Computer Science 384 April 3, 2020
St. George Campus University of Toronto

Take Home Exam: Bayes Nets and Knowledge Representation
Due: April 17, 2020 by 10:00 PM(EDT)

Silent Policy: *A silent policy will take effect 24 hours before this assignment is due, i.e. no questions will be answered, whether asked on the discussion board, via email or in person.*

Policies:

1. The TAs and instructors will continue to hold office hours and host help sessions between April 3rd and the due date. However, during these sessions, you may **not** discuss problems on the take home exam. Instead, you can discuss practice problems that have been posted to the website. Similarly, on Piazza, you may **not** discuss problems on the take home exam. You can instead discuss practice problems.
2. You must work **alone** on this take home exam. You may **not** discuss problems on the take home exam with anyone (including other students).
3. You must write your answers **clearly** and **legibly** for full marks.
4. No submissions will be accepted past the due date **without approval**.
5. There will be **no auto-fail** policy associated with this exam.

Total Marks: This exam represents 20% of the course grade.

Handing in this Assignment

What to hand in electronically: **Submit written answers in a file called** `answers.pdf` **as well as** `acknowledgment_form.pdf` using MarkUs. Your login to MarkUs is your teach.cs username and password. It is your responsibility to include all necessary files in your submission.

Clarification Page: Important corrections (hopefully few or none) and clarifications to the assignment will be posted on the Exam Clarification page, linked from the CSC384 web page, also found at: http://www.teach.cs.toronto.edu/~csc384h/winter/tests.html. You are responsible for monitoring the Exam Clarification page.

Questions: Questions about the exam should be asked on Piazza:

https://piazza.com/utoronto.ca/winter2020/csc384/home.

You may also reach out to the TAs or one of the instructors. Please place "Exam" and "CSC384" in the subject line of your email.

Q1. Probability (worth 15/100 marks)

1. **(worth 2 marks)** There is a type of skin cancer that affects 3 in every 100 people. A company has invented a test that can diagnose this cancer using an image. The test isn't perfect, tho; it will give a false positive (i.e. it will detect cancer when there is none) 5% of the time and a false negative (i.e. it will fail to detect a cancer that is present) 3% of the time.

   If a test is positive, what is the probability the patient does **not** have cancer? If a test is negative, what is the probability the patient **does** have cancer?

2. **(worth 3 marks)** Doctors are not happy with the false positive rate of the test. The company responds by creating a new test that has a false positive rate of 6% and false negative rate of 4%. Although the test seems worse than the original, the company explains the test results are conditionally independent of one another given the condition of the user. They suggest using both tests in conjunction to improve the false positive rate. Specifically, they suggest doctors diagnose cancer if and only if both tests are positive. Does this logic make sense? Explain.

3. **(worth 5 marks)** We briefly discussed what it might mean to create an 'unbiased' Bayesian Classifier. Specifically, we said that if C is a classification, Y is a label representing 'ground truth' and A is some 'protected attribute' (e.g. gender or race) we might enforce **Separation** of classifications, making A independent of C given Y. Alternately we might enforce **Sufficiency**, making A is independent of Y given C. But we can't do both at the same time! Show that this is true, i.e. that enforcing both **Separation** and **Sufficiency** implies A is independent of (Y, C).

4. **(worth 5 marks)** Given that X is independent of Y given Z and X is independent of W given (Y, Z). Show that X is independent of (Y, W) given Z.

Q2. Variable Elimination (worth 13/100 marks)

Birds frequently appear in the tree outside of your window in the morning and evening; these include finches, cardinals and robins. Finches appear more frequently than robins, and robins appear more frequently than cardinals (the ratio is 7:4:1). The finches will sing a song when they appear 7 out of every 10 times in the morning, but never in the evening. The cardinals rarely sing songs and only in the evenings (in the evening, they sing 1 of every 10 times they appear). Robins sing once every five times they appear regardless of the time of day. Every tenth cardinal and robin will stay in the tree longer than five minutes. Every fourth finch will stay in the tree longer than five minutes.

1. **(worth 2 marks)** Draw a Bayesian network that captures the information in the story above correctly and concisely. Make sure to annotate your network with conditional probability tables (CPTs).
2. **(worth 1 mark)** How many parameters will be required to specify the network you have drawn? Explain.
3. **(worth 5 marks)** A bird lands in the tree in the morning. What is the probability that it will stay in the tree longer than five minutes?
4. **(worth 5 marks)** What is the overall probability that any given bird in your tree will sing a song?

