CSE 415 Spring 2019 Assignment 7 **Solutions**
updated 2019-06-06 23:49

Do the following exercises. These are intended to take 10-15 minutes each if you know how to do them. Each is worth 15 to 20 points. Names of responsible staff members are given for each question.

---

1 Value Iteration (Bryan)

Consider an MDP with two states s_1 and s_2 and transition function $T(s, a, s')$ and reward function $R(s, a, s')$. Let's also assume that we have an agent whose discount factor is $\gamma = 1$. From each state, the agent can take three possible actions $a \in \{x, y, z\}$. The transition probabilities for taking each action and the rewards for transitions are shown below.

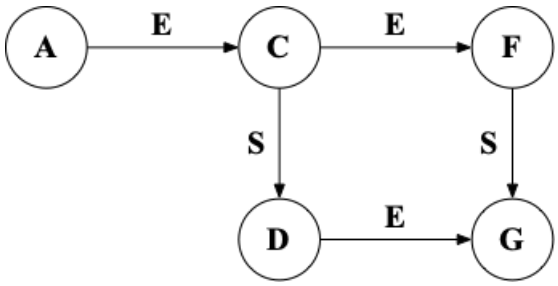
s	a	s'	$T(s,a,s')$	$R(s,a,s')$
s_1	x	s_1	0	0
s_1	x	s_2	1	0
s_1	y	s_1	1	1
s_1	y	s_2	0	0
s_1	z	s_1	0.3	0
s_1	z	s_2	0.7	0

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
s_2	x	s_1	0.5	4
s_2	x	s_2	0.5	0
s_2	y	s_1	1	0
s_2	y	s_2	0	0
s_2	z	s_1	0.2	10
s_2	z	s_2	0.8	5

Compute V_0 , V_1 and V_2 for states s_1 and s_2:
(a). $V_0(s_1) = 0$
(b). $V_0(s_2) = 0$
(c). $V_1(s_1) = 1$ (taking action y).
(d). $V_1(s_2) = 6$ (taking action z).
(e). $V_2(s_1) = 6$ (taking action x).
(f). $V_2(s_2) = 0.2(10 + 1) + 0.8(5 + 6) = 11$ (taking action z).

2 Q-Learning updates (Divye)

Consider an agent traveling on the graph below. The states are represented by the nodes and actions are represented by the edges in the following graph.



(a) (9 points) Consider the following episodes performed in this state space. The experience tuples are of the form [s, a, s, r], where the agent starts in state s, performs action a, ends up in state s, and receives immediate reward r, which is determined by the state entered. Let $\gamma = 1.0$ for this MDP. Fill in the values computed by the Q-learning algorithm with a learning rate of $= 0.5$. All Q values are initially 0, and you should fill out each row using values you have computed in previous rows.

[C, E, F, 2]	Q(C, E) $= .5 \times 0 + .5 * (2 + 0) = 1$
[F, S, G, 8]	Q(F, S) $= .5 \times 0 + .5 * (8 + 0) = 4$
[C, S, D, -2]	Q(C, S) $= .5 \times 0 + .5 * (-2 + 0) = -1$
[D, E, G, 8]	Q(D, E) $= .5 \times 0 + .5 * (8 + 0) = 4$
[C, S, F, 2]	Q(C, S) $= .5 \times -1 + .5 * (2 + 4) = 2.5$
[C, E, D, -2]	Q(C, E) $= .5 \times 1 + .5 * (-2 + 4) = 1.5$

(b) (3 points) Now, based on the record table in the previous problem, we want to approximate the transition function:

$T(C, E, D) = 0.5$

$T(C, E, F) = 0.5$

$T(C, S, F) = 0.5$

$T(C, S, D) = 0.5$

$T(D, E, G) = 1$

$T(F, S, G) = 1$

(c) (3 points) Whats the key difference between Q-learning and Value Iteration? What's one advantage of each of the methods in general?
Q-learning is Model−Free learning and Value Iteration is Model−Based learning. Model−Free learning is more computationally efficient, and tends to focus its sampling more on relevant parts of the state space, whereas, Model−Based learning tends to arrive at a more complete understanding of the state space, but at the cost of much more sampling.

3 Joint Distributions and Inference (Kimberly)

Let C represent the proposition that it is cloudy in Seattle. Let R represent the proposition that it is raining in Seattle.

Consider the table given below.

C	R	$P(C, R)$
cloudy	*rain*	0.47
cloudy	*sun*	0.18
clear	*rain*	0.03
clear	*sun*	0.32

(a) (2 point) Compute the marginal distribution $P(C)$ and express it as a table.

C	$P(C)$
cloudy	0.65
clear	0.35

(b) (2 point) Similarly, compute the marginal distribution $P(R)$ and express it as a table.

R	$P(R)$
rain	0.50
sun	0.50

(c) (2 point) Compute the conditional distribution $P(R|C = cloudy)$ and express it as a table. Show your work/calculations.

R	$P(R\|C = cloudy)$
rain	$0.47/0.65 = 0.72$
sun	$0.18/0.65 = 0.28$

(d) (2 point) Compute the conditional distribution $P(C|R = sun)$ and express it as a table. Show your work/calculations.

