

Literature review: Reinforcement Learning

Daniel Hernandez

1 Introduction

subfiles

1.1 Markov Decision Processes

subfiles

A Markov Decision Process (MDP) is fully defined by the tuple $\langle S, A, P, R \rangle$. Whereby S is the set of states $s \in S$ spawned by the Markov chain used in the MDP, A is the set of actions $a \in A$ that can be taken in the MDP. P is a transition kernel which maps a state-action pair to a future state, $P : S \times A \Rightarrow S$. If the environment is stochastic, as opposed to deterministic, the function P maps a state-action pair to a distribution over states in S .

From here we can introduce the notion of an agent. (link to control theory?), An agent is an entity that on every state $s_t \in S$ it can take an action $a_t \in A$ in an environment transforming the environment from s_t to s_{t+1} . The behaviour of an agent is fully defined by a policy π . A policy π is a mapping from states to actions, $\pi : S \Rightarrow A$. The agent chooses which action a_t to take in every state s_t by querying its policy such that $a_t = \pi(s_t)$. If the policy is stochastic, π will map an action to a distribution over action $a_t \sim \pi(s_t)$.

Note that in general it is not the case that all actions $a \in A$ can be taken on every state $s_t \in S$.

References