hw4_mdps_q11_policies

Question 11: Policies

5/5 points (ungraded)

John, James, Alvin and Michael all get to act in an MDP $(S, A, T, \gamma, R, s_0)$.

- John runs value iteration until he finds V^* which satisfies $\forall s \in S : V^*(s) = \max_{a \in A} \sum_{s'} T(s, a, s') \left(R(s, a, s') + \gamma V^*(s')\right)$ and acts according to $\pi_{\text{John}} = \arg\max_{a \in A} \sum_{s'} T(s, a, s') \left(R(s, a, s') + \gamma V^*(s')\right)$.
- James acts according to an arbitrary policy π_{James}.
- Alvin takes James's policy π_{James} and runs one round of policy iteration to find his policy π_{Alvin}.
- Michael takes John's policy and runs one round of policy iteration to find his policy π_{Michael}.

Note: One round of policy iteration = performing policy evaluation followed by performing policy improvement. Mark all of the following that are guaranteed to be true:

- [ ] It is guaranteed that $\forall s \in S : V^{\pi_{\text{James}}}(s) \geq V^{\pi_{\text{Alvin}}}(s)$
- [x] It is guaranteed that $\forall s \in S : V^{\pi_{\text{Michael}}}(s) \geq V^{\pi_{\text{Alvin}}}(s)$
- [ ] It is guaranteed that $\forall s \in S : V^{\pi_{\text{Michael}}}(s) > V^{\pi_{\text{John}}}(s)$
- [ ] It is guaranteed that $\forall s \in S : V^{\pi_{\text{James}}}(s) > V^{\pi_{\text{John}}}(s)$
- [ ] None of the above.

✔

Submit

✔ Correct (5/5 points)