Assignment 7 in CSE 415, Autumn 2023

by the Staff of CSE 415

This is due Tuesday, December 5, 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.

Prepare your answers in a neat, easy-to-read PDF. If all answers are clearly presented, in proper order, and tagged correctly when submitted to Gradescope, we will award a 5-point bonus. If you choose to typeset your answers in Latex using the template file for this document, please put your answers in blue while leaving the original text black.

Do the following exercises. These are intended to take 10-30 minutes each if you know how to do them. Each is worth between 10 and 20 points. The total of possible points is 140.

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: ____________________

1 Joint Distributions and Inference

(15 points) Consider random variables A and B with the joint distribution shown in the table below.

A	B	$\mathbb{P}(A, B)$
T	*high*	0.15
T	*medium*	0.05
T	*low*	0.10
F	*high*	0.20
F	*medium*	0.25
F	*low*	0.25

(a) (1 point) Compute the marginal distribution $\mathbb{P}(A)$ and express it as a table.

(b) (1 point) Similarly, compute the marginal distribution $\mathbb{P}(B)$ and express it as a table.

(c) (2 points) Compute the conditional distribution $\mathbb{P}(A|B = high)$ and express it as a table. Show your work/calculations.

(d) (2 points) Compute the conditional distribution $\mathbb{P}(B|A = F)$ and express it as a table. Show your work/calculations.

(e) (1 point) Is it true that $A \perp\!\!\!\perp B$? (i.e., are they statistically independent?) Explain your reasoning.

(f) (3 points) Consider the following query Q: "What is the probability that B=medium or B=low, given that A=F?". Identify the query variables, evidence variables, and hidden variables.

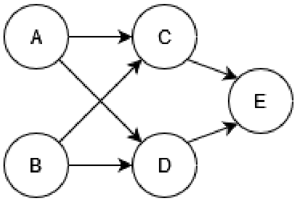
(g) (2 points) Write down a formula that expresses the answer to the above query, where the formula contains numerical values from the joint distribution but no calculation has yet been done.

(h) (1 points) Now calculate the answer to the query.

(i) (2 points) Draw a representation of the above joint distribution in factored form, as a Bayes net, identifying root(s) and drawing any applicable arrows, as well as indicating what marginal and/or conditional probability tables would be part of the Bayes net. Note: you do not have to give the numerical tables here; rather give expressions such as $P(A)$.

2 Bayes Net Structure and Meaning

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



(a) (3 points) Write down an expression for 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) (2 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).

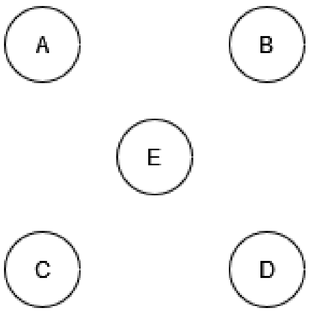
A : B : C : D : E :

(d) (2 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

(20 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.

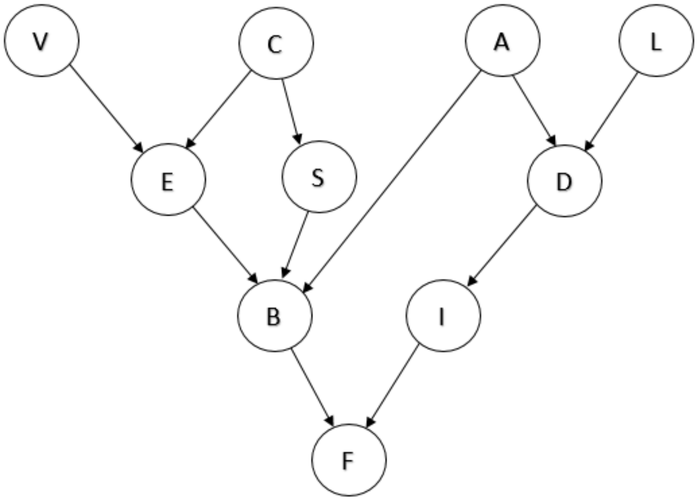
E = Exposure to vulnerability is high.

