

Contents

Preface xix

Acknowledgments xxi

1 *Introduction* 1

1.1 Decision Making 1

1.2 Applications 2

1.3 Methods 5

1.4 History 7

1.5 Societal Impact 12

1.6 Overview 14

PART I PROBABILISTIC REASONING

2 *Representation* 19

2.1 Degrees of Belief and Probability 19

2.2 Probability Distributions 20

2.3 Joint Distributions 24

2.4 Conditional Distributions 29

2.5 Bayesian Networks 32

2.6 Conditional Independence 35

2.7 Summary 36

2.8 Exercises 38

3	<i>Inference</i>	43	
3.1	Inference in Bayesian Networks	43	
3.2	Inference in Naive Bayes Models	48	
3.3	Sum-Product Variable Elimination	49	
3.4	Belief Propagation	53	
3.5	Computational Complexity	53	
3.6	Direct Sampling	54	
3.7	Likelihood Weighted Sampling	57	
3.8	Gibbs Sampling	60	
3.9	Inference in Gaussian Models	63	
3.10	Summary	65	
3.11	Exercises	66	
4	<i>Parameter Learning</i>	71	
4.1	Maximum Likelihood Parameter Learning	71	
4.2	Bayesian Parameter Learning	75	
4.3	Nonparametric Learning	82	
4.4	Learning with Missing Data	82	
4.5	Summary	89	
4.6	Exercises	89	
5	<i>Structure Learning</i>	97	
5.1	Bayesian Network Scoring	97	
5.2	Directed Graph Search	99	
5.3	Markov Equivalence Classes	103	
5.4	Partially Directed Graph Search	104	
5.5	Summary	106	
5.6	Exercises	107	
6	<i>Simple Decisions</i>	111	
6.1	Constraints on Rational Preferences	111	
6.2	Utility Functions	112	
6.3	Utility Elicitation	114	
6.4	Maximum Expected Utility Principle	116	
6.5	Decision Networks	116	
6.6	Value of Information	119	
6.7	Irrationality	122	
6.8	Summary	125	
6.9	Exercises	125	

PART II SEQUENTIAL PROBLEMS

7	<i>Exact Solution Methods</i>	133
7.1	Markov Decision Processes	133
7.2	Policy Evaluation	136
7.3	Value Function Policies	139
7.4	Policy Iteration	140
7.5	Value Iteration	141
7.6	Asynchronous Value Iteration	145
7.7	Linear Program Formulation	147
7.8	Linear Systems with Quadratic Reward	147
7.9	Summary	150
7.10	Exercises	151
8	<i>Approximate Value Functions</i>	161
8.1	Parametric Representations	161
8.2	Nearest Neighbor	163
8.3	Kernel Smoothing	164
8.4	Linear Interpolation	167
8.5	Simplex Interpolation	168
8.6	Linear Regression	172
8.7	Neural Network Regression	174
8.8	Summary	175
8.9	Exercises	177
9	<i>Online Planning</i>	181
9.1	Receding Horizon Planning	181
9.2	Lookahead with Rollouts	183
9.3	Forward Search	183
9.4	Branch and Bound	185
9.5	Sparse Sampling	187
9.6	Monte Carlo Tree Search	187
9.7	Heuristic Search	197
9.8	Labeled Heuristic Search	197
9.9	Open-Loop Planning	200
9.10	Summary	208
9.11	Exercises	209

10	<i>Policy Search</i>	213
10.1	Approximate Policy Evaluation	213
10.2	Local Search	215
10.3	Genetic Algorithms	215
10.4	Cross Entropy Method	218
10.5	Evolution Strategies	219
10.6	Isotropic Evolutionary Strategies	224
10.7	Summary	226
10.8	Exercises	226
11	<i>Policy Gradient Estimation</i>	231
11.1	Finite Difference	231
11.2	Regression Gradient	234
11.3	Likelihood Ratio	234
11.4	Reward-to-Go	237
11.5	Baseline Subtraction	241
11.6	Summary	245
11.7	Exercises	246
12	<i>Policy Gradient Optimization</i>	249
12.1	Gradient Ascent Update	249
12.2	Restricted Gradient Update	251
12.3	Natural Gradient Update	253
12.4	Trust Region Update	254
12.5	Clamped Surrogate Objective	257
12.6	Summary	263
12.7	Exercises	264
13	<i>Actor-Critic Methods</i>	267
13.1	Actor-Critic	267
13.2	Generalized Advantage Estimation	269
13.3	Deterministic Policy Gradient	272
13.4	Actor-Critic with Monte Carlo Tree Search	274
13.5	Summary	277
13.6	Exercises	277

