hw4_mdps_q9_wrong_discount_factor

Question 9: Wrong Discount Factor

0.0/5.0 points (graded)

Bob notices value iteration converges more quickly with smaller γ and rather than using the true discount factor γ, he decides to use a discount factor of $\alpha\gamma$ with $0 < \alpha < 1$ when running value iteration. Mark each of the following that are guaranteed to be true:

☐ While Bob will not find the optimal value function, he could simply rescale the values he finds by $\frac{1-\gamma}{1-\alpha}$ to find the optimal value function.

☑ If the MDP's transition model is deterministic and the MDP has zero rewards everywhere, except for a single transition at the goal with a positive reward, then Bob will still find the optimal policy. ✓

☐ If the MDP's transition model is deterministic, then Bob will still find the optimal policy.

☑ Bob's policy will tend to more heavily favor short-term rewards over long-term rewards compared to the optimal policy. ✓

☐ None of the above.

Option 1: False. If Bob simply rescales all the values, the policy that he finds will be exactly the same.

Option 2: True. In this case, the value of every state will be scaled down by a factor of α. Therefore, the optimal policy will not change.

Option 3: False. Consider the MDP from Question 7 for $\alpha = .1$ and $\gamma = .9$. The optimal policy in this MDP is to go left at state d. However, Bob's policy will tell you to go right at state d.

Option 4: True. Bob's policy is the optimal policy from running value iteration with a lower discount factor. The lower discount factor favors short term rewards.

Submit

ℹ Answers are displayed within the problem