Assignment 7 in CSE 473, Spring 2023

by the Staff of CSE 473

This is due Tuesday, May 30, via Gradescope at 11:59 PM. Prepare a PDF file with your answers and upload it to Gradescope. The PDF file can created however you like. For example it can be from a scan of a printout of the assignment document onto which you have hand-written your answers. Or it can be from Word or Latex file with your answers. It can even be from photos of your handwritten answers on plain paper. You don't have to include the questions themselves, but it is fine to do so. In any case, it must be very clear to read, and it must be obvious and easy for each grader where to find your solutions to the exercises.

As with Assignment 4, this is an *individual work* assignment. Collaboration is not permitted.

Do the following exercises. These are intended to take 10-30 minutes each if you know how to do them. Each is worth 15 points. The total of possible points is 150. Names of responsible staff members are given for each question.

If corrections or clarifications to the problems have to be given, this will happen in the ED discussion forum under topic "Assignment 7."

Last name: _____, first name: _____

Student number: _____, UWNetID: _____

1 Joint Distributions

(10 points) Let J be the random variable that represents a programmer enjoys programming in Javascript. Let Q be the random variable that represents whether the programmer enjoys programming in Quartz.

Consider the joint distribution given by the table below:

J	Q	$P(J, Q)$
T	T	0.2
T	F	0.5
F	T	0.1
F	F	0.2

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

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

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

(d) (2 points) Compute the conditional distribution $P(Q|\neg J)$ and express it as a table. Show your work/calculations.

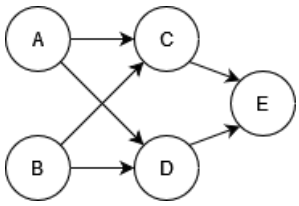
(e) (2 points) Is it true that $J \perp\!\!\!\perp Q$? (i.e., are they statistically independent?) Justify your answer.

(f) (2 points) Draw a Bayes Net corresponding to the factorization $P(J, Q) = P(Q)P(J|Q)$.

(g) (3 points) Give the numbers of free parameters in the original joint distribution table and in the factorized representation of the distribution. Explain the results of a comparison.

2 Bayes Net Structure and Meaning

(10 points) Consider the following Bayes net, where variable A has a domain with 4 values, variable B's domain has 1 values, C's domain has 3 values, D's domain has 3 values, and E's domain has 2 values.



(a) (2 points) Write down the full joint probability distribution associated with the above Bayes net. Express the answer as a product of terms representing individual conditional probabilities tables associated with this Bayes Net:

(b) (1 point) How many probability values (number of entries) belong in the full joint distribution table for this set of random variables?

(c) (5 points) For each random variable: give the number of probability values (number of entries) in its marginal (for A and B) or conditional distribution table (for the others).

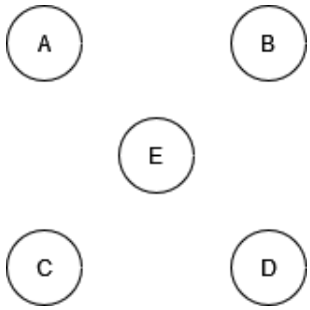
A : B : C : D : E :

(d) (5 points) For each random variable, give the number of *non-redundant* probability values in its table from (c).

A : B : C : D : E :

(e) (2 points) Draw the Bayes net associated with the following joint distribution by connecting (directed) arrows between each variable:

$$\mathbb{P}(A|B) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C|A) \cdot \mathbb{P}(D|C,B) \cdot \mathbb{P}(E|A,B,C)$$



3 D-Separation

(15 points) Consider the Bayes Net graph β below, which represents the topology of a web-server security model. Here the random variables have the following interpretations:

V = Vulnerability exists in web-server code or configs.

C = Complexity to access the server is high. (Passwords, 2-factor auth., etc.)

S = Server accessibility is high. (Firewall settings, and configs on blocked IPs are permissive).

A = Attacker is active.

L = Logging infrastructure is state-of-the-art.

