hw4_mdps_q9_wrong_discount_factor

Question 9: Wrong Discount Factor

5/5 points (ungraded)

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.
- [x] 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.
- [x] 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.

✔

Submit

✔ Correct (5/5 points)