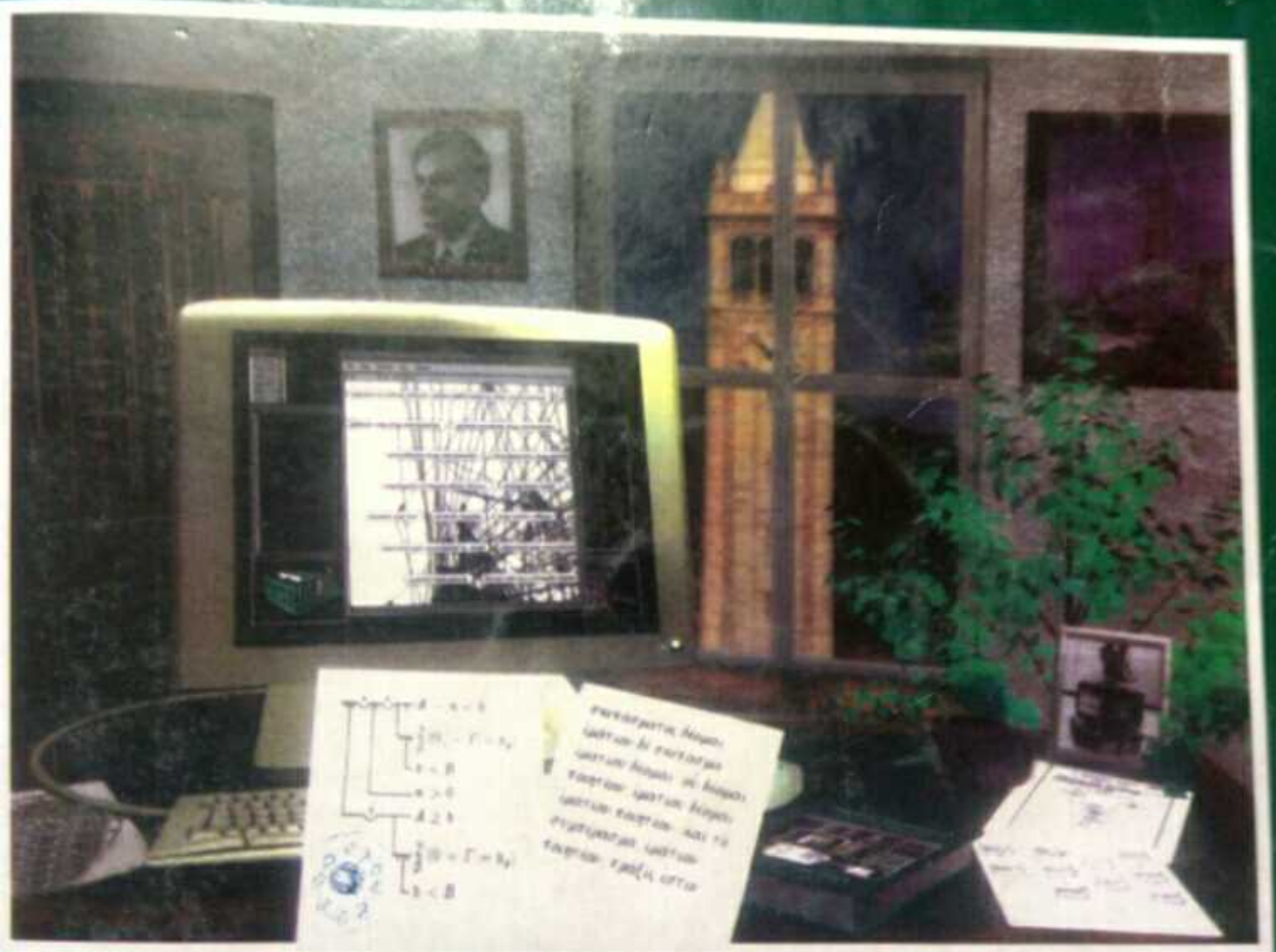


Artificial Intelligence A Modern Approach

SECOND EDITION



Stuart Russell • Peter Norving

Contents

1 Artificial Intelligence

1	Introduction	1
1.1	What is AI?	1
	Acting humanly: The Turing Test approach	2
	Thinking humanly: The cognitive modeling approach	3
	Thinking rationally: The "laws of thought" approach	4
	Acting rationally: The rational agent approach	4
1.2	The Foundations of Artificial Intelligence	5
	Philosophy (428 B.C. - present)	5
	Mathematics (c. 800 - present)	7
	Economics (1776 - present)	9
	Neuroscience (1861 - present)	10
	Psychology (1879 - present)	12
	Computer engineering (1940 - present)	14
	Control theory and Cybernetics (1948 - present)	15
	Linguistics (1957 - present)	16
1.3	The History of Artificial Intelligence	16
	The gestation of artificial intelligence (1943 - 1955)	16
	The birth of artificial intelligence (1956)	17
	Early enthusiasm, great expectations (1952 - 1969)	18
	A dose of reality (1966 - 1973)	21
	Knowledge-based systems: The key to power? (1969 - 1979)	22
	AI becomes an industry (1980 - present)	24
	The return of neural networks (1986 - present)	25
	AI becomes a science (1987 - present)	25
	The emergence of intelligent agents (1995 - present)	27
1.4	The State of the Art	27
1.5	Summary	28
	Bibliographical and Historical Notes	29
	Exercises	30

2 Intelligent Agents

2.1	Agents and Environments	32
2.2	Good Behavior: The Concept of Rationality	34
	Performance measures	35
	Rationality	35
	Omniscience, learning, and autonomy	36
2.3	The Nature of Environments	38
	Specifying the task environment	38
	Properties of task environments	40
2.4	The Structure of Agents	44
	Agent programs	44
	Simple reflex agents	46
	Model-based reflex agents	48

Goal-based agents	49
Utility-based agents	51
Learning agents	51
2.5 Summary	51
Bibliographical and Historical Notes	51
Exercises	51

II Problem-solving

3 Solving Problems by Searching	59
3.1 Problem-Solving Agents	59
Well-defined problems and solutions	61
Formulating problems	61
3.2 Example Problems	61
Toy problems	61
Real-world problems	61
3.3 Searching for Solutions	69
Measuring problem-solving performance	71
3.4 Uninformed Search Strategies	71