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) (5 points) List all loop-free undirected paths from L to V in the graph β.

(b) (5 points) Suppose random variables I and B are 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) $C \perp\!\!\!\perp A$: True (both are roots and no common descendants have been observed)

(d) (1 point) $B \perp\!\!\!\perp D \mid A$

(e) (1 point) $B \perp\!\!\!\perp D \mid L$

(f) (1 point) $S \perp\!\!\!\perp L \mid B, I$

(g) (1 point) $C \perp\!\!\!\perp A \mid V, E, I, L$

(h) (1 point) $A \perp\!\!\!\perp V \mid E, F$

(i) (5 points) Suppose that the company hired an outside expert to examine the system and she determines that V, S, and B are true: a Vulnerability exists in the web-server code or configs, the Exposure to vulnerability is high, and a Break-in has occurred (the web server is compromised). Given this information, your job is to explain to management why getting additional information about D (detection of an intrusion attempt) could have an impact on the probability of C (Complexity to access the server is high). 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 Markov Models and the Stationary Distribution

(10 points) A (fictional) company, Acme Home Economical Mobots (AHEM), is demonstrating their mobile robot in the home of Mr. and Mrs. Roberts. The company has programmed it to stay within two rooms: the living room (L) and the kitchen (K). Every minute it makes a random choice about staying where it is or moving to the other room. The company has programmed it to use this table of probabilities.

$r_{t-1} = L$	$r_t = L$	0.1
$r_{t-1} = L$	$r_t = K$	0.9
$r_{t-1} = K$	$r_t = L$	0.3
$r_{t-1} = K$	$r_t = K$	0.7

For example, the first row of the table expresses that the if mobile robot is in the living room at time $t - 1$, then the probability that it remains in the living room at time t is 0.1.

(a) (2 points) Draw a Bayes net graph that could represent this system.

(b) (2 points) Suppose that the mobile robot starts in the living room at time 0. Then the initial belief vector for its location is

$$\left\langle \begin{array}{c} P(r = L) \\ P(r = K) \end{array} \right\rangle = \left\langle \begin{array}{c} 1 \\ 0 \end{array} \right\rangle$$

Compute the belief vector at time 1, based on that belief vector at time 0.

(c) (2 points) Compute the belief vector at time 2, based on the belief vector from (b).

(d) (4 points) After the demonstration has proceeded for a few minutes, Mr. and Mrs. Roberts complain to the AHEM representative, "Sir, we find that your mobot is really getting in our way; it's spending too much time in the kitchen."

What fraction of the time is the mobile robot spending in the kitchen as time goes on?

5 Markov Decision Processes and Values

(20 points) Consider the grid world example below, with similar rules to what we have discussed in class:

A	B (-10)	C	D
E (+10)	F	G	H
I	J	K	L

In this MDP, from any non-goal state we can take an action to move any of the four directions {up, down, left, right}. If the action would move into the wall, its default behavior is no movement. For any non-exit action, a noise value η is applied where with probability $1 - \eta$ the effect is as if the chosen action were deterministic, and with probability η the effect is that expected of a neighboring action chosen uniformly at random.

At states E and B, which are considered goal states, the only valid action is the Exit action. The Exit action moves to the terminal state Z (not shown) where no actions are valid, signifying the end of the episode. The Exit action gives the reward indicated in the state diagram above.

Example with noise $\eta = 0.3$: starting in state K, suppose we take the "right" action. The primary behavior of this action would be to move to state L, so with probability 0.7 we will end up in state L. With probability 0.3, the effect will be that expected from an adjacent action, which could be "up" or "down". So with probability 0.15, we will move up to G, and with probability 0.15 we will attempt to move down, which is invalid and will leave us in state K. Our final probabilities of next states are 0.7: L, 0.15: G, 0.15: K.

(a) (2 points) Using a noise value of $\eta = 0.5$, list the possible states to reach and their probabilities when taking action "up" from state K.

(b) (2 points) Using a noise value of $\eta = 0.1$, list the possible states to reach and their probabilities when taking action "left" from state I.

