DEPARTMENT OF
COMPUTING SCIENCE
UNIVERSITY OF ALBERTA

CMPUT 366: Assignment #3

Due at 10pm on November 21, 2019

Abstract

For this assignment use the following consultation model:

- you can discuss assignment questions and exchange ideas with other CMPUT 366 students;
- you must list all members of the discussion in your solution;
- you may **not** share/exchange/discuss written material and/or code;
- you must write up your solutions individually;
- you must fully understand and be able to explain your solution in any amount of detail as requested by the instructor and/or the TAs.

Anything that you use in your work and that is not your own creation must be properly cited by listing the original source. Failing to cite others' work is plagiarism and will be dealt with as an academic offence.

First name: Shaoyang

Last name: Zhou

CCID: Shaoyang@ualberta.ca

Collaborators: Chris Lo

Your mark: ______ **out of 100**

Submission

The assignment you downloaded from eClass is a single ZIP archive which includes this document as a PDF file as well as its LaTeX source. Your answers are to be submitted electronically via eClass. Your submission must be a single PDF file with all of your answers. To generate the PDF file you can do any of the following:

1. insert your answers into the provided LaTeX source file between `\begin{answer}` and `\end{answer}`. Then run the source through LaTeX to produce a PDF file;

2. write your answers in the blank spaces under each question. Make sure you write as legibly as possible for we cannot give you any points if we cannot read your hand-writing;

3. use your favourite text processor and type up your answers there. Make sure you number your answers in the same way as the questions are numbered in this assignment.

1 Reinforcement Learning

Consider an MDP with actions $\mathcal{A} = \{a, b, c\}$, states $\mathcal{S} = \{W, X, Y\}$ and unknown dynamics. Suppose that you have previously computed the following estimated action values:

$$
\begin{array}{ccc}
Q(W,a) = 0 & Q(X,a) = 8 & Q(Y,a) = 0 \\
Q(W,b) = 0 & Q(X,b) = 7 & Q(Y,b) = 0 \\
Q(W,c) = 0 & Q(X,c) = 16 & Q(Y,c) = 0
\end{array}
$$

Now suppose that starting from state $S_t = W$, the current behaviour policy selects action $A_t = a$, leading to reward $R_{t+1} = 4$ and a transition to state $S_{t+1} = X$. The behaviour policy then selects action $A_{t+1} = a$.

1.1

[6 points] What is the updated estimate for $Q(W, a)$ according to the Q-learning algorithm? Assume a step size of $\alpha = 0.5$ and a discount rate of $\gamma = 1$. Show your work.

* Following Q-Learning in Slides

Init | $S=W$

Choose A | $A=a$

observe R,S' | $R=4, S'=X$

Update Q(W,a) | $Q(S,A) = Q(S,A) + \frac{1}{2}\left[R + \max_{a' \text{ of } S'}(Q(S',a')) - Q(S,A)\right]$

$= 0 + \frac{1}{2}[4 + 16 - 0]$

$= 0 + \frac{1}{2}[20]$

$Q(S,A) = 10$

1.2

[6 points] What is the updated estimate for $Q(W, a)$ according to the Sarsa algorithm? Assume a step size of $\alpha = 0.5$ and a discount rate of $\gamma = 1$. Show your work.

* Following SARSA in Slides

Init | $S=W$

Choose a | $A=a$

observe R,S' | $R=4, S'=X$

choose a' | $A'=a$

Update Q(S,a) | $Q(S,A) = Q(S,A) + \frac{1}{2}[R + Q(S',A') - Q(S,A)]$

$= 0 + \frac{1}{2}[4 + 8 - 0]$

$= \frac{1}{2} \cdot 12$

$Q(S,A) = 6$

2 Uncertainty

2.1

[4 points] Suppose Alice and Bob independently read a clock. Unlike the clock in our lectures, this clock does have markings on its face. However, being a 12-hour clock, it does not distinguish between before noon and after noon times. For instance a 10:00 display can be understood as 10am or 10pm. Alice's answer is the random variable A. Bob's answer is the random variable B. Assume that Alice and Bob both read the time perfectly but guess randomly whether it is before noon or after noon. For instance, given that the clock shows 10:00, A can be 10am while B can be 10pm. Are Alice and Bobs opinions independent? In other words, is it true that $P(A|B) = P(A)$ and $P(B|A) = P(B)$? Explain your answer.