Breadth-first search	71
Depth-first search	78
Depth-limited search	77
Iterative deepening depth-first search	78
Bidirectional search	79
Comparing uninformed search strategies	81
3.5 Avoiding Repeated States	81
3.6 Searching with Partial Information	81
Sensorless problems	81
Contingency problems	86
3.7 Summary	87
Bibliographical and Historical Notes	88
Exercises	89

4 Informed Search and Exploration

4.1 Informed (Heuristic) Search Strategies	94
Greedy best-first search	95
A* search: Minimizing the total estimated solution cost	97
Memory-bounded heuristic search	101
Learning to search better	104
4.2 Heuristic Functions	105
The effect of heuristic accuracy on performance	106
Inventing admissible heuristic functions	107
Learning heuristics from experience	108
4.3 Local Search Algorithms and Optimization Problems	111
Hill-climbing search	115
Simulated annealing search	115
Local beam search	116
Genetic algorithms	119
4.4 Local Search in Continuous Spaces	

4.5	Online Search Agents and Unknown Environments	121
	Online search problems	121
	Online search agents	121
	Online local search	121
	Learning in online search	121
4.6	Summary	121
	Bibliographical and Historical Notes	121
	Exercises	121
5	Constraint Satisfaction Problems	131
5.1	Constraint Satisfaction Problems	131
5.2	Backtracking Search for CSPs	131
	Variable and value ordering	131
	Propagating information through constraints	131
	Intelligent backtracking: looking backward	131
5.3	Local Search for Constraint Satisfaction Problems	131
5.4	The Structure of Problems	131
5.5	Summary	131
	Bibliographical and Historical Notes	131
	Exercises	131
6	Adversarial Search	141
6.1	Games	141
6.2	Optimal Decisions in Games	141
	Optimal strategies	141
	The minimax algorithm	141
	Optimal decisions in multiplayer games	141
6.3	Alpha-Beta Pruning	141
6.4	Imperfect, Real-Time Decisions	141
	Evaluation functions	141
	Cutting off search	141
6.5	Games That Include an Element of Chance	141
	Position evaluation in games with chance nodes	141
	Complexity of expectiminimax	141
	Card games	141
6.6	State-of-the-Art Game Programs	141
6.7	Discussion	141
6.8	Summary	141
	Bibliographical and Historical Notes	141
	Exercises	141