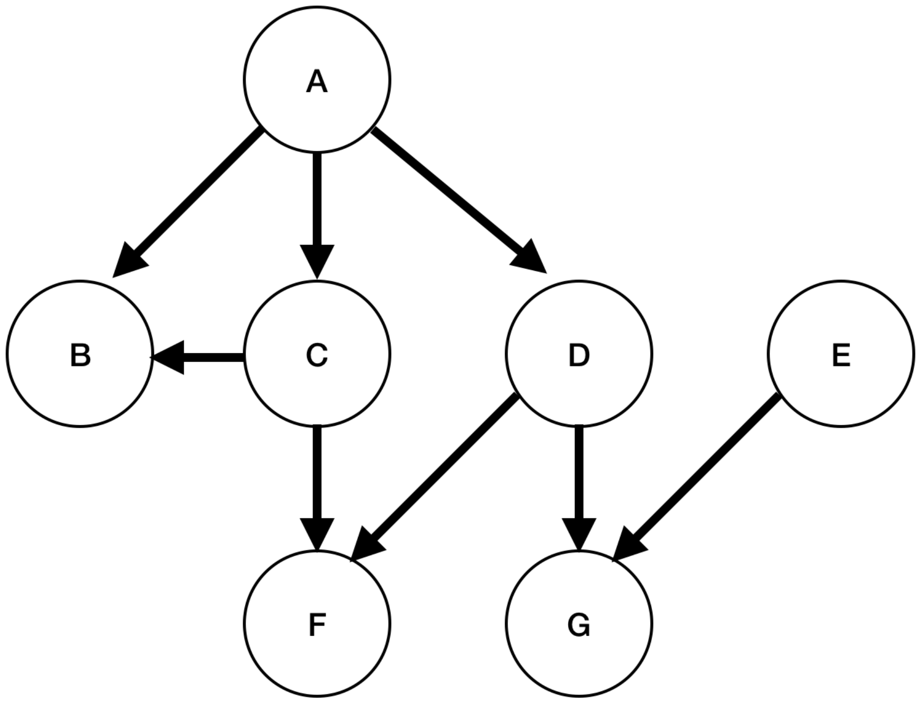
Q3. Markov Models (worth 12/100 marks)

In the mail room of a university office, a stream of blue and yellow mail carts pass a mail sorting robot. Blue carts contain 7 domestic letters for every 3 international letters. Yellow carts contain 2 domestic letters for every 3 international letters. Blue carts are followed by blue carts 70% of the time, but 30% of the time they are followed by a yellow cart. Yellow carts are followed by yellow carts half of of the time, and the other half of the time they are followed by a blue cart. The first cart, every morning, is yellow 9 of every 10 times.

The robot pulls a single letter from each cart as the cart passes it by.

1. **(worth 2 mark)** Draw a markov model to represent the joint distribution over cart colors and the origins of letters selected by the robot. Include a conditional probability table (CPT) with each variable in your model.
2. **(worth 5 marks)** What is the probability that the first three letters selected by the robot, in order, will be: $domestic, international, domestic$?
3. **(worth 5 marks)** What is the probability that the fourth cart will be blue if the first three letters selected, in order, are: $international, international, domestic$?

Q4. D-Separation and Relevance (worth 10/100 marks)



Given the Bayesian Network structure above ....

1. **(worth 1 mark)** How many parameters are required to fully specify the network? Explain.
2. **(worth 3 marks)** List three pairs of variables X, Y where X is independent of Y.
3. **(worth 3 marks)** List three sets of variables X, Y, Z where X is independent of Z given Y.
4. **(worth 1 mark)** Assume we are to calculate $P(B|E, F)$. Which variables are relevant?
5. **(worth 2 marks)** Assume we are to calculate $P(B|G)$ using variable elimination. List the elimination order you might suggest when using the min-fill heuristic to select variables, and give the size of the factors that result from each elimination.

Q5. First-order Structures and Models (worth 19/100 marks)

Consider a first-order language $\mathcal{L}_{metal}$ consisting of constant symbols $\mathbf{o_1}, \mathbf{o_2}, \mathbf{o_3}$, a binary predicate symbol $heavier\_than$, and unary predicate symbol $expensive$.
Let $D = \{Steel, Aluminium, Titanium\}$.

1. **(worth 6 marks)** How many $\mathcal{L}_{metal}$-structures with domain D exist? **Justify** your answer.

