

## Contents

Preface	xi
<b>PART ONE: DATA ANALYSIS FOUNDATIONS</b>	<b>1</b>
<b>1 Data Matrix</b> . . . . .	<b>3</b>
1.1 Data Matrix	3
1.2 Attributes	4
1.3 Data: Algebraic and Geometric View	5
1.4 Data: Probabilistic View	16
1.5 Further Reading	28
1.6 Exercises	28
<b>2 Numeric Attributes</b> . . . . .	<b>29</b>
2.1 Univariate Analysis	29
2.2 Bivariate Analysis	40
2.3 Multivariate Analysis	46
2.4 Data Normalization	50
2.5 Normal Distribution	52
2.6 Further Reading	58
2.7 Exercises	58
<b>3 Categorical Attributes</b> . . . . .	<b>61</b>
3.1 Univariate Analysis	61
3.2 Bivariate Analysis	70
3.3 Multivariate Analysis	81
3.4 Distance and Angle	86
3.5 Discretization	87
3.6 Further Reading	89
3.7 Exercises	90
<b>4 Graph Data</b> . . . . .	<b>92</b>
4.1 Graph Concepts	92
4.2 Topological Attributes	96
4.3 Centrality Analysis	101
	v

vi	Contents
4.4	Graph Models 111
4.5	Further Reading 131
4.6	Exercises 132
<b>5</b>	<b>Kernel Methods . . . . . 134</b>
5.1	Kernel Matrix 138
5.2	Vector Kernels 144
5.3	Basic Kernel Operations in Feature Space 149
5.4	Kernels for Complex Objects 155
5.5	Further Reading 161
5.6	Exercises 161
<b>6</b>	<b>High-dimensional Data . . . . . 163</b>
6.1	High-dimensional Objects 163
6.2	High-dimensional Volumes 167
6.3	Hypersphere Inscribed within Hypercube 170
6.4	Volume of Thin Hypersphere Shell 171
6.5	Diagonals in Hyperspace 172
6.6	Density of the Multivariate Normal 173
6.7	Appendix: Derivation of Hypersphere Volume 177
6.8	Further Reading 181
6.9	Exercises 181
<b>7</b>	<b>Dimensionality Reduction . . . . . 184</b>
7.1	Background 184
7.2	Principal Component Analysis 188
7.3	Kernel Principal Component Analysis 203
7.4	Singular Value Decomposition 210
7.5	Further Reading 215
7.6	Exercises 215
	<b>PART TWO: FREQUENT PATTERN MINING 217</b>
<b>8</b>	<b>Itemset Mining . . . . . 219</b>
8.1	Frequent Itemsets and Association Rules 219
8.2	Itemset Mining Algorithms 223
8.3	Generating Association Rules 237
8.4	Further Reading 238
8.5	Exercises 239
<b>9</b>	<b>Summarizing Itemsets . . . . . 244</b>
9.1	Maximal and Closed Frequent Itemsets 244
9.2	Mining Maximal Frequent Itemsets: GenMax Algorithm 247
9.3	Mining Closed Frequent Itemsets: Charm Algorithm 250
9.4	Nonderivable Itemsets 252
9.5	Further Reading 258
9.6	Exercises 258

Contents	vii
<b>10 Sequence Mining</b> . . . . .	261
10.1 Frequent Sequences	261
10.2 Mining Frequent Sequences	262
10.3 Substring Mining via Suffix Trees	269
10.4 Further Reading	279
10.5 Exercises	279
<b>11 Graph Pattern Mining</b> . . . . .	282
11.1 Isomorphism and Support	282
11.2 Candidate Generation	286
11.3 The gSpan Algorithm	290
11.4 Further Reading	298
11.5 Exercises	299
<b>12 Pattern and Rule Assessment</b> . . . . .	303
12.1 Rule and Pattern Assessment Measures	303
12.2 Significance Testing and Confidence Intervals	318
12.3 Further Reading	330
12.4 Exercises	330
<b>PART THREE: CLUSTERING</b>	332
<b>13 Representative-based Clustering</b> . . . . .	334
13.1 K-means Algorithm	334
13.2 Kernel K-means	339
13.3 Expectation-Maximization Clustering	343
13.4 Further Reading	360
13.5 Exercises	361
<b>14 Hierarchical Clustering</b> . . . . .	364
14.1 Preliminaries	364
14.2 Agglomerative Hierarchical Clustering	366
14.3 Further Reading	372
14.4 Exercises	373
<b>15 Density-based Clustering</b> . . . . .	375
15.1 The DBSCAN Algorithm	375
15.2 Kernel Density Estimation	379
15.3 Density-based Clustering: DENCLUE	385
15.4 Further Reading	390
15.5 Exercises	391
<b>16 Spectral and Graph Clustering</b> . . . . .	394
16.1 Graphs and Matrices	394
16.2 Clustering as Graph Cuts	401
16.3 Markov Clustering	417
16.4 Further Reading	422
16.5 Exercises	424

