

Trajectory $\tau = \{s_1, o_1, a_1, s_2, o_2, a_2, \dots, s_T, o_T, a_T\}$

Envrionment

Observation

Action

Cost or -Reward

