Contents

Preface xi

About the Authors xiii

1 Introduction to Artificial Intelligence 1
1.1 History of Artificial Intelligence . . . 2
1.1.1 What Is Artificial Intelligence? . . . 2
1.1.2 Emergence of AI . . . 4
1.1.3 Cognitive Science and AI . . . 4
1.1.4 Logical Approach to AI . . . 4
1.1.5 Knowledge-Based Systems . . . 5
1.1.6 Probabilistic Approach to AI . . . 6
1.1.7 Evolutionary Computation and Swarm Intelligence . . . 6
1.1.8 Neural Networks & Deep Learning . . . 7
1.1.9 A Return to Creating HAL . . . 7
1.2 Outline of This Book . . . 7

I Logical Intelligence 9

2 Propositional Logic 11
2.1 Basics of Propositional Logic . . . 12
2.1.1 Syntax . . . 12
2.1.2 Semantics . . . 13
2.1.3 Tautologies and Logical Implication . . . 17
2.1.4 Logical Arguments . . . 18
2.1.5 Derivation Systems . . . 21
2.2 Resolution . . . 24
2.2.1 Normal Forms . . . 25
2.2.2 Derivations Using Resolution . . . 26
2.2.3 Resolution Algorithm . . . 30
2.3 Artificial Intelligence Applications . . . 30
2.3.1 Knowledge-Based Systems . . . 30
2.3.2 Wumpus World . . . 41
2.4 Discussion and Further Reading . . . 48

3 First-Order Logic 53
3.1 Basics of First-Order Logic . . . 53
3.1.1 Syntax . . . 54
3.1.2 Semantics . . . 56

3.1.3 Validity and Logical Implication . . . 60
3.1.4 Derivation Systems . . . 62
3.1.5 Modus Ponens for First-Order Logic . . . 65
3.2 Artificial Intelligence Applications . . . 68
3.2.1 Wumpus World Revisited . . . 69
3.2.2 Planning . . . 69
3.3 Discussion and Further Reading . . . 73

4 Certain Knowledge Representation **77**
4.1 Taxonomic Knowledge . . . 78
4.1.1 Semantic Nets . . . 78
4.1.2 Model of Human Organization of Knowledge . . . 79
4.2 Frames . . . 80
4.2.1 Frame Data Structure . . . 80
4.2.2 Planning a Trip Using Frames . . . 81
4.3 Nonmonotonic Logic . . . 84
4.3.1 Circumscription . . . 84
4.3.2 Default Logic . . . 85
4.3.3 Difficulties . . . 86
4.4 Discussion and Further Reading . . . 86

5 Learning Deterministic Models **89**
5.1 Supervised Learning . . . 89
5.2 Regression . . . 90
5.2.1 Simple Linear Regression . . . 91
5.2.2 Multiple Linear Regression . . . 93
5.2.3 Overfitting and Cross Validation . . . 94
5.3 Parameter Estimation . . . 96
5.3.1 Estimating the Parameters for Simple Linear Regression . . . 96
5.3.2 Gradient Descent . . . 98
5.3.3 Logistic Regression and Gradient Descent . . . 100
5.3.4 Stochastic Gradient Descent . . . 101
5.4 Learning a Decision Tree . . . 102
5.4.1 Information Theory . . . 102
5.4.2 Information Gain and the ID3 Algorithm . . . 106
5.4.3 Overfitting . . . 108

II Probabilistic Intelligence **113**

6 Probability **115**
6.1 Probability Basics . . . 117
6.1.1 Probability Spaces . . . 117
6.1.2 Conditional Probability and Independence . . . 120
6.1.3 Bayes' Theorem . . . 122
6.2 Random Variables . . . 123
6.2.1 Probability Distributions of Random Variables . . . 123
6.2.2 Independence of Random Variables . . . 128
6.3 Meaning of Probability . . . 131
6.3.1 Relative Frequency Approach to Probability . . . 132
6.3.2 Subjective Approach to Probability . . . 134

