KOÇ UNIVERSITY
College of Engineering

COMP 341: INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Midterm 2

Fall 2019, 02/12/2019
Duration: 110 Minutes

Name: Solutions

ID: 00110001 00110000 00110000

- This exam contains 9 pages including this cover page and 5 questions. Check to see if any pages are missing. Put your initials on the top of every page, in case the pages become separated.
- By submitting this exam, you **agree** to fully comply with Koç University Student Code of Conduct, and accept any punishment in case of failure to comply. If you do not submit this exam on time, you will receive 0 credits.
- The exam is **closed book** and **closed notes**. You are **not** allowed to use any additional material including any electronic equipment such as computers and mobile phones.
- You are expected to be able provide clear and concise answers. Gibberish will not receive any credit. Your answers need to be readable by a human, illegible writing is not gradable.
- Read each question carefully and make sure to follow instructions. The written instructions take precedence over an answer that the instructor might give you during the exam, unless the instructor makes a class wide announcement.
- Do not write in the table below.

Question:	1	2	3	4	5	Total
Points:	16	16	16	17	35	100
Score:						

1. (16 points) True or False :

False — If X and Y are conditionally independent given Z, then the following is always holds, $P(Z|X,Y) = P(Z|X)P(Z|Y)$

True — The normalization procedure is possible since all probabilities need to add up to 1.

True — The joint distribution can be used to calculate any query about its random variables.

False — Variable elimination in Bayesian Networks is more efficient due to it always having better computational complexity than enumeration.

False — In decision networks, we need to specify a conditional probability table for the action nodes.

True — Value of perfect information is always non-negative.

False — Samples obtained from Gibbs sampling need to be weighted with their probabilities.

True — We can still answer queries in Hidden Markov Models even if we do not get emissions at each time step.

2. (16 points) Probability. Answer the questions below. Note that $\cap$ means intersection (events happening together), $\cup$ means union (either one event happens or both events happen).

(a) (5 points) The events, A and B, are independent. Write the below probabilities in terms of $P(A)$ and $P(B)$

- $P(A \cap B) = P(A)P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$

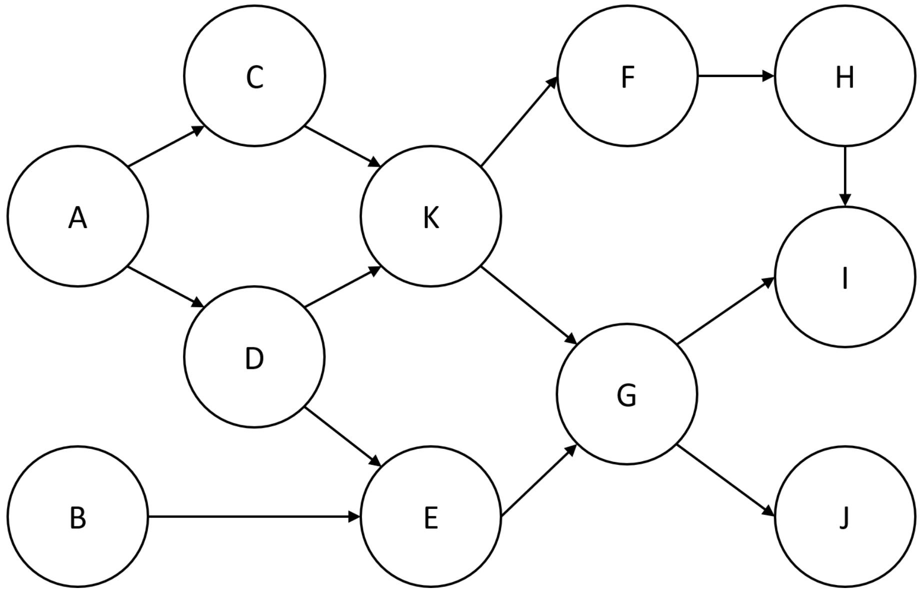
(b) (5 points) The events, A and B, are disjoint (or mutually exlusive). Write the below probabilities in terms of $P(A)$ and $P(B)$.

