OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG — INF FAKULTÄT FÜR INFORMATIK

Prof. Dr.-Ing. habil. Sanaz Mostaghim — Magdeburg, Winter 2021/22

Exam of the Course: Computational Intelligence in Games

Family Name:	First Name:
Matriculation Number:	Signature:

1. Please keep your student ID ready in front of you.
2. Please fill your *Name*, *First Name* and *Matr.-Nr.* and sign the exam.
3. The exam has **7 Assignments** and consists of **11 pages**.
   Please check if your exam is complete before starting.
5. No further tools are allowed. No calculators are allowed.
6. Attempts of cheating lead to failing the exam.
7. Please use only black or blue pens and write clearly.
   Unreadable solutions may be excluded from evaluation.
8. The time for the exam is 120 minutes.

Assignment	1	2	3	4	5	6	Sum
Points	12	12	12	12	4	8	60

Assignment 1 Game Theory $(2 + 3 + 3 + 4 = 12$ points$)$

Consider the prisoners-dilemma from the lecture. However, the two “Players” are given an additional choice of “rejecting the deal”. Consequently, the “game” contains the following rewards:

Variable	Meaning	Value
R	Cooperation PayOff	6
T	Temptation PayOff	9
P	Punishment PayOff	0
S	Sucker’s PayOff	9
L	Loner’s PayOff	3

All PayOff values are expected sentences of the prisoners in years. The *Loner’s Payoff* is awarded to both players if any of the “players” abstain (“reject the deal”) from the “game”. The *Temptation Payoff* is awarded if one player defects the other. The *Punishment PayOff* is given if both players defect each other. The *Sucker’s PayOff* is given to the player who cooperated, but was defected by the other.

a) What is the payoff matrix of the first player?

b) What is the expected payoff of the first player playing *Cooperate* with a second player using a random strategy?

c) What is the best pure strategy of the first player to play with another player with the following mixed strategy *Cooperate*: 81%, *Defect*: 10% and *Abstain*: 9%. What payoff can the first player expect?

d) Now consider the second player to be an adaptive player, which chooses the *Cooperate* strategy with exactly the same percentage as the first player chooses the *Defect* strategy. Both players never want to consider the *Abstain* strategy. What is the best mixed strategy for player one and what payoff can he/she expect?

Assignment 2 Evolutionary Game Theory ($4 + 6 + 2 = 12$ points)

Consider the following generalized 2-Player games:

$$A = \begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{bmatrix} 2,2 & 3,0 \\ 0,3 & 4,4 \end{bmatrix} \end{array} \qquad B = \begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{bmatrix} 3,3 & 0,3 \\ 3,0 & 5,5 \end{bmatrix} \end{array}$$

a) In which of the payoff matrices is C an Evolutionary Stable Strategy? Why?

b) Use replicator equations to calculate the set of fixed points $\mathcal{F}$, for the number of cooperators x_C for game A.

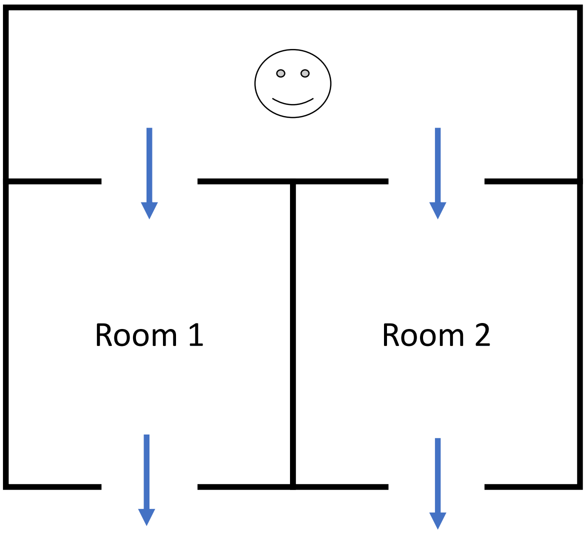
c) Determine the fix points $\mathcal{F}$ from the previous task, to which the cooperators population $x_C(t)$ of game A converges to.

Assignment 3 **Dynamic Programming** **($4 + 6 + 2 = 12$ points)**

Consider the following scenario. An agent is in state S_0 and wants to leave the house. In S_0 it can take the action a_1 (go to room 1 = S_1) or action a_2 (go to room 2 = S_2). In each room, it has two actions a_s (stay) and a_l (leave). Since the goal is to leave the house, for a_s, it gets a reward of 0 and for a_l it gets a reward of 1.

The agent uses the following policy π:

- $\pi(a_1|S_0) = 0.4$ and $\pi(a_2|S_0) = 0.6$.
- In room 1, the probability for a_s is 0.8.
- In room 2 the probability for a_s is 0.



a) Plot the given scenario as a Markov Decision Process.

b) Compute two iterations of iterative policy evaluation to estimate the expected value of each state. Consider $\gamma = 1$.

c) What is the optimal policy π in (b)? Explain.

Assignment 4 Application of Reinforcement Learning ($6 + 6 = 12$ points)

You are given the task to estimate the time it takes the security service to visit a set of departments. Therefore, you follow the security guard during one night on his visits of each building. After one night you created the following records.