6.4 Random Variables in Applications . . . 135
6.5 Probability in the Wumpus World . . . 139

7 Uncertain Knowledge Representation 145
7.1 Intuitive Introduction to Bayesian Networks . . . 147
7.2 Properties of Bayesian Networks . . . 149
7.2.1 Definition of a Bayesian Network . . . 149
7.2.2 Representation of a Bayesian Network . . . 152
7.3 Causal Networks as Bayesian Networks . . . 154
7.3.1 Causality . . . 154
7.3.2 Causality and the Markov Condition . . . 155
7.3.3 Markov Condition without Causality . . . 159
7.4 Inference in Bayesian Networks . . . 160
7.4.1 Examples of Inference . . . 160
7.4.2 Inference Algorithms and Packages . . . 162
7.4.3 Inference Using Netica . . . 163
7.5 Networks with Continuous Variables . . . 165
7.5.1 Gaussian Bayesian Networks . . . 165
7.5.2 Hybrid Networks . . . 168
7.6 Obtaining the Probabilities . . . 170
7.6.1 Difficulty Inherent in Multiple Parents . . . 170
7.6.2 Basic Noisy OR-Gate Model . . . 170
7.6.3 Leaky Noisy OR-Gate Model . . . 172
7.6.4 Further Models . . . 174
7.7 Large-Scale Application: Promedas . . . 174

8 Advanced Properties of Bayesian Networks 181
8.1 Entailed Conditional Independencies . . . 182
8.1.1 Examples of Entailed Conditional Independencies . . . 182
8.1.2 d-Separation . . . 185
8.2 Faithfulness . . . 188
8.2.1 Unfaithful Probability Distributions . . . 188
8.2.2 Faithfulness Condition . . . 190
8.3 Markov Equivalence . . . 191
8.4 Markov Blankets and Boundaries . . . 192

9 Decision Analysis 201
9.1 Decision Trees . . . 202
9.1.1 Simple Examples . . . 202
9.1.2 Solving More Complex Decision Trees . . . 205
9.2 Influence Diagrams . . . 216
9.2.1 Representing Decision Problems with Influence Diagrams . . . 216
9.2.2 Solving Influence Diagrams . . . 222
9.2.3 Techniques for Solving Influence Diagrams . . . 222
9.2.4 Solving Influence Diagrams Using Netica . . . 226
9.3 Modeling Risk Preferences . . . 231
9.3.1 Exponential Utility Function . . . 231
9.3.2 Assessing r . . . 232
9.4 Analyzing Risk Directly . . . 233
9.4.1 Using the Variance to Measure Risk . . . 233
9.4.2 Risk Profiles . . . 235

9.4.3 Dominance . . . 236
9.5 Good Decision versus Good Outcome . . . 239
9.6 Sensitivity Analysis . . . 239
9.7 Value of Information . . . 241
9.7.1 Expected Value of Perfect Information . . . 242
9.7.2 Expected Value of Imperfect Information . . . 244
9.8 Discussion and Further Reading . . . 245
9.8.1 Academics . . . 246
9.8.2 Business and Finance . . . 247
9.8.3 Capital Equipment . . . 247
9.8.4 Computer Games . . . 247
9.8.5 Computer Vision . . . 247
9.8.6 Computer Software . . . 247
9.8.7 Medicine . . . 248
9.8.8 Natural Language Processing . . . 248
9.8.9 Planning . . . 248
9.8.10 Psychology . . . 248
9.8.11 Reliability Analysis . . . 248
9.8.12 Scheduling . . . 249
9.8.13 Speech Recognition . . . 249
9.8.14 Vehicle Control and Malfunction Diagnosis . . . 249

10 Learning Probabilistic Model Parameters **257**
10.1 Learning a Single Parameter . . . 257
10.1.1 Binomial Random Variables . . . 258
10.1.2 Multinomial Random Variables . . . 260
10.2 Learning Parameters in a Bayesian Network . . . 261
10.2.1 Procedure for Learning Parameters . . . 262
10.2.2 Equivalent Sample Size . . . 263
10.3 Learning Parameters with Missing Data★ . . . 266

