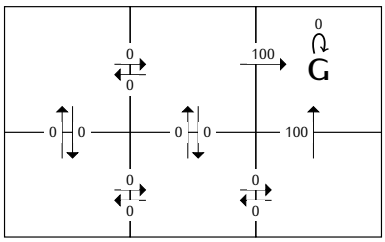
Reinforcement learning

Exercise 1



Immediate reward function r

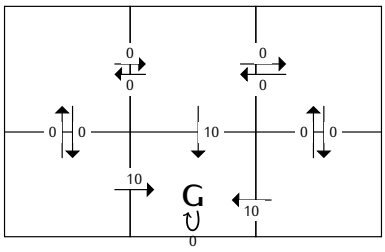
We use a discounted factor of $\gamma = 0.9$.

Compute the function $Q(s, a)$ for each pair of state and action.

Compute the function $V^\star$ for each state.

Provide multiple optimal strategies.

Exercise 2



Immediate reward function r

1. We repeat the exercise for the above situation with a discount factor $\gamma = 0.8$: provide $Q(s, a)$ for each transition, $V^\star$ for each state, and an optimal policy.
2. Suggest a change to the reward function that alters the Q values but does not alter the optimal policy.
3. Suggest a change to the reward function that alters Q but not $V^\star$
4. Consider we use Q-learning to this environment, assuming we initialise the table of $\widehat{Q}$ values to zero. Assume the agent begins at the bottom left and travels clockwise around the grid until it reaches G. Describe which $\widehat{Q}$ are modified and what is their values after one episode. Do the same for a second, third episode.

When an agent chooses action a in state s, it moves to state s', the update rule is

$$\widehat{Q}(s, a) \leftarrow r(s, a) + \gamma \max_{a'} \widehat{Q}(s', a')$$