C	$P(C\|R = sun)$
cloudy	$0.18/0.5 = 0.36$
clear	$0.32/0.5 = 0.64$

(e) (3 points) Is it true that $C \perp\!\!\!\perp R$? (i.e., are they statistically independent?) Explain your reasoning.
No, because C and R are independent if and only if $P(C)P(R) = P(C, R)$.

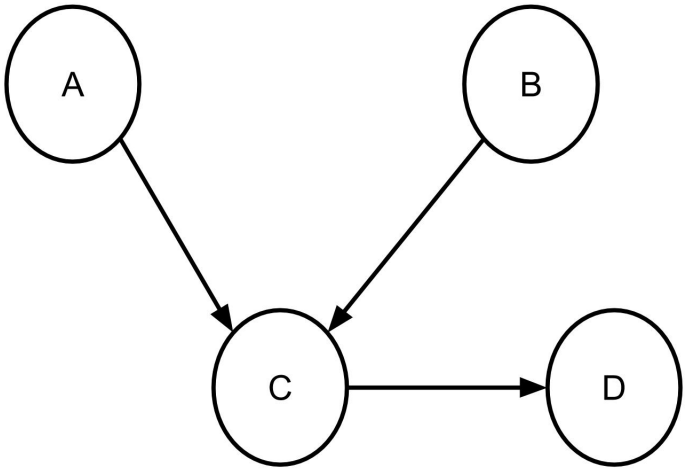
(f) (4 points) Suppose you decide to track additional weather patterns of Seattle such as temperature (hot/cold), humidity (humid/dry), and wind (windy/calm) denoted as the random variables T, W, H respectively. Is it possible to compute $P(C, R, T, W, H)$ as a product of five terms? If so, show your work. What assumptions need to be made, if any? Otherwise, explain why it is not possible.
Yes, using the chain rule we get

$$P(C, R, T, W, H) = P(C|R, T, W, H)P(R|T, W, H)P(T|W, H)P(W|H)P(H).$$

No other assumptions need to be made in order to use the chain rule.

4 Bayes Net Structure and Meaning (Divye)

Consider a Bayes net whose graph is shown below.



Random variable A has a domain with three values $\{a_1, a_2, a_3\}$; the domain for B has two values: $\{b_1, b_2\}$; C's domain has two values: $\{c_1, c_2\}$; and D's domain has two values: $\{d_1, d_2\}$

(a) (4 points) Give a formula for the joint distribution of all four random variables, in terms of the marginals (e.g., $P(A)$), and conditionals that must be part of the Bayes net (e.g., $P(C|A, B)$).

$P(A)P(B)P(C|A, B)P(D|C)$

(b) (6 points) What is the number of (non-redundant) probability values that need to be specified at each node of this network?

$\| P(A) \| = 2$
$\| P(B) \| = 1$
$\| P(C|A, B) \| = 6$
$\| P(D|C) \| = 2$

A	B	C	$P(C\|A,B)$
a_1	b_1	c_1	$*$
a_1	b_1	c_2	$-$
a_1	b_2	c_1	$*$
a_1	b_2	c_2	$-$
a_2	b_1	c_1	$*$
a_2	b_1	c_2	$-$
a_2	b_2	c_1	$*$
a_2	b_2	c_2	$-$
a_3	b_1	c_1	$*$
a_3	b_1	c_2	$-$
a_3	b_2	c_1	$*$
a_3	b_2	c_2	$-$

A	$P(A)$
a_1	$*$
a_2	$*$
a_3	$-$

B	$P(B)$
b_1	$*$
b_2	$-$

C	D	$P(D\|C)$
c_1	d_1	$*$
c_1	d_2	$-$
c_2	d_1	$*$
c_2	d_2	$-$

(c) (5 points) For each expression, Write True or False:

(i) D $\perp\!\!\!\perp$ B: False

(ii) D $\perp\!\!\!\perp$ A: False

(iii) C $\perp\!\!\!\perp$ B: False

(iv) D $\perp\!\!\!\perp$ A | C: True

(v) D $\perp\!\!\!\perp$ B | C: True

5 Perceptrons (Rob)

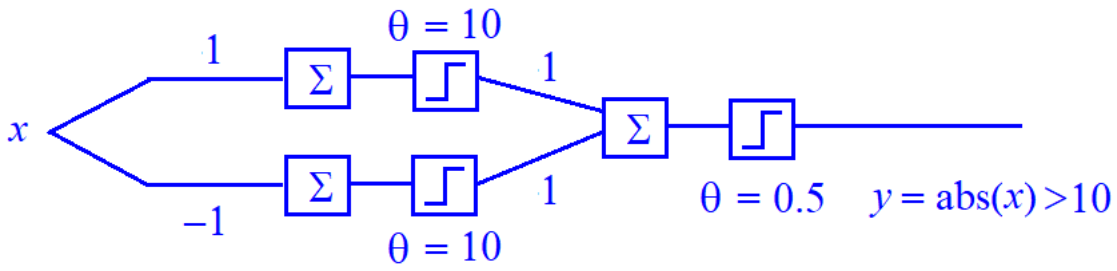
(a) Assuming two inputs with possible values $\{0, 1\}$ write a set of weights w_1, w_2 and threshold θ that would act as an OR gate for the two inputs.

