

Contents

Acknowledgments	xiii
CHAPTER 1 Introduction, overview, and applications	1
1.1 Introduction	1
1.2 Why is this book important?	2
1.3 Organization of the book	3
1.4 Informatics	4
1.5 Statistics for analytics	5
1.5.1 Value and variance	5
1.5.2 Sample and population tests	6
1.5.3 Regression and estimation	10
1.6 Algorithms for analytics	16
1.6.1 k -Means and k -nearest neighbor clustering	16
1.6.2 Unclustering	19
1.6.3 Markov models	20
1.7 Machine learning	22
1.7.1 Entropy	23
1.7.2 SVM and kernels	24
1.7.3 Probability	25
1.7.4 Dimensionality reduction and information gain	27
1.7.5 Optimization and search	28
1.7.6 Data mining and knowledge discovery	30
1.7.7 Recognition	31
1.7.8 Ensemble learning	33
1.8 Artificial intelligence	36
1.8.1 Genetic algorithms	36
1.8.2 Neural networks	42
1.8.3 Immunological algorithms	49
1.9 A platform for building a classifier from the ground up (binary case)	52
1.10 A platform for building a classifier from the ground up (general case)	61
1.10.1 Training and validation	61
1.10.2 Testing and deployment	71
1.10.3 Comparing training and testing data set results	95

1.11	Summary	96
	References.....	98
	Further reading	98
CHAPTER 2	Ground truthing	99
2.1	Introduction	99
2.2	Pre-validation	100
2.3	Optimizing settings from training data	108
2.4	Learning how to Learn	114
2.5	Deep learning to deep unlearning	122
2.6	Summary	123
	References.....	124
CHAPTER 3	Experimental design	125
3.1	Introduction	125
3.2	Data normalization.....	126
3.2.1	Simple (unambiguous) normalization	126
3.2.2	Bias normalization.....	127
3.2.3	Normalization and experimental design tables	132
3.3	Designs for the pruning of aging data.....	133
3.4	Systems of systems	135
3.4.1	Systems	135
3.4.2	Hybrid systems	136
3.4.3	Dynamically-updated systems	138
3.4.4	Interfaces.....	138
3.4.5	Gain.....	139
3.4.6	Domain normalization	142
3.4.7	Sensitivity analysis	143
3.5	Summary	144
	References.....	144
CHAPTER 4	Meta-analytic design patterns	147
4.1	Introduction	147
4.2	Cumulative response patterns	148
4.2.1	Identifying zones of interest.....	149
4.2.2	Zones of interest for sequence-dependent predictive selection	152
4.2.3	Traditional cumulative gain curves, or lift curves.....	154
4.3	Optimization of analytics.....	161
4.3.1	Decision trees	161
4.3.2	Putative-identity triggered patterns	162
4.3.3	Expectation-maximization and maximum-minimum patterns.....	164

4.4	Model agreement patterns	168
4.4.1	Hybrid regression.....	169
4.4.2	Modeling and model fitting.....	170
4.5	Co-occurrence and similarity patterns.....	171
4.6	Sensitivity analysis patterns.....	173
4.7	Confusion matrix patterns.....	173
4.8	Entropy patterns	175
4.9	Independence pattern	178
4.10	Functional NLP patterns (macro-feedback)	182
4.11	Summary	183
	References.....	184
CHAPTER 5	Sensitivity analysis and big system engineering	187
5.1	Introduction.....	187
5.2	Sensitivity analysis of the data set itself	190
5.3	Sensitivity analysis of the solution model	194
5.4	Sensitivity analysis of the individual algorithms	194
5.5	Sensitivity analysis of the hybrid algorithmics.....	196
5.6	Sensitivity analysis of the path to the current state	198
5.7	Summary	200
	References.....	201
CHAPTER 6	Multipatch predictive selection	203
6.1	Introduction.....	203
6.2	Predictive selection	204
6.3	Means of predicting	204
6.4	Means of selecting	206
6.5	Multi-path approach.....	212
6.6	Applications	214
6.7	Sensitivity analysis	215
6.8	Summary	215
	Reference	216
CHAPTER 7	Modeling and model fitting	217
7.1	Introduction.....	217
7.2	Chemistry analogues for analytics	218
7.3	Organic chemistry analogues for analytics	220
7.4	Immunological and biological analogues for analytics	222
7.5	Anonymization analogues for model design and fitting.....	224
7.6	LSE, error variance, and entropy: Goodness of fit.....	225

7.7	Make mine multiple models!	226
7.8	Summary	227
	References	228
CHAPTER 8	Synonym-antonym and reinforce-void patterns	229
8.1	Introduction	229
8.2	Synonym-antonym patterns	230
8.3	Reinforce-void patterns	231
8.4	Broader applicability of these patterns	235
8.5	Summary	236
	References	236
	Further reading	236
CHAPTER 9	Analytics around analytics	237
9.1	Introduction	237
9.2	Analytics around analytics	238
9.2.1	Entropy and occurrence vectors	238
9.2.2	Functional metrics	242
9.2.3	E-M (expectation-maximization) approaches	244
9.2.4	System design concerns	245
9.3	Optimizing settings from training data	246
9.4	Hybrid methods	247
9.5	Other areas for investigation around the analytics	249
9.6	Summary	249
	References	250
	Further reading	251
CHAPTER 10	System design optimization	253
10.1	Introduction	253
10.1.1	System considerations—Revisiting the system gains	254
10.1.2	System gains—Revisiting and expanding the system biases	256
10.1.3	Nothing ventured, nothing gained	260
10.2	Module optimization	260
10.3	Clustering and regularization	261
10.3.1	Sum of squares regularization	265
10.3.2	Variance regularization	266
10.3.3	Cluster size regularization	267
10.3.4	Small cluster regularization	268
10.3.5	Number of clusters regularization	268
10.3.6	Discussion of regularization methods	269

10.4	Analytic system optimization	270
10.5	Summary	271
	References.....	271
CHAPTER 11	Aleatory and expert system techniques	273
11.1	Introduction.....	273
11.2	Revisiting two earlier aleatory patterns.....	275
11.2.1	Sequential removal of features aleatory pattern	275
11.2.2	Sequential variation of feature output aleatory pattern	278
11.3	Adding random elements for testing	280
11.4	Hyperspectral aleatory approaches.....	282
11.5	Other aleatory applications in machine and statistical learning.....	283
11.6	Expert system techniques	284
11.7	Summary	284
	References.....	285
	Further reading	285
CHAPTER 12	Application I: Topics and challenges in machine translation, robotics, and biological sciences.....	287
12.1	Introduction.....	287
12.2	Machine translation.....	288
12.3	Robotics.....	291
12.4	Biological sciences	295
12.5	Summary	297
	References.....	298
CHAPTER 13	Application II: Medical and health-care informatics, economics, business, and finance	299
13.1	Introduction.....	299
13.2	Healthcare	300
13.3	Economics	302
13.4	Business and finance.....	305
13.5	Summary	308
13.6	Postscript: Psychology	308
	References.....	310
CHAPTER 14	Discussion, conclusions, and the future of data	311
14.1	Chapter 1	311
14.2	Chapter 2.....	312
14.3	Chapter 3.....	313

14.4	Chapter 4.....	313
14.5	Chapter 5.....	314
14.6	Chapter 6.....	315
14.7	Chapter 7.....	315
14.8	Chapter 8.....	316
14.9	Chapter 9.....	316
14.10	Chapter 10.....	317
14.11	Chapter 11.....	318
14.12	Chapter 12.....	318
14.13	Chapter 13.....	319
14.14	The future of meta-analytics.....	319
Index	321