- $P(A \cap B) = 0$
- $P(A \cup B) = P(A) + P(B)$

(c) (6 points) Your favorite politician has a social media account. However, he only writes 25% of his posts, whereas the remaining 75% is written by his assistant. When he writes a post, there is a 40% chance that there is a typo. The posts written by his assistant never has a typo. Calculate the following probabilities.

- Probability that a given post doesn't have a typo. You may leave your answer as a fraction.
  $0.25 \cdot 0.6 + 0.75 = 0.9$
- Given a post without a typo, probability of it being written by the politician. You may leave your answer as a fraction.
  $P(W = p|T = false) = P(W = p, T = f)/P(T = f) = (0.25 \cdot 0.6)/(0.9) = 1/6$

3. (16 points) Bayesian Networks. Answer the questions using the Bayesian Network below.



(a) (12 points) Answer the questions about conditional independence below with Yes/No (considering all possible CPTs).

Yes Are A and B conditionally independent given D?

No Are A and B conditionally independent given E?

No Are A and I conditionally independent given K?

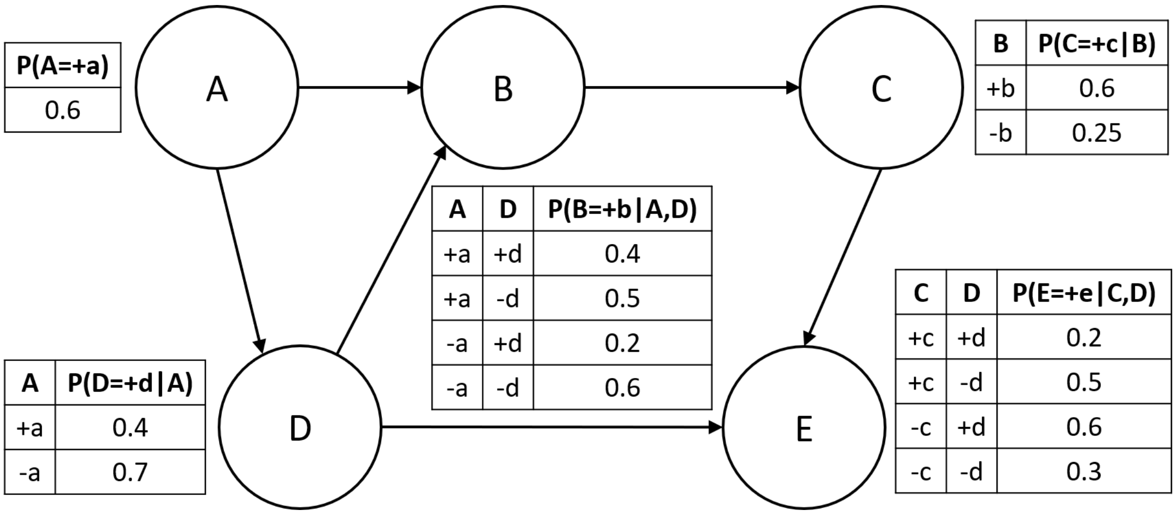
No Are A and I conditionally independent given G?

No Are B and I conditionally independent given G?

No Are B and I conditionally independent given E and G?

(b) (4 points) Write down all the variables that are in the Markov Blanket of K.
C,D,E,F,G

4. (17 points) Bayesian Networks. You are given the Bayesian Network (BN) below. All random variables are binary. The CPTs for the variables are given only for their *positive* values (subtract from 1 to get the probabilities for the *negative* values). Answer the questions using this BN.



P(A=+a)
0.6

A	P(D=+d|A)
+a	0.4
-a	0.7

A	D	P(B=+b|A,D)
+a	+d	0.4
+a	-d	0.5
-a	+d	0.2
-a	-d	0.6

B	P(C=+c|B)
+b	0.6
-b	0.25

C	D	P(E=+e|C,D)
+c	+d	0.2
+c	-d	0.5
-c	+d	0.6
-c	-d	0.3

(a) (2 points) What is the joint distribution, $P(A,B,C,D,E)$, of this BN?

