Assignment 6 in CSE 415, Winter 2019

by the Staff of CSE 415

March 9, 2019

This is due March 15 via Gradescope at 11:59 PM. Prepare a PDF file with your answers and upload it to Gradescope. Additional details about ways to prepare your PDF file will be posted in Piazza.

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

Last name: McNulty, first name: Zachary

Student number: 1636402

1 Value Iteration (Hamid)

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	0.5	0
s_1	z	s_2	0.5	0

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
s_2	x	s_1	0.5	10
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	0.5	2
s_2	z	s_2	0.5	4

Compute V_0 , V_1 and V_2 for states s_1 and s_2. (The first 2 are worth 1 point each. The others are worth 2 points each.)

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

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

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

(d). $V_1(s_2) =$ 5 ?

(e). $V_2(s_1) =$ 5 ?

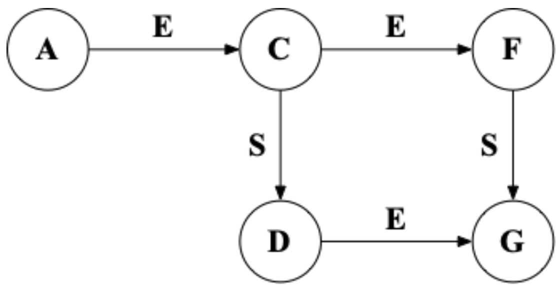
best option is x as you go directly to state s2 which has value 5

(f). $V_2(s_2) =$ 8 ?

best option as x again; half the time you get 10 as a reward for going to state s1 + 1 from ending in state 1, and the other half the time you dont get a reward but you go to state 2 which has valu 5

2 Q-Learning updates (Jifan)

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.



Qnew(s,a) = alpha Qold(s,a) + alpha (r + maxQ(s', a'))

(a) 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.

all Q values = 0

[A, E, C, 2]	Q(A, E) = 1	
[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	-1 (1/2) + 1/2(2 + 4) where 4 is Q(F,S)
[C, E, D, -2]	Q(C, E) = 1.5	1 (1/2) + 1/2(-2 + 4) where 4 is Q(D,E)

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

$T(A, E, C) =$ 1

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

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

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

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

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

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

$R(*, *, C) =$ 2

$R(*, *, D) =$ -2

$R(*, *, F) =$ 2

$R(*, *, G) =$ 8

T(A, E, C) was calculated based on prob(C I A, E) and this is the manner in which the other transition probabilities were calculated

(ii) Assume the above transition and reward functions correctly approximates the true transition and reward functions. Also, assume the discount factor $\gamma = 1.0$ What are the optimal value functions V^*?

$V^*(A) =$ 10

$V^*(C) =$ 8 = max(choose action E, choose action S) = 0.5 * (2 + 1*8) + 0.5(-2 + 1*8) = 8

$V^*(D) =$ 8

$V^*(F) =$ 8

(c) (i) The method we use in Problem 2.1 is Model– Free learning.

(ii) The method we use in Problem 2.2 is Model– Based learning.

(iii) What's one advantage of each of the methods in general?

Model free learning is useful as often the transition probabilities and rewards are not known ahead of time

Model based learning is useful if the transition probabilities and rewards are known because you can directly calculate the optimal plan without having to run through any actions. This is often much faster and requires far less steps to converge than a typical model free learning method.

Text

3 Joint Distributions and Inference (Sam, Steve)

Let X represent the proposition that Tesla stock is low. Let Y represent the proposition that it is raining today in Seattle.

Consider the table given below.

X	Y	$P(X,Y)$
T	T	0.12
T	F	0.18
F	T	0.28
F	F	0.42

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

(b) (1 point) Similarly, compute $P(Y)$.

(c) (2 points) Compute $P(X) \cdot P(Y)$.

(d) (1 point) Is it true that $X \perp\!\!\!\perp Y$? (i.e., are they statistically independent?)

(e) (2 points) Without using any numbers from the table, give a formula for computing $P(X|Y)$ from $P(Y|X)$, $P(Y|\neg X)$, and $P(X)$.