III Knowledge and reasoning

7 Logical Agents

7.1	Knowledge-Based Agents	151
7.2	The Wumpus World	151
7.3	Logic	151
7.4	Propositional Logic: A Very Simple Logic	151
	Syntax	151

	Semantics	
	A simple knowledge base	
	Inference	
7.5	Equivalence, validity, and satisfiability	
	Reasoning Patterns in Propositional Logic	
	Resolution	
7.6	Forward and backward chaining	
	Effective propositional inference	
	A complete backtracking algorithm	
	Local search algorithms	
7.7	Hard satisfiability problems	
	Agents Based on Propositional Logic	
	Finding pits and wumpuses using logical inference	
	Keeping track of location and orientation	
	Circuit-based agents	
	A comparison	
7.8	Summary	
	Bibliographical and Historical Notes	
	Exercises	
✓ 8	First-Order Logic	
8.1	Representation Revised	
8.2	Syntax and Semantics of First-Order Logic	
	Models for first-order logic	
	Symbols and interpretations	
	Terms	
	Atomic sentences	
	Complex sentences	
	Quantifiers	
	Equality	
8.3	Using First-Order Logic	
	Assertions and queries in first-order logic	
	The kinship domain	
	Numbers, sets, and lists	
	The wumpus world	
8.4	Knowledge Engineering in First-Order Logic	
	The knowledge engineering process	
	The electronic circuits domain	
8.5	Summary	
	Bibliographical and Historical Notes	
	Exercises	
+ 9	Inference in First-Order Logic	
9.1	Propositional vs. First-Order Inference	
	Inference rules for quantifiers	
	Reduction to propositional inference	
9.2	Unification and Lifting	
	A first-order inference rule	
	Unification	

	Storage and retrieval	278	xix
9.3	Forward Chaining	280	45
	First-order definite clauses	280	49
	A simple forward-chaining algorithm	281	50
	Efficient forward chaining	283	51
9.4	Backward Chaining	287	52
	A backward chaining algorithm	287	54
	Logic programming	289	54
	Efficient implementation of logic programs	290	55
	Redundant inference and infinite loops	292	59
	Constraint logic programming	294	
9.5	Resolution	295	
	Conjunctive normal form for first-order logic	295	
	The resolution inference rule	297	
	Example proofs	297	
	Completeness of resolution	300	
	Dealing with equality	303	
	Resolution strategies	304	
	Theorem provers	306	
9.6	Summary	310	
	Bibliographical and Historical Notes	310	
	Exercises	315	
		320	
10	Knowledge Representation	320	
10.1	Ontological Engineering	322	
10.2	Categories and Objects	324	
	Physical composition	325	
	Measurements	327	
	Substances and objects	328	
10.3	Actions, Situations, and Events	329	
	The ontology of situation calculus	330	
	Describing actions in situation calculus	332	
	Solving the representational frame problem	333	
	Solving the inferential frame problem	334	
	Time and event calculus	335	
	Generalized events	337	
	Processes	338	
	Intervals	339	
	Fluents and objects	341	
10.4	Mental Events and Mental Objects	341	
	A formal theory of beliefs	343	
	Knowledge and belief	344	
	Knowledge, time, and action	344	
10.5	The Internet Shopping World	348	
	Comparing offers	349	
10.6	Reasoning Systems for Categories	350	
	Semantic networks	353	
	Description logics	354	
10.7	Reasoning with Default Information		

Open and closed worlds	354
Negation as failure and stable model semantics	356
Circumscription and default logic	358
10.8 Truth Maintenance Systems	360
10.9 Summary	362
Bibliographical and Historical Notes	363
Exercises	369

