CSC 246/446 - Project 4 (http://www.cs.rochester.edu/~stefanko/Teaching/09CS446)

4.1 (**due Apr. 23, 2009**) Consider the following MDP

- the states are $S = \{1, 2, \ldots, 11\} \times \{1, 2, \ldots, 11\}$.
- the reward function is given by $R((6, 6)) = 1$, and $R((x, y)) = 0$ for $(x, y) \neq (6, 6)$
- the discount factor is $\gamma = 0.99$.
- in each state there are two actions:
  - horizontal: if the state is (x, y) then the next state is $(\max\{1, x - 1\}, y)$ with probability $1/2$ and $(\min\{11, x + 1\}, y)$ with probability $1/2$.
  - vertical: if the state is (x, y) then the next state is $(x, \max\{1, y - 1\})$ with probability $1/2$ and $(x, \min\{11, y + 1\})$ with probability $1/2$.
- compute the optimal policy using both value iteration and policy iteration (solving a system of linear equations in the "value-computation phase").
- Give an 11×11 matrix filled with H's and V's containing the optimal policy.
