Department of Computer Engineering

Artificial Intelligence

Assignment 7 Part 2

Dr. Rohban

Parsa Mohammadian — 98102284

January 7, 2022

Contents

1

1.1

V-value is state-value function. It means that $V^{\pi}(s)$ is the value of starting from state s and obeying policy π. At the other hand, Q-value is action-value function. So $Q^{\pi}(s,a)$ is the value of taking action a from state s and obeying policy π. Since R and P are not available, we can only use Q-value for this kind of problems (model-free).

1.2

In ϵ-greedy method, we choose currently best action with probability $1-\epsilon$ and randomly choose other actions with probability ϵ. For the normal policy iteration ($\epsilon = 1$ or always choose the current best action), we only exploit our current knowledge and never explore. In this fashion we might stuck in local optimum. But when ϵ is not 1, we explore with probability $1-\epsilon$ and seek for better local optimum even if we have already found one, and finally we might get to the global optimum.

1.3

$$
\begin{gathered}
\forall s : \mathbb{E}_{a\sim\pi'}[Q^{\pi}(s,a)] \geq \mathbb{E}_{a\sim\pi}[Q^{\pi}(s,a)] \\
\mathbb{E}[X] = \sum_{x\in X} p(x)x \\
\Rightarrow \sum_{a\in A} P(a)Q^{\pi^{'}}(s,a) \geq \sum_{a\in A} P(a)Q^{\pi}(s,a) \\
P(a) \propto \pi(a|s) \\
\Rightarrow \sum_{a\in A} \pi(a|s)Q^{\pi^{'}}(s,a) \geq \sum_{a\in A} \pi(a|s)Q^{\pi}(s,a) \\
V^{\pi}(s) = \sum_{a\in A(s)} \pi(a|s)Q^{\pi}(s,a) \\
\Rightarrow V^{\pi^{'}}(s) \geq V^{\pi}(s)
\end{gathered}
$$