Homework

- Modify the gridworld you created last week to create a 4x4 version
  - No obstacles
  - Rewards of -1 on all transitions
  - Goal in the top left corner – entering the goal state ends the episode
- Given this environment and a uniform random policy, implement 2 versions of policy evaluation
  - The in-place version, as presented in the book
  - A two-array version, which only updates the value function after looping through all states (see pg 75)
  - Use a threshold value of $\theta = 0.01$
  - NB: Careful of how you handle the terminal state!!
- For a given γ, record the number of iterations of policy evaluation until convergence

By next week's lecture, submit on Moodle (groups of up to 4):

1. A 2d heatmap plot of the value function for $\gamma = 1$
2. A combined plot of both versions of policy evaluation for different discount rates
   1. The x-axis should be the discount rate. The range of discounts should be specified by np.logspace(-0.2, 0, num=20)
   2. The y-axis should be the number of iterations to convergence
3. Your code