Quiz 5

Nam Nguyen

October 2020

1 Introduction

Once the states of a robot are partially observable, we apply Q-learning
False. Q-learning attempts to update the value of $Q(s, a)$ at state s and action a. In the case of having partial observation, the agent only estimate the value of current state by probability distribution. Thus, it is not recommended to use Q-learning for partially observable state.
Once facing delayed reward, we apply TD-learning
True. TD-*lambda* utilize the temporal discount factor $0 \leq \gamma \leq 1$, yielding the accumulated reward over time

$$V(X) = \sum \gamma^i r_i^{\pi(X)}$$

2 References

Advance Robotics Fall 2012, University of California at Berkeley.