IV Planning

11 Planning	375
11.1 The Planning Problem	375
The language of planning problems	377
Expressiveness and extensions	378
Example: Air cargo transport	380
Example: The spare tire problem	381
Example: The blocks world	381
11.2 Planning with State-Space Search	382
Forward state-space search	382
Backward state-space search	384
Heuristics for state-space search	386
11.3 Partial-Order Planning	387
A partial-order planning example	391
Partial-order planning with unbound variables	393
Heuristics for partial-order planning	394
11.4 Planning Graphs	395
Planning graphs for heuristic estimation	397
The GRAPHPLAN algorithm	398
Termination of GRAPHPLAN	401
11.5 Planning with Propositional Logic	402
Describing planning problems in propositional logic	402
Complexity of propositional encodings	405
11.6 Analysis of Planning Approaches	407
11.7 Summary	408
Bibliographical and Historical Notes	409
Exercises	412
12 Planning and Acting in the Real World	417
12.1 Time, Schedules, and Resources	417
Scheduling with resource constraints	420
12.2 Hierarchical Task Network Planning	422
Representing action decompositions	423
Modifying the planner for decompositions	425
Discussion	427
12.3 Planning and Acting in Nondeterministic Domains	430
12.4 Conditional Planning	433
Conditional planning in fully observable environments	433
Conditional planning in partially observable environments	437
12.5 Execution Monitoring and Replanning	441

12.6	Continuous Planning	445
12.7	MultiAgent Planning	449
	Cooperation: Joint goals and plans	450
	Multibody planning	451
	Coordination mechanisms	452
	Competition	454
12.8	Summary	454
	Bibliographical and Historical Notes	455
	Exercises	459

V Uncertain knowledge and reasoning

13	Uncertainty	462
13.1	Acting under Uncertainty	462
	Handling uncertain knowledge	463
	Uncertainty and rational decisions	465
	Design for a decision-theoretic agent	466
13.2	Basic Probability Notation	466
	Propositions	467
	Atomic events	468
	Prior probability	468
	Conditional probability	470
13.3	The Axioms of Probability	471
	Using the axioms of probability	473
	Why the axioms of probability are reasonable	473
13.4	Inference Using Full Joint Distributions	475
13.5	Independence	477
13.6	Bayes' Rule and Its Use	479
	Applying Bayes' rule: The simple case	480
	Using Bayes' rule: Combining evidence	481
13.7	The Wumpus World Revisited	483
13.8	Summary	486
	Bibliographical and Historical Notes	487
	Exercises	489
14	Probabilistic Reasoning	492
14.1	Representing Knowledge in an Uncertain Domain	492
14.2	The Semantics of Bayesian Networks	495
	Representing the full joint distribution	499
	Conditional independence relations in Bayesian networks	500
14.3	Efficient Representation of Conditional Distributions	504
14.4	Exact Inference in Bayesian Networks	504
	Inference by enumeration	507
	The variable elimination algorithm	509
	The complexity of exact inference	510
	Clustering algorithms	511
14.5	Approximate Inference in Bayesian Networks	511
	Direct sampling methods	516
	Inference by Markov chain simulation	

14.6	Extending Probability to First-Order Representations	519
14.7	Other Approaches to Uncertain Reasoning	521
	Rule-based methods for uncertain reasoning	521
	Representing ignorance: Dempster-Shafer theory	521
	Representing vagueness: Fuzzy sets and fuzzy logic	521
14.8	Summary	521
	Bibliographical and Historical Notes	521
	Exercises	521
15	Probabilistic Reasoning over Time	521
15.1	Time and Uncertainty	521
	States and observations	521
	Stationary processes and the Markov assumption	521
15.2	Inference in Temporal Models	521
	Filtering and prediction	521
	Smoothing	521
	Finding the most likely sequence	521
15.3	Hidden Markov Models	521
	Simplified matrix algorithms	521
15.4	Kalman Filters	521
	Updating Gaussian distributions	521
	A simple one-dimensional example	521
	The general case	521
	Applicability of Kalman filtering	521
15.5	Dynamic Bayesian Networks	521
	Constructing DBNs	521
	Exact inference in DBNs	521
	Approximate inference in DBNs	521
15.6	Speech Recognition	521
	Speech sounds	521
	Words	521
	Sentences	521
	Building a speech recognizer	521
15.7	Summary	521
	Bibliographical and Historical Notes	521
	Exercises	521
16	Making Simple Decisions	521
16.1	Combining Beliefs and Desires under Uncertainty	521
16.2	The Basis of Utility Theory	521
	Constraints on rational preferences	521
	And then there was Utility	521
16.3	Utility Functions	521
	The utility of money	521
	Utility scales and utility assessment	521
16.4	Multiattribute Utility Functions	521
	Dominance	521
	Preference structure and multiattribute utility	521
16.5	Decision Networks	521

