Machine Learning - Assignment 6

Posted Friday, November 15, 5:30pm
Due Friday, November 21, 5:30pm

1. [30 points] (Adapted from Puterman). When we go out to dinner, we always like to park as close as possible to the restaurant. Assume the restaurant is situated on a very long street running east to west, which allows parking on one side only. The street is divided into one-car-length sections. We approach the restaurant from the east, starting D units away. The probability that a parking spot at distance s from the restaurant is unoccupied is p_s, independent of all other spots. Formulate this problem as a Markov Decision Process. Be sure to specify all the elements of your MDP!

2. [10 points] Sutton and Barto book, Section 3.7, Exercise 3.8

3. [15 points] Sutton and Barto book, Section 3.7, Exercises 3.10 and 3.11.

4. [10 points] Suppose that you have an MDP M_1 with discount factor $\gamma < 1$. An arbitrary policy π has value V_1^π in this MDP. Construct an MDP M_2 with discount factor 1, such that $V_2^\pi = V_1^\pi$.

5. [15 points] Suppose you observe the following 11 episodes generated by an unknown Markov process, where A and B are states and the numbers are rewards:

   - A,0,B,1
   - A,2
   - A,0,B,0,A,2
   - A,0,B,0,A,0,B,1
   - B,0,A,2
   - B,0,A,0,B,1
   - B,0,A,2
   - A,0,B,1
   - B,1

- B,1
- B,1

(a) Give the values for states A and B that would be obtained by the Monte Carlo method discussed in class on this data set.

(b) If you formed a model of the Markov process based on this data, what would it be? Explain your answer.

(c) Give the values for states A and B that would be obtained by the batch TD(0) method. Show the way you computed your answer.

6. [10 points] As we discussed in class, state aggregation is a special case of generalizing function approximation, in which states are grouped together and one table entry (value estimate) is used for the whole group.

   (a) Show that state aggregation is a special case of linear function approximation.

   (b) Are tile codings (CMACs) creating just a state aggregation, or not? Explain the similarities and differences.

7. [10 points] Consider the value iteration update algorithm applied to action values:

$$Q_{n+1}(s,a) \leftarrow r_s^a + \gamma \sum_{s'} p_{ss'}^a \max_{a'} Q_n(s',a')$$

   Let us write $Q_{n+1} = B(Q_n)$, where Q_n, Q_{n+1} are vectors and B is the Belamn update operator (defined through the equation above). Prove that, for any vectors Q, Q', we have

$$||B(Q) - B(Q')||_\infty \leq \gamma ||Q - Q'||_\infty$$

   where $||Q||_\infty = \max_{s,a} |Q(s,a)|$.