DS 598 HW1

Write your name here

January 24, 2024

Instructions: Please write your solution in latex and submit the compiled PDF to the blackboard submission portal. You can use the latex source file for each HW assignment as a starting point.

Problem 1. Calculate the V^* function for the following MDP. Each grid represent a state and there are 4 actions in each state traveling to the 4 adjacent states respectively. Numbers in the grid represent reward for getting into the grid. Let $\gamma = 0.9$ be the discounting factor.

		4	
	8		1
3			

Figure 1: Problem 1 MDP.

Problem 2. During the lecture, we derived the Bellman Equation (BE) for V^π and Bellman Optimality Equation (BOE) for V^*. Derive the Bellman equation for Q^π and Q^*. You can use the BE and BOE for V functions as a starting point.

Problem 3. In the lecture we've proved that $V^* = V^{\pi^*}$, where π^* is defined as

$$\pi^*(s) = \arg\max_a \left[r(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)} V^*(s') \right]$$

Prove that $Q^* = Q^{\pi^*}$. You can use the established equality for V as a starting point.

Problem 4. Prove that the Value Iteration (VI) algorithm converges, i.e. show that the Bellman Optimality Equation satisfies the contraction property.

Problem 5. Calculate the occupancy measure d^{π}_{μ} for the following MDP: There are 4 states S_1, S_2, S_3, S_4. μ is a point-mass on S_1, i.e $\mu(S_1) = 1$ and $\mu(S_2) = \mu(S_3) = \mu(S_4) = 0$. The transition probability following π is indicated by the arrow and numbers in the plot. The discounting factor is set to $\gamma = 0.9$.

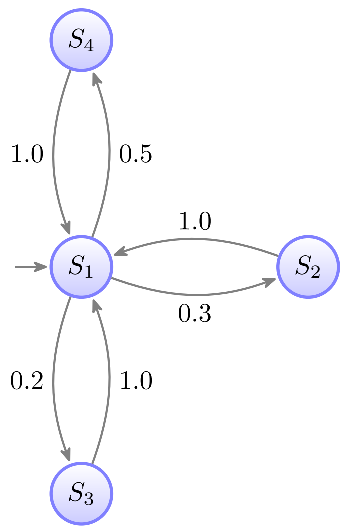

Figure 2: Problem 5 MDP.