COMP3702/COMP7702 Artificial Intelligence
Semester 2, 2020
Tutorial 8

Before you begin, please note:

- Tutorial exercises are provided to help you understand the materials discussed in class, and to improve your skills in solving AI problems.
- Tutorial exercises will not be graded. However, you are highly encouraged to do them for your own learning. Moreover, we hope you get the satisfaction from solving these problems.
- The skills you acquire in completing the tutorial exercises will help you complete the assignments.
- You'll get the best learning outcome when you try to solve these exercises on your own first (before your tutorial session), and use your tutorial session to ask about the difficulties you face when trying to solve this set of exercises.

Exercises

Exercise 8.1. UQCarRental is opening its business!!! Their first order of business is to buy cars that customers can rent. To this end, they have a choice between buying 2 Tesla Model X or 5 GenCar.

A **GenCar** costs \$40,000. Since UQCarRental already has a wide customer base for this type of car, they know that all GenCar will be rented out for \$175 per day per car for 330 days per year with certainty. When a customer rents the car, the cost for maintenance and fuel that UQCarRental must pay is \$25 per day per car. When the car is not being rented (i.e., 35 days in a year), the GenCar cars would be in a mechanic repair shop for minor repairs, which cost \$30 per car per day.

A **Tesla Model S** costs \$120,000, UQCarRental does not have any information about customer's desirability yet. However, based on the market price, they plan to rent out each Tesla for \$500 per day. When a customer rents a Tesla, the daily maintenance cost charged to UQCarRental is \$10 per day per car. When it is not being rented out, the cost is \$5 per day per car. UQCarRental is expecting each Tesla to be in a mechanic repair shop and undergo minor repairs for 35 days in a year, and costs \$30 per car per day.

Suppose UQCarRental is deciding which cars to buy based only on the first year profit after buying the cars. Using the concept of **Maximum Expected Utility** and conservative estimate of the probability that customers will rent Tesla Model X, please answer the following question:

Suppose a survey reveals that the probability that each Tesla is rented for 330 days per year is 75%, and for a conservative estimate on expected utility, assume that the rest of the probability mass will be a Tesla car is not rented for the entire of the 330 days. Also, you learn that Tesla offers an upgrade for \$20,000 that allows you to rent out the Tesla car for \$600 per car per day with no effect on its demand. Should you buy: (i) 5 GenCars, (ii) 2 Tesla Model X without modification, or (iii) 2 Tesla Model X with the modification?

Exercise 8.2. In general, an MDP's **objective** is:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right].$$

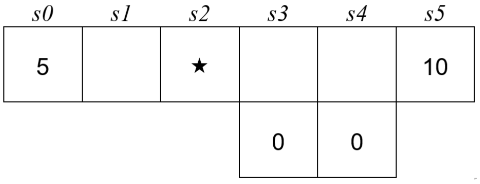
The **value function** of an MDP, $V^\pi(s)$, is the expected future cost of following an (arbitrary) policy, π, starting from state, s, given by:

$$V^\pi(s) = \sum_{s' \in \mathcal{S}} P(s' \mid \pi(s), s)\left[R(s, \pi(s), s') + \gamma V^\pi(s')\right].$$

where the policy $\pi(s)$ determines that action taken in state s. Here we have dropped the time index, as it is redundant, but note that $a_t = \pi(s_t)$. Also note that $R(s,a) = \sum_{s'} P(s' \mid s, a) R(s, a, s')$.

Question: Derive V^π from the MDP objective function.

Exercise 8.3. Consider the gridworld below:



An agent is currently on grid cell s_2, as indicated by the star, and would like to collect the rewards that lie on both sides of it. If the agent is on a numbered square (0, 5 or 10), the instance terminates and the agent receives a reward equal to the number on the square. On any other (non-numbered) square, its available actions are to move Left and Right. Note that Up and Down are never available actions. If the agent is in a square with an adjacent square below it, it does not always move successfully: when the agent is in one of these squares and takes a move action, it will only succeed with probability p. With probability $1 - p$, the move action will fail and the agent will instead fall downwards into a trap. If the agent is not in a square with an adjacent space below, it will always move successfully.

a) Consider the policy π_R, which is to always move right when possible. For each state $s \in \{s_1, s_2, s_3, s_4\}$ in the diagram above, give the value function V^{π_R} in terms of $\gamma \in [0, 1]$ and $p \in [0, 1]$.

b) Consider the policy π_L, which is to always move left when possible. For each state $s \in \{s_1, s_2, s_3, s_4\}$ in the diagram above, give the value function V^{π_L} in terms of γ and p.

Exercise 8.4. UQFruit is a very small fruit stall at UQ. It sells two types of fruit: **apples** and **bananas**. It buys its stock each morning, but it can only buy at most 3 pieces of fruit per day, and can only store and sell 4 pieces of fruit per day. Despite its size, UQFruit wants to build a good reputation by stocking the fruit such that it can minimize the number of customers who could not get their choice of fruit. You can assume that:

1. Apples and bananas are the only two types of fruit that the customer may want to buy.
2. The customer will not change its preference and will leave UQFruit without buying anything if his/her choice of fruit is not available.
3. The fruit will always be in a good condition.

Suppose the customer's preference depends on the amount of each type of fruit available at the beginning of the day (right after UQFruit buys its stock), and are modeled by the following probability transition function: P_a for apples and P_b for bananas.

$$P_a = \begin{bmatrix} 0.3 & 0.2 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.1 & 0.2 \end{bmatrix}$$

$$P_b = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.2 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.2 & 0.2 & 0.1 & 0.3 \\ 0.2 & 0.2 & 0.2 & 0.1 & 0.3 \end{bmatrix}$$

The element at row-i and column-j, where $i, j \in 0, \ldots, 4$, of the matrix P_a represents the probability that there will be j customers that want to buy an apples given there are i apples at the beginning of the day. This meaning also applies respectively to the elements of P_b. **Question**: Please define an MDP problem to represent the stocking problem that UQFruit faces each morning, by specifying its: state space, action space, transition function, reward function, and γ (discount factor).

(*Next week we will work through solving this MDP problem using value iteration, policy iteration and MCTS.*)