Foundations of Artificial Intelligence

Prof. Dr. B. Nebel, Prof. Dr. W. Burgard — University of Freiburg
Dr. A. Kleiner, R. Mattmüller — Department of Computer Science
Summer Term 2008

Exercise Sheet 11

Due: Friday, July 18, 2008

Exercise 11.1 (Complexity of exact inference)

Prove that the 3-SAT problem can be reduced to exact inference in Bayesian networks and hence that exact inference in NP-hard.
Hint: Consider a network with one variable for each proposition symbol, one for each clause, and one for the conjunction of clauses.

Exercise 11.2 (Value iteration algorithm)

Consider the following grid world. The u values specify the utilities after convergence of the value iteration and r is the reward associated with a state. Assume that $\gamma = 1$ and that an agent can perform four possible actions: **North**, **South**, **East** und **West**. With probability 0.7 the agent reaches the intended state, with probability 0.2 it moves to the right of the intended direction, and with probability 0.1 to the left.

$u = 8$	$u = 15$	$u = 12$
$u = 2$	$r = 2$	$u = 10$
$u = 7$	$u = 16$	$u = 11$

Which is the best action an agent can execute if he is currently in the center state of the grid world? Justify your answer. Which utility does the center state have?

Exercise 11.3 (Convergence of value iteration)

Consider the Bellman update as an operator B which, if applied to a vector of utilities (one value for each state), returns a new vector of utilities. Then the Bellman update can be written as $U_{i+1} \leftarrow BU_i$. Let $||\cdot||$ be the max norm, i.e. $||U|| = \max_s |U(s)|$. Then the true utilities U satisfy the following inequality (without proof):

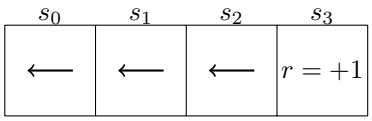
$$||BU_i - U|| \leq \gamma ||U_i - U||.$$

Additionally, we have that (again without proof) $|U(s)| \leq R_{\max}/(1-\gamma)$ for each state s, where $R_{\max}$ is the maximal reward of all states. Calculate how many iterations N of the value iteration algorithm are at most necessary until the

error $||U_N - U||$ becomes less than or equal to a given $\epsilon > 0$. You may assume $\gamma < 1$.

Exercise 11.4 (Policy iteration algorithm)

Let $\gamma = 0.5$ and let there be only the actions **East** and **West**. With probability 0.9 the agent reaches the intended state (or stays where he was, if the action would move him out of the grid), and with probability 0.1 he moves in the opposite direction. The reward in the three western states is -0.05 each.



Perform one step of the policy iteration algorithm. The initial policy is given by the arrows in the states. Give the linear system of equations for the first policy evaluation, a solution to the system as well as the first improved policy π_1.

The exercise sheets may and should be worked on in groups of three (3) students. Please fill the cover sheet[1] and attach it to your solution.

[1] `http://www.informatik.uni-freiburg.de/~ki/teaching/ss08/gki/coverSheet-english.pdf`