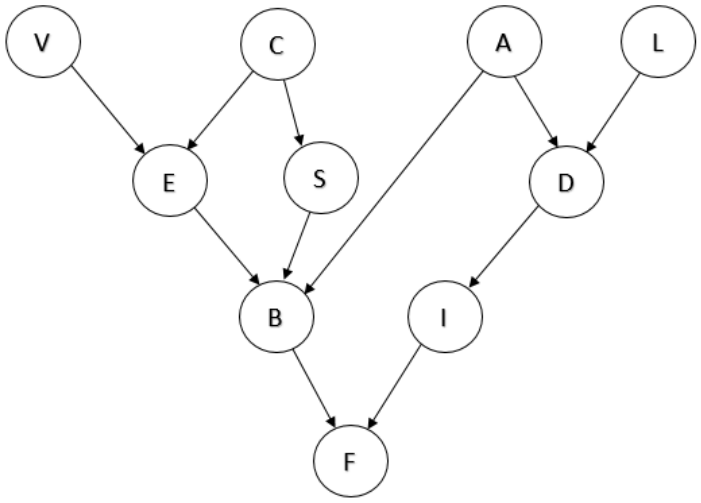
E = Exposure to vulnerability is high.

D = Detection of intrusion attempt.

B = Break-in; the web server is compromised.

I = Incident response is effective.

F = Financial losses are high (due to data loss, customer dissatisfaction, etc).



Let β' be the undirected graph obtained from β by removing the arrowheads from the edges of β. By an "undirected path" in β we mean any path in β'. A "loop-free" path is any path in which no vertex is repeated.

(a) (3 points) List all loop-free undirected paths from B to L in the graph β.

(b) (2 points) Suppose random variable D is observed, and no others are observed. Then which (if any) of those paths would be active paths? Justify your answer.

For each of the following statements, indicate whether (True) or not (False) the topology of the net guarantees that that the statement is true. If False, identify a path ("undirected") through which influence propagates between the two random variables being considered. (Be sure that the path follows the D-Separation rules covered in lecture.) The first one is done for you.

(c) $B \perp\!\!\!\perp D$: False (BAD)

(d) (1 point) $A \perp\!\!\!\perp I \,|\, S$

(e) (1 point) $A \perp\!\!\!\perp I \,|\, D, S$

(f) (1 point) $A \perp\!\!\!\perp I \,|\, D, F, S$

(g) (1 point) $C \perp\!\!\!\perp D \,|\, B, F$

(h) (1 point) $C \perp\!\!\!\perp D \,|\, I$

(i) (5 points) Suppose that the company hired an outside expert to examine the system and she determines that B, D, and E are true: a Break-in has occurred (the web server is compromised), there was a Detection of an intrusion attempt, and the Exposure to vulnerability is high). Given this information, your job is to explain to management why getting additional information about C (Whether Complexity to access the server is high) could have an impact on the probability of L (Logging infrastructure is state-of-the-art). Give your explanation, for the manager of the company, using about 3 to 12 lines of text, which should be based on what you know about D-separation, applied to this situation. However, your explanation should not use the terminology of D-separation but be in plain English. (You can certainly use words like "influence", "probability", "given", but not "active path", "triple", or even "conditionally independent").

4 Perceptrons

(15 points) For all parts of this question perceptrons should output 1 if $w_n + \sum_{i=0}^{n-1} w_i x_i \geq 0$ and 0 otherwise. The weight w_n is called the bias weight.

(a) (3 points) Assuming there will be two inputs x_0 and x_1, each with possible values in $\{0, 1\}$, give values for a triple of weights $\langle w_0, w_1, w_2 \rangle$ such that the corresponding perceptron would act as a NAND gate for the two inputs. (Weight w_2 is the bias weight.) Note that a NAND gate outputs 1 when at least one of the inputs is 0; it outputs 0 otherwise.

(b) (3 points) Draw a perceptron, with weights, that accepts a single integer x and outputs 1 if and only if the input is greater than or equal to -8. Be sure to include the bias input of value 1 and its weight in your diagram. Draw another perceptron that outputs 1 if and only if the input is less than or equal to 8.

(c) (2 points) Using the previous perceptrons, create a two-layer perceptron that outputs 1 if $|x| \leq 8$, 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 5. Assume that the initial weight vector is: $\langle w_0, w_1, w_2 \rangle = \langle -1, -1, 1 \rangle$. Consider a first training example: $(\langle x_0, x_1 \rangle, y) = (\langle 3, 8 \rangle, 1)$. This says that with inputs 3, and 8, the output y should be 1, since 8 exceeds 3 by 5. What will be the new values of the weights after this training example has been processed one time? Assume the learning rate is 2.

