.3 What else can you do? 368
- 9.4 How do you get it? 368

10 The Explorer 369

- 10.1 Getting started 369
 - Preparing the data* 370
 - Loading the data into the Explorer* 370
 - Building a decision tree* 373
 - Examining the output* 373
 - Doing it again* 377
 - Working with models* 377
 - When things go wrong* 378
- 10.2 Exploring the Explorer 380
 - Loading and filtering files* 380
 - Training and testing learning schemes* 384
 - Do it yourself: The User Classifier* 388
 - Using a metalearner* 389
 - Clustering and association rules* 391
 - Attribute selection* 392
 - Visualization* 393
- 10.3 Filtering algorithms 393
 - Unsupervised attribute filters* 395
 - Unsupervised instance filters* 400
 - Supervised filters* 401

- 10.4 Learning algorithms 403**
 - Bayesian classifiers* 403
 - Trees* 406
 - Rules* 408
 - Functions* 409
 - Lazy classifiers* 413
 - Miscellaneous classifiers* 414
- 10.5 Metalearning algorithms 414**
 - Bagging and randomization* 414
 - Boosting* 416
 - Combining classifiers* 417
 - Cost-sensitive learning* 417
 - Optimizing performance* 417
 - Retargeting classifiers for different tasks* 418
- 10.6 Clustering algorithms 418**
- 10.7 Association-rule learners 419**
- 10.8 Attribute selection 420**
 - Attribute subset evaluators* 422
 - Single-attribute evaluators* 422
 - Search methods* 423

- 11 The Knowledge Flow interface 427**
 - 11.1 Getting started 427**
 - 11.2 The Knowledge Flow components 430**
 - 11.3 Configuring and connecting the components 431**
 - 11.4 Incremental learning 433**

- 12 The Experimenter 437**
 - 12.1 Getting started 438**
 - Running an experiment* 439
 - Analyzing the results* 440
 - 12.2 Simple setup 441**
 - 12.3 Advanced setup 442**
 - 12.4 The Analyze panel 443**
 - 12.5 Distributing processing over several machines 445**

13	The command-line interface	449
13.1	Getting started	449
13.2	The structure of Weka	450
	<i>Classes, instances, and packages</i>	450
	<i>The weka.core package</i>	451
	<i>The weka.classifiers package</i>	453
	<i>Other packages</i>	455
	<i>Javadoc indices</i>	456
13.3	Command-line options	456
	<i>Generic options</i>	456
	<i>Scheme-specific options</i>	458
14	Embedded machine learning	461
14.1	A simple data mining application	461
14.2	Going through the code	462
	<i>main()</i>	462
	<i>MessageClassifier()</i>	462
	<i>updateData()</i>	468
	<i>classifyMessage()</i>	468
15	Writing new learning schemes	471
15.1	An example classifier	471
	<i>buildClassifier()</i>	472
	<i>makeTree()</i>	472
	<i>computeInfoGain()</i>	480
	<i>classifyInstance()</i>	480
	<i>main()</i>	481
15.2	Conventions for implementing classifiers	483
	References	485
	Index	505
	About the authors	525