CMPSCI 687 Homework 1

Due September 26, 2017, 11pm Eastern Time

Instructions: This homework assignment consists of a written portion and a programming portion. Collaboration is not allowed on any part of this assignment. Submissions must be typed (hand written and scanned submissions will not be accepted). We recommend that you use LaTeX. The assignment should be submitted as a single .pdf on Moodle. The automated system will not accept assignments after 11:55pm on September 26.

Part One: Written (30 Points Total)

1. (15 Points) Given an MDP $M = (\mathcal{S}, \mathcal{A}, P, R, d_0, \gamma)$ and a fixed policy, π, the probability that the action at time $t = 0$ is $a \in \mathcal{A}$ is:

$$\Pr(A_0 = a) = \sum_{s \in \mathcal{S}} d_0(s)\pi(s, a).$$

Write similar expressions (using only the terms defined in M) for the following:

- The probability that the state at time $t = 3$ is either $s \in \mathcal{S}$ or $s' \in \mathcal{S}$.

$$\begin{aligned}\Pr(S_4 = s \cup S_4 = s') = &\sum_{s_0} d_0(s_0) \sum_{a_0} \pi(s_0, a_0) \sum_{s_1} P(s_0, a_0, s_1) \\ &\times \sum_{a_1} \pi(s_1, a_1) \sum_{s_2} P(s_1, a_1, s_2) \\ &\times \sum_{a_2} \pi(s_2, a_2)(P(s_2, a_2, s) + P(s_2, a_2, s'))\end{aligned}$$

- The probability that the action at time $t = 16$ is $a' \in \mathcal{A}$ given that the action at time $t = 15$ is $a \in \mathcal{A}$ and the state at time $t = 14$ is s.

$$\begin{aligned}&\Pr(A_{16} = a' | S_{14} = s, A_{15} = a) \\ &= \frac{\Pr(A_{16} = a', A_{15} = a | S_{14} = s)}{\Pr(A_{15} = a | S_{14} = s)} \\ &= \frac{\sum_{a_{14}} \pi(s, a_{14}) \sum_{s_{15}} P(s, a_{14}, s_{15})\pi(s_{15}, a) \sum_{s_{16}} P(s_{15}, a, s_{16})\pi(s_{16}, a')}{\sum_{a_{14}} \pi(s, a_{14}) \sum_{s_{15}} P(s, a_{14}, s_{15})\pi(s_{15}, a)}\end{aligned}$$

- The expected reward at time $t = 6$ given that the action at time $t = 3$ is $a \in \mathcal{A}$, and the state at time $t = 5$ is $s \in \mathcal{S}$.

$$\begin{aligned}\mathbf{E}\left[R_6|A_3 = a, S_5 = s\right] = &\sum_{a_5} \pi(s, a_5) \sum_{s_6} P(s, a_5, s_6) \\ &\times \sum_{a_6} \pi(s_6, a_6) \sum_{s_7} P(s_6, a_6, s_7) R(s_6, a_6, s_7)\end{aligned}$$

- The probability that the initial state was $s \in \mathcal{S}$ given that the state at time $t = 1$ is $s' \in \mathcal{S}$.

$$\begin{aligned}\Pr(S_0 = s|S_1 = s') = &\frac{\Pr(S_0 = s, S_1 = s')}{\Pr(S_1 = s')} \\ = &\frac{d_0(s) \sum_{a_0} \pi(s, a_0) P(s, a_0, s')}{\sum_{s_0} d_0(s_0) \sum_{a_0} \pi(s, a_0) P(s_0, a_0, s')}\end{aligned}$$

- The probability that the action at time $t = 5$ is $a \in \mathcal{A}$ given that the initial state is $s \in \mathcal{S}$, the state at time $t = 5$ is $s' \in \mathcal{S}$, and the action at time $t = 6$ is $a' \in \mathcal{A}$.

$$\begin{aligned}\Pr(A_5 = a|S_0 = s, S_5 = s', A_6 = a') = &\Pr(A_5 = a|s_5 = s', A_6 = a') \\ = &\frac{\Pr(A_5 = a, A_6 = a'|s_5 = s')}{\Pr(A_6 = a'|S_5 = s')} \\ = &\frac{\pi(s', a) \sum_{s_6} P(s', a, s_6) \pi(s_6, a')}{\sum_{a_5} \pi(s', a_5) \sum_{s_6} P(s', a_5, s_6) \pi(s_6, a')}\end{aligned}$$

2. (2 Points) How many deterministic policies are there for an MDP with $|\mathcal{S}| < \infty$ and $|\mathcal{A}| < \infty$? (You may write your answer in terms of $|\mathcal{S}|$ and $|\mathcal{A}|$).

$$|\mathcal{S}|^{|\mathcal{A}|}$$