Assignment 7 in CSE 415, Spring 2019

by the Staff of CSE 415

This is due June 6 via Gradescope at 11:59 PM. (No late days allowed.) Prepare a PDF file with your answers and upload it to Gradescope.

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.

Last name: Wang, first name: Ziyuan

Student number: 1869306

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. (The first 2 are worth 2 points each. The others are worth 4 points each.)

(a). $V_0(s_1) =$ 0 ?

(b). $V_0(s_2) =$ 0 ?

(c). $V_1(s_1) =$ 1 ?

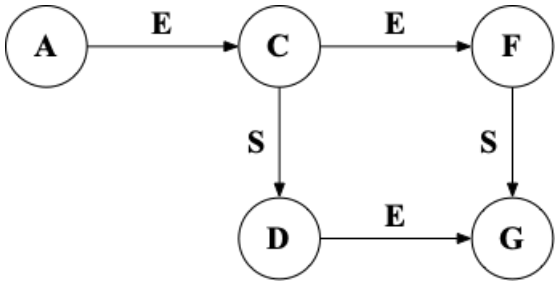
(d). $V_1(s_2) =$ 6 ?

(e). $V_2(s_1) =$ 6 ?

(f). $V_2(s_2) =$ 11 ?

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) = 1
[F, S, G, 8]	Q(F, S) = 4
[C, S, D, -2]	Q(C, S) = -1
[D, E, G, 8]	Q(D, E) = 4
[C, S, F, 2]	Q(C, S) = 2.5
[C, E, D, -2]	Q(C, E) = 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 genera?

Q-Learning can turn values into a (new) policy, but value Iteration can not.

Q-Learning advantage: Converge to optimal policy.

Value Iteration: Converge faster than Q-Learning.

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

cloudy	0.65
clear	0.35

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

rain	0.5
sun	0.5

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

R	P(R|C=cloudy)
rain	$\frac{47}{65}$
sun	$\frac{18}{65}$

$$\frac{P(cloudy, rain)}{P(C=cloudy)} = \frac{0.47}{0.65}$$

$$\frac{P(cloudy, sun)}{P(C=cloudy)} = \frac{0.18}{0.65}$$

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

C	P(C|R=sun)
cloudy	$\frac{9}{25}$
clear	$\frac{16}{25}$

$$\frac{P(cloudy, sun)}{P(sun)} = \frac{0.18}{0.5}$$

$$\frac{P(clear, sun)}{P(sun)} = \frac{0.32}{0.5}$$

(e) (3 points) Is it true that $C \perp\!\!\!\perp R$? (i.e., are they statistically independent?) Explain your reasoning.

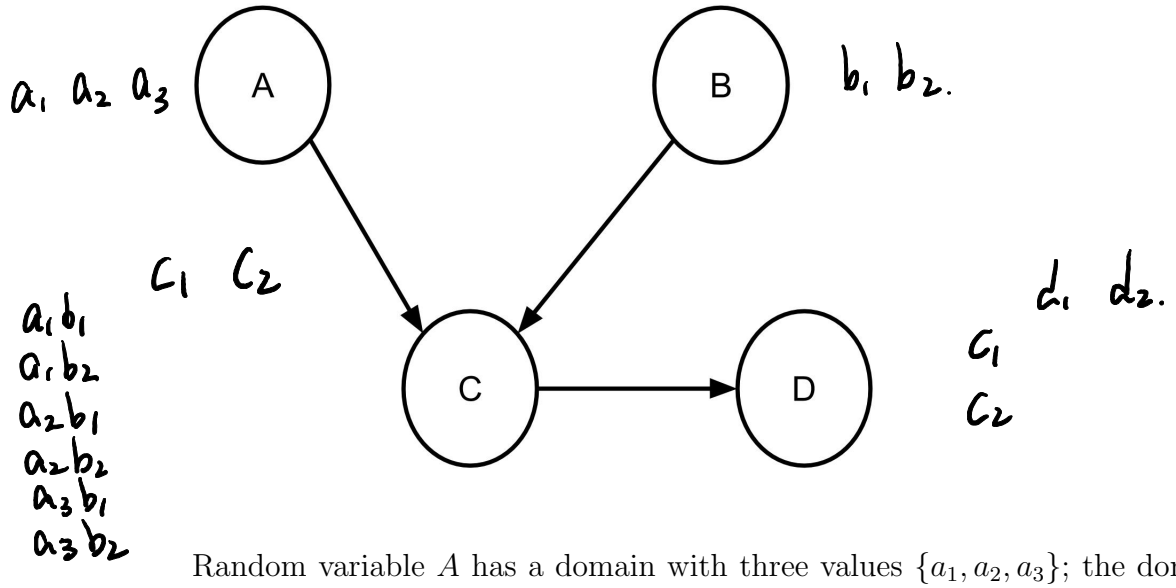
Not true, since $P(C) \cdot P(R) \neq P(C, R)$

