

**Hint: L: Left, R: Right, U: Up, and D:Down**

Q1) The following table shows the reward function for a 3x3 square environment.

-5	0	10
0	-5	0
0	0	-5

Assume the agent is using the shortest path policy to the top right corner (value 10). Assume that the transition function is always leading to one state. For example: if an agent moves up from any location, it will always go up. and the following table shows the possible actions at each state.

D   R	D   R	-
U   R	D   L	D   L
U   R	U   R	U   L

Calculate the Value function using the bellman equation for the following 3x3 environment using a discount factor of 0.9.

Q2) An agent is set in the bottom left corner and moving in a 3x3 square environment.

0	0	+1
0	-1	0
Start	0	-1

The above table is the Reward function for a certain environment. Assume that the agent actions happen 100% of the time. For example: if the agent moves up from the start, it will always go up. The agent performed the following actions (R, R, U, L, U, R), assume the agent is using TD to update its Value function using 0.1 learning rate and 0.9 discount factor. Update the following V function state in the following table.

0.8	0	0.9
0.4	0.2	0.3
0.2	0.5	0.1