(f) (2 points) For someone who wants to buy Tesla stock when it is low, would it be better to buy on a rainy day than not on a rainy day, according to the given model? Explain.

a)

P(X = T) = P(X I Y)P(Y) + P(X I NOT Y)P(NOT Y) = P(X and Y) + P(X and NOT Y) = 0.3
P(X = F) = P(NOT X I Y)P(Y) + P(NOT X I NOT Y)P(NOT Y) = P(NOT X and Y) + P(NOT X and NOT Y) = 0.7

X	T	F
P(X)	0.3	0.7

b)

P(Y = T) = P(X and Y) + P(NOT X and Y) = 0.4
P(Y = F) = P(X and NOT Y) + P(NOT X and NOT Y) = 0.6

X	T	F
P(Y)	0.4	0.6

c)

X	Y	P(X)	P(Y)	P(X)*P(Y)
T	T	0.3	0.4	0.12
T	F	0.3	0.6	0.18
F	T	0.7	0.4	0.28
F	F	0.7	0.6	0.42

d)

$P(X) = 0.3$ while $P(X \mid Y) = P(X, Y) / P(Y) = 0.12 / 0.4 = 0.3$
As $P(X) = P(X \mid Y)$ then X and Y are statistically independent. $P(X \text{ and } Y) = P(X \mid Y)P(Y) \rightarrow P(X)P(Y)$

i.e. from the table we generated in part c) we can see that P(X AND Y) = P(X) * P(Y) for any possible values X and Y can take on.

e)

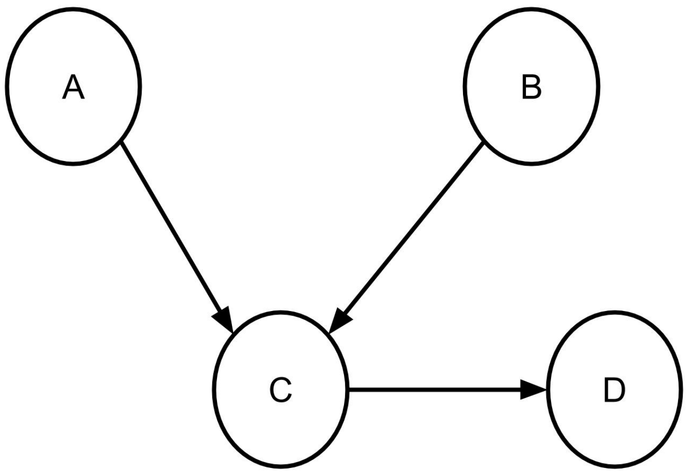
$P(X \mid Y) = P(Y \mid X)P(X) / P(Y) = P(Y \mid X)P(X) \;/\; [\; P(Y \mid X)\, P(X) + P(Y \mid \text{NOT } X)\, (1 - P(X)) \;]$

f)

X and Y are independent as we saw in exercise 3d so whether or not it is raining in Seattle tells us nothing abou the state of Tesla stock. So it does not matter whether or not its raining if we are trying to buy Tesla's stock low.

4 Bayes Net Structure and Meaning (Divye, Steve)

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

dont count last one

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

a)

P(A, B, C, D) = P(D | C) * P(C | A,B) * P(A)*P(B)

b)

For A, just P(a1) and P(a2) will suffice as then P(a3) = 1 - P(a2) - P(a1) —> two probability values
For B, just P(b1) will suffice as then P(b2) = 1 - P(b1) —> one probability value
For C, for all 3*2 choices of A,B respectively, simply P(c1 | A,B) will suffice as P(c2 | A,B) = 1 - P(c1 | A,B) yielding 6 probability values total
For D, for all 2 choices of C, simply P(d1 | C) will suffice as P(d2 | C) = 1 - P(d1 | C) —> two probability values

This yields a total of 2 + 1 + 6 + 2 = 11 probability values for whole network.

5 Markov Models (Sam and Steve)

Assume we model the stock market with a two-state Markov model. On each day, the market is either rising (R) or falling (F). The transition matrix is this:

P(X_{t+1} | X_t)

X_t	X_{t+1}	$\cancel{P(X_{t+1} = R)}$
R	R	0.6
R	F	0.4
F	R	0.8
F	F	0.2

(a) (2 points) Assuming the market is falling today, determine the probability that it will be rising tomorrow.

