Maximum Entropy IRL

① Maximum entropy IRL

Background: Agent behavior: ζ / trajectory

state: s_i → feature $f_{s_i} \in \mathbb{R}^k$

Action: a_i

Goal: optimizing some function f_{s_i} → Reward value

$f_\zeta = \sum_{s_i \in \zeta} f_{s_i}$ > reward for all the path.

$\text{Reward}(f_\zeta) = \theta^T f_\zeta = \sum_{s_i \in \zeta} \theta^T f_{s_i}$

feature expectation → $\sum_{\text{Path } \zeta_i} P(\zeta_i) f_{\zeta_i} = \bar{f}$ // Probabilistic problem.

Deterministic Path distribution: $P(\zeta_i|\theta) = \frac{1}{Z(\theta)} e^{\theta^T f_{\zeta_i}}$ → Partition function

Distribution

[Plans with higher reward is preferred

Nondeterministic path: $P(\zeta|\theta,T) = \sum_{\sigma \in \zeta} P_T(\sigma) \frac{e^{\theta^T f_\zeta}}{Z(\theta,\sigma)} \mathbb{I}_{\zeta \in \sigma}$ ← Identity

$\approx \frac{e^{\theta^T f_\zeta}}{Z(\theta,T)} \prod_{s_{t+1}, a_t, s_t \in \zeta} P_T(s_{t+1}|a_t, s_t)$ ← given??

Stochastic policies: $P(\text{action}|\theta,T) \propto \sum_{\zeta, a \in \zeta_{t=0}} P(\zeta|\theta,T)$

Learning from Demonstration:

$\theta^* = \arg\max_\theta L(\theta) = \arg\max_\theta \sum_{\text{examples}} \log P(\bar{\zeta}|\theta,T)$

$\nabla L(\theta) = \bar{f} - \sum_\zeta P(\zeta|\theta,T) f_\zeta = \bar{f} - \sum_{s_i} D_s f_{s_i}$ → state visitation freq. (Alg 1)