CS47100 Assignment 3

Due date: Wednesday November 16, 2022 (11:59pm)

This assignment will involve written and programming exercises.

Part 1: Written Assignment (50 pts)

Probability and Uncertainty

1. (20 pts) Basic Probability

   (a) (4 pts) Which of the following expressions is guaranteed to be 1, given two random variables X and Y, and no independence or conditional independence assumptions between them are given? For each expression, show why and why not it is guaranteed to be 1.

   - $\sum_y P(X = x|Y = y)$
   - $\sum_x P(X = x|Y = y)$
   - $\sum_x \sum_y P(X = x|Y = y)$
   - $\sum_x P(X = x) + \sum_y P(Y = y)$

   (b) (4 pts) Given two Boolean random variables A and B, where $P(A) = \frac{1}{2}, P(B) = \frac{1}{4}$, and $P(A|\neg B) = \frac{1}{3}$, apply Bayes's rule and calculate the value of $P(A|B)$.

   (c) You are a witness of a night-time hit-and-run accident involving a car in West Lafayette. Assuming that all cars in West Lafayette are gray or black. You swear, under oath, that the car you saw was black. Extensive testing shows that under the dim lighting conditions, discrimination between black and gray is 60% reliable. In other words, if a car is indeed black, with a 60% chance it will appear black under the dim lighting conditions. Similarly, if a car is indeed grey, the chance for it to appear grey under the dim lighting conditions is 60%.

      i. (6 pts) Is it possible to calculate the most likely color for the fled car to help the police find the suspect? Show why or why not. (*Hint*: distinguish carefully between the proposition that the car is black and the proposition that it appears black.)

      ii. (6 pts) Now consider the case when you are given the additional information that 57 out of 100 cars in West Lafayette are gray. Show whether or not it is possible to to calculate the most likely color for the fled car.

2. (11 pts) Inference and Independence

   (a) (5 pts) Let X_1, X_2 and Y be three discrete random variables. We want to calculate $P(Y|X_1, X_2)$ but we do not possess any independence/conditional independence information about them. For each of the following five cases, suppose we know the probability distributions listed in it. Please show whether it is possible for us to calculate $P(Y|X_1, X_2)$ and if yes, how to calculate.

      i. $P(X_1, X_2)$, $P(Y)$, $P(X_1|Y)$ and $P(X_2|Y)$

      ii. $P(X_1, X_2)$, $P(Y)$ and $P(X_1, X_2|Y)$

      iii. $P(X_1|Y)$, $P(X_2|Y)$ and $P(Y)$

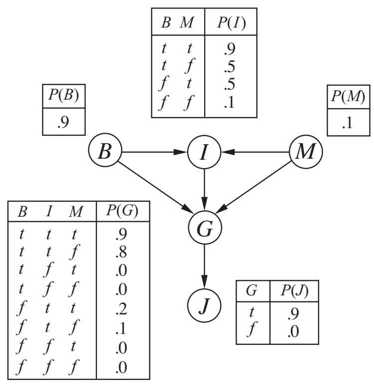
      iv. $P(X_1)$, $P(X_2)$ and $P(X_1, X_2|Y)$

v. $P(X_1)$, $P(X_2)$, $P(X_1|Y)$ and $P(X_2|Y)$

(b) (6 pts) Now suppose you know that $X_1 \perp\!\!\!\perp X_2|Y$ (i.e., X_1 and X_2 are conditionally independent with each other given Y). Now, which of the above cases are sufficient for us to calcuate $P(Y|X_1, X_2)$? Justify your answers.

Bayesian Networks

3. (19 pts) Consider the following Bayesian network (each variable is a Boolean variable, and the conditional probability table encodes the probability/conditional probability for a variable to take the value of “true”, e.g., for variable B, we have $P(B = \mathit{true}) = 0.9$):



P(B)
.9

B	M	P(I)
t	t	.9
t	f	.5
f	t	.5
f	f	.1

P(M)
.1

B	I	M	P(G)
t	t	t	.9
t	t	f	.8
t	f	t	.0
t	f	f	.0
f	t	t	.2
f	t	f	.1
f	f	t	.0
f	f	f	.0

G	P(J)
t	.9
f	.0

(a) (6 pts) How many free parameters are needed to specify an arbitrary joint probability distribution over 5 Boolean variables? How many free parameters are needed to specify the joint probability distribution over 5 Boolean variables given the Bayesian network above? (*Hint*: For a single random variable with n possible values, you need $n - 1$ free parameters to specify its probability distribution.)

(b) (6 pts) Calculate the value of $P(J = t|B, I, M)$ (i.e., list $P(J = t|B = b, I = i, M = m)$ for all combinations of b, i, m) using inference by enumeration.

(c) (7 pts) Calculate the value of $P(J, B)$ using variable elimination.

Submission

Upload your answers to the **written** questions as a pdf format file in Gradescope:

- For your pdf file, use the naming convention `username_hw#.pdf`. For example, your TA with username *mmostafi* would name his pdf file for HW3 as `mmostafi_hw3.pdf`.
- To make grading easier, please start a new page in your pdf file for each subquestion. Hint: use a `\newpage` command in LaTeX after every question ends. For example, use a `\newpage` command after each of part (a)-(b) of Question 1.
- After uploading to Gradescope, mark each page to identify which question is answered on the page. (Gradescope will facilitate this.)
- Follow the above convention and instruction for future assignments as well.

Part 2: Programming Assignment (50 pts)

For the programming assignments we will use the Pacman project designed for the course CS188 at UC Berkeley: `https://inst.eecs.berkeley.edu/~cs188/fa20/project4/`

Please remember that solutions to any assignment should be your own. Using other people so- lutions, within or outside Purdue goes against the course academic honesty policy. The TAs will be using code similarity measures to detect plagiarism cases when grading the assignment.

In this assignment, we will use the Pacman projects 4. In project 4, you will design inference methods for Pacman agents to use to locate and eat invisible ghosts. As in HW1 and HW2, the assignment includes an autograder for you to grade your answers on your machine.

TODO:

1. Complete Project 4, Questions 0-4 described on the Berkeley site. Submit your modified versions of `inference.py` and `bustersAgents.py` for grading. We will multiply your original score returned by `autograder.py` (10 pts in total) by 5.

Submission

Please upload the following files: `bustersAgents.py` and `inference.py` to Gradescope.