Introduction to AI

Assignment 2

March 28, 2023

1. Please recall the robot runner in the grid world example with tiger and food cells on right last column. The states are grid cells, i.e., $(1,1),(1,2),\ldots$. First component of a state is the row number, second component is the column number. The actions available are North, East, West, South. Rewards are given as follows:

   - $R(*,*,*) = -0.1$ (Penalty of movement, assuming result state is not food or tiger)
   - $R(*,*,(3,4)) = +2$ (Moving to food cell)
   - $R(*,*,(2,4)) = -1$ (Moving to tiger cell)

   Transition probability is 0.5 to the state in the direction of the action and 0.25 in states perpendicular to the direction of the action. If agent hits a wall, the agent moves back to its original location. Terminating states are when agent is in the food cell or in the tiger cell. Discount factor is 0.95. For value iteration method, the values of different states $V^t(.)$ at an iteration **t** are given by:

3	0.653	1.059	1.381	Food
2	0.400		0.434	Tiger
1	0.082	−0.11	−0.00	−0.35
	1	2	3	4

   (a) Compute the values of $V^{t+1}(.)$ for different states at iteration $t+1$. The V^{t+1} values for some states are given in the table below. You need to compute the values for states with "?" entry. Write numerical answers for such cells in the table below.

3	0.653	1.059	1.381	Food
2	0.400		0.434	Tiger
1	0.082	?	?	?
	1	2	3	4

   Please show analytical expressions denoting all computations without using any python code.

   For state $S = (1,3)$,

$$\begin{aligned} Q^{t+1}((1,3),N) &= \sum_{s'} P(s'|(1,3),N)[R((1,3),N,s') + \gamma V^t(s')] \\ &= .5(-.1 + .95(.434)) + .25(-.1 + .95(-.11)) + .25(-.1 + .95(-.35)) \\ &= -0.00435 \\ Q^{t+1}((1,3),W) &= \sum_{s'} P(s'|(1,3),W)[R((1,3),W,s') + \gamma V^t(s')] \\ &= .5(-.1 + .95(-.11)) + .25(-.1 + .95(.434)) + .25(-.1 + .95(0)) \\ &= -0.0875 \end{aligned}$$

   Taking action S and E are trivial, as S is worse than N because the successful action is greater for N than S and the unsuccessful action are the same. By the same logic, action W is worse

than E. Hence,

$$V^{t+1}(1,3) = -0.00435$$

For state $S = (1,4)$,

$$\begin{aligned} Q^{t+1}((1,4),W) &= \sum_{s'} P(s'|(1,4),W)[R((1,4),W,s') + \gamma V^t(s')] \\ &= .5(-.1 + .95(0)) + .25(-1) + .25(-.1 + .95(-.35)) \\ &= -0.4275 \\ Q^{t+1}((1,4),S) &= \sum_{s'} P(s'|(1,4),S)[R((1,4),S,s') + \gamma V^t(s')] \\ &= .5(-.1 + .95(-.35)) + .25(-.1 + .95(0)) + .25(-.1 + .95(-.35)) \\ &= -0.2075 \end{aligned}$$

Taking action N and E are trivial, as N is the worst move as it goes to the Tiger. Since going W is worst than S, we can also conclude that E is worse than S as E has a chance of meeting the Tiger which automatically decreases our utility by 0.25. Hence,

$$V^{t+1}(1,4) = -0.2075$$

For state $S = (1,2)$,

$$\begin{aligned} Q^{t+1}((1,2),W) &= \sum_{s'} P(s'|(1,2),W)[R((1,2),W,s') + \gamma V^t(s')] \\ &= .5(-.1 + .95(0.082)) + .25(-.1 + .95(-.11)) + .25(-.1 + .95(-.11)) \\ &= -0.1236 \end{aligned}$$

Taking action W is the only action with a positive successful state. The other actions have negative successful states.

$$V^{t+1}(1,2) = -0.1236$$

Hence, at V^{t+1}

3				Food
2		■		Tiger
1		−0.1236	−0.00435	−0.2075
	1	2	3	4

(b) What is the policy for each state as per the function $Q^{t+1}(..)$. Note it down below for each entry with "?" mark.

3	E	E	E	Food
2	N	■	N	Tiger
1	N	W	N	S
	1	2	3	4

From the Q^{t+1} I've calculated in the previous part, the optimal policy is shown in the diagram above.