(b) (3 points) Assuming the market is falling today, determine the probability that it will be rising the day after tomorrow.

(c) (5 points) Find the stationary probability $r = P_\infty(R)$ that the market is rising.

a) Since these are Markov Models, the probability distribution of X_{t+1} depends entirely on P(X_t)

P(X_{t+1} = R | X_t = F) = 0.8, just reading out of the table

b)

P(X_{t+2} = R | X_t = F)
= P(X_{t+2} = R | X_{t+1} = F) * P(X_{t+1} = F | X_t = F) + P(X_{t+2} = R | X_{t+1} = R) * P(X_{t+1} = R | X_t = F)
= 0.8*0.2 + 0.6*0.8
= 0.64

c)

P_{\infty}(R) = P(R | F)*P_{\infty} (F) + P(R | R)*P_{\infty} (R) = 0.8 * P_i(F) + 0.6*P_i(R)
P_i(R) = 2*P_i(F)   AND P_i(R) + P_i(F) = 1
P_i(R) = 2/3,   P_i(F) = 1/3

6 HMMs (Divye)

In this exercise, we will run the forward algorithm for 2 steps, for the following given HMM.

Transition from/to	To S_1	To S_2
From S_1	0.4	0.6
From S_2	0.2	0.8

Emission/Output Probabilities	E_1	E_2	E_3
S_1	0.2	0.6	0.2
S_2	0.2	0.4	0.4

Lets say the initial probability for S_1 is $P_0(S_1) = 0.3$ (at $t = 0$), and we observed E_2 at t=1 and E_3 at t=2.

(a) (3 points) What's $P_0(S_2)$ for t=0?

(b) (3 points) What's $P_1(S_1)$ and $P_1(S_2)$ (for t=1)? This is after the observation at time $t = 1$.

(c) (4 points) What's $P_2(S_1)$ and $P_2(S_2)$ (for t=2)? This is after the observation at time $t = 2$.

a)

P0 (S2) = 1 - P0(S1) = 0.7

b)

B1(S1) = P0(S1)*P(E2 | S1) = 0.3 * 0.6 = 0.18
B1(S2) = P0(S2)*P(E2 | S2) = 0.7*0.4 = 0.28

Thus P1(S1) = 0.18 / (0.18 + 0.28) —> 0.39
Thus P1(S2) = 0.28 / (0.18 + 0.28) —> 0.61

c)

B2(S1) = P(E3 | S1) * [P(S1 | S1) * B1(S1) + P(S1 | S2) * B1(S2)] = 0.2 * [0.4*0.18 + 0.2 * 0.28] = 0.0256
B2(S2) = P(E3 | S2) * [P(S2 | S1) * B1(S1) + P(S2 | S2) * B1(S2)] = 0.4 * [0.6* 0.18 + 0.8 * 0.28] = 0.1328

Thus P2(S1) = 0.0256 / (0.0256 + 0.1328) = 0.162
Thus P2(S2) = 0.838

7 Perceptrons (Rob)

(a) (2 point) 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.

(b) (2 points) 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.

(c) (2 points) Using the previous perceptrons, create a two-layer perceptron that outputs 1 if abs(input) > 10 and 0, otherwise.

(d) (2 points) Suppose we want to train a perceptron to compare two number x_1 and x_2 and output a 1 provided that x_2 exceeds x_1 by at least 2. We will include a bias input $x_0 = 1$, so the perceptron can effectively learn its threshold, as part of the standard perceptron training. Assume that the initial weight vector is all zeros: $\langle w_0, w_1, w_2 \rangle = \langle 0, 0, 0 \rangle$. Assume that the actual threshold is 0, and which will not actually change during training. Consider a first training example: $(\langle x_1, x_2 \rangle, y) = (\langle 2, 5 \rangle, 1)$. This says that with inputs 1 (which is the bias and not shown in the training example), 2, and 5, the output y should be 1, since 5 exceeds 2 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. c_k

(e) (2 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 is 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.

