ECS 170 Homework 7

Hardy Jones
999397426
Professor Davidson
Winter 2014

1. Consider a cumulative discount reward with $\gamma = 0$ and $\gamma = 1$. What type of behavior would these reward functions encourage?

   The reward function with $\gamma = 0$ would reward short term goals. The reward function with $\gamma = 1$ would reward long term goals.

2. Why would you prefer to use Q-learning to implement a black-jack player rather than the mini-max algorithm?

   BlackJack is a partially observable game. Minimax would have to search the entire state space for every possible card combination can could occur. This is realistically impossible. The only viable option is to create an evaluation function that will most likely not be optimal.

   Q-learning can learn the optimal choice to make given enough time to learn.

3. One Q-learning algorithm described in class assumed that the δ and r functions were deterministic. Can this algorithm be used for

   (a) Learning to control an inverted pendulum

   Yes.

   (b) Play chess

   Yes.

4. Evaluate the table.

Step	Start State	Subsequent State	Update to Q(s, a)
1	B1	A1	Q(B1, Up) = 0
2	A1	A2	Q(A1, Right) = 9
3	A2	A3	Q(A2, Right) = 7.5
4	A3	B3	Q(A3, Down) = 14
5	B3	B2	Q(B3, Left) = 10
6	B1	A1	Q(B1, Up) = 9
7	A1	A2	Q(A1, Right) = 16.5
8	A2	A3	Q(A2, Right) = 16
9	A3	B3	Q(A3, Down) = 14
10	B3	B2	Q(B3, Left) = 19