CS 188 Introduction to
Summer 2015 Artificial Intelligence

Midterm 2

- You have approximately 80 minutes.

- The exam is closed book, closed calculator, and closed notes except your one-page crib sheet.

- Mark your answers ON THE EXAM ITSELF. If you are not sure of your answer you may wish to provide a *brief* explanation. All short answer sections can be successfully answered in a few sentences AT MOST.

First name	
Last name	
SID	
edX username	
Name of person on your left	
Name of person on your right	

For staff use only:

Q1.	Probability and Bayes Nets	/10
Q2.	Factors	/5
Q3.	Moral Graphs	/9
Q4.	Hearthstone Decisions	/6
Q5.	Sampling	/12
Q6.	I Heard You Like Markov Chains	/6
	Total	/48

THIS PAGE IS INTENTIONALLY LEFT BLANK

Q1. [10 pts] Probability and Bayes Nets

(a) [3 pts] A, B, and C are random variables with binary domains. How many entries are in the following probability tables and what is the sum of the values in each table? Write a "?" in the box if there is not enough information given.

Table	Size	Sum
$P(A, B|C)$	8	2
$P(A|+b, +c)$	2	1
$P(+a|B)$	2	?

(b) [4 pts] Circle true if the following probability equalities are valid and circle false if they are invalid (leave it blank if you don't wish to risk a guess). Each True/False question is worth 1 points. Leaving a question blank is worth 0 points. **Answering incorrectly is worth −1 points.**

No independence assumptions are made.

(i) [1 pt] [*true* or *false*] $P(A, B) = P(A|B)P(A)$
False. $P(A, B) = P(A|B)P(B)$ would be a valid example.

(ii) [1 pt] [*true* or *false*] $P(A|B)P(C|B) = P(A, C|B)$
False. This assumes that A and C are conditionally independent given B.

(iii) [1 pt] [*true* or *false*] $P(B, C) = \sum_{a \in A} P(B, C|A)$
False. $P(B, C) = \sum_{a \in A} P(A, B, C)$ would be a valid example.

(iv) [1 pt] [*true* or *false*] $P(A, B, C, D) = P(C)P(D|C)P(A|C, D)P(B|A, C, D)$
True. This is a valid application of the chain rule.

(c) Space Complexity of Bayes Nets

Consider a joint distribution over N variables. Let k be the domain size for all of these variables, and let d be the maximum indegree of any node in a Bayes net that encodes this distribution.

(i) [1 pt] What is the space complexity of storing the entire joint distribution? Give an answer of the form $O(\cdot)$.
$O(k^N)$ was the intended answer. Because of the potentially misleading wording, we also allowed $O(Nk^{d+1})$, one possible bound on the space complexity of storing the Bayes net ($O((N-d)k^{d+1})$ is an asymptotically tighter bound, but this requires considerably more effort to prove).

(ii) [1 pt] Draw an example of a Bayes net over four binary variables such that it takes less space to store the Bayes net than to store the joint distribution.
A simple Markov chain works. Size $2+4+4+4 = 14$, which is less than $2^4 = 16$. Less edges, less inbound edges (v-shape), or no edges would work too.
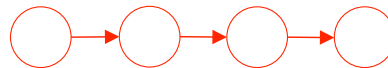
(iii) [1 pt] Draw an example of a Bayes net over four binary variables such that it takes more space to store the Bayes net than to store the joint distribution.
Size $2 + 2 + 2 + 2^4 = 22$, which is more than $2^4 = 16$. Other configurations could work too, especially any with a node with indegree 3.
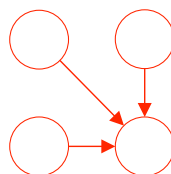
Q2. [5 pts] Factors

Consider the probability tables below for two factors $P(A|+b,C)$ and $P(C|+b)$.