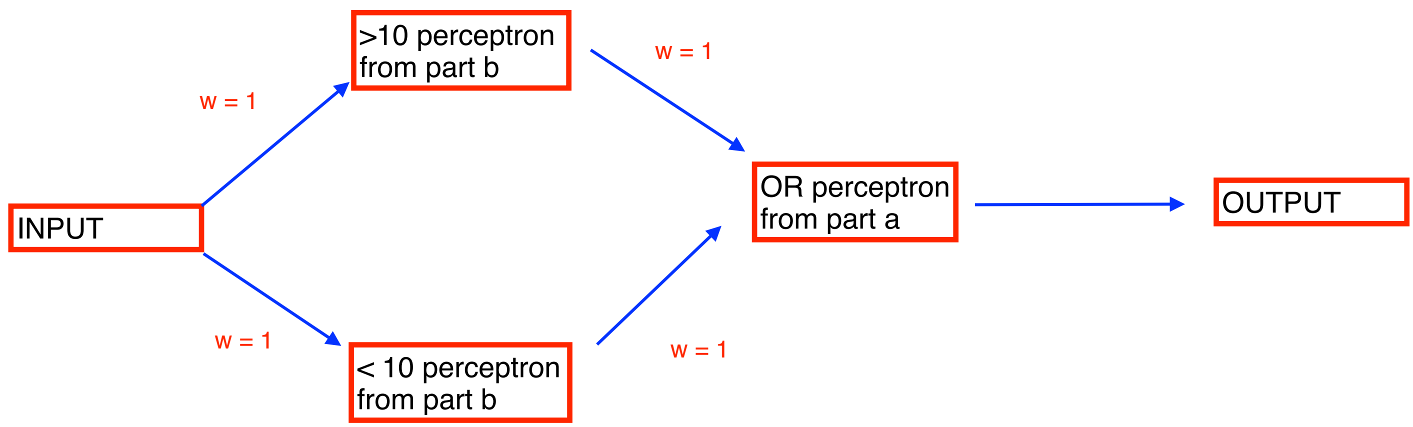
a)

w1 = 1, w2 = 1, theta = 0.5

b)

p2: w = 1, theta = 10.5   p2: w = -1, theta = 10.5

c)



d)

We will attempt to classify the given input <x0, x1, x2> = <1, 2, 5> using our current weights w_0 = <0, 0, 0>. This yields a weighted sum of 1*0 + 2*0 + 5*0 = 0 which is equal to our threshold of 0. Thus, in this case the perceptron fires. As this is the correct response as 5 is at least 2 greater than 2,
the weights will not be updated during this round of the learning.

Thus, w_1 = <0, 0, 0>

e)

We will attempt to classify the given input <x0, x1, x2> = <1, 3, 4> using our weights from the previous problem w_1 = <0, 0, 0>. Thus yields a weighted sum of 1*0 + 3*0 + 4*0 = 0 which is equal to our threshold 0. Thus, in this case the perceptron fires. However, this is an incorrect positive response: 4 is not at least 2 greater than 3. Thus, we will update our weights to be w_2 = w_1 - X_2 * c_2 = <0, 0, 0> - <1,3,4> * 1 = <-1, -3, -4>

w_2 = <-1, -3, -4>

8 NLP: PCFGs (Steve)

There is a famous sentence "Time flies like an arrow." Sometimes it is extended to "Time flies like and arrow, but fruit flies like a banana." We'll use the short version. With the given probabilistic context-free grammar, find two legal parses, and compute a score for each one. Then identify the most probable parse using the scores. The grammar is given below. Consider the number at the right of a production to be the conditional probability of applying that production given that the symbol to be expanded is the symbol on the left-hand side of the production. Convert each probability into a score by taking score $= -\log_2(p)$.