viii	Contents
<b>17 Clustering Validation</b> . . . . .	426
17.1 External Measures	426
17.2 Internal Measures	441
17.3 Relative Measures	450
17.4 Further Reading	464
17.5 Exercises	465
<b>PART FOUR: CLASSIFICATION</b>	467
<b>18 Probabilistic Classification</b> . . . . .	469
18.1 Bayes Classifier	469
18.2 Naive Bayes Classifier	475
18.3 $K$ Nearest Neighbors Classifier	479
18.4 Further Reading	480
18.5 Exercises	482
<b>19 Decision Tree Classifier</b> . . . . .	483
19.1 Decision Trees	485
19.2 Decision Tree Algorithm	487
19.3 Further Reading	498
19.4 Exercises	499
<b>20 Linear Discriminant Analysis</b> . . . . .	501
20.1 Optimal Linear Discriminant	501
20.2 Kernel Discriminant Analysis	508
20.3 Further Reading	515
20.4 Exercises	515
<b>21 Support Vector Machines</b> . . . . .	517
21.1 Support Vectors and Margins	517
21.2 SVM: Linear and Separable Case	523
21.3 Soft Margin SVM: Linear and Nonseparable Case	527
21.4 Kernel SVM: Nonlinear Case	533
21.5 SVM Training: Stochastic Gradient Ascent	537
21.6 Further Reading	543
21.7 Exercises	544
<b>22 Classification Assessment</b> . . . . .	546
22.1 Classification Performance Measures	546
22.2 Classifier Evaluation	560
22.3 Bias-Variance Decomposition	570
22.4 Ensemble Classifiers	574
22.5 Further Reading	584
22.6 Exercises	585
<b>PART FIVE: REGRESSION</b>	587
<b>23 Linear Regression</b> . . . . .	589
23.1 Linear Regression Model	589

<b>Contents</b>	ix
23.2 Bivariate Regression	590
23.3 Multiple Regression	596
23.4 Ridge Regression	606
23.5 Kernel Regression	611
23.6 $L_1$ Regression: Lasso	615
23.7 Further Reading	621
23.8 Exercises	621
<b>24 Logistic Regression</b> . . . . .	<b>623</b>
24.1 Binary Logistic Regression	623
24.2 Multiclass Logistic Regression	630
24.3 Further Reading	635
24.4 Exercises	635
<b>25 Neural Networks</b> . . . . .	<b>637</b>
25.1 Artificial Neuron: Activation Functions	637
25.2 Neural Networks: Regression and Classification	642
25.3 Multilayer Perceptron: One Hidden Layer	648
25.4 Deep Multilayer Perceptrons	660
25.5 Further Reading	670
25.6 Exercises	670
<b>26 Deep Learning</b> . . . . .	<b>672</b>
26.1 Recurrent Neural Networks	672
26.2 Gated RNNs: Long Short-Term Memory Networks	682
26.3 Convolutional Neural Networks	694
26.4 Regularization	712
26.5 Further Reading	717
26.6 Exercises	718
<b>27 Regression Evaluation</b> . . . . .	<b>720</b>
27.1 Univariate Regression	721
27.2 Multiple Regression	735
27.3 Further Reading	752
27.4 Exercises	752
Index	755