	A	B	C	Value
	$+a$	$+b$	$+c$	w
$P(A|+b,C)$	$+a$	$+b$	$-c$	x
	$-a$	$+b$	$+c$	y
	$-a$	$+b$	$-c$	z

	B	C	Value
$P(C|+b)$	$+b$	$+c$	r
	$+b$	$-c$	s

(a) [1 pt] What probability distribution results from multiplying these two factors?

$$f_1 = P(A, C|+b)$$

(b) [3 pts] Write the complete probability table for the resulting factor f_1, including the computed values (in terms of the letters r, s, w, x, y, z).

	A	B	C	Value
	$+a$	$+b$	$+c$	wr
$P(A, C|+b)$	$+a$	$+b$	$-c$	xs
	$-a$	$+b$	$+c$	yr
	$-a$	$+b$	$-c$	zs

(c) [1 pt] Assuming the given tables for $P(A|+b,C)$ and $P(C|+b)$ were normalized. Do we need to normalize the values in f_1 to generate valid probablities?

No. No evidence was introduced, just multiplying doesn't require normalization.

Q3. [9 pts] Moral Graphs

(a) [2 pts] For each of the following queries, we want to preprocess the Bayes net before performing variable elimination. Query variables are double-circled and evidence variables are shaded. Cross off all the variables that we can ignore in performing the query. If no variables can be ignored in one of the Bayes nets, write "None" under that Bayes net.
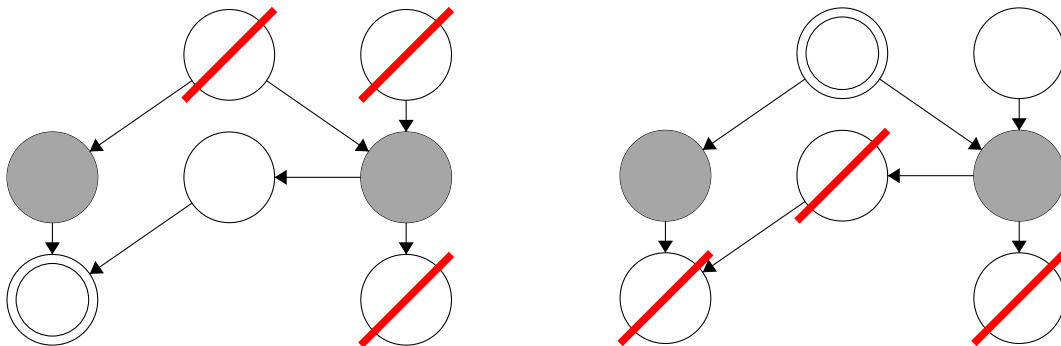

Let B be a Bayes net with a set of variables V. The **Markov blanket** of a variable $v \in V$ is the smallest set of variables $S \subset V$ such that for any variables $v' \in V$ such that $v \neq v'$ and $v' \notin S$, $v \perp\!\!\!\perp v'|S$. Less formally, v is independent from the entire Bayes net given all the variables in S.

(b) [2 pts] In each of the following Bayes nets, shade in the Markov blanket of the double-circled variable.
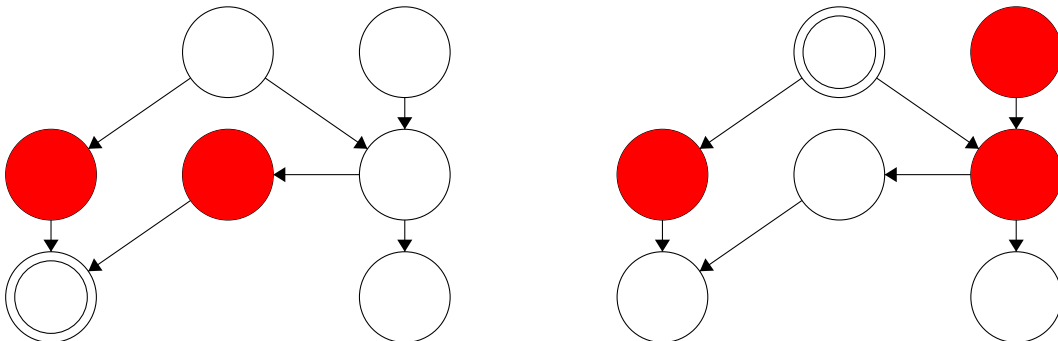

The **moral graph** of a Bayes net is an **undirected** graph with the same vertices as the Bayes net (i.e. one vertex corresponding to each variable) such that each variable has an edge connecting it to every variable in its Markov blanket.