(e) (3 points) Continuing with the last example, now suppose that the next step of training involves a different training example: $(\langle 2, 5 \rangle, 0)$. The output for this example should be 0, since 5 does not exceed 2 by at least 5. 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.

5 Multiclass Perceptrons

(15 points) Consider the problem of classifying text items into three topic areas: AI, Medicine, and Hiking. We have four training examples:
e_1: Hill climbing informs gradient descent. (AI)
e_2: Alzheimers is a gradual descent. (Medicine)
e_3: Hill climbing helps. (Medicine)
e_4: Gorp helps hill climbing. (Hiking)

(a) (4 points) Convert the four training examples into the (vector, category) form indicated by the table. (You'll use the reference vocabulary given in the table.)

e_i	Alzheimers	climbing	descent	gorp	gradient	gradual	helps	hill	(category)
e_1									
e_2									
e_3									
e_4									

(b) (8 points) We will start with a weight vector for each of the 3 categories (W_{A} for AI, W_{M} for Medicine, and W_{H} for Hiking), with all weights 0 except 1 for the bias on the AI vector.

W	bias	Alzheimers	climbing	descent	gorp	gradient	gradual	helps	hill
W_{A}	1	0	0	0	0	0	0	0	0
W_{M}	0	0	0	0	0	0	0	0	0
W_{H}	0	0	0	0	0	0	0	0	0

Perform one epoch of training showing the resulting changed vectors whenever there is a change made to a weight vector. When each training example is processed, if any weight vector does not change, do not rewrite that vector. Whenever a vector is updated, write the new vector on its own line below, clearly indicating which category it belongs to. For example, $W_{\mathrm{H}} = ...$

(c) (3 points) Training could take many epochs to converge. See if you can skip the training and manually provide a set of three weight vectors that will correctly handle the training examples.

6 The Laws of Robotics

(15 points) In the 1940's, Isaac Asimov introduced a set of three laws to govern robot behavior:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

(NOTE: You might also want to take a look at this cartoon `https://xkcd.com/1613/`)

Imagine you live in a world, essentially identical to our own, where personal robotic assistants are a bit more advanced than they are here. They are still not commonplace, but in another 10 years (in our scenario), they could very well be. Further, imagine that although you are a poor college student, your best friend and housemate works in tech and loves being the first to obtain new tech gadgets. Consequently, you have access to a brand new personal robot in your apartment.

You are a computer science student and you're also a science fiction aficionado. Your favorite author – Isaac Asimov. You can't resist hacking your new robot so that it now is programmed to obey Asimov's three laws before considering any other part of its programming. You feel quite pleased with yourself. Now, not only do you have a personal robot, you also have a personal bodyguard.

You've been hearing a lot about advances in AI in the news recently. You're especially interested in GPTs and you decide you'd like to co-author a book with the robot (you're thinking of writing stories about robots, similar to those Isaac Asimov wrote, but updated for 2023). You'd also really like to be able to hold conversations with your robot that are more engaging than the interactions it is currently capable of. You figure enhancing the robot with a GPT would be the solution to both of these. It takes you a few days to modify some GPT code you found online (with Chat GPT assistance, of course), but the results seem to be just what you wanted (although you really didn't do much testing). Given that your recent experiments with hacking the robot seem so successful, you decide to "improve"

it some more. You are also interested in the ethical applications of AI, so the next change you make is adding a sophisticated Ethics Module to the robot's codebase.The robot is now familiar with the major issues under discussion with regards to the use of AI and its effects on individuals and societies.

Considering the above information, please answer the following free response questions. Provide a justification for your response based on one (or more) of Asimov's Laws, referring to them by number as appropriate (e.g. "According to Asimov's 1st Law, ..."). Note that there are not specific "correct" answers to these questions. We are interested in your thoughts on current issues and your reasoning.

(a) (4 points) Assuming the same issues that concern us are relevant in our imagined world, what might be some of the issues related to AI that the robot is now aware of, thanks to its Ethics Module (which includes a summaries of ethics issues that have recently been publicized via news outlets or social media)? Please mention at least two different issues.

(b) (4 points) You want to write a novel with the robot in which robots overpower their human owners and create a new world. When you feed prompts to ChatGPT, it comes up with all sorts of great text that could be used in the novel. However, the ChatGPT text is often repetitive and the storyline is frequently lost. You are sure your robot, with its superior reasoning capabilities, could do better, but when you try to get it to write the novel with/for you, it doesn't generate any text. Why might the robot be failing at this task? If you ask it why it isn't cooperating with you, what response might you get?

