CMSC 478 Machine Learning - Spring 2019
Homework Assignment 5
Due at the start of class on April 9^{th}

1. (Reinforcement Learning) Implement Q-learning (with the programming language of your choice) and use it to find an optimal policy for traversing an NxN grid world with 4 actions that move one step in the standard compass directions. Each episode starts with the learner in the upper left corner state and ends when the learner enters the lower right corner state. Attempts to move off the edges of the grid world should result in the agent staying in the same state. Choose any reward structure that causes the learner to minimize the number of steps required to end the episode. Turn in the following:

    - Your code
    - List of the parameters used in learning, such as rewards, discount factor, learning rate, exploration probability, and anything else that seems relevant
    - A learning curve that shows the number of steps until the goal state is reached on the vertical axis and the episode number on the horizontal axis. Generate learning curves for $N = 5, 10, 15$. Write a brief description of what happens to the curves as N increases and explain why it happens.
    - Try a second reward structure, describe it, and generate the learning curve using that structure for $N = 15$. Which of the two that you tried is best and why?