1. Problem Statement

We have been given a maze environment. The transition model for which is as follows: the intended outcome occurs with probability 0.8, and with probability 0.1 the agent moves at either right angle to the intended direction. If the move would make the agent walk into a wall, the agent stays in the same place as before. The rewards for the white squares are -0.04, for the green squares are +1, and for the brown squares are -1. Note that there are no terminal states; the agent’s state sequence is infinite.

With the given transition model and the reward function, the optimal policy, and the utilities of all the (non-wall) states using both value iteration and policy iteration must be calculated. In addition to this, the optimal policy, and the utilities of all the states are to be displayed, and the utility estimates as a function of the number of iterations are to be plotted. The discount factor of 0.99 is to be used for the purpose of this question.

1. Organization of the code
2. Assumption

For the purpose of this assignment, it has been assumed that the agent can intend to move towards a wall or the boundary of gird. In these cases, it will try to move towards the wall (or boundary) with probability of 0.8 and to the left or right with probabilities of 0.1 each.

1. Value Iteration
   1. Set up
   2. Running Value Iteration
   3. Results
   4. Plots
2. Policy Iteration
   1. Set up
   2. Running policy iteration
   3. Results
   4. Plots
3. Complex grid environment