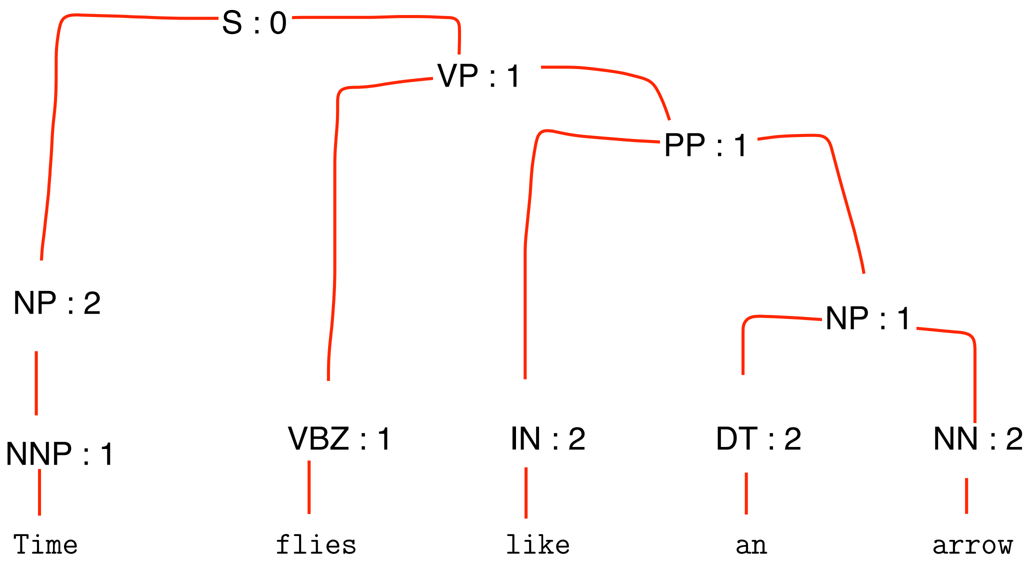
each score is just the -1*log2(prob) or log2(1/prob)

				scores
S	::=	NP VP	1.00	0
NP	::=	DT NN	0.50	1
NP	::=	JJ NNS	0.25	2
NP	::=	NNP	0.25	2
VP	::=	VBZ PP	0.50	1
VP	::=	VBP NP	0.50	1
DT	::=	an	0.25	2
PP	::=	IN NP	0.50	1
NN	::=	arrow	0.25	2
NNP	::=	Time	0.50	1
JJ	::=	Time	0.125	3
NNS	::=	flies	0.25	2
VBZ	::=	flies	0.50	1
VBP	::=	like	0.50	1
IN	::=	like	0.25	2

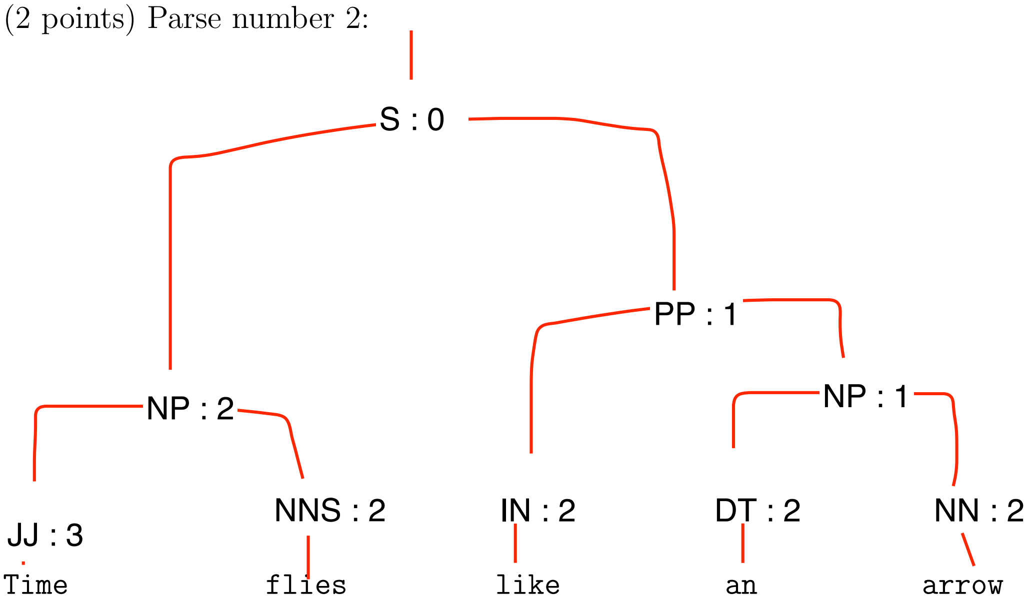
(a) (1 point) Scores for each production rule (give a list of the 15 numbers):