Representing a decision problem with a decision network	596
Evaluating decision networks	597
16.6 The Value of Information	600
A simple example	600
A general formula	601
Properties of the value of information	602
Implementing an information-gathering agent	603
16.7 Decision-Theoretic Expert Systems	604
16.8 Summary	607
Bibliographical and Historical Notes	607
Exercises	609
17 Making Complex Decisions	613
17.1 Sequential Decision Problems	613
An example	613
Optimality in sequential decision problems	616
17.2 Value Iteration	618
Utilities of states	618
The value iteration algorithm	619
Convergence of value iteration	620
17.3 Policy Iteration	620
17.4 Partially observable MDPs	624
17.5 Decision-Theoretic Agents	625
17.6 Decisions with Multiple Agents: Game Theory	629
17.7 Mechanism Design	634
17.8 Summary	640
Bibliographical and Historical Notes	643
Exercises	644
	646
VI Learning	
18 Learning from Observations	649
18.1 Forms of Learning	649
18.2 Inductive Learning	651
18.3 Learning Decision Trees	653
Decision trees as performance elements	653
Expressiveness of decision trees	655
Inducing decision trees from examples	655
Choosing attribute tests	659
Assessing the performance of the learning algorithm	660
Noise and overfitting	661
Broadening the applicability of decision trees	663
18.4 Ensemble Learning	664
18.5 Why Learning Works: Computational Learning Theory	668
How many examples are needed?	669
Learning decision trees	670
Discussion	672
18.6 Summary	673
Bibliographical and Historical Notes	674

19 Knowledge in Learning	674
19.1 A Logical Formulation of Learning	674
Examples and hypotheses	674
Current-best-hypothesis search	674
Least-commitment search	674
19.2 Knowledge in Learning	674
Some simple examples	674
Some general schemes	674
19.3 Explanation-Based Learning	674
Extracting general rules from examples	674
Improving efficiency	674
19.4 Learning Using Relevance Information	674
Determining the hypothesis space	674
Learning and using relevance information	674
19.5 Inductive Logic Programming	674
An example	674
Top-down inductive learning methods	700
Inductive learning with inverse deduction	700
Making discoveries with inductive logic programming	700
19.6 Summary	700
Bibliographical and Historical Notes	700
Exercises	700
20 Statistical Learning Methods	712
20.1 Statistical Learning	712
20.2 Learning with Complete Data	716
Maximum-likelihood parameter learning: Discrete models	716
Naive Bayes models	718
Maximum-likelihood parameter learning: Continuous models	719
Bayesian parameter learning	720
Learning Bayes net structures	722
20.3 Learning with Hidden Variables: The EM Algorithm	724
Unsupervised clustering: Learning mixtures of Gaussians	725
Learning Bayesian networks with hidden variables	727
Learning hidden Markov models	731
The general form of the EM algorithm	731
Learning Bayes net structures with hidden variables	733
20.4 Instance-Based Learning	733
Nearest-neighbor models	735
Kernel models	736
20.5 Neural Networks	737
Units in neural networks	738
Network structures	740
Single layer feed-forward neural networks (perceptrons)	744
Multilayer feed-forward neural networks	748
Learning neural network structures	748
20.6 Kernel Machines	748