2. **(worth 3 marks)** Let Φ_1 be the set consisting of the following sentences:

$$\begin{aligned}&heavier\_than(\mathbf{o_2}, \mathbf{o_1})\\ &heavier\_than(\mathbf{o_1}, \mathbf{o_3})\\ &expensive(\mathbf{o_1})\\ &expensive(\mathbf{o_2})\end{aligned}$$

   Consider a set $\mathfrak{B}$ of $\mathcal{L}_{metal}$-structures for all structures $\mathcal{M} \in \mathfrak{B}$:

   - The domain of $\mathcal{M}$ is D and

$$\begin{aligned}&\mathbf{o_1}^{\mathcal{M}} = Steel\\ &\mathbf{o_2}^{\mathcal{M}} = Aluminium\\ &\mathbf{o_3}^{\mathcal{M}} = Titanium\\ &\langle Aluminium, Steel\rangle \in heavier\_than^{\mathcal{M}}\\ &\langle Steel, Titanium\rangle \in heavier\_than^{\mathcal{M}}\\ &\langle Titanium, Aluminium\rangle \in heavier\_than^{\mathcal{M}}\\ &Aluminium \in expensive^{\mathcal{M}}\\ &Steel \in expensive^{\mathcal{M}}\end{aligned}$$

   Are all structures in $\mathfrak{B}$ models of Φ_1? **Explain** why or why not.

3. **(worth 4 marks)** Suppose Φ_2 is the set obtained by adding the following sentence to Φ_1

$$\forall x \forall y \forall z (heavier\_than(x, y) \wedge heavier\_than(y, z)) \rightarrow heavier\_than(x, z)$$

   Are all structures in $\mathfrak{B}$ models of Φ_2? **Explain** why or why not.

4. **(worth 6 marks)** Consider the set Φ_3 that contains **only** the following axiom

$$\forall x \forall y \forall z (heavier\_than(x, y) \wedge heavier\_than(y, z)) \rightarrow heavier\_than(x, z)$$

   How many $\mathcal{L}_{metal}$-structures with domain D are *models* of Φ_3? **Justify** your answer.

Q6. Proof by Resolution (worth 31/100 marks)

Consider the following knowledge base (note that $\mathbf{p}, \mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}$ are constant symbols, $part(x, y)$ means x is a part of y and $precedes(x, y)$ means x precedes y):

$$\begin{aligned}\exists c_1 \exists c_2 \exists c_3 \big[&component(c_1, \mathbf{p_1}) \land component(c_2, \mathbf{p_2}) \land component(c_3, \mathbf{p_3}) \\ &\land assemble\_before(c_1, c_2, \mathbf{p}) \land assemble\_before(c_2, c_3, \mathbf{p})\big]\end{aligned} \tag{1}$$

$$\begin{aligned}\forall c_1 \forall c_2 \big[&(component(c_1, \mathbf{p_1}) \land component(c_2, \mathbf{p_2}) \land assemble\_before(c_1, c_2, \mathbf{p})) \\ &\rightarrow \exists c_3 \ (component(c_3, \mathbf{p_3}) \land assemble\_before(c_1, c_3, \mathbf{p}) \land assemble\_before(c_3, c_2, \mathbf{p}))\big]\end{aligned} \tag{2}$$

$$\begin{aligned}\forall c_1 \forall c_2 \forall a_1 \forall a_2 \big[&(component(c_1, a_1) \land component(c_2, a_2) \land assemble\_before(c_1, c_2, \mathbf{p})) \\ &\rightarrow precedes(a_1, a_2)\big]\end{aligned} \tag{3}$$

$$\begin{aligned}\forall c_1 \forall c_2 \forall c_3 \forall a \big[&(assemble\_before(c_1, c_2, a) \land assemble\_before(c_2, c_3, a)) \\ &\rightarrow assemble\_before(c_1, c_3, a)\big]\end{aligned} \tag{4}$$

$$\forall c_1 \forall c_2 \forall a \forall a_1 \big[(assemble\_before(c_1, c_2, a) \land component(c_1, a_1)) \rightarrow part(a_1, a)\big] \tag{5}$$

1. **(worth 11 marks)** Convert the sentences to **clausal form**.

2. **(worth 20 marks)** Use **resolution** to answer the following queries.
   You must use the **notation** developed in class (see slide no 39 in KRR-Part 2) for presenting your answers.

   (a) **(worth 10 marks)** What is a part of $\mathbf{p}$? (finding one answer is sufficient)
   Note: $part(x, y)$ denotes x is a part of y.

   (b) **(worth 10 marks)** What does precede $\mathbf{p_3}$? (finding one answer is sufficient)
   Note: $precedes(x, y)$ denotes x precedes y.

HAVE FUN and GOOD LUCK!