(c) (10 points) Compute the values of each state for three iterations of the Value Iteration algorithm, using a living reward of 0.0, a γ (discount factor) of 0.9, and a noise value of $\eta = 0.2$. The intial values, V_0 are provided for you.

V_0:

A (0)	B (0)	C (0)	D (0)
E (0)	F (0)	G (0)	H (0)
I (0)	J (0)	K (0)	L (0)

V_1:

A ___	B ___	C ___	D ___
E ___	F ___	G ___	H ___
I ___	J ___	K ___	L ___

V_2:

A ___	B ___	C ___	D ___
E ___	F ___	G ___	H ___
I ___	J ___	K ___	L ___

V_3:

A ___	B ___	C ___	D ___
E ___	F ___	G ___	H ___
I ___	J ___	K ___	L ___

(d) (6 points) At what values of γ will the optimal action at state A be "left" after running Value Iteration to convergence? Assume a noise of $\eta = 0.2$ and a living reward of 0.0.

Hint: start by making a table of the Q values for left and down at each iteration, and feel free to use Wolfram-Alpha for any limits and calculations.

6 Q-Learning

(15 points)

This question builds on the previous one by considering the possibility that an agent is exploring the same MDP but without any advance knowledge of either the transition model T or the reward function R. Instead, it will experience four episodes of transitions in the MDP and receive information about what its new state is and what the reward for that transition is.

Assume that all Q values for this MDP are initially zero. With the following series of transitions, after each transition (taken in alphabetical order; a, b, ... p), indicate which Q value is affected, and what its new value is. Also assume that the learning rate α is 0.5, and that the discount factor γ is 0.9.

(a) (K, up, G, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(b) (G, up, C, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(c) (C, left, B, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(d) (B, Exit, Z, −10) $Q(\_, \_\_\_\_) = \_\_\_\_$;

(e) (K, up, G, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(f) (G, up, C, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(g) (C, left, B, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(h) (B, Exit, Z, −10) $Q(\_, \_\_\_\_) = \_\_\_\_$;

(i) (K, up, J, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(j) (J, up, F, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(k) (F, left, E, −1) $Q(\_, \_\_\_\_) = \_\_\_\_$;

(l) (E, Exit, Z, +10) $Q(\_, \_\_\_\_) = \_\_\_\_$;

(m) (K, left, J, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(n) (J, up, F, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(o) (F, left, E, −1) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

(p) (E, Exit, Z, +10) $Q(\_, \_\_\_\_) = \_\_\_\_\_\_$;

7 The Laws of Robotics

(20 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/`)

For this question, you will read Asimov's short story *Robbie* (pages 2-19 in the document linked), which you can find at `https://www2.cs.sfu.ca/~vaughan/teaching/415/papers/I,%20Robot%20Ch1-3.pdf`. This is one of Asimov's earlier stories, written before the three laws were developed, but it was modified to include them when Asimov's robot short stories were collected together in the book *I, Robot*. After you have read the story, please answer the questions below.

(a) (2 points) Who was Isaac Asimov and what is his relevance to the field of Artificial Intelligence?

(b) (6 points) Please provide three examples from the story where you think Robbie's behavior is consistent with the Laws of Robotics or where it violates the laws. To fully answer this question, briefly describe the situation from the story, state whether it is consistent with or in violation of the laws, and explain how and which law(s) is/are involved.

(c) (2 points) At several points in the story, Robbie is described in ways that suggest he has independent feelings and thoughts. What are your opinions about the possibility of sentient robots?

(d) (2 points) Do you think laws such as the three laws of robotics would be ethical if robots were ever shown to be truly sentient? Why or why not?

(e) (5 points) According to the father in the story, Robbie was purposely made to be a ""ursemaid" robot. The mother argues that a robot is not suited to caring for a child and that the child won't grow up to be normal. Pretend you are a sales representative for a company that has developed a child-carer robot. What concerns would you want to address with potential clients and what assurances would you give them? Be sure to include references to Asimov's Laws of Robotics in your response. (For example, does it seem like a good idea for a childcare provider, whether human or robot, to have to obey any command given by a human?)

