Foundations of Artificial Intelligence

Prof. Dr. J. Boedecker, Prof. Dr. W. Burgard, Prof. Dr. F. Hutter, Prof. Dr. B. Nebel, Dr. rer. nat. M. Tangermann
T. Schulte, M. Krawez, R. Rajan, S. Adriaensen, K. Sirohi
Summer Term 2020

University of Freiburg
Department of Computer Science

Exercise Sheet 9 — Solutions

Exercise 9.1 (Markov Decision Processes)

Consider the Markov decision process underlying the sequence decision problem for the grid world in exercise 8.2.

(a) Perform a policy-improvement step for the initial state **S** and the (non-optimal) policy E. Which is the optimal action in this state? The utility of all non-terminal states is assumed to be zero and rewards are assumed to be additive (i.e. not discounted).

Solution:

Let s be the Starting position. We have to calculate the expected utility for each action and then choose the action with the maximum expected utility. The expected utility for an action, with no immediate reward, is calculated as

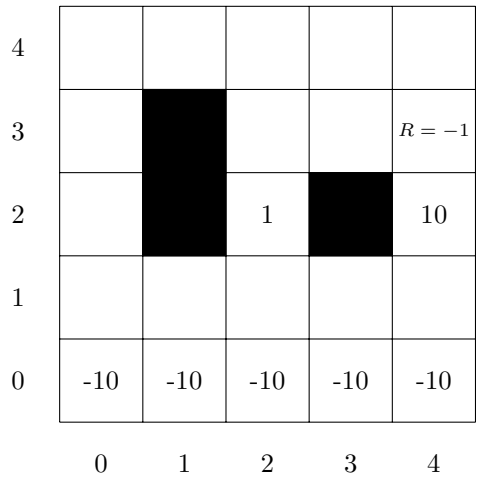
$$\sum_{s'} P(s, a, s')U(s').$$

So:

$$\begin{aligned}
\sum_{s'} P(s, \textbf{North}, s')U(s') &= 0.8 * 0 + 0.1 * 0 + 0.1 * 0 = 0 \\
\sum_{s'} P(s, \textbf{East}, s')U(s') &= 0.8 * 0 + 0.1 * (-10) + 0.1 * 0 = -1 \\
\sum_{s'} P(s, \textbf{South}, s')U(s') &= 0.8 * (-10) + 0.1 * 0 + 0.1 * 0 = -8 \\
\sum_{s'} P(s, \textbf{West}, s')U(s') &= 0.8 * 0 + 0.1 * (-10) + 0.1 * 0 = -1
\end{aligned}$$

The best action is $\arg\max_a \sum_{s'} P(s, a, s')U(s')$, i.e., **North**.

(b) Perform a first step of the value iteration algorithm for state (4,3) for the slightly **modified grid world** below. The immediate reward of this state is $R = -1$. Assume discounted rewards and a discount factor of $\gamma = 0.5$. Initially, all non-terminal states have a utility of zero.



Solution:

$$U'(s) = R(s) + \gamma \max_{a \in A(s)} \sum_{s'} P(s'|s,a)U(s')$$

$$0.5 * \begin{cases} 0.8*0 + 0.1*0 + 0.1*10 & (\rightarrow) \\ 0.8*0 + 0.1*0 + 0.1*0 & (\uparrow) \\ 0.8*0 + 0.1*0 + 0.1*10 & (\leftarrow) \\ 0.8*10 + 0.1*0 + 0.1*0 & (\downarrow) \end{cases}$$

$$= -1 + 0.5 * 8 = 3$$

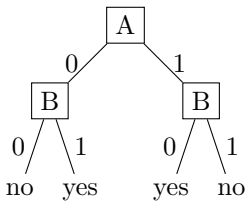
Exercise 9.2 (Decision Trees)
Specify decision trees representing the following boolean functions:

(a) A XOR B

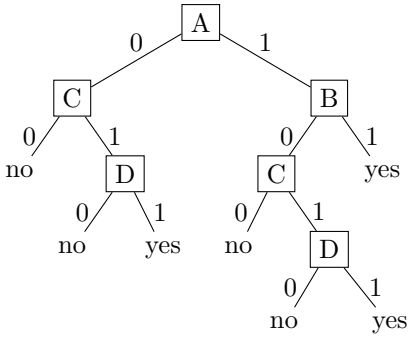
(b) $(A \wedge B) \vee (C \wedge D)$

Solution:

(a)



(b)


A
0
1
C
B
0
1
0
1
no
D
C
yes
0
1
0
1
no
yes
no
D
0
1
no
yes