Department	1st Walk	2nd Walk	3rd Walk
Start at office	22:30	1:00	4:00
Computer Science Department	22:35	1:10	4:03
Physics Department	22:45	1:15	4:10
Math Department	22:50	1:20	4:15
Business Department	23:20	1:45	5:30
Back at the office	23:30	1:50	5:35

Use the travelled minutes as reward and the remaining time until you arrive at home as your return function.

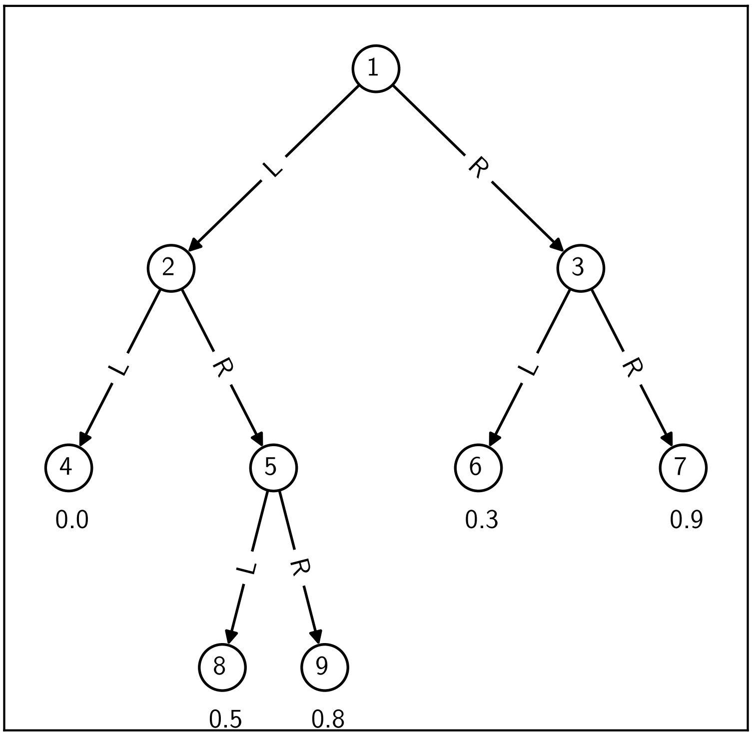
a) Use the Monte Carlo Method to update the value estimate for each episode. Use the first column to generate initial values for $V(s)$. Write down the update rule.

b) Apply Temporal Difference Learning ($TD(0)$) with $\alpha = 0.5$ and $\gamma = 1$ for the same data set.

Assignment 5 Flat Monte Carlo (4 points)

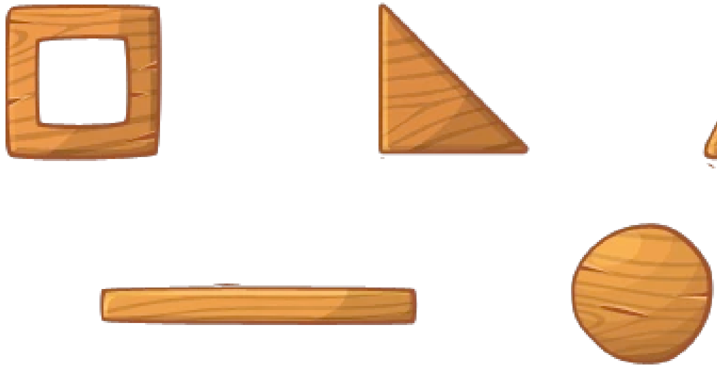
Consider the full decision tree of an example game below. The edges show the actions an agent may take and the state it will end up in. Additionally, below the terminal states the rewards are visible. The current game state is 1. The random policy is a *RightLeftPolicy*, which chooses for every odd iteration (episode) the right branch in the tree and for every even iteration (episode) the left branch of the tree.

Compute $k = 2$ episodes per action of *Flat Monte Carlo* on the given tree.

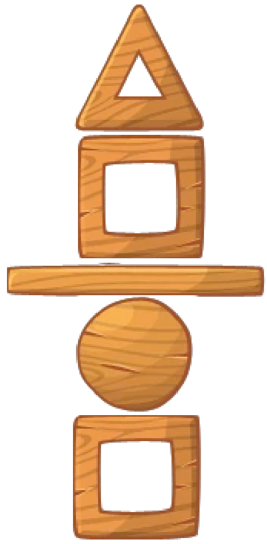


Assignment 6 Procedural Content Generation (2 + 2 + 2 + 2 = 8 points)

Suppose you want to generate a stack of blocks that does not topple over. The following picture shows the five available tiles and an example stack:



Blocks to be used to build a stack.



Example of a stack.

We plan to procedurally generate stacks of block using an evolutionary algorithm. Each stack should consist of at least 5 but not more than 10 blocks.

a) Describe a representation (encoding of solution candidates) to be used in such an EA. Show how your representation works by encoding the stack shown above.

b) Describe a mutation operator that suits your representation. Provide a graphical example by showing how a stack is mutated.

c) Describe a crossover operator that suits your representation. Provide a graphical example showing the result of a crossover of two stacks.

d) To make the stacks more interesting, each stack should consist of at least 3 different block types. Identify which of the EA's components can be adapted to ensure this constraint.