hw5_rl_q6_qlearning_properties

Question 6: Q-Learning Properties

0.0/4.0 points (graded)

In general, for Q-Learning to converge to the optimal Q-values...

- [x] It is necessary that every state-action pair is visited infinitely often. ✔
- [x] It is necessary that the learning rate α (weight given to new samples) is decreased to 0 over time. ✔
- [ ] It is necessary that the discount γ is less than 0.5.
- [ ] It is necessary that actions get chosen according to $\arg\max_a Q(s, a)$.

a) In order to ensure convergence in general for Q learning, this has to be true. In practice, we generally care about the policy, which converges well before the values do, so it is not necessary to run it infinitely often.

b) In order to ensure convergence in general for Q learning, this has to be true.

c) The discount factor must be greater than 0 and less than 1, not 0.5.

d) This would actually do rather poorly, because it is purely exploiting based on the Q-values learned thus far, and not exploring other states to try and find a better policy.

Submit

Answers are displayed within the problem