Exam Dynamic Programming & Reinforcement Learning
February 2023

This exam consists of **4** problems, each consisting of several questions.
All answers should be motivated, including calculations, formulas used, etc.
The minimal grade is 1. All questions give 0.5 points when answered correctly.
You are only allowed to use pen and paper.

1a. Formulate the dynamic programming recursion for solving the shortest path problem and give an interpretation of the value function.
Consider the following instance: There are 5 nodes, numbered 1 to 5, between every node there is a connection, with length given by $d(i,j) = 2 + (|i-j|-1)^2$.
b. Make a drawing of the graph with the lengths of the connections.
c. Find the shortest path from 1 to 5 using dynamic programming and give a table with all intermediate results. What is the shortest path?

2. A company has 2 identical machines of which it wants at least 1 to be operational at any time. Machines can fail and it takes a geometrically number of time units with success probability 1/2 to repair a machine, meaning that a machine which is being repaired turns to the working state with probability 1/2 at the end of the time period. The operational machine breaks at the end of a period with probability 1/2. There is a single repairman who works when there is a broken machine. We formulate this as a discrete-time Markov reward chain.
a. Choose an appropriate state space and determine the transition probabilities.
b. Formulate and solve the equations to determine the stationary distribution.
A time period without a working machine costs 10, repairing costs 1 per time period.
c. Determine the long-run average costs using the solution of a.
d. Formulate the Poisson equation and use it to determine the long-run average costs.
e. At any moment the company can replace the repairman by another which has success probability 2/3 but costs 2 per time unit. Formulate the Bellman equation for the problem that determines when to use which repairman.

3a. Give the definition of the Gittins index.
A Markov-reward chain has 3 states, transitions $p(2|1) = p(3|2) = p(3|3) = 1$ and rewards $r(1) = r(2) = 0$ and $r(3) = 1$.
b. Calculate the Gittins index using the definition for each state and discount factor.
Another Markov-reward chain also has 3 states, transitions $p(2|1) = p(3|1) = 0.5$, $p(2|2) = p(3|3) = 1$ and rewards $r(1) = r(3) = 0$ and $r(2) = 2$.
c. Calculate the Gittins index using the method with the restart option for state 1 and discount factor 0.5.
d. Consider a multi-armed bandit with one arm of each type, both starting in state 1, and discount factor 0.5. Describe the optimal policy.
e. What is its total expected discounted reward?

4a. Cite 3 elements that makes the RL setting general in the sense that it can represent basically all sequential decision-making problems (tip: think about the simplest MDP possible and all the elements that can make it more complex)
b. Why is it not possible to solve the game of chess with simple MCTS algorithms (without any value function estimator)?
c. The Bellman equation that is at the core of reinforcement learning makes use of the fact that the Q-function can be written in a recursive form. Can you write this Bellman equation? (Make sure that you define all the terms).
d. Describe in words the ”optimistic values” exploration strategy.
e. What is the reinforce algorithm?