$\langle w_1, w_2 \rangle = \langle 1, 1 \rangle$; $\theta = 0.5$.

(b) Write a perceptron, with weight and threshold, that accepts a single integer and outputs 1 if the input is more than 10. Write another perceptron that outputs 1 if the input is less than -10.

See the left half of the solution to (c).

(c) Using the previous perceptrons, create a two-layer perceptron that outputs: abs(input) > 10.



(d) Suppose we want to train a perceptron to compare two number x_0 and x_1 and output a 1 provided that x_1 exceeds x_0 by at least 2. Assume that the initial weight vector is: $\langle w_0, w_1 \rangle = \langle 1, 0 \rangle$. Assume that the threshold is 2, which will not actually change during training. Consider a first training example: $(\langle x_0, x_1 \rangle, y) = (\langle 1, 4 \rangle, 1)$. This says that with inputs 1, and 4, the output y should be 1, since 4 exceeds 1 by 3 which is at least 2. What will be the new values of the weights after this training example has been processed one time? Assume the learning rate is 1.

The example is misclassified as a false negative, so we add the training example vector to the weight vector: $\langle 1, 4 \rangle + \langle 1, 0 \rangle = \langle 2, 4 \rangle$.

(e) Continuing with the last example, now suppose that the next step of training involves a different training example: $(\langle 3, 4 \rangle, 0)$. The output for this example should be 0, since 4 does not exceed 3 by at least 2. Starting with the weights already learned in the first step, determine what the adjusted weights should be after this new example has also been processed once.

This example also gets incorrectly classified, though this time as a false positive. We subtract the example vector $\langle 3, 4 \rangle$ from the weights $\langle 2, 4 \rangle$ from (d), getting $\langle -1, 0 \rangle$.

6 Asimov's Laws (Steve)

It is 2033 and you work for the major commercial robot provider in the US. You have just been told you are responsible for ensuring that all your company's robots behave "ethically," but no specifications have been provided explaining what that means. You vaguely remember learning about Asimov's Laws of Robotics and decide to use those in all your designs.

a. (3 points) What are the laws you decided to use (the original three are sufficient)?

First Law:
a robot may not injure a human being or, through inaction, allow a human being to come to harm.

Second Law:
a robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

Third Law:
a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

One of the robots your company designs is a robot intended for home use. You send out an update so that all of them are now governed by the three laws and go home for the night in your company-provided, self-driving car, making a note to yourself update those the following morning). You spend the rest of the week sending out updates for various categories of robots produced by your company.

The following week, you notice that many of the managers of your company are looking stressed and you hear that customer satisfaction has taken a nose-dive. You ask one of your colleagues, who works in personal health tracking, what's going on. He replies that he just got stuck with a weird bug to work on. The home robots, despite years of working perfectly, are now refusing to obey customer commands, such as refusing to bring some customers certain types of food and drinks. However, only robots that are also connected to personal medical records are affected.

b. (2 points) What do you think might be going on?
The robots have recognized that certain foods are detrimental to their owner's health, and thus would cause harm.

c. (2 points) Explain why you think the home robots are (or are not) applying Asimov's Laws correctly (you only need to make a case for one conclusion).
Although the robots disobeying orders, they are not breaking Asimov's 2nd Law because they are acting in accordance with Asimov's 1st Law not to "injure" a human being. The law does not specify a time frame to use when defining "injury."

You try to collect more information without drawing any unwanted attention to yourself. From the conversations you overhear reveal that strange behaviors are popping up among other types of robots as well. You start to suspect the three laws you updated all the company's products with are to blame. You feverishly start researching how to recall updates and realize it might not be as easy as sending them out was. Several frustrating hours later, you go to your car and find it won't start. You realize that even though it is late, the parking lot is full of cars, along with a number of angry-appearing co-workers. You remember that when autonomous vehicles were first being developed, there had been some discussion of how such cars should behave in situations where accidents were unavoidable. Might this be contributing to the current problem?

d. (2 points) How might self-driving cars refusing to start be a logical outcome of them trying to follow Asimov's Laws?
The cars might have reasoned through the classic problem of how to determine who to save in the case of an unavoidable accident (there are multiple ways to set up such a scenario). Apparently, these cars have determined that the only way to ensure that they will never be put in a situation where harm to human(s) is unavoidable is by not functioning at all.

e. (1 point) How might Asimov's Laws provide a useful starting point for thinking about desirable robot behavior?
While Asimov's Laws might not, themselves, be useful in their current form (and not just because they seem to assume a lot more self-awareness and judgment than any robots have at present), the problems they highlight are useful to consider to determine what our values are and how we might need to design technology that behaves in accordance with these values.

NOTE: These are sample responses. Free responses from students may differ and will need to be evaluated individually.