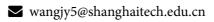
Reinforcement Learning

Notes

Jingye Wang



Spring 2020

Contents

1	Intr	oduction	3	
2	Review of Basic Probability			
	2.1	Interpretation of Probability	5	
	2.2	Transformations	5	
	2.3	Limit Theorem	5	
	2.4	Sampling & Monte Carlo Methods	6	
	2.5	Basic Inequalities	8	
	2.6	Concentration Inequalities	10	
	2.7	Conditional Expectation	12	
3	Bandit Algorithms			
	3.1	Bandit Models	14	
	3.2	Stochastic Bandits	14	
	3.3	Greedy Algorithms	15	
	3.4	UCB Algorithms	16	
	3.5	Thompson Sampling Algorithms	17	
	3.6	Gradient Bandit Algorithms	18	
4	Markov Chains			
	4.1	Markov Model	20	
	4.2	Basic Computations	20	
	4.3	Classifications	21	

CONTENTS	2

	4.4	Stationary Distribution	22		
	4.5	Reversibility	22		
	4.6	Markov chain Monte Carlo	23		
5	Mar	kov Decision Process	25		
	5.1	Markov Process	25		
	5.2	Markov Reward Process	25		
	5.3	Markov Decision Process	26		
	5.4	Dynamic Programming	28		
6	Model-free Prediction				
	6.1	Monte-Carlo Policy Evaluation	30		
	6.2	Temporal-Difference Learning	31		
7	Model-free Control				
	7.1	On Policy Monte-Carlo Control	33		
	7.2	On Policy Temporal-Difference Control	35		
	7.3	Off-Policy Q-Learning Control	36		
	7.4	Off-Policy Importance Sampling Control	36		
8	Valu	e Function Approximation	37		
	8.1	Introduction on Function Approximation	37		
	8.2	Incremental Method	38		
	8.3	Batch Methods	40		
	8.4	Deep Q-Learning	41		
9	Policy Optimization				
	9.1	Policy Optimization	42		
	9.2	Monte-Carlo Policy Gradient	43		
	9.3	Actor-Critic Policy Gradient	46		
	9.4	Extension of Policy Gradient	47		