# CSE 4309/5361 - *Artificial Intelligence II*

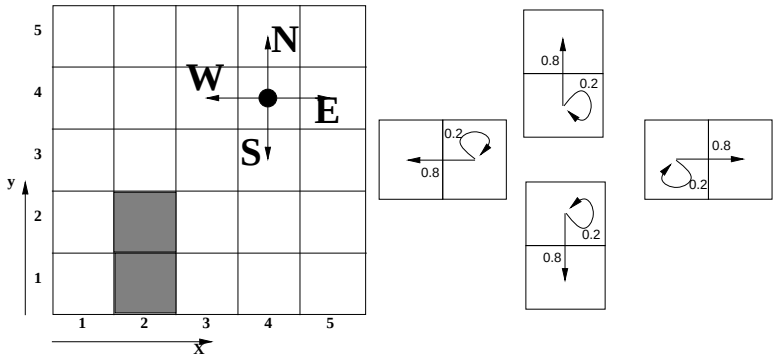
Homework 3- Spring 2013

Due Date: April 11, 2013

Note: Problems marked with * are required only for students enrolled in CSE 5361. They will be graded for students enrolled in CSE 4309 for extra credit.

## Markov Decision Problems

Consider a similar navigation scenario as in the second assignment where an agent moves in a grid world and the agent's actions succeed only probabilistically. In particular, as in the second assignment, the agent's $North$, $South$, $East$ and $West$ actions each succeed with probability $0.8$ and with probability $0.2$ the agent slips and as a result stays in the same place (without noticing that it did so). In contrast to the second assignment, however, the agent here has a GPS-like sensor that tells it exactly which grid location it is in, and there are obstacles in the environment that the agent can not cross.




1. Assuming that the agent knows where the obstacles are and where the goal is located, it can use Value iteration to compute the optimal utility function for a given reward structure and discount factor. For this problem, consider the reward function where the agent receives a reward of $100$ when it reaches the goal and a reward of $-10$ when it hits an obstacle.

   a) Implement Value iteration to compute the optimal utility function, $U^*(s)$, for the given reward function and a given obstacle and goal configuration in a 20x20 grid world. Show the resulting utility function for three different discount factors.

   b) Implement a rational agent that uses the utility function derived using value iteration to optimally navigate to the goal.

   c)* Change the obstacles such that they can be crossed (e.g. as in the case of water obstacles that can be swum through - but at an additional cost indicated by the negative reward). Use Value iteration to compute the corresponding utility function for discount factors of 0.5 and 0.99 in environments with significant numbers of obstacles and show the resulting utility functions and optimal policies. Briefly discuss what difference the treatment of the obstacles and the discount factor make in terms of the learned policy.

## Q-Learning

2. Considering the same world as in Problem 1 a) and b) but consider an agent that initially does not know where the goal and the obstacles are.

   a) Implement a Q-learning agent that can learn the optimal Q-value function for an arbitrary, fixed obstacle and goal configuration under these conditions.

   b)* Apply the same agent to the scenario of Problem 1 c).