14	<i>Policy Validation</i>	281
14.1	Performance Metric Evaluation	281
14.2	Rare Event Simulation	285
14.3	Robustness Analysis	288
14.4	Trade Analysis	289
14.5	Adversarial Analysis	291
14.6	Summary	295
14.7	Exercises	295
PART III MODEL UNCERTAINTY		
15	<i>Exploration and Exploitation</i>	299
15.1	Bandit Problems	299
15.2	Bayesian Model Estimation	301
15.3	Undirected Exploration Strategies	301
15.4	Directed Exploration Strategies	303
15.5	Optimal Exploration Strategies	306
15.6	Exploration with Multiple States	309
15.7	Summary	309
15.8	Exercises	311
16	<i>Model-Based Methods</i>	317
16.1	Maximum Likelihood Models	317
16.2	Update Schemes	318
16.3	Exploration	321
16.4	Bayesian Methods	326
16.5	Bayes-Adaptive Markov Decision Processes	329
16.6	Posterior Sampling	330
16.7	Summary	332
16.8	Exercises	332
17	<i>Model-Free Methods</i>	335
17.1	Incremental Estimation of the Mean	335
17.2	Q-Learning	336
17.3	Sarsa	338
17.4	Eligibility Traces	341

17.5	Reward Shaping	343
17.6	Action Value Function Approximation	343
17.7	Experience Replay	345
17.8	Summary	348
17.9	Exercises	351
18	<i>Imitation Learning</i>	355
18.1	Behavioral Cloning	355
18.2	Data Set Aggregation	358
18.3	Stochastic Mixing Iterative Learning	358
18.4	Maximum Margin Inverse Reinforcement Learning	361
18.5	Maximum Entropy Inverse Reinforcement Learning	365
18.6	Generative Adversarial Imitation Learning	369
18.7	Summary	371
18.8	Exercises	372

PART IV STATE UNCERTAINTY

19	<i>Beliefs</i>	379
19.1	Belief Initialization	379
19.2	Discrete State Filter	380
19.3	Kalman Filter	383
19.4	Extended Kalman Filter	385
19.5	Unscented Kalman Filter	387
19.6	Particle Filter	390
19.7	Particle Injection	394
19.8	Summary	395
19.9	Exercises	397
20	<i>Exact Belief State Planning</i>	407
20.1	Belief-State Markov Decision Processes	407
20.2	Conditional Plans	408
20.3	Alpha Vectors	411
20.4	Pruning	412
20.5	Value Iteration	416
20.6	Linear Policies	419
20.7	Summary	419
20.8	Exercises	422

21	<i>Offline Belief State Planning</i>	427	
21.1	Fully Observable Value Approximation	427	
21.2	Fast Informed Bound	429	
21.3	Fast Lower Bounds	430	
21.4	Point-Based Value Iteration	431	
21.5	Randomized Point-Based Value Iteration	433	
21.6	Sawtooth Upper Bound	436	
21.7	Point Selection	440	
21.8	Sawtooth Heuristic Search	442	
21.9	Triangulated Value Functions	445	
21.10	Summary	447	
21.11	Exercises	448	
22	<i>Online Belief State Planning</i>	453	
22.1	Lookahead with Rollouts	453	
22.2	Forward Search	453	
22.3	Branch and Bound	456	
22.4	Sparse Sampling	456	
22.5	Monte Carlo Tree Search	457	
22.6	Determinized Sparse Tree Search	459	
22.7	Gap Heuristic Search	460	
22.8	Summary	464	
22.9	Exercises	467	
23	<i>Controller Abstractions</i>	471	
23.1	Controllers	471	
23.2	Policy Iteration	475	
23.3	Nonlinear Programming	478	
23.4	Gradient Ascent	481	
23.5	Summary	486	
23.6	Exercises	486	