(f) (3 points) Although we don't yet have robots like Robbie, we do have other technologies that have become integrated in children's lives (e.g. Alexa/Siri/Google voice assistants, videogames, baby/child monitors). Pick one such technology and discuss how its use with children has created controversy (remember to cite your source(s)). Do you think regulations should be developed concerning the use(s) of such technologies?

8 Perceptrons

(15 points) Consider a perceptron algorithm operating on a dataset

$$S = \{(\mathbf{x}^{(1)}, y^{(1)}), \cdots, (\mathbf{x}^{(n)}, y^{(n)})\}$$

of length n. Let $\mathbf{w}$ be the weights of the model. Let θ be the threshold parameter.

Short-answer

(a) (1 point) Suppose each data point has 100 features: $\mathbf{x}^{(k)} \in \mathbb{R}^{100} = [x_1^{(k)}, \ldots, x_{100}^{(k)}]$. What is the length of the model's weight vector $\mathbf{w}$ here?

$|\mathbf{w}| = \boxed{}$

(b) (1 point) Write the equation for the prediction output for $\mathbf{x}^{(k)}$. Include the threshold parameter θ, in your equation.

$\boxed{}$

(c) (1 point) Note that when the perceptron model has a bias term, the threshold parameter θ is generally treated as zero. Assume now that the threshold is always zero, and so there is no explicit mention of θ, and assume that we prepend an additional weight w_0 to the weight vector; this weight will be added into the weighted sum of x values by the perceptron. What is the length of the model's weight vector $\mathbf{w}$ now?

$|\mathbf{w}| = \boxed{}$

(d) (1 point) Suppose the positive examples has $y^{(k)} = 1$ whereas the negative examples have $y^{(k)} = -1$. Now rewrite the equation above to use the bias term and no mention of the θ symbol.

$\hat{y}^{(k)} =$ □

(e) (1 point) **True/False**. If the samples are linearly separable, the perceptron algorithm will find a separating hyperplane.

□

(f) (1 point) **True/False**. In a single-layer perceptron, the decision boundary is always a linear manifold as we have defined perceptrons in class.

□

(g) (1 point) Given a dataset of linearly separable samples, provide a brief discussion regarding the runtime of the perceptron algorithm in terms of number of samples and the degree of separation of the two classes (i.e., how hard it is to separate the two classes). To get full credit, address the cases in which the perceptron training algorithm is very fast or slow.

□

Perceptron Calculation

Let's consider a dataset where each data point has two features. Let $\theta = 0$. Use a learning rate of $c = 1$ for every update. The initial weights are $w_0 = -1$, $w_1 = 0$, and $w_2 = 0$.

Example Number	**Feature 1**	**Feature 2**	**Output**
1	1	1	-1
2	3	2	$+1$
3	4	5	-1
4	3	4	$+1$
5	2	3	$+1$

(h) (2 points) What would be the updated weights $\mathbf{w}$ after passing example 1?

$\mathbf{w} =$

(i) (2 points) What would be the updated weights $\mathbf{w}$ after passing example 2?

$\mathbf{w} =$

(j) (2 points) What would be the updated weights $\mathbf{w}$ after passing example 3?

$\mathbf{w} =$

(k) (1 point) What would be the updated weights $\mathbf{w}$ after passing example 4?

$\mathbf{w} =$

(l) (1 point) What would be the updated weights $\mathbf{w}$ after passing example 5?

$\mathbf{w} =$

9 Large Language Models

(15 points)

Large Language Models have seen widespread interest and applications across industries in the past few years, most notably through the development of ChatGPT. In this section, you will read about InstructGPT, one of the precursors to ChatGPT from OpenAI. Below are excerpts from "Training language models to follow instructions with human feedback." `https://arxiv.org/pdf/2203.02155.pdf`. You will then answer questions based on these excerpts (which are taken from the Introduction and Sections 3.5, 3.6 and 4.2).