(c) [3 pts] Add edges to the graph on the right so that it is the moral graph of the Bayes net on the left.
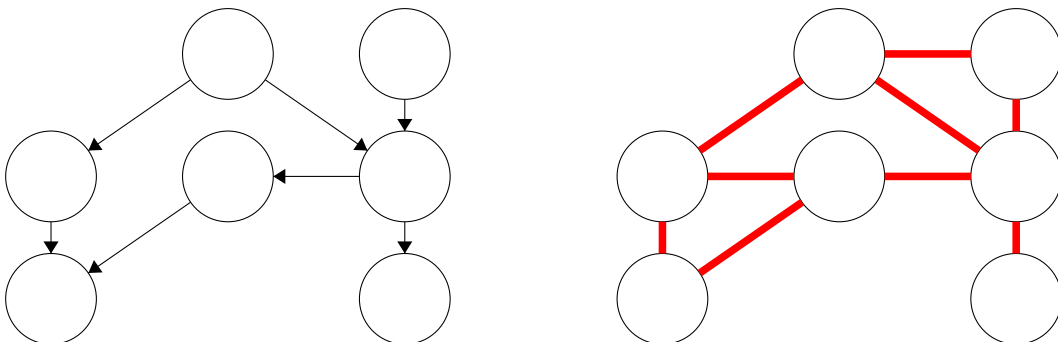
(d) [2 pts] The following is a query in a moral graph for a larger Bayes net (the Bayes net is not shown). Cross off all the variables that we can ignore in performing the query.
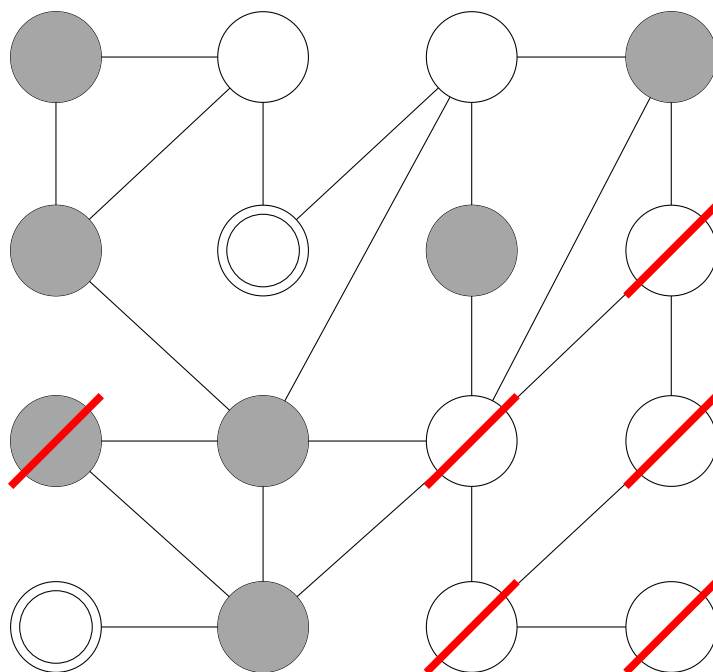
Q4. [6 pts] Hearthstone Decisions

You are playing the game Hearthstone. You are up against the famous player Trump.

On your turn, you can choose between playing 0, 1, or 2 minions. You realize Trump might be holding up an Area of Effect (AoE) card, which is more devastating the more minions you play.

