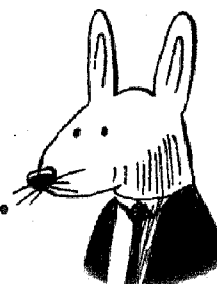


contents



foreword	xi
preface	xiii
acknowledgments	xv
about this book	xvii
about the author	xxi

1 Introduction to deep reinforcement learning **1**

.....

What is deep reinforcement learning?	2
The past, present, and future of deep reinforcement learning	15
The suitability of deep reinforcement learning	22
Setting clear two-way expectations	25

2 Mathematical foundations of reinforcement learning **31**

.....

Components of reinforcement learning	33
MDPs: The engine of the environment	45

3	Balancing immediate and long-term goals	65
		
	The objective of a decision-making agent	66
	Planning optimal sequences of actions	77
4	Balancing the gathering and use of information	97
		
	The challenge of interpreting evaluative feedback	99
	Strategic exploration	117
5	Evaluating agents' behaviors	131
		
	Learning to estimate the value of policies	133
	Learning to estimate from multiple steps	150
6	Improving agents' behaviors	167
		
	The anatomy of reinforcement learning agents	168
	Learning to improve policies of behavior	176
	Decoupling behavior from learning	187
7	Achieving goals more effectively and efficiently	203
		
	Learning to improve policies using robust targets	205
	Agents that interact, learn, and plan	217
8	Introduction to value-based deep reinforcement learning	239
		
	The kind of feedback deep reinforcement learning agents use	241
	Introduction to function approximation for reinforcement learning	248
	NFQ: The first attempt at value-based deep reinforcement learning	253

9	More stable value-based methods	275
~~~~~		
	DQN: Making reinforcement learning more like supervised learning	276
	Double DQN: Mitigating the overestimation of action-value functions	293
<b>10</b>	<b>Sample-efficient value-based methods</b>	<b>309</b>
~~~~~		
	Dueling DDQN: A reinforcement-learning-aware neural network architecture	310
	PER: Prioritizing the replay of meaningful experiences	323
11	Policy-gradient and actor-critic methods	339
~~~~~		
	REINFORCE: Outcome-based policy learning	340
	VPG: Learning a value function	350
	A3C: Parallel policy updates	355
	GAE: Robust advantage estimation	362
	A2C: Synchronous policy updates	365
<b>12</b>	<b>Advanced actor-critic methods</b>	<b>375</b>
~~~~~		
	DDPG: Approximating a deterministic policy	377
	TD3: State-of-the-art improvements over DDPG	384
	SAC: Maximizing the expected return and entropy	391
	PPO: Restricting optimization steps	398
13	Toward artificial general intelligence	411
~~~~~		
	What was covered and what notably wasn't?	413
	More advanced concepts toward AGI	424
	What happens next?	432
	index	437