Homework 4: Reinforcement Learning and Statistical Learning

Due December 1st, 2010.

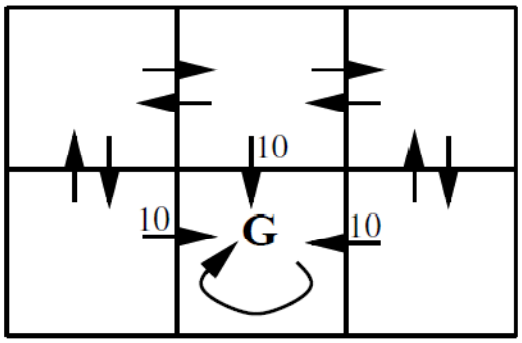
1. (Adapted From T. M. Mitchell, *Machine Learning*, problem 13.2.) Consider the deterministic grid world shown below with the absorbing goal state **G**. Here the immediate rewards are 10 for the labeled transitions and 0 for all unlabeled transitions. Let the discounting factor $\gamma = 0.8$.

(a) Give the V^* value for every state in this grid world.

(b) Give the $Q(s, a)$ value for every transition in this grid world.

(c) Show an optimal policy π^*—i.e., the optimal action from each state.

(d) Now consider applying the Q-learning algorithm to this grid world, assuming the table of $\hat{Q}$ values is initialized to zeros. Assume the agent begins in the bottom left grid square and then travels clockwise around the perimeter of the grid until it reaches the absorbing goal state, completing the first training episode. Describe which $\hat{Q}$ values are modified as a result of this episode, and give their revised values. Answer the question again assuming the agent now performs a second identical episode. Answer it for a third episode.



. Consider the following Bayesian network: $A \rightarrow B \rightarrow C$. And the following data table, with entries '?1' and '?2' missing at random:

A	B	C
F	F	F
F	F	?1
F	T	F
T	T	T
T	?2	T
T	F	T

(a) Use the data to estimate initial parameters for this network, using maximum likelihood estimation for simplicity.

(b) Apply the EM algorithm (by hand) to estimate the values of the missing data, reestimate the parameters, etc. until convergence. Show your calculations.

(c) How many iterations does EM take to converge? Will this always be the case? Explain.

3. Show that adding additional edges to a bayes network does not decrease the likelihood of the data. (If you can't show this for general case, you can try comparing likelihood of a specific data with the network in the problem above, and a version of it with a directed link from A to C)