ROIL – Robust Offline Imitation Learning

Gersi Doko

Dept. of Computer Science, University of New Hampshire

2024

Introduction

IRL is a learning paradigm where an agent learns a policy from expert demonstrations.

Common applications include robotics, autonomous vehicles, and medicine.

Domains where typically the reward function is hard to represent, but demonstrations are plentiful.

Introduction

We aim to learn a policy that performs well in the MDP, without access to the true reward function $r^\star$.

Preliminaries

$\mathcal{S}$ is the state space.

$\mathcal{A}$ is the action space.

$\mathcal{P}$ is the probability transition matrix.

p_0 is the initial state distribution.

r is the reward function.

γ is the discount factor.

Preliminaries

We are given a dataset of state, action pairs D_e generated by some expert policy π_e.

$$D_e = (s_i, \pi_e(s_i))_{i=1}^{N}$$

We aim to learn a policy π that performs well in the MDP, without access to the true reward function $r^\star$, that π_e follows.

$$\mathcal{W} = \{w \in \mathbb{R}^k \mid ||w||_1 \leq 1\}$$

We assume that $\exists w \in \mathcal{W} \mid r^\star = \Phi w$.

Preliminaries

$$\rho(\pi, r) = \lim_{t \to \infty} \mathbb{E}^{\pi, \mathcal{P}}[\gamma^t r(s_t, \pi(s_t))]$$

$$\min_{\pi \in \Pi} \max_{r \in \mathcal{R}} \rho(\pi_e, r) - \rho(\pi, r)$$