$P(A,B,C,D,E) = P(A)P(D|A)P(B|A,D)P(C|B)P(E|C,D)$

(b) (7 points) You want to use likelihood weighting to calculate $P(E|+d,-b)$. Fill in the table with the corresponding weights. Then calculate $P(E|+d,-b)$ using the counts. You may leave the result as a mathematical expression.

A	B	C	D	E	counts	weights
+a	-b	+c	+d	+e	2	$0.4 \times 0.6 = 0.24$
+a	-b	+c	+d	-e	14	$0.4 \times 0.6 = 0.24$
+a	-b	-c	+d	+e	23	$0.4 \times 0.6 = 0.24$
+a	-b	-c	+d	-e	19	$0.4 \times 0.6 = 0.24$
-a	-b	+c	+d	+e	2	$0.7 \times (1-0.2) = 0.56$
-a	-b	+c	+d	-e	6	$0.7 \times (1-0.2) = 0.56$
-a	-b	-c	+d	+e	23	$0.7 \times (1-0.2) = 0.56$
-a	-b	-c	+d	-e	11	$0.7 \times (1-0.2) = 0.56$

$P(E=+e|+d,-b) = \frac{(2+23)(0.24)+(2+23)(0.56)}{(2+23)(0.24)+(2+23)(0.56)+(14+19)(0.24)+(6+11)(0.56)} = \frac{20}{37.44} \approx 0.53$

0.5 for weights, 3 for the expression

(c) (8 points) Calculate $P(E|+d,-b)$ with any exact inference method.

This solution will be based on variable elimination. Given the evidence, A and E are conditionally independent so we do not need any factor including A. Solution:

Initial Factors: $f_1(-b,C) = P(C|-b)$ and $f_2(C,+d,E) = P(E|C,+d)$

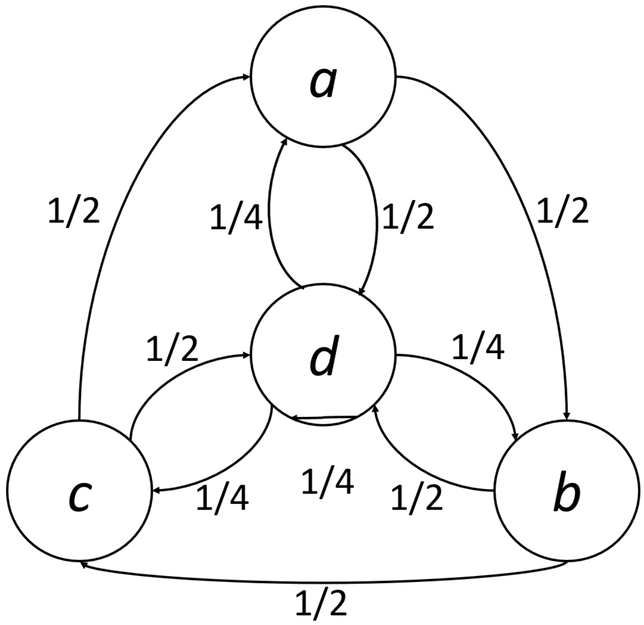
Step 1: Merge $f_1(-b,C)$ and $f_2(C,+d,E)$ to get $f_3(-b,C,+d,E)$

C	E	$f_3(-b,C,+d,E)$
$+c$	+e	$0.25 \times 0.2 = 0.05$
$+c$	-e	$0.25 \cdot (1-0.2) = 0.2$
$-c$	+e	$(1-0.25) \cdot 0.6 = 0.45$
$-c$	-e	$(1-0.25) \cdot (1-0.6) = 0.3$

Step 2: Sum out C from $f_3(-b,C,+d,E)$ to get $f_4(-b,+d,E)$
$f_4(-b,+d,+e) = 0.05 + 0.45 = 0.5$
$f_4(-b,+d,-e) = 0.2 + 0.3 = 0.5$

No need to normalize since all the values sum up to 1. Thus $P(+e|+d,-b) = 0.5$ and $P(-e|+d,-b) = 0.5$

5. (35 points) Hidden Markov Models. You are given the underlying Markov Chain of an HMM below, along with the emission probabilities. Answer the questions below usign this HMM.



