AIL 722: Reinforcement Learning

Lecture 20: Problem Session

Raunak Bhattacharyya

ScAI YARDI SCHOOL OF ARTIFICIAL INTELLIGENCE
INDIAN INSTITUTE OF TECHNOLOGY DELHI

Question

- Suppose MDP with 100 states and 4 actions (up, down, left, right)
- Policy improvement step:

$$\pi^{(k+1)}(s) = \arg\max_a \left[r(s,a) + \gamma \sum_{s'} T(s' \mid s,a) V^{\pi^{(k)}}(s') \right]$$

- Assume if there are ties between actions, they are broken in order.

Is it possible that $\pi^{(2)} \neq \pi^{(3)}$ but $\pi^{(2)} = \pi^{(4)}$?

Question

- Suppose MDP with discrete state space (size n) and action space (size m)
- What is the time complexity of the policy improvement step?
- If we know transition probs belong to {0,1} can we give a tighter complexity bound?

Question

- Suppose MDP with 10 possible states and 5 possible actions
- Suppose every state-action pair has a non-zero probability of transitioning to every state
- For each iteration of VI, where a single iteration corresponds to all the states being updated, how many times will we need to evaluate the transition function?

Question

- MDP with 5 states ($s^{1:5}$) and 2 actions: stay and continue. We know that:

$$
\begin{aligned}
T(s_i \mid s_i, a_S) &= 1 \text{ for } i \in \{1,2,3,4\} \\
T(s_{i+1} \mid s_i, a_C) &= 1 \text{ for } i \in \{1,2,3,4\} \\
T(s_5 \mid s_5, a) &= 1 \text{ for all actions } a \\
R(s_i, a) &= 0 \text{ for } i \in \{1,2,3,5\} \text{ and for all actions } a \\
R(s_4, a_S) &= 0 \\
R(s_4, a_C) &= 10
\end{aligned}
$$

What is the discount factor γ if the optimal value $V^*(s^1) = 1$?