Exercises Lecture 4
Introduction to Artificial Intelligence (IAI)

Exercise 1: Bellman equation

Equation 16.5 in Chapter 16 of RN21 shows the Bellman equation $U(s)$, which defines agent-environment interaction in Markov decision Processes. There are, however, several variations of the Bellman equation.

a) Write the Bellman equation for $U^{\pi}(s)$ that defines the utility of state s when following policy π. Explain the difference between $U(s)$ and $U^{\pi}(s)$.
b) Explain how the Bellman equation for $U(s)$ would change if the Markov property would not hold.
c) Explain the difference between the utility function $U(s)$ and the Q-function $Q(s,a)$.

In the lecture, we focused on reward functions on the form $R(s,a,s')$. This is, however, not the only way to model rewards. We could, for instance, just make the reward depend on the current state and applied action $R(s,a)$. This would change the Bellman equation.

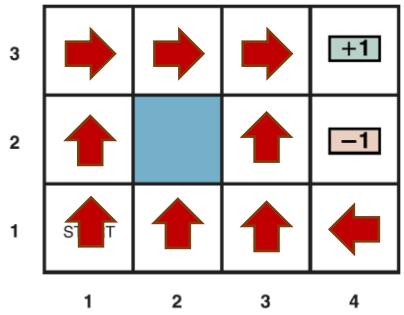
d) Show that the following two expressions are equal (hint: this only requires math):

$$\max_{a \in A(s)} \sum_{s'} P(s'|s,a)[R(s,a) + \gamma U(s')]$$

$$\max_{a \in A(s)} R(s,a) + \gamma \sum_{s'} P(s'|s,a) U(s')$$

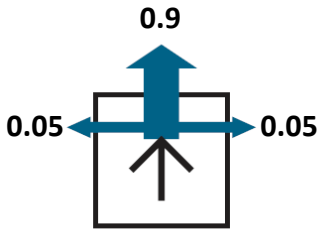
Exercise 2: Grid world (adapted from RN13)

Consider the state and action space of the 4×3 grid world of Figure 16.1 in RN21 Assume a deterministic transition function. Assume initial state (1,1), and the terminal states (4,3) and (4,2) with reward +1 and -1 respectively. Any other state has a reward of -0.05.

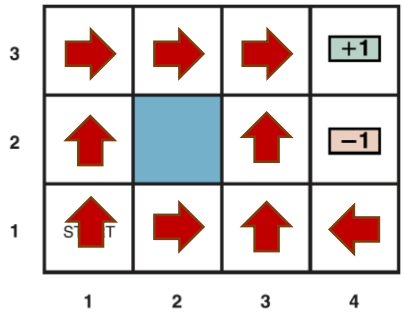


a) Given the policy shown above. Calculate the utility function of every state with $\gamma = 1$.
b) Given the policy shown above. Calculate the utility function of every state with $\gamma = 0.99$.

Replace the deterministic transition function by a stochastic transition function. The agent executes the intended action with probability 0.9, but with probability 0.1 the agent moves at right angles to the intended direction. See the illustration below. Note that moving into a wall causes no movement at all.



c) Given the policy shown above. Let the agent perform two transitions. What are the reachable states? What is the probability of reaching each of these states? (Hint: make a complete tree of reachable states two time steps deep and track the probability of each state in level 2).
d) (Coding task) Given the stochastic transition function and the new policy for each state shown below. Calculate the value function for all states with $\gamma = 0.99$.



Mandatory exercise: Markov decision process (adapted from SB18 example 4.2)

Jack manages two locations for a car rental company. Each day, customers arrive at each location to rent cars. If Jack has a car available, he rents it out to earn $10. If he is out of cars at that location, then he loses revenue. Cars become available for renting the day after they are returned. To help ensure that cars are available where they are needed, Jack can move them between the two locations overnight, at a cost of $2 per car moved. Let's assume that the daily number of rentals is 3 and 4 cars at location 1 and 2 respectively, while the daily number of returns is 3 and 2 cars at location 1 and 2 respectively.

The maximum number of cars in each location is 20.

Express the car rental company as a Markov decision process.

a) Formulate the state and action space (Hint: both the state and the action are a tuple)
b) Define the transition function and reward function.
c) Explain whether it is likely that cars will be moved from location 2 to 1, and whether it is likely that Jack can keep up with total rental demand without losing business.