Transitions Probabilities:

- $P(s_{t+1} = i|s_t = d) = \frac{1}{4}, i = a, b, c, d$
- $P(s_{t+1} = d|s_t = j) = \frac{1}{2}, j = a, b, c$
- $P(s_{t+1} = b|s_t = a) = \frac{1}{2}$
- $P(s_{t+1} = c|s_t = b) = \frac{1}{2}$
- $P(s_{t+1} = a|s_t = c) = \frac{1}{2}$

Emission Probabilities:

- $P(e_t = d|s = d) = 7/10$
- $P(e_t = j|s = d) = 1/10, j = a, b, c$
- $P(e_t = i|s = i) = 3/4, i = a, b, c$
- $P(e_t = d|s = i) = 1/4, i = a, b, c$

Probabilities not shown are equal to 0

(a) (5 points) You are given that $s_0 = d$. What is the state distribution after a single time step?
Just follow the transition probabilities for 1 step. Note that the questions states $P(s_0 = d) = 1$, thus we do not need to look at any other state for s_0.

$P(s_1 = i|s_0 = d) = 1/4$ for $i = a, b, c, d$ OR

$P(s_1 = a|s_0 = d) = 1/4$
$P(s_1 = b|s_0 = d) = 1/4$
$P(s_1 = c|s_0 = d) = 1/4$
$P(s_1 = d|s_0 = d) = 1/4$

Note that $P(s_1 = i|s_0 = d) = B'(s_1 = i)$

(b) (8 points) You get an emission at the first time step as $e_1 = a$. What is the state distribution now?

By looking at the emission probabilities and the evidence, we can see that the state can either be a or d. All remaining options would have 0 emission probabilities. As such:

$P(s_1 = a|e_1 = a) \propto P(e_1 = a|s_1 = a)P(s_1 = a|s_0 = 1)$
(OR $B_1(s = a) = P(e_1 = a|s_1 = a)B'(s_1 = a)$)
$P(s_1 = a|e_1 = a) \propto 3/4 \cdot 1/4 = 3/16 = 15/80$

Similarly;
$P(s_1 = d|e_1 = a) \propto 1/10 \cdot 1/4 = 1/40 = 2/80$

After normalization, $P(s_1 = a|e_1 = a) = 15/17$ and $P(s_1 = a|e_1 = a) = 2/17$

(c) (8 points) You get another emission at the second time step as $e_2 = c$. What is the state distribution now?

By looking at the emission probabilities and the evidence, we can see that the state can either be c or d. All remaining options would have 0 emission probabilities so associated probabilities will be ignored.

$P(s_2 = c|e_1 = a) \propto \Sigma_{s\in\{a,d\}} P(s_2 = c|s_1)P(s_1 = s|e_1 = a) = 0 \cdot 15/17 + 1/4 \cdot 2/17 = 1/34$
$P(s_2 = d|e_1 = a) \propto \Sigma_{s\in\{a,d\}} P(s_2 = d|s_1)P(s_1 = s|e_1 = a) = 1/2 \cdot 15/17 + 1/4 \cdot 2/17 = 16/34$

$P(s_2 = c|e_1 = a, e_2 = c) \propto P(e_2 = c|s_2 = c)P(s_2 = c|e_1 = a) = 3/4 \cdot 1/34 = 3/(4 \cdot 34) =$
$P(s_2 = d|e_1 = a, e_2 = c) \propto P(e_2 = d|s_2 = c)P(s_2 = d|e_1 = a) = 1/10 \cdot 16/34 = 16/(10 \cdot 34)$

After normalization, $P(s_2 = c|e_1 = a, e_2 = c) = 15/47$ and $P(s_2 = d|e_1 = a, e_2 = c) = 32/47$

(d) (4 points) Comment on what s_1 was after these two time steps.

