

Project overview: model-free and model-based agents

SAC

Actor $\pi_{\theta}(s) \rightarrow a$

Critic $Q_{\phi}(s, a)$ (x2)

Entropy α, \mathcal{H}

Replay buffer

Model-free, max entropy

TD3

Actor $\pi_{\theta}(s) \rightarrow a$

Critic $Q_{\phi}(s, a) + \text{target}$

Delayed policy, clipped Q

Replay buffer

Model-free, deterministic

TD-MPC2

Encoder $s \rightarrow \vec{z}$

Dynamics $d(\vec{z}, a)$

Reward, Q , policy heads

Opponent model

Model-based, latent planning

Hockey Environment s_t, a_t, r_t, s_{t+1}