- If Trump has the AoE, then your chances of winning are:

  - 60% if you play 0 minions
  - 50% if you play 1 minion
  - 20% if you play 2 minions

- If Trump does NOT have the AoE, then your chances of winning are:

  - 20% if you play 0 minions
  - 60% if you play 1 minion
  - 90% if you play 2 minions
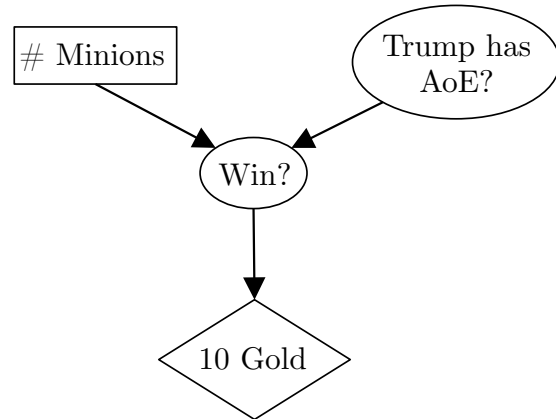
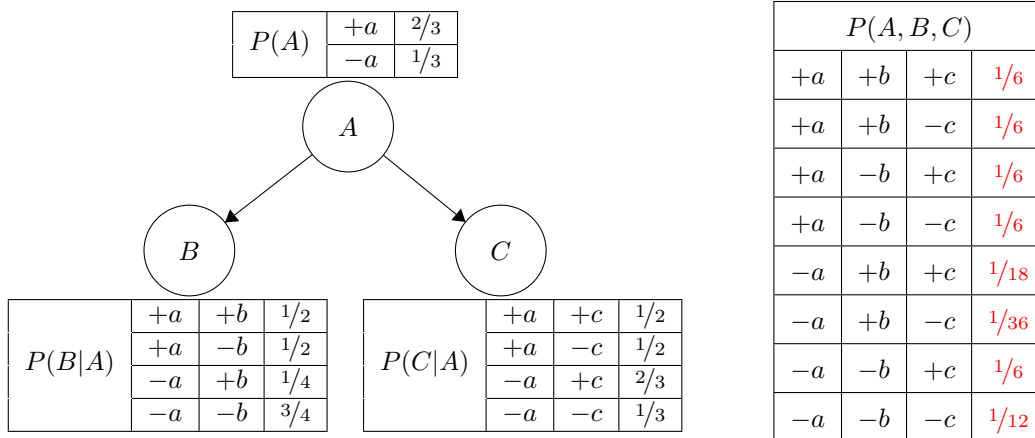You know that there is a 50% chance that Trump has an AoE.

Winning this game is worth 10 gold and losing is worth 0.

Solution notation: A: Trump has AoE?, W: Win?, M: Number of minions

(a) [1 pt] How much gold would you expect to win choosing 0 minions?

$\sum_w \sum_a (P(w|Minion = 0, a)P(a)R(w) = 10\sum_a (P(w|Minion = 0, a)P(a) = 10(.6 \cdot .5 + .2 \cdot .5) = 4$

(b) [1 pt] How much gold would you expect to win choosing 1 minion?

$\sum_w \sum_a (P(w|Minion = 1, a)P(a)R(w) = 10\sum_a (P(w|Minion = 1, a)P(a) = 10(.5 \cdot .5 + .6 \cdot .5) = 5.5$

(c) [1 pt] How much gold would you expect to win choosing 2 minions?

$\sum_w \sum_a (P(w|Minion = 2, a)P(a)R(w) = 10\sum_a (P(w|Minion = 2, a)P(a) = 10(.2 \cdot .5 + .9 \cdot .5) = 5.5$

(d) [1 pt] How much gold would you expect to win if you know the AoE is in Trump's hand?

$\max_m \sum_w P(w|m, +a)R(w) = 10\max_m P(w|m, +a) = 10\max\{.6, .5, .2\} = 6$

(e) [1 pt] How much gold would you expect to win if you know the AoE is NOT in Trump's hand?

$\max_m \sum_w P(w|m, -a)R(w) = 10\max_m P(w|m, -a) = 10\max\{.2, .6, .9\} = 9$

(f) [1 pt] How much gold would you be willing to pay for to know whether or not the AoE is in Trump's hand? (Assume your utility of gold is the same as the amount of gold.)

Two. The difference between $MEU(\{\}) = 5.5$ and $MEU(\{A\}