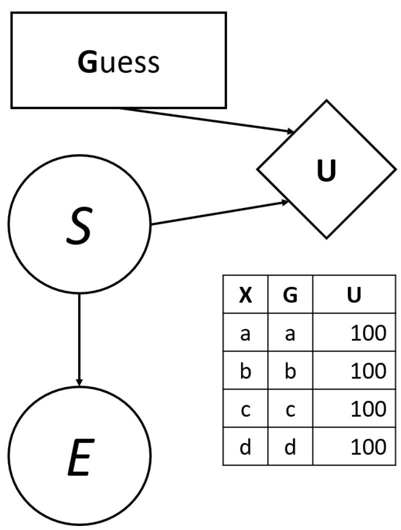
The main thing I was looking for was the fact that the HMM being in the state d at $t = 1$ is higher than what we calculated initially at part b. In other words $P(s_1 = d|e_1 = a) < P(s_1|e_1 = a, e_2 = c)$. If you actually did the math, there is a bonus waiting for you!

(e) (10 points) The HMM is reset and you have no idea of the state distribution. To be on the safe side, you assume that the hidden state can be anything, i.e., $P(s_0 = i) = 1/4, i = a, b, c, d$. At this point, correctly guessing the state gives you a bonus of 100. However, getting an emission also has a cost of 10. Would you guess without getting an emission or after getting an emission?

This is no longer a reasoning over time question but a decision network question. There are two ways to answer this question. One way is the "simple reasoning" and the other is the decision network.

Simple Reasoning:
Level 1: If we get one of a, b, c, we have $3/4$ of being correct. Thus expected utility is 75. If we get d, then our EU is 70. Both (minus 25) are larger then 10, so it is worth to pay to get an emission. (This is wrong since it misinterprets the emission probabilities). This gets 3 points.
Level 2: Let's assume that $P(e|x) = P(x|e)$. This implies that the probability of getting any emission is also uniformly distributed. Then follow the logic of level 1. This gets 5 points.

Decision Network Answer (This gets 10 points but requires much more work):
I wanted you to convert the HMM with a uniform prior distribution to a BN for one time step and then add the necessary action and decision nodes. It is fine if this was done implicitly. Then the next step was to calculate the value of getting an emission.
We have four actions; guessing one of the states. We get 100 for a correct guess. It is implied that all the other utilities are zeros.



X	G	U
a	a	100
b	b	100
c	c	100
d	d	100

Making a guess without any observation assuming a uniform distribution:
$MEU(\emptyset) = 100 \cdot 1/4 = 25$

To calculate the rest, we need $P(s = i|e = i) = P(e = i|s = i)P(s = i)/P(e = i)$. Let's keep it like this for now.

It should be obvious that MEU for whatever observation is achieved when we guess what is observed. Noting that $U(g=i,e=j)=0$ for $i \neq j$:
$MEU(e=j) = P(s=j|e=j)U(g=j,s=j) = 100P(s=j|e=j)$

Value of getting an emission (Noting $P(s)=1/4$ from the uniform prior assumption):
$VPI(e) = \Sigma_j P(e=j)MEU(e=j) - 25$
$P(e=j)MEU(e=j) = 100P(e=j)P(s=j|e=j)$
$= 100P(e=j)P(e=j|s=j)P(s=j)/P(e=j)$
$= 100P(e=j|s=j)P(s=j) = 25P(e=j|s=j)$.
Then
$VPI(e) = \Sigma_j P(e=j)MEU(e=j) - 25 = 25 \cdot (3 \cdot 3/4 + 7/10) - 25 = 73.65 - 25 = 48.75$
$VPI(e) > 10$, thus I would be willing to pay for the emission.