(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.

It is not possible, because these 5 terms are not independent.

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

$P(A,B,C,D) = P(D|C) \cdot P(C|A,B) \cdot P(A) \cdot P(B)$

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

A: 3 B: 2 C: 12 D: 4

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

(i) D ⊥⊥ B False

(ii) D ⊥⊥ A False

(iii) C ⊥⊥ B False

(iv) D ⊥⊥ A | C True

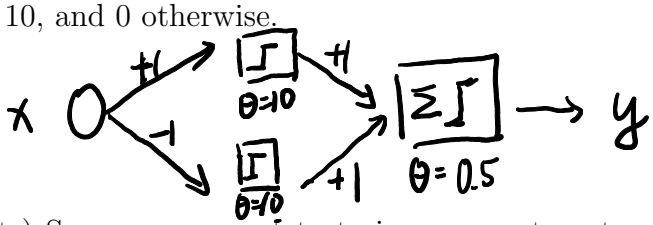
(v) D ⊥⊥ B | C True.

5 Perceptrons (Rob)

(a) (4 points) Assuming two inputs x_1 and x_2 with possible values $\{0, 1\}$ give values for a pair of weights w_1, w_2 and threshold θ such that the corresponding perceptron would act as an OR gate for the two inputs. $w_1=1$ $w_2=1$ $\theta=0.5$

(b) (4 points) Draw a perceptron, with weight and threshold, that accepts a single integer x and outputs 1 if the input is greater than 10. Draw another perceptron that outputs 1 if the input is less than -10.

(c) (4 points) Using the previous perceptrons, create a two-layer perceptron that outputs 1 if $|x| > 10$, and 0 otherwise.



(d) (4 points) Suppose we want to train a perceptron to compare two numbers x_0 and x_1 and produce output $y = 1$ 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 $\theta = 2$, 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.

$1\times1 + 4\times0 = 1 < 2$ $y_i(1) = 0$

$$\begin{bmatrix} w_0(2) \\ w_1(2) \end{bmatrix} = \begin{bmatrix} w_0(1) \\ w_1(1) \end{bmatrix} + 1 \cdot (1-0) \cdot \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

(e) (4 points) 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.

$3\times2 + 4\times4 = 22 > 2$ $y_i(2) = 1$

$$\begin{bmatrix} w_0(3) \\ w_1(3) \end{bmatrix} = \begin{bmatrix} w_0(2) \\ w_1(2) \end{bmatrix} + 1 \cdot (0-1) \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

6 Asimov's Laws (Steve)

It is 2031 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 being 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 first and second law.

One of the robots your company designs is a home-bartender robot. 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-bartender robots, despite years of working perfectly, are now refusing to obey customer commands, such as refusing to bring some customers certain types of drinks and snacks. However, only robots that are also connected to insurance-claims records are affected.

b. (3 points) What do you think might be going on?

Customers may return the robots and engineers in the company will work for debugging.

c. (3 points) Explain why you think the bartender robots are (or are not) applying Asimov's Laws correctly (you only need to make a case for one conclusion).

The reason for this bug is that the robots find out that these types of drinks and snacks are bad for human-beng's life. The robots obey the laws

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. (3 points) How might self-driving cars refusing to start be a logical outcome of them trying to follow Asimov's Laws?

It is possible because the robots find out that driving car may injure human-beng's life.

e. (3 point) How might Asimov's Laws provide a useful starting point for thinking about desirable robot behavior?

The laws are not suitable for some certain situations, so it would do some undesirable behavior.

The technology singularity is a hypothetical future point in time, at which technogical growth becomes uncontrollable and irreversable, resulting in unfathomable changes to human civilization.