No

- ~~Yes~~, assuming they both recall their observation truthfully
  - Since outside of the time unit they are effectively randomly guessing a binary variable, this is in essence two independent coin tosses. If they are both recalling and observing the unit of time perfectly, then this bit of information should ~~[illegible]~~. confirm each other.

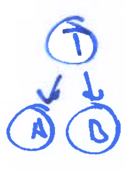
    ie $P(A=12pm) = \frac{1}{24}$, counting only hours

    but $P(A=12pm | B=12am) = \frac{1}{2}$, counting only hours

2.2

[4 points] Are Alice and Bobs opinions independent given the clock display? n other words, is it true that $P(A|B,T) = P(A|T)$ and $P(B|A,T) = P(B|T)$? Explain your answer.

- Yes, assuming they both truthfully recall their observation



  - Given T, the information from the other variable then has no further meaning, it would be redundant. Hence, we would expect that once we condition on T, any other ~~[illegible]~~ variable would not change our probability given our knowledge of their condition.

2.3

[20 points] Consider the following fictional scenario (all numbers are completely made up). Imagine 2% of the people who walk through a specific metal detector at an airport are carrying a gun. 30% of the people who walk through the same metal detector are carrying coins. The remaining 68% are carrying nothing made of metal. Everyone carries *only one* of the following items: nothing, coins, a gun.

If someone carries a gun through this metal detector it will beep with probability 95%. If someone carries coins through this same metal detector it will beep with probability 80%. If someone carries nothing made of metal through the detector it will beep 25% of the time.

Suppose that the metal detector beeps when someone walks through it. What is the probability that the person is carrying a gun? Show how you calculated your answer.

• Via bayes's rule

$$P(\text{Gun} \mid \text{Beep}) = \frac{P(\text{Beep} \mid \text{Gun})\, P(\text{Gun})}{P(\text{Beep})}$$ Find components

$P(\text{Beep} \mid \text{Gun}) = 95\% \times 2\%$, $P(\text{Gun}) = 2\%$

$$P(\text{Beep}) = \sum_{t \in \text{Type}} P(\text{Beep} \mid t)$$

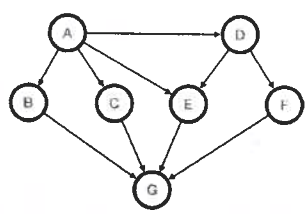
$$= 2\% \cdot 95\% + 10\% \cdot 90\% + 68\% \cdot 25\% = 0.4290$$

$$P(\text{Gun} \mid \text{Beep}) = \frac{\cancel{0.02} \times 0.95 \times 0.02}{0.4290} = \frac{\cancel{0.00038 \times}}{0.4290}\ 0.95 \times 0.02 = \cancel{0.000866} \quad \frac{0.95 \times 0.02}{0.4290}$$

3 Belief Networks: Basics

3.1

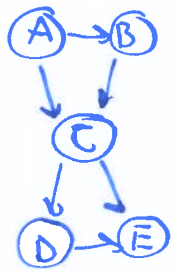
[5 points] What factorization of the joint distribution $P(A, B, C, D, E, F, G)$ does the network below represent?



$P(A \ldots G) = P(A) \cdot P(B \mid A) \cdot P(C \mid A) \cdot P(D \mid A) \cdot P(E \mid A, D) \cdot P(F \mid D) \cdot \ldots$
$P(G, B, C, E, F)$

3.2

[5 points] Draw a belief network that is consistent with a joint distribution that factors as $P(A, B, C, D, E) = P(E|C, D)P(D|C)P(C|A, B)P(B|A)P(A)$.



3.3

[3 points] Suppose that every random variable in the joint distribution of Question 3.2 has a domain containing 10 elements. How many rows are needed to list the full joint distribution in an explicit table?

10^5 to cover $A \times B \times C \times D \times E$ combinations.

* Could remove one case to establish base case

3.4

[7 points] Suppose that every random variable in the joint distribution of Question 3.2 has a domain containing 10 elements. How many rows in total are needed to list the conditional probability tables for your belief network representation?

