Advanced AI Techniques (WS05/06)

Exercise sheet 8
Deadline: Tuesday, 17 Jan 06

Exercise 1 (4 points)

Let optimistic average sum evaluation formula be the formula $E(a) = \frac{k \cdot r(a) + s}{k+1}$ if action a was chosen k times with reward $r(a)$, whereby s is the optimistic initialization. What condition for s must be satisfied so that for n-armed bandits with constant deterministic reward and "optimistic average sum" evaluation formula, the greedy action selection method becomes stable, i. e., within finite time approaching the optimal strategy without deviating from it once it is established.

Exercise 2 (2 points)

Derive the Bellmann equation for Q^π without the expectation operator using the equation

$$Q^\pi(s,a) = E_\pi\{\Sigma_{k=0}^{\infty}\gamma^k r_{t+k+1} | s_t = s, a_t = a\}$$

with expectation operator. This derivation can be done analogously to the one for $V^\pi(s)$ presented in the lecture.

Exercise 3 (6 points)

a. Draw the diagram of all states and probabilistic transitions (MDP diagram) for the following example of a skiing trip:

You plan to go skiing in a skiing region with three skiing slopes A (red), B (red) and C (blue), and three lifts X (to the top of slope A), Y (to slope B), Z (to slope C). Successfully going down a red slope, you get twice as much pleasure as going down the blue one. From A, you can either run down to lift Y or Z. From B, you can either go to lift X or to lift Y. From C, you can go down to lift X or Z. There is always the option of waiting on top of a slope (without getting any pleasure from it). Assume that you are a bad skier: The transition probabilities are 0.6 for slope B, and 0.75 for slopes A and C. In all other cases, you end in hospital for the rest of the time (having an accident and each time step in hospital rewards the negative of the pleasure of running down the blue slope).

b. What are the state-value Bellmann equations for the random policy that selects either one of the three actions in each state with probability $\frac{1}{3}$ *? (Choose* $\gamma = \frac{2}{3}$*.)*

c. Calculate the state-value functions for each of the states for the random policy. (You do not have to derive the result, you can use an equation solver.)

d. For one of the states (being at the top of A, or of B, or of C), verify by inserting the corresponding values from the solution (c.) that the Bellmann equation is solved.