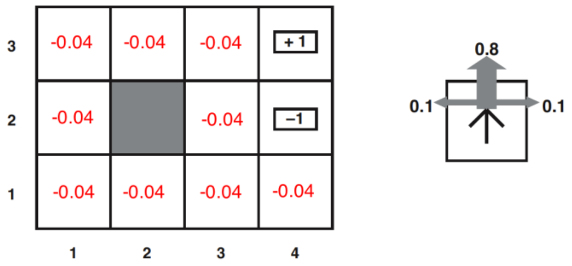
Worksheet 8 - copied from Martha Lewis

Q1: Gridworld and Value Iteration

Consider the gridworld in Figure 1.



3	-0.04	-0.04	-0.04	+1
2	-0.04		-0.04	−1
1	-0.04	-0.04	-0.04	-0.04
	1	2	3	4

Figure 1: A gridworld!

(a) Do we need to wait until the algorithm has converged until we know the utility of each state, or are there some states whose utility we already know?

(b) Suppose we initialise the utility of every state to 0, and then perform one iteration of the value iteration algorithm. What is the utility of each state?

Q2: 3x3 Gridworld Policy

Consider the 3x3 gridworld shown in Table 1.

r	-1	10
-1	-1	-1
-1	-1	-1

Table 1: 3x3 Gridworld

The transition model is as follows: 80% of the time the agent moves in the intended direction; the remaining 20% of the time, it moves perpendicularly (10% left, 10% right). Given different values of r, determine the optimal policy using discounted rewards with $\gamma = 0.99$.

(a) $r = -3$

(b) $r = +3$

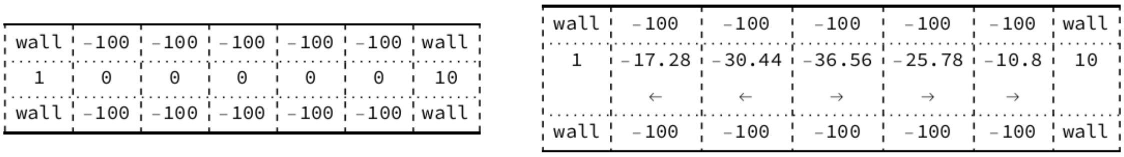
Q3: Bridge Crossing Problem

Figure 2 shows a narrow bridge gridworld where a robot must cross safely.

(a) Using a discount value of 0.9, calculate the utility of each non-terminal grid square after one and two moves.

(b) The optimal policy in Figure 2 fails to cross the bridge. What is the effect of decreasing the discount value?

(c) What is the effect of increasing the utility of the goal? Choose a new value so that the optimal policy is to cross the bridge, and show the utility of each grid square after three iterations.



wall	−100	−100	−100	−100	−100	wall
1	0	0	0	0	0	10
wall	−100	−100	−100	−100	−100	wall

wall	−100	−100	−100	−100	−100	wall
1	−17.28 ←	−30.44 ←	−36.56 →	−25.78 →	−10.8 →	10
wall	−100	−100	−100	−100	−100	wall

Figure 2: Bridge crossing(a) rewards for the bridge-crossing problem in gridworld. (b) utilities after 5 iterations, and the corresponding optimal policy