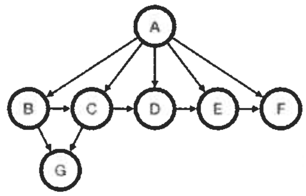
- Represent each variable w/ its own lookup table, mapping inputs × outputs combinations

Node	Input combinations	Output combinations	Total combinations (Rows)
A	0	10	10
B	10	10	10^2
C	10^2	10	10^3
D	10	10	10^2
E	10^2	10^2	10^3

≈ 22,010 Rows

4 Belief Networks: Variable Elimination

Consider the belief network below.



4.1

[15 points] List the factors that would be created, and the operations used to create them, by running the variable elimination algorithm on this belief network to answer the query $P(B|G,E)$. Use the variable ordering G, E, A, B, C, D, F.

Obs = G, E, Query = B

Initial
f(A)
f2(B,A)
f3(C,B,A)
f4(E,D,A)
f5(F,E,A)
f6(G,B,C)
f(D,C,A)

G: Condition
remove f(G,B,C)
replace w/ f(B,C
f1(A)
f2(B,A)
f3(C,B,A)
f4(E,D,A)
f5(F,E,A)
f7(B,C)
f(D,C,A)

E: Condition
remove f(E,D,A), f(F,E,A)
replace w/ f(D,A), f(F,A)
f1(A)
f2(B,A)
f3(C,B,A)
f8(D,A)
f9(F,A)
f7(B,C)
f(D,C,A)

A: Sum out A
$f_{10}(B,C,D,F) = \sum_A f(A) \times f(B,A) \times f(C,B,A) \times f(D,C,A) \times f(D,A) \times f(F,A)$
f(B,C,D,F)
f(B,C)

B: Skip

C: Sum out C $f(B,D,F) = \sum_C f(B,C,D,F) \times f(B,C)$
f(B,F

F: Sum out
f(B)

4.2

[15 points] List the factors that would be created, and the operations used to create them, by running the variable elimination algorithm on this belief network to answer the query $P(B|G, E)$. Use the variable ordering G, E, F, D, C, B, A.

Init

f(A)
f(B,A)
f(C,B,A)
f(D,C,A)
f(E,D,A)
f(F,E,A)
f(G,B,C)

G: Condition

replace f(G,B,C)
w/ f(B,C)

f(A)
f(B,A)
f(C,B,A)
f(D,C,A)
f(E,D,A)
f(F,E,A)
f(B,C)

E: Condition

replace f(E,D,A), f(F,E,A)
w/ f(D,A), f(F,A)

f(A)
f(B,A)
f(C,B,A)
f(D,C,A)
f(D,A)
f(F,A)
f(B,C)

F: Sum out

replace f(F,A)
w/ f(A)

f(A)
f(B,A)
f(C,B,A)
f(D,C,A)
f(D,A)
f(A)
f(B,C)

D: Sum out

replace Σ_D f(D,C,A) × f(D,A)
w/ f(C,A)

f(A)
f(B,A)
f(C,B,A)
f(C,A)
f(A)
f(B,C)

C: Sum out

replace Σ_C f(C,B,A) × f(C,A) × f(B,C)
w/ f(B,A)

f(A)
f(B,A)
f(B,A)
f(A)

B: skip

A: Sum out

replace Σ_A f(A) × f(B,A) × f(B,A) × f(A)
w/ f(B)

f(B)

4.3

[10 points] Which of the two given variable orderings is more efficient for this query? Explain your answer.

- Second ordering is more efficient
  - Variable ordering generates more efficient sums of products
  - Cross products in summation terms should be as small as possible.
  - Hence consider the different ordering of variable "A"
    - In 4.1, summing A early means generating a large cross product of

      $f(A) \times f(B,A) \times f(D,A,E) \times f(D,C,B) \times f(E,A) \times f(F,A)$

      $10 \quad 10^2 \quad 10^3 \quad 10^3 \quad 10^2 \quad 10^2$

    - In 4.2, summing A late generates considerable savings

      $f(A) \times f(A,D) \times f(A,B) \times f(A)$

      $10 \quad 10^2 \quad 10^2 \quad 10$

  - 4.2 computes much smaller cross products in general as it sums out the critical variable A late, hence it does a much better job in keeping small, efficient products