Artificial Intelligence 2 – SS 2020
Assignment 9: Markov Decision Procedures
– Given June 12., Due June 21. –

Hint: Exercises need to be handed in via StudOn at 23:59 on the day they are due or earlier. Please use only the exercise group of your tutor to hand in your work.

If any concepts here seem unfamiliar to you or you have no idea how to proceed, consult the lecture materials, ask a fellow student, your tutor, or on the Forum.

If a problem asks for code, comment it or make it otherwise self-explanatory. You do not need to write a lot, but it should be enough to convince your tutor that you understand what the code does. We may deduct up to 30% for uncommented and unclear code, but would prefer not to.

Problems with no points (0pt) will not be graded, but might appear on the exam in a similar form. For these, we will provide a reference solution after the submission deadline. If you find the reference solution unclear, ask about it on the forum or in in a tutorial.

Problem 9.1 (Long term vs short term)

Consider the following world: 0pt

+50	-1	-1	-1	. . .	-1	-1	-1	$\boxed{-1}$
Start	▨	▨	▨	▨ . . .	▨	▨	▨	▨
-50	+1	+1	+1	. . .	+1	+1	+1	$\boxed{+1}$

In the *Start* state an agent has two possible actions, *Up* and *Down*. They can't return to *Start* though and the can't pass grey fields, so after the first move the only possible action is *Right*.

Assuming a discounted reward function and that the world is 101 fields wide, for what values of the discount γ should the agent choose *Up* and for which *Down*? Compute the utiliy of each action as a function of γ.

Problem 9.2 (Markov Games)

60pt

We want to apply Markov Decision Procedures to two-player games as considered in KI1. We assume two players A, B and model everything from A's point of view - in particular, we have a reward function $R(s)$ for our states and our *states* are *only those game states, where it's A's move.* Hence, the transition function $T(s, a, s')$ models the probability that, if player A does action a in state s, player B will play such that we end up in state s'. If the game ends immediately with A's move, we take that to be a successor state, too (assuming that B just makes an empty move).

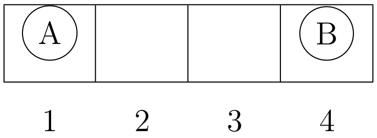
Let $N(s, a)$ be the set of possible successor states of s after picking move a.

20 pt

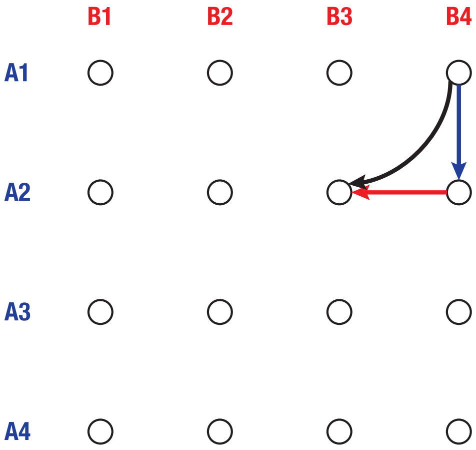
1. Suppose that B moves so that the higher A's utility $U(s')$, the smaller the probability that B's going to pick a move resulting in s'. In a state s, if A picks given action a, then the probability that B picks a move ending in s' is $P(s') = \frac{c}{U(s')}$ for some constant c. Compute c and write down the corresponding Bellman equation so that it does not contain T or c.

10 pt

2. Consider a game with four fields F_1, F_2, F_3, F_4. Player A starts in F_1, player B in F_4. Each turn, a player has to move to an adjacent field (i.e. if the player is in F_i, he can move to F_{i+1} or F_{i-1}). If an adjacent field contains the other player, one may jump over him. E.g. assume player A is in F_2 and player B in F_3, then player A may jump over B to end up in F_4. The game ends when one player reaches the opposite end of the board, i.e. A wins if he ends up in F_4, B wins if he ends up in F_1. player A starts. The reward for player A winning is 100, the reward for losing is 1 (as to avoid utility 0 anywhere).



Draw the state space, that is, *only those game states, where it's A's move*, and transitions between them, starting with the state A at F_1, B at F_4, and ending with the winning states for A and B. To get an idea of what is going on, you can also draw A's and B's moves as arrows in different colors/styles. Arrange the states (s_A, s_B) on a two-dimensional grid, using s_A and s_B as coordinates, like in the picture below with the first step drawn in (one move by A and one move by B).



Above, A's only possible move at the beginning of the game is marked in blue, and B's possible only possible move after that is marked in red. The corresponding transition is marked in black. Note that if either of the players had more than one possible move from some state, you would draw multiple arrows from that state. You do not really need this to figure out the next two steps, but it might help.

10 pt

3. Using the Bellman equation from the first task, which utilities change in the first step of value iteration, if we initialise with $U(s) = 50$ for all of the states? (Remember, the update is applied simultaneously to all the states at each iteration.) Take the reward $R(s)$ to be 0 except for at terminal states, and the discount factor to be 1.

   What are their new values?

20 pt

4. Compute the two utilities that change in the next step (again, they will be states where it is A's move). If you could not solve the first task, use the following Bellman equation:

$$U(s) = R(s) + \max_a \left(\sum_{s' \in N(s,a)} \frac{U(s')}{\mid N(s,a) \mid} \right)$$