COMP-424: Artificial intelligence

Homework 5

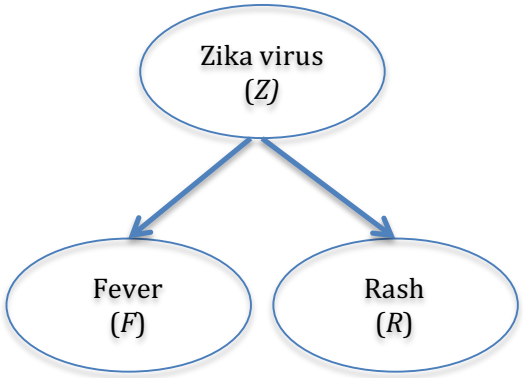
Due on *myCourses* April 15, 11:59pm.

General instructions.

- This is an individual assignment. You can discuss solutions with your classmates, but should only exchange information orally, or else if in writing through the discussion board on *myCourses*. All other forms of written exchange are prohibited.
- Unless otherwise mentioned, the only sources you should need to answer these questions are your course notes and the textbook. Any other source used should be acknowledged with proper referencing style.
- Submit a single pdf document containing all pages of your solution on your McGill's *myCourses* account.

Question 1: Utility

Consider the Bayes Net defined in Question 1 of Homework 4, with all Bernoulli variables. In this case we are going to ignore the Mosquito (*M*) variable. Assume $P(Z)=0.05$, $P(F|Z)=0.8$, $P(F|\neg Z)=0.1$, $P(R|Z)=0.6$, $P(R|\neg Z)=0.2$. Add the fact that having the Zika virus (*Z*) has a utility of -200 if detected and properly treated, but has a utility of -1000 if left untreated (the Utility doesn't change as a function of Fever or Rash). Applying treatment when there is no virus has a utility of -100, and applying no treatment if there is no virus has a utility of 0. Use the principle of Maximum Expected Utility and Value of Information to answer the following questions. For parts (a)-(c) assume the Zika diagnostic test has 100% accuracy.



(a) Given no information on fever or rash, how much should you pay to find out whether or not you have the Zika virus?

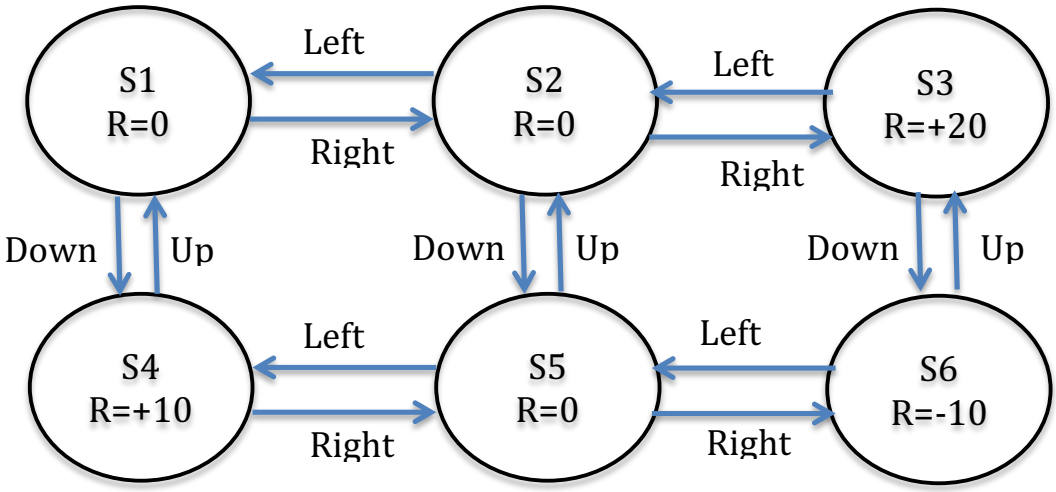
(b) How much should you pay to find out whether or not you have the Zika virus if you observe a fever (*F=True*) and but no rash (*R=False*)?

(c) How much should you pay to find out whether or not you have the Zika virus if you observe a rash (*R=True*), but don't know whether or not you have a fever?

(d) A company has developed a new test for Zika, however this test is not perfectly accuracy: If Zika virus is present, it correctly finds it with 80% accuracy but fails to find it in 20% of tests, and if Zika virus is not present, it reports a false positive in 20% of cases. How much should they charge for this test, to make it competitive with the diagnostic test that has 100% accuracy? (*Hint*: Set the cost of the new test to have the same MEU as the other test.)

Question 2: Markov Decision Processes

Consider the MDP shown below. It has 6 states and 4 actions. As shown on the figure, the transitions for all actions have a Pr=0.8 of succeeding (and leading to the state shown by the arrow) and Pr=0.2 of failing (in which case the agent stays in place). For other transitions that are not shown, assume that they cause the state to stay the same (e.g. *T(S1,Left,S1)=1*). The rewards depend on state only and are shown in each node (state); rewards are the same for all actions (e.g. $R(S4,a)=+10, \forall a$). Assume a discount factor of $\gamma=0.95$.



(a) Describe the space of all possible policies for this MDP. How many are there?
(b) Assuming an initial policy $\pi^0(s)=Left, \forall s,$ perform policy evaluation to get the initial value function for each state, $V^0(s)$, $\forall s$.
(c) Given the initial estimate, V^0, if you run an iteration of policy improvement, what will be the new policy at each state? If necessary, break ties alphabetically, e.g. "Down" before "Left", etc.)
(d) What is the optimal value function at each state for this domain?
(e) Is the optimal value function unique? Explain.
(f) What is the optimal policy at each state for this domain?
(g) Is the optimal policy unique? Explain.
(h) Suggest a change to the reward function that changes the value function but does not change the optimal policy.

Question 3: Reinforcement learning

For the domain above, assume you do not know the transition probabilities or rewards. Without further knowledge, you should assume transition probabilities can be stochastic.

You observe the following trajectories:

	T=1	T=2	T=3	T=4
Trajectory #1	S1, r=0, Down	S4, r=10, Down	S4, r=10, Right	S5, r=0, Up
Trajectory #2	S1, r=0, Down	S4, r=10, Right	S5, r=0, Up	S2, r=0, Right
Trajectory #3	S1, r=0, Right	S2, r=0, Right	S3, r=20, Left	S2, r=0, Down

(a) Using Monte-Carlo estimation, what is the estimate of the value function at each state after seeing trajectory 1? After trajectory 2? After trajectory 3?
(b) Using TD-learning, with a learning rate of α=0.1. What is your estimate of the value function at each state after seeing trajectory 1? After trajectory 2? After trajectory 3?

Question 4: RL in games

Find and read the following paper:
Silver et al. "Mastering the game of Go with deep neural networks and tree search". *Nature*, 529, 484–489.

Summarize the main points by answering the following questions:

(a) What are the motivations for this work (e.g. specific technical problems being solved)?
(b) What methods and algorithms are proposed to solve the problem?
(c) What are the main experimental results presented to evaluate the proposed solution?
(d) What are the primary contributions (novel ideas/applications) of this research?
(e) What are the limitations of the approach and/or results?
(f) What are future directions for this research?

For further insights into how to read technical papers, see:
http://www.cs.mcgill.ca/~martin/teaching/comp762-fall-2005/howto.html