

Reinforcement Learning: An Introduction  
学习笔记

黎雷蕾

2017 年 11 月 3 日

### 摘要

学医三年，自谓天下无不治之症；  
行医三年，始信世间无可用之方。——孙思邈

纸上得来终觉浅，绝知此事要躬行。——陆游

绳锯木断，水滴石穿。——罗大经

# 目录

<b>1</b>	<b>Multi-armed Bandits(多臂老虎机问题)</b>	<b>2</b>
1.1	A $k$ -armed Bandit Problem( $k$ 臂老虎机问题) . . . . .	2

## Chapter 1

# Multi-armed Bandits(多臂老虎机问题)

### 1.1 A $k$ -armed Bandit Problem( $k$ 臂老虎机问题)

$k$  臂老虎机问题指的是每一步可以有  $k$  个选择，每进行一个选择，就可以获得一个回馈 (reward)，算法的目的是在  $n$  步后获取最大的收益。

定义第  $t$  步的选择是  $A_t$ ，相应的回馈是  $R_t$ ，那么第  $t$  步选择  $a$  动作对应的期望回馈就可定义为：

$$q_*(a) = \mathbb{E}[R_t | A_t = a] \quad (1.1)$$