The Homework of CS285

Deep Reinforcement Learning

Harry Zhang

2025 年 8 月 9 日

目录

Chapter 1: Homework 1

1.1 Analysis

1.1.1 Part A

这个作业相当于是 slide 里条件的弱化版本，slides 里的条件是每个状态不等于专家状态的概率都为 ϵ，这里只是期望小于 ϵ。

假设如下条件成立：

$$\mathbb{E}_{p_{\pi^*(s)}}\left[\pi_\theta(a \neq \pi^*(s) \mid s)\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{p_{\pi^*(s_t)}}\left[\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right] \leq \epsilon \tag{1.1}$$

在 t 时刻，s_t 的状态分布为：

$$\begin{aligned}p_\theta(s_t) =&(1 - Pr[\cup_{t=1}^{t}(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)])p_{\pi^*}(s_t)+\\ &Pr[\cup_{t=1}^{t}(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)])p_{\text{mistake}}(s_t)\end{aligned} \tag{1.2}$$

从两边同时减去 $p_{\pi^*}(s_t)$，得到：

$$\begin{aligned}|p_\theta(s_t) - p_{\pi^*}(s_t)| &= \Pr\left[\cup_{t'=1}^{t}\left(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right)\right] \cdot |p_{\text{mistake}}(s_t) - p_{\pi^*}(s_t)|\\ &\leq 2\sum_{t=1}^{T}(\pi_\theta(a_{t'} \neq \pi^*(s_{t'}) \mid s_{t'}))\end{aligned} \tag{1.3}$$

所以：

$$\begin{aligned}\sum_{s_t}|p_\theta(s_t) - p_{\pi^*}(s_t)| &\leq 2\sum_{t=1}^{T}\sum_{s_t}p_{\pi^*}(s_t)(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t))\\ &= 2\sum_{t=1}^{T}E_{p_{\pi^*(s_t)}}[\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)]\\ &= 2T\epsilon\end{aligned} \tag{1.4}$$

得证。

1.1.2 Part B

这个作业相当于是 slide 里条件的弱化版本，slides 里的条件是每个状态不等于专家状态的概率都为 ϵ，这里只是期望小于 ϵ。

假设如下条件成立：

$$\mathbb{E}_{p_{\pi^*(s)}}\left[\pi_\theta(a \neq \pi^*(s) \mid s)\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{p_{\pi^*(s_t)}}\left[\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right] \leq \epsilon \tag{1.5}$$

在 t 时刻，s_t 的状态分布为：

$$\begin{aligned} p_\theta(s_t) =& (1 - Pr[\cup_{t=1}^{t}(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)])p_{\pi^*}(s_t) + \\ & Pr[\cup_{t=1}^{t}(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)])p_{\text{mistake}}(s_t) \end{aligned} \tag{1.6}$$

从两边同时减去 $p_{\pi^*}(s_t)$，得到：

$$\begin{aligned} |p_\theta(s_t) - p_{\pi^*}(s_t)| &= \Pr\left[\cup_{t'=1}^{t}\left(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right)\right] \cdot |p_{\text{mistake}}(s_t) - p_{\pi^*}(s_t)| \\ &\leq 2\sum_{t=1}^{T}(\pi_\theta(a_{t'} \neq \pi^*(s_{t'}) \mid s_{t'})) \end{aligned} \tag{1.7}$$

所以：

$$\begin{aligned} \sum_{s_t}|p_\theta(s_t) - p_{\pi^*}(s_t)| &\leq 2\sum_{t=1}^{T}\sum_{s_t}p_{\pi^*}(s_t)(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)) \\ &= 2\sum_{t=1}^{T}E_{p_{\pi^*(s_t)}}[\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)] \\ &= 2T\epsilon \end{aligned} \tag{1.8}$$

Your solution here.

1.2 Problem 2

Your solution here.

Chapter 2: Homework 2

2.1 Introduction

This is the second homework assignment for CS285.

2.2 Problem 1

Your solution here.

2.3 Problem 2

Your solution here.

Chapter 3: Homework 3

3.1 Introduction

This is the third homework assignment for CS285.

3.2 Problem 1

Your solution here.

3.3 Problem 2

Your solution here.

Chapter 4: Homework 4

4.1 Introduction

This is the fourth homework assignment for CS285.

4.2 Problem 1

Your solution here.

4.3 Problem 2

Your solution here.

Chapter 5: Homework 5

5.1 Introduction

This is the fifth homework assignment for CS285.

5.2 Problem 1

Your solution here.

5.3 Problem 2

Your solution here.