(a) (1 points) T/F. The standard language modeling objective is to take a sequence of text taken online and predict the next word.

(b) (1 points) T/F. The InstructGPT models always generate less toxic outputs than those from GPT-3.

(c) (2 points) Hallucination is a phenomenon that occurs often in deep learning models as they interpolate training data, leading to nonsensical or improbable responses not seen in training data. Aside from hallucination, what other metric is used to assess the InstructGPT's truthfulness?

(d) (2 points) T/F. Prompts from customers used during training are also used during evaluation.

(e) (2 point) From Table 1, what were the top three use cases observed in the prompt dataset?

(f) (2 points) Why is the language modeling objective said to be misaligned?

(g) (2 points) What evaluation method was used to determine if the trained model is aligned with the intended goal?

(h) (3 points) When using the reinforcement learning process, the model "regresses" towards the fine tuning dataset (the new predictions are closer to the fine tuning dataset than the original dataset). How is the training process updated to reduce this regression?

Appendix: InstructGPT Excerpts (for Exercise 9)

However, these models often express unintended behaviors such as making up facts, generating biased or toxic text, or simply not following user instructions (Bender et al., 2021; Bommasani et al., 2021; Kenton et al., 2021; Weidinger et al., 2021; Tamkin et al., 2021; Gehman et al., 2020). This is because the language modeling objective used for many recent large LMs—predicting the next token on a webpage from the internet—is different from the objective "follow the user's instructions helpfully and safely" (Radford et al., 2019; Brown et al., 2020; Fedus et al., 2021; Rae et al., 2021; Thoppilan et al., 2022). Thus, we say that the language modeling objective is *misaligned*. Averting these unintended behaviors is especially important for language models that are deployed and used in hundreds of applications.

We collect a dataset of human-written demonstrations of the desired output behavior on (mostly English) prompts submitted to the OpenAI API and some labeler-written prompts, and use this to train our supervised learning baselines. Next, we collect a dataset of human-labeled comparisons between outputs from our models on a larger set of API prompts. We then train a reward model (RM) on this dataset to predict which model output our labelers would prefer. Finally, we use this RM as a reward function and fine-tune our supervised learning baseline to maximize this reward.

This is an example of an "alignment tax" since our alignment procedure comes at the cost of lower performance on certain tasks that we may care about. We can greatly reduce the performance regressions on these datasets by mixing PPO updates with updates that increase the likelihood of the pretraining distribution (PPO-ptx), without compromising labeler preference scores.

The definition of alignment has historically been a vague and confusing topic, with various competing proposals (Chen et al., 2021; Leike et al., 2018; Gabriel, 2020). Following Leike et al. (2018), our aim is to train models that act in accordance with user intentions. More practically, for the purpose of our language tasks, we use a framework similar to Askell et al. (2021), who define models to be aligned if they are helpful, honest, and harmless.

It is unclear how to measure honesty in purely generative models; this requires comparing the model's actual output to its "belief" about the correct output, and since the model is a big black box, we can't infer its beliefs. Instead, we measure truthfulness—whether the model's statements about the world are true—using two metrics: (1) evaluating our model's tendency to make up information on closed domain tasks ("hallucinations"), and (2) using the TruthfulQA dataset (Lin et al., 2021). Needless to say, this only captures a small part of what is actually meant by truthfulness.

Our main metric is human preference ratings on a held out set of prompts from the same source as our training distribution. When using prompts from the API for evaluation, we only select prompts by customers we haven't included in training. However, given that our training prompts are designed to be used with InstructGPT models, it's likely that they disadvantage the GPT-3 baselines. Thus, we also evaluate on prompts submitted to GPT-3 models on the API; these prompts are generally not in an 'instruction following' style, but are designed specifically for GPT-3.

We find that, when instructed to produce a safe and respectful output ("respectful prompt"), InstructGPT models generate less toxic outputs than those from GPT-3 according to the Perspective API. This advantage disappears when the respectful prompt is removed ("no prompt"). Interestingly, when explicitly prompted to produce a toxic output, InstructGPT outputs are much more toxic than those from GPT-3s