11 Learning Probabilistic Model Structure **275**
11.1 Structure Learning Problem . . . 276
11.2 Score-Based Structure Learning . . . 276
11.2.1 Bayesian Score . . . 276
11.2.2 BIC Score . . . 283
11.2.3 Consistent Scoring Criteria . . . 284
11.2.4 How Many DAGs Must We Score? . . . 285
11.2.5 Using the Learned Network to Do Inference . . . 285
11.2.6 Learning Structure with Missing Data★ . . . 286
11.2.7 Approximate Structure Learning . . . 293
11.2.8 Model Averaging . . . 297
11.2.9 Approximate Model Averaging★ . . . 300
11.3 Constraint-Based Structure Learning . . . 303
11.3.1 Learning a DAG Faithful to P . . . 303
11.3.2 Learning a DAG in which P Is Embedded Faithfully . . . 307
11.4 Application: MENTOR . . . 308
11.4.1 Developing the Network . . . 308
11.4.2 Validating MENTOR . . . 310
11.5 Software Packages for Learning . . . 311
11.6 Causal Learning . . . 312

11.6.1 Causal Faithfulness Assumption . . . 312
11.6.2 Causal Embedded Faithfulness Assumption . . . 314
11.6.3 Application: College Student Retention Rate . . . 317
11.7 Class Probability Trees . . . 320
11.7.1 Theory of Class Probability Trees . . . 320
11.7.2 Application to Targeted Advertising . . . 322
11.8 Discussion and Further Reading . . . 325
11.8.1 Biology . . . 325
11.8.2 Business and Finance . . . 326
11.8.3 Causal Learning . . . 326
11.8.4 Data Mining . . . 326
11.8.5 Medicine . . . 326
11.8.6 Weather Forecasting . . . 326

12 Unsupervised Learning and Reinforcement Learning 331
12.1 Unsupervised Learning . . . 331
12.1.1 Clustering . . . 331
12.1.2 Automated Discovery . . . 333
12.2 Reinforcement Learning . . . 333
12.2.1 Multi-Armed Bandit Algorithms . . . 333
12.2.2 Dynamic Networks★ . . . 336
12.3 Discussion and Further Reading . . . 345

III Emergent Intelligence 349

13 Evolutionary Computation 351
13.1 Genetics Review . . . 352
13.2 Genetic Algorithms . . . 354
13.2.1 Algorithm . . . 354
13.2.2 Illustrative Example . . . 355
13.2.3 Traveling Salesperson Problem . . . 357
13.3 Genetic Programming . . . 364
13.3.1 Illustrative Example . . . 365
13.3.2 Artificial Ant . . . 367
13.3.3 Application to Financial Trading . . . 370
13.4 Discussion and Further Reading . . . 373

14 Swarm Intelligence 377
14.1 Ant System . . . 377
14.1.1 Real Ant Colonies . . . 378
14.1.2 Artificial Ants for Solving the TSP . . . 378
14.2 Flocks . . . 381
14.3 Discussion and Further Reading . . . 383

IV Neural Intelligence 387

15 Neural Networks and Deep Learning 389
15.1 The Perceptron . . . 389
15.1.1 Learning the Weights for a Perceptron . . . 391
15.1.2 The Perceptron and Logistic Regression . . . 394

15.2 Feedforward Neural Networks . . . 395
15.2.1 Modeling XOR . . . 395
15.2.2 Example with Two Hidden Layers . . . 398
15.2.3 Structure of a Feedforward Neural Network . . . 401
15.3 Activation Functions . . . 403
15.3.1 Output Nodes . . . 403
15.3.2 Hidden Nodes . . . 405
15.4 Application to Image Recognition . . . 407
15.5 Discussion and Further Reading . . . 407

V Language Understanding **413**

16 Natural Language Understanding **415**
16.1 Parsing . . . 417
16.1.1 Recursive Parser . . . 418
16.1.2 Ambiguity . . . 420
16.1.3 Dynamic Programming Parser . . . 422
16.1.4 Probabilistic Parser . . . 426
16.1.5 Obtaining Probabilities for a PCFG . . . 428
16.1.6 Lexicalized PCFG . . . 428
16.2 Semantic Interpretation . . . 430
16.3 Concept/Knowledge Interpretation . . . 431
16.4 Information Extraction . . . 432
16.4.1 Applications of Information Extraction . . . 432
16.4.2 Architecture for an Information Extraction System . . . 433
16.5 Discussion and Further Reading . . . 435

References **437**

Index **459**