0 1 2 2 1 1 2 1 2 1 3 2 1 1 2

(b) (2 points) Parse number 1:



(c) (2 points) Parse number 2:



(d) (2 points) Total score and overall probability for parse number 1: score: 13, prob: 1/2^13

(e) (2 points) Total score and overall probability for parse number 1: score: 15 , prob: 1/2^15

(f) (1 point) Which parse is more probable?

Parse 1 (it has a lower score and hence higher probability)

9 NLP: Document Similarity (Steve)

Consider the sentences A, B, and C below. First, stem any words having morphological inflections. Next eliminate stopwords in each sentence. Finally determine the minimal reference vocabulary, sort it alphabetically, and encode each sentence as a vector of occurrence counts according to the reference vocabulary.

A. "Computer~~s~~ ~~were~~ often thought ~~of as~~ mechanical ~~and~~ ~~without~~ creativity."

B. "~~The~~ computer ~~is~~ creative ~~not~~ mechanical, ~~and~~ ~~it~~ ~~can~~ think."

C. "~~A~~ mechanical computer ~~has~~ ~~no~~ ~~moving~~ part~~s~~."

move part

Treat as stopwords: all prepositions, pronouns, articles, words of length three or less, and any form of the verbs to have or to be.

(a) (3 points) The sentences after stemming and stopword elimination:

A. Computer often think mechanical creative

B. Computer creative mechanical think

C. mechanical computer move part

(b) (1 point) Reference vocabulary:

computer, creative, mechanical, move, often, part, think

(c) (3 points) Vectors representing the bag-of-words descriptions of the sentences:

A. < 1, 1, 1, 0, 1, 0, 1>

B. < 1, 1, 1 , 0, 0, 0, 1>

C. <1, 0, 1, 1, 0, 1, 0>

(d) (2 points) Suppose we are trying to decide whether A is more similar to B than to C. Compute the cosine similarity for A-B and for A-C.

A-B : cos(x) = 4 / sqrt(20) = 2/sqrt(5) —> x ~ 26.57 degrees

A-C : cos(x) = 2 / sqrt(20) = 1/sqrt(5) —> x ~ 63.43 degrees

(e) (1 point) Which classification is implied by the results?

A is more similar to B than to C

10 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 harm a human either directly or through inaction

Second Law: A robot must follow the orders given to it by a human unless they violate the first law

Third Law: A robot must protect its own existence as long as doing so would not violate one of the first two 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?

They see the food they are serving may have a negative effect on their human's health due to some of the human's pre-existing medical conditions. Thus, following that order to bring food is a violation of the first law.

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).

By providing that food, the robot is knowing performing actions that will deteriorate the health of the human, causing them harm. Thus, such an action would be a violation of the first law so the robot is correctly applying 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. (2 points) How might self-driving cars refusing to start be a logical outcome of them trying to follow Asimov's Laws?

They know that by starting, they are possibly putting themselves in a situation where they will be forced to injure/kill a human in a crash: i.e. choosing between the life of a pedestrian and the life of the passengers. This would be an unavoidable violation of the first law. The robots only option to avoid and not violate any laws is to just not start.

e. (1 point) How might Asimov's Laws provide a useful starting point for thinking about desirable robot behavior?

These kind of weird paradoxes force us into thinking more about how to balance simple, understandable constraints on robots and the ability of robots to actually perform their desired tasks without being constantly stalled by hypotheticals surrounding these rules. In order to properly answer these questions, however, these paradoxes show us we have to think more fundamentally about what it is ethical for a robot to do. Bringing back the first example, do robots have an obligation to look out for our well-being , regardless of what we as humans want? How much do we allow robots to intervene with the lifestyle choices we make, and where do we draw the line?