Homework 3: Gridworld

劉安得

106061151

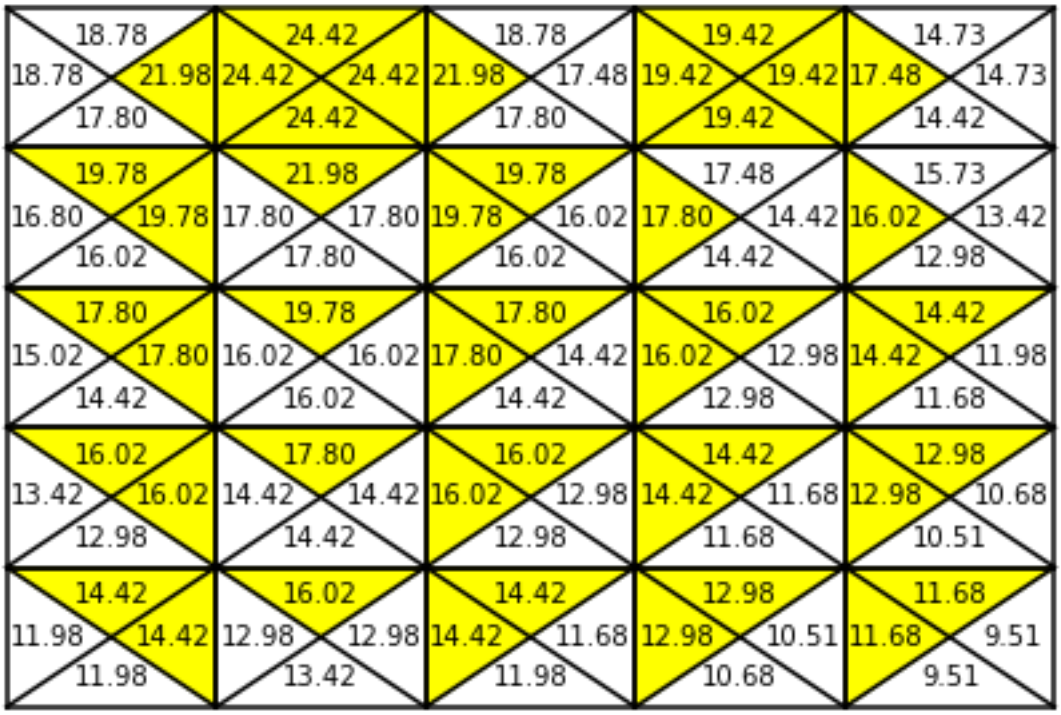
1. Implementation
   - evaluateActionValue()

     Implement bellman equation for q*(*s*, *a*)

$$
\begin{aligned}
q_*(s,a) &= \mathbb{E}\left[R_{t+1} + \gamma \max_{a'} q_*(S_{t+1}, a') \;\middle|\; S_t = s, A_t = a\right] \\
&= \sum_{s',r} p(s', r \mid s, a)\left[r + \gamma \max_{a'} q_*(s', a')\right],
\end{aligned}
$$

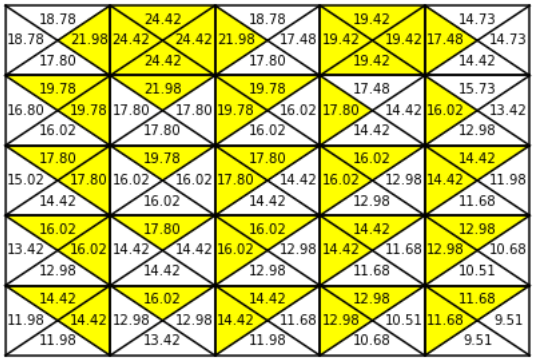
```
for i in range(WORLD_SIZE):
    for j in range(WORLD_SIZE):
        for k in range(len(ACTIONS)):
            next_state, reward = step([i, j], ACTIONS[k])
            new_q_value[i, j, k] = reward + DISCOUNT *
np.max(q_value[next_state[0], next_state[1]])
```

2. Experiments and Analysis



I think q_values is reasonable because A and B has the highest action value and other grids seems that it inclines moving forward to A or B.

- Compare w/ HW2.2.



	1	2	3	4	5
1	21.98	24.42	21.98	19.42	17.48
2	19.78	21.98	19.78	17.8	16.02
3	17.8	19.78	17.8	16.02	14.42
4	16.02	17.8	16.02	14.42	12.98
5	14.42	16.02	14.42	12.98	11.68

In HW2.2, we plot the table of state value function; and in this homework, we plot the table of action value function. The two results match with each other. The highest action value in each state in action value table is the state value in state value function. And the highest action value indicates which action is the optimal policy.