PART V MULTIAGENT SYSTEMS

24	<i>Multiagent Reasoning</i>	493
24.1	Simple Games	493
24.2	Response Models	494
24.3	Dominant Strategy Equilibrium	497
24.4	Nash Equilibrium	498
24.5	Correlated Equilibrium	498
24.6	Iterated Best Response	503
24.7	Hierarchical Softmax	504
24.8	Fictitious Play	505
24.9	Gradient Ascent	509
24.10	Summary	509
24.11	Exercises	511
25	<i>Sequential Problems</i>	517
25.1	Markov Games	517
25.2	Response Models	519
25.3	Nash Equilibrium	520
25.4	Fictitious Play	521
25.5	Gradient Ascent	526
25.6	Nash Q-Learning	526
25.7	Summary	528
25.8	Exercises	530
26	<i>State Uncertainty</i>	533
26.1	Partially Observable Markov Games	533
26.2	Policy Evaluation	535
26.3	Nash Equilibrium	537
26.4	Dynamic Programming	540
26.5	Summary	542
26.6	Exercises	542

27	<i>Collaborative Agents</i>	545
27.1	Decentralized Partially Observable Markov Decision Processes	545
27.2	Subclasses	546
27.3	Dynamic Programming	549
27.4	Iterated Best Response	550
27.5	Heuristic Search	550
27.6	Nonlinear Programming	551
27.7	Summary	554
27.8	Exercises	556

APPENDICES

A	<i>Mathematical Concepts</i>	561
A.1	Measure Spaces	561
A.2	Probability Spaces	562
A.3	Metric Spaces	562
A.4	Normed Vector Spaces	562
A.5	Positive Definiteness	564
A.6	Convexity	564
A.7	Information Content	565
A.8	Entropy	566
A.9	Cross Entropy	566
A.10	Relative Entropy	567
A.11	Gradient Ascent	567
A.12	Taylor Expansion	568
A.13	Monte Carlo Estimation	569
A.14	Importance Sampling	570
A.15	Contraction Mappings	570
A.16	Graphs	572
B	<i>Probability Distributions</i>	573
C	<i>Computational Complexity</i>	575
C.1	Asymptotic Notation	575
C.2	Time Complexity Classes	577
C.3	Space Complexity Classes	577
C.4	Decidability	579

<i>D</i>	<i>Neural Representations</i>	581
D.1	Neural Networks	581
D.2	Feedforward Networks	582
D.3	Parameter Regularization	585
D.4	Convolutional Neural Networks	587
D.5	Recurrent Networks	588
D.6	Autoencoder Networks	592
D.7	Adversarial Networks	594
<i>E</i>	<i>Search Algorithms</i>	599
E.1	Search Problems	599
E.2	Search Graphs	600
E.3	Forward Search	600
E.4	Branch and Bound	601
E.5	Dynamic Programming	604
E.6	Heuristic Search	604
<i>F</i>	<i>Problems</i>	609
F.1	Hex World	609
F.2	2048	610
F.3	Cart-Pole	611
F.4	Mountain Car	612
F.5	Simple Regulator	613
F.6	Aircraft Collision Avoidance	614
F.7	Crying Baby	615
F.8	Machine Replacement	617
F.9	Catch	619
F.10	Prisoner's Dilemma	621
F.11	Rock-Paper-Scissors	621
F.12	Traveler's Dilemma	622
F.13	Predator-Prey Hex World	623
F.14	Multicaregiver Crying Baby	624
F.15	Collaborative Predator-Prey Hex World	625

<i>G Julia</i>	627
G.1 Types	627
G.2 Functions	640
G.3 Control Flow	643
G.4 Packages	645
G.5 Convenience Functions	648
<i>References</i>	651
<i>Index</i>	671