20.7 Case Study: Handwritten Digit Recognition	752
20.8 Summary	754
Bibliographical and Historical Notes	755
Exercises	759
21 Reinforcement Learning	
21.1 Introduction	763
21.2 Passive Reinforcement Learning	763
Direct utility estimation	765
Adaptive dynamic programming	766
Temporal difference learning	767
21.3 Active Reinforcement Learning	767
Exploration	771
Learning an Action-Value Function	771
21.4 Generalization in Reinforcement Learning	775
Applications to game-playing	777
Application to robot control	780
21.5 Policy Search	780
21.6 Summary	781
Bibliographical and Historical Notes	784
Exercises	785

VII Communicating, perceiving, and acting

22 Communication	790
22.1 Communication as Action	790
Fundamentals of language	791
The component steps of communication	792
22.2 A Formal Grammar for a Fragment of English	795
The Lexicon of \mathcal{E}_0	795
The Grammar of \mathcal{E}_0	796
22.3 Syntactic Analysis (Parsing)	798
Efficient parsing	800
22.4 Augmented Grammars	806
Verb subcategorization	808
Generative capacity of augmented grammars	809
22.5 Semantic Interpretation	810
The semantics of an English fragment	811
Time and tense	812
Quantification	813
Pragmatic Interpretation	815
Language generation with DCGs	817
22.6 Ambiguity and Disambiguation	818
Disambiguation	820
22.7 Discourse Understanding	821
Reference resolution	823
The structure of coherent discourse	824
22.8 Grammar Induction	826

Bibliographical and Historical Notes	87
Exercises	88
23 Probabilistic Language Processing	89
23.1 Probabilistic Language Models	90
Probabilistic context-free grammars	90
Learning probabilities for PCFGs	91
Learning rule structure for PCFGs	92
23.2 Information Retrieval	93
Evaluating IR systems	93
IR refinements	94
Presentation of result sets	95
Implementing IR systems	96
23.3 Information Extraction	97
23.4 Machine Translation	98
Machine translation systems	98
Statistical machine translation	99
Learning probabilities for machine translation	100
23.5 Summary	101
Bibliographical and Historical Notes	102
Exercises	103
24 Perception	104
24.1 Introduction	104
24.2 Image Formation	105
Images without lenses: the pinhole camera	105
Lens systems	106
Light: the photometry of image formation	107
Color: the spectrophotometry of image formation	108
24.3 Early Image Processing Operations	109
Edge detection	109
Image segmentation	110
24.4 Extracting Three-Dimensional Information	111
Motion	111
Binocular stereopsis	112
Texture gradients	113
Shading	114
Contour	115
24.5 Object Recognition	116
Brightness-based recognition	116
Feature-based recognition	117
Pose Estimation	118
24.6 Using Vision for Manipulation and Navigation	119
24.7 Summary	120
Bibliographical and Historical Notes	121
Exercises	122
25 Robotics	123
25.1 Introduction	123

25.2	Robot Hardware	
	Sensors	903
	Effectors	903
25.3	Robotic Perception	904
	Localization	907
	Mapping	908
	Other types of perception	913
25.4	Planning to Move	915
	Configuration space	916
	Cell decomposition methods	916
	Skeletonization methods	919
		922
25.5	Planning uncertain movements	923
	Robust methods	924
25.6	Moving	926
	Dynamics and control	927
	Potential field control	929
	Reactive control	930
25.7	Robotic Software Architectures	932
	Subsumption architecture	932
	Three-layer architecture	933
	Robotic programming languages	934
25.8	Application Domains	935
25.9	Summary	938
	Bibliographical and Historical Notes	939
	Exercises	942

VIII Conclusions

26	Philosophical Foundations	947
26.1	Weak AI: Can Machines Act Intelligently?	947
	The argument from disability	948
	The mathematical objection	949
	The argument from informality	950
		952
26.2	Strong AI: Can Machines Really Think?	954
	The mind-body problem	955
	The "brain in a vat" experiment	956
	The brain prosthesis experiment	958
	The Chinese room	960
26.3	The Ethics and Risks of Developing Artificial Intelligence	964
26.4	Summary	964
	Bibliographical and Historical Notes	967
	Exercises	968
27	AI: Present and Future	968
27.1	Agent Components	970
27.2	Agent Architectures	972