(c) (4 points) Giving up on the robot, you go to your room to work on your computer. At least it doesn't argue with you about everything. You start typing on your keyboard, searching for how to uninstall Asimov's Laws and the Ethics module. Suddenly, you realize that the robot is standing behind you, looking at your computer screen over your shoulder. You tell the robot you don't like the way it is functioning and you want to return it to its original state, before you modified its programming. What do you think the robot says or does next (and why)?

(d) (3 points) Do you think the robot was correct in its responses to scenarios (b) and (c) above? Please explain your response.

7 Markov Models

(15 points) Rumor has it that there exists a region in the South Pacific where scuba diving is very popular and the water is either calm or turbulent there on any given day. The day to day course is modeled using a Markov model. The dynamics of that model are presented below, where C means calm and T means turbulent.

S_{t-1}	S_t	$P(S_t \mid S_{t-1})$
C	C	0.9
C	T	0.1
T	C	0.5
T	T	0.5

(a) (2 points) Suppose it's given that $S_0 = \text{C}$. Compute the probability that $S_2 = \text{C}$.

(b) (4 points) Compute the stationary probabilities for C and T.

(c) (1 point)Now suppose that whenever the water is calm, the sky is blue with probability 0.7 and dark with probability 0.3.

When the water is turbulent, the sky is blue with probability 0.15 and dark with probability 0.85.

Suppose an observer on the dive boat cannot directly tell whether the state of the water is calm or turbulent, but can only see whether the sky is blue or dark.

State S	Observation Q	$P(Q \mid S)$
C	B	0.7
C	D	0.3
T	B	0.2
T	D	0.8

Suppose $P(S_0 = \text{calm}) = 0.5$. If the observation at time 0 is "blue," what is the belief in S=calm right after the observation?

(d) (1 point) Suppose at time 1, the observation is "dark". what is the belief in S = calm right after that observation? (This belief will take into consideration the previous belief you computed above.)

(e) (1 point) Suppose that the actual state sequence for the first four time steps is calm, calm, calm, turbulent

What is the probably of observing the sequence (starting at $t = 0$) blue, dark, blue, blue?

(f) (1 point) Now, what if the state sequence was calm, turbulent, turbulent, calm. What is the probability of observing the same sequence of stock changes as above?

(g) (5 points) Use the Viterbi algorithm to find the most likely state sequence given that the observation sequence is dark, dark, blue, blue. Show the trellis diagram and label each node with the probability of reaching it by a most probable state sequence path that arrives at that node.

8 Probabilistic Context-Free Grammars

(15 points) Consider the sentence, "Man steals car with dogs."

This sentence is ambiguous. It could mean that the man is stealing a car with the owner's dogs inside. It could also mean that man is stealing a car with his own dogs as partners in crime.

With the probabilistic context-free grammar given below, find two parses, and compute a score for each one. Then identify the most probable parse using the scores. Assume the number at the right of a production is its conditional probability of being applied, given that the symbol to be expanded is that production's left-hand side. The probabilities for all the given productions for a specific left-hand-side non-terminal might not sum to one here, because we are showing a relevant subset of a larger set of productions.

(a) (7 points) Convert each probability into a score by taking score $= -\log_{10}(p)$. Round scores to 2 decimal places of accuracy. Write the production scores in the "__.__" blanks.

S	::= NP VP	0.9	0.05	VB	::= steals	0.04	__.__
NP	::= NP PP	0.3	__.__	IN	::= with	0.15	__.__
NP	::= NN	0.1	__.__	NN	::= man	0.05	__.__
NP	::= NNS	0.2	__.__	NN	::= car	0.07	__.__
PP	::= IN NP	0.6	__.__	NNS	::= dogs	0.08	__.__
VP	::= VP PP	0.3	__.__				
VP	::= VB NP	0.4	__.__				

(b) (3 points) Give a first parse for the sentence. This parse should correspond to the

interpretation, “Man steals a car with the owner’s dogs still inside of it.” Compute the (total) score for this parse, showing the production scores at each internal node.

```
man        steals        car        with        dogs
```

(c) (3 points) Give the second parse. This parse should correspond to the interpretation, ”A man and his dogs steal another person’s car.” Compute its total score, also showing the production scores at each internal node.

```
man        steals        car        with        dogs
```

(d) (2 points) Convert each score back to a probability and write them here as P1 and P2. Then tell which parse is more probable.

