Exam Dynamic Programming & Reinforcement Learning
December 2022

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.

1. Consider the knapsack problem: for a finite set of items S, each with a reward r_i and a weight w_i, $i \in S$, you have to find the subset that maximises the reward while the sum of the weights remains below a given upper bound W.
a. Formulate a dynamic programming recursion to solve this problem. Motivate your choice of state space.
b. Use this recursion to solve the following instance: S consists of 5 items with $w = (1, 2, 3, 4, 5)$, $r = (2, 3, 4, 5, 6)$, and $W = 8$.
c. Suppose that, next to a constraint on the sum of the weights, each item has a volume and there is also a constraint on the sum of the volumes. Formulate a dynamic programming for this problem. Motivate your choice of state space.

2. Consider a Markov chain with states $\{1, 2, 3\}$ and transition probabilities $p(3|1) = p(3|2) = 1$ and $p(1|3) = 1 - p(2|3) = 1/3$. (Remember that $p(y|x)$ is the probability of going from x to y.)
a. Is this chain communicating and/or periodic? Does the limiting distribution exist? Motivate your answer.
b. Formulate the set of equations to find the stationary distribution and solve them.
c. Let $r(x) = 3 - x$. Formulate the Poisson equation for the average case and give all solutions.
d. Let $r(x) = 3 - x$. Formulate the Poisson equation for the discounted case, for arbitrary discount factor. (Do not solve this set of equations.)
Now transition times are random: they are all exponential with $\tau_x = x$.
e. Does the time-limiting distribution exist? Motivate your answer.
f. Give the stationary distribution using an appropriate formula.

3. Consider a Markov decision chain with states $\{1, 2, 3, 4\}$ and 2 actions in every state. The transition probabilities are $p(2|1, 1) = p(3|2, 1) = p(4|3, 1) = p(4|4, 1) = 1$, $p(1|1, 2) = p(1|2, 2) = p(2|3, 2) = p(3|4, 2) = 1$ and the direct rewards $r(1, 1) = 3$, $r(2, 1) = 2$, $r(3, 1) = 1$, $r(4, 1) = 0$, $r(1, 2) = 0$, $r(2, 2) = 0$, $r(3, 2) = 2$, and $r(4, 2) = 2$. (Remember that $p(y|x, a)$ is the probability of going from x to y under action a, $r(x, a)$ is the direct reward in x under action a.)
a. Draw the state-transition diagram.
b. Solve the average-reward Poisson equation for the policy that applies action 1 everywhere.
c. Apply policy iteration starting from the solution found under b.
d. Formulate the Bellman equation and show that the final answer of c is a solution.
e. Give the Q-values of the solution of d.

4a. Cite three different elements that can play a role in the generalization of RL algorithms.
b. How can heuristic evaluators be used to improve simple MCTS algorithms (2 elements)?
c. In the finite state space and finite action space setting, what are the (two) conditions that ensure convergence?
d. Describe the difference between off-policy and on-policy learning algorithms.