

# CPSC 422

## Practice Midterm Exam Questions

March 2007

This is meant to give some example questions. This is much longer than the exam. The exam may be nothing like this. But if you can do this, you should have no problems with the exam.

**You may bring in one letter sized piece of paper with anything written on it. You may not use calculators, PDAs, robotic assistants or other electronic aids.**

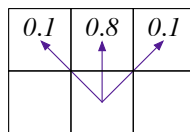
Some important points (that students often forget):

- Read and answer the question. You will not get marks for writing things (whether they are true or not) that are not relevant to the question.
- Use proper English in full sentences. You will not get marks if we cannot work out what you are saying.
- If a question asks about a particular instance of a problem, make sure your answer refers to that instance. Writing a general formula that you may have copied from the sheet you can bring in, is not worth any marks. (The questions are usually asking to apply that formula to a particular case, to make sure you understand it).

1. Answer the following questions. Use proper English. Be concise in your answers. (You will lose marks for stating irrelevant facts). You must use your own words (text from the textbook or another source copied onto your crib sheet will not get any marks).
  - (a) Give the intuition behind the notion of a transduction? Why do we define causal transductions?
  - (b) Why does an agent have a belief state?
  - (c) Explain why we need a command function and a state transition function in a robot controller but not other functions.

2. Suppose our Q-learning agent, with fixed  $\alpha$ , and discount gamma, was in state 34 did action 7, received reward 3 and ended up in state 65. What value(s) get updated? Give an expression for the new value. (You need to be as specific as possible)
3. In temporal difference learning (e.g. Q-learning), to get the average of a sequence of k values, we let  $\alpha_k = 1/k$ . Explain why it may be advantageous to keep  $\alpha_k$  fixed in the context of reinforcement learning.
4. Explain what happens in reinforcement learning if the agent always chooses the action that maximizes the Q-value. Suggest two ways that can force the agent to explore.
5. In MDPs and reinforcement learning explain why we often use discounting of future rewards.
6. What is the main difference between asynchronous value iteration and standard value iteration? Why does asynchronous value iteration often work better than standard value iteration?
7. What is the relationship between asynchronous value iteration and Q-learning.
8. In feature-based reinforcement learning, what new data point does the experience  $\langle s, a, r, s', a' \rangle$  give for the linear regression? (I.e., what is the new data point for what value?)
9. Consider a  $5 \times 5$  grid game similar to the simple game used in assignments. The agent can be at one of the 25 locations, and there can be a treasure at one of the corners or no treasure.

In this game the “up” action has the dynamics given by

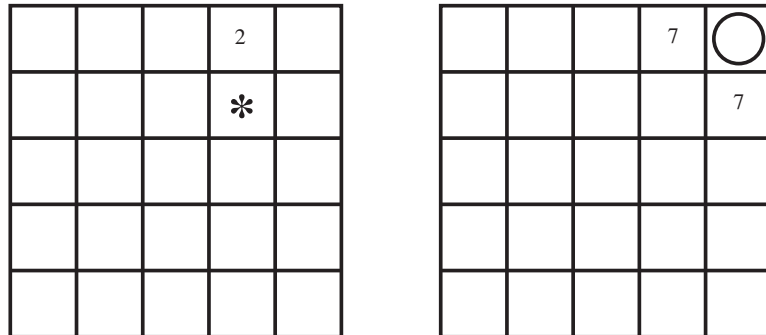


That is the agent goes up with probability 0.8 and goes up-left with probability 0.1 and up-right with probability 0.1.

If there is no treasure, a treasure can appear with probability 0.2. When it appears, it appears randomly at one of the corners, and each corner has an equal probability of treasure appearing. The treasure stays where it is until the agent lands on the square where the treasure is, and the agent gets

an immediate reward of +10, and the treasure disappears in the next state transition. The agent and the treasure move simultaneously so that if the agent arrives at a square at the same time the treasure appears at the same time, it gets the reward.

Suppose we are doing asynchronous value iteration and have the following value for each state:



where the left grid shows the values for the states where there is no treasure and the right grid shows the values of the states when there is a treasure at the top-right.

Consider the next step of asynchronous value iteration. For state  $s_{13}$ , which is marked by \* in the above figure, and the action  $a_2$  which is “up”, what value is assigned to  $Q[s_{13}, a_2]$  on the next iteration of value iteration? You need to show all working, but don't need to do any arithmetic (i.e., leave it as an expression). Explain each terms in your expression.

10. In learning under uncertainty, when the the EM algorithm used? What is the E-step? What is the M-step?
11. Why don't we use the empirical frequencies when learning probabilities from data? That is, if we observe  $n$  occurrences of  $A$  out of  $m$  cases, why shouldn't we just use  $n/m$  as our probability estimate?

You should also expect some questions about what you learned from doing your assignment (e.g., about designing features).