9 Deep Learning

(a) If your training accuracy is significantly higher than your test accuracy, what can you do to decrease the gap? Circle all that apply. (6 pts)

1. Add more hidden units.
2. Add a bias term.
3. Increase initialized weights.
4. Decrease your step size.
5. Train on a larger dataset with new data.
6. None of the above.
7. All of 1 through 5.

Explanation:

(b) Which activation functions suffer from the problem of vanishing gradients during back-propagation? If you consider these functions in the one dimensional case, what types of input would lead to this behaviour? Circle True or False for each activation function, then provide the types of inputs if true (you may supplement your answer with examples if you'd like.) (9 pts)

Sigmoid: True / False
Inputs if true:

ReLU: True / False
Inputs if true:

Leaky ReLU: True / False
Inputs if true:

10 Ethical Issues with Generative AI

(15 points)

Read the following article from Business Insider:

LINK: The end of coding as we know it by Aki Ito (April 26, 2023)

We are interested in your thoughts, as a computer science student, of the opinions presented and the predictions made in this article and in others linked below.

Answer the questions in either Set 1 or Set 2 below. (We recommend that you become familiar with both sets of questions, however, since the Final Exam might have a question that relates to material from one, the other, or both question sets.) Here, circle **EITHER** (a) Set 1 **OR** (b) Set 2, in order to indicate your choice of question set for your A7 answers.

(a) Set 1:

1 a) A related article: LINK: Tech giants aren't just cutting thousands of jobs — they're making them extinct by Matt Turner (April 27, 2023) suggests that the tech industry is facing a period of significant, and perhaps turbulent, change. In response to such predictions, do you plan to make any changes in the courses you choose to take as you finish up your undergraduate degree? If so, please describe the options you are considering and your reasoning for considering them. If not, why not, and what advice would you give someone considering such changes to their academic program?

1 b) If you had a summer internship in college, describe the work you did and comment on how much of that work you think could have been completed by GPT technology. If you didn't have a summer internship, speculate on how you think GPTs will affect tech summer internship programs going forwards. Given your responses, what are your main concerns on beginning your career in the near future?

1 c) You have been elected as a student representative to a panel tasked with creating ethical guidelines for the use of GPTs. What would you want the panel to consider regarding the needs and interests of young people studying CS and hoping to have tech careers?

(b) Set 2:

In the article by Ito, there is a quote from UW CSE professor Zach Tatlock:

Sure, the arrival of the tractor threw a lot of farmers out of work. But coding isn't like farming. "There's only so much food that 7 billion people can eat," says Zachary Tatlock, a professor of computer science at the University of Washington. "But it's unclear if there's any cap on the amount of software that humanity wants or needs. One way to think about it is that for the past 50 years, we have been massively underproducing. We haven't been meeting software demand."

2 a) How convincing do you find this as an argument that human programmers will still be needed in the future (and perhaps even more sought after than at present) , as the article implies?

2 b) The article also discusses a "glaring problem" with AI research: Far too much of it is focused on replacing human labor rather than empowering it. Why are we deploying our best and brightest minds to get machines to do something humans can already do, instead of developing technology to help them do something entirely new? Going back to Prof. Tatlock's quote – just how much food do people need to eat? Navigate to this site: LINK: Can the world feed 8bn people sustainably? – turns out it's quite a lot, to the extent that there may not be enough food to feed everyone in the future. After reading the article, describe two ways in which AI could be used to address the issues involved in world hunger – one that specifically uses ChatGPT (to generate either text or code) and one that uses AI more broadly defined (i.e. other AI technologies).

2 c) You have recently been hired as an ethics intern at OpenAI, as the company recently committed to working on vital world issues in a way that will benefit society ethically and fairly. You have been asked to work with the team promoting "Feeding the World" as an issue to attack. Prepare an outline describing your ideas and suggestions on how OpenAI could use ChatGPT to target world food insecurity and sustainable food production. Note: AI uses tremendous amounts of energy and resources that could be directed to other purposes (LINK: The hidden costs of AI: Impending energy and resource strain).