CPSC 422

Practice Midterm Exam Questions

March 2011

The exam covers up to and including Q-learning and SARSA in Reinforcement learning.

This set of practice questions is meant to give some example questions to make sure you understand the material. This is 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, phones, 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 a formula to a particular case, to make sure you understand it).

This practice midterm concentrates on what was not covered in assignments. You should also expect some questions about what you should have learned from doing your assignment (e.g., HMMs, and learning decision trees and naive Bayesian classifiers).

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 an agent controller needs a command function and a state transition function, but not other functions.

2. Suppose a Q-learning agent, with fixed α, and discount γ, 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 explicit as possible.)

3. In temporal difference learning (e.q. 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 α_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 to force the agent to explore.

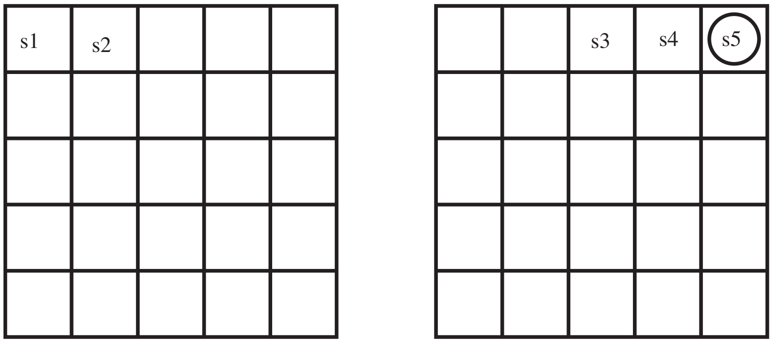
5. In MDPs and reinforcement learning, explain why we often discount future rewards.

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. Consider a grid world, similar (but simpler) to the simple game used in the class. The agent can move up, right, down or left. A prize can appear at one of the corners. Monsters can (stochastically) attack at certain locations.

   Suppose that the agent steps through the state space in the order of steps given in the diagram below, (i.e., going from $s1$ to $s2$ to $s3$ to $s4$ to $s5$), each time doing a “right” action. In this diagram the right grid represents those states where there is a treasure in the top-right position and the states on the left represent states where there is no treasure.



You can assume that this is the first time the robot has visited any of these states. All Q-values are initialized to zero. Assume that the discount is 0.9. Do not include variable names in the arithmetic expressions and do not evaluate the arithmetic. Explain the answers.

(a) Suppose the agent received a reward of -10 entering state s_3 and received a reward of $+10$ on entering the state $s5$, and no other rewards. What Q-values are updated during Q-learning based on this sequence of experiences? Explain what values they get assigned. You should assume that $\alpha_k = 1/k$.

(b) Suppose that, at some later time in the same run of Q-learning, the robot revisits the same states: $s1$ to $s2$ to $s3$ to $s4$ to $s5$, and hasn't visited any of these states in between (i.e, this is the second time visiting any of these states). Suppose this time, the agent receives a reward of $+10$ on entering the state $s5$, and every other reward was zero. Assume that $\alpha_k = 1/k$. What Q-values have their values changed? What are their new values?

(c) How would your answers to the previous parts change if it was using SARSA instead of Q-learning?

9. In learning under uncertainty, when is the EM algorithm used? What is the E-step? What is the M-step?

10. 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?

11. Suppose you get a job where the boss is interested in localization of a robot with multiple laser sensors in a factory. The boss has heard of variable elimination, rejection sampling, and particle filtering and wants to know which

would be most suitable for this task. You must write a report for your boss (using proper English sentences), explaining which one of these technologies would be most suitable. For the one that is most suitable, explain what information it requires to be used for localization. For the two technologies that are not the most suitable, explain why you rejected them.

(a) variable elimination (i.e., exact inference as used in HMMs)

(b) rejection sampling

(c) particle filtering

12. Consider learning a decision tree from data, where there are probabilities at the leaves, and the aim is to minimize the sum-of-squares error. Suppose tree t_1 is the same as tree t_0, but t_1 has one extra split where there is a leaf in t_0. Which of the following are true, and explain why:

    (a) If the leaves predict the empirical proportion on the training data, t_1 is never worse than t_0 in terms of error on the training data.

    (b) If the leaves predict the empirical proportion on the training data, t_1 is always better than t_0 in terms of error on the training data.

    (c) If the leaves predict the empirical proportion on the training data plus a pseudo-count of 1, t_1 is never worse than t_0 in terms of error on the training data.

    (d) If the leaves predict the empirical proportion on the training data, t_1 is never worse than t_0 in terms of error on the test data.

    (e) If the leaves predict the empirical proportion on the training data plus a pseudo-count of 1, t_1 is never worse than t_0 in terms of error on the test data.

    Explain *overfitting* in terms of t_0 and t_1.

13. It is possible that a naive Bayesian classifier can have a worse error on the training data than just predicting a point probability of the class (ignoring all of the input features)? If so, when would this occur. If not, explain why.