Artificial Intelligence: Second Partial test

04 June 2018

1 Exercise 1 (Points 30)

Consider an undiscounted MDP having three states (1,2,3). State 3 is terminal. In state 1 and 2 there are two possible actions A and B. The transition and reward models are as follow:

- In state 1, action A moves the agent to state 2 with probability .8 and a reward of -2 while it leaves the agent in state 1 with probability .2 and a reward of -1. In state 1, action B moves the agent to state 3 with probability .8 and a reward of 0 while it leaves the agent in state 1 with probability .2 and a reward of -1.
- In state 2, action A moves the agent to state 1 with probability .8 and a reward of -1 while it leaves the agent in state 2 with probability .2 and a reward of -2. In state 2, action B moves the agent to state 3 with probability .8 and a reward of 0 while it leaves the agent in state 2 with probability .2 and a reward of -2.

Answer the following questions:

1. Provide a transition diagram for the MDP described above.
2. Provide a qualitative discussion about the optimal policy for this MDP.
3. Show the value of $v(1)$ for the first two iterations of a Value Iteration algorithm. Assume $v(s) = 0 \ \forall s$.
4. Is it necessary to compute the value of $v(2)$ to answer to the previous question ? Motivate your answer.

2 Exercise 2 (Points 30)

Consider the Bayesian Network in Figure 1. Answer the following questions:

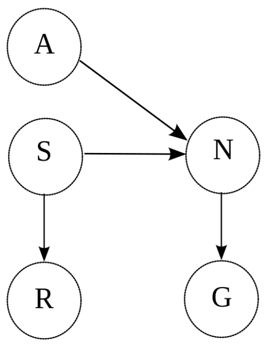

Figure 1: Bayesian Network.

1. State whether we can guarantee that N is conditionally independent from R given (S, A). Motivate your answer.

2. State whether we can guarantee that A is conditionally independent from G given S. Motivate your answer.

3. Assume variables $\{S, N\}$ can take three different values while $\{A, R, G\}$ are binary. State how many parameters must be provided to compute the joint probability table for this Bayesian Network.

3 Exercise 3 (Points 30)

Consider an environment with states $\{A, B, L, R\}$, actions $\{r, l\}$ where states $\{L, R\}$ are terminal. Assume $\gamma = 1$ and $\alpha = 0.5$. Consider the following sequence of learning episodes:

E1 $(A, r, B, 0)(B, r, R, +1)$

E2 $(A, r, B, 0)(B, l, A, 0)$

E3 $(A, r, B, 0)(B, l, R, +1)$

1. Build an estimate for $\hat{T}$ and $\hat{R}$

2. Compute $v(s)$ for all non-terminal states by using a direct evaluation approach

3. Compute $v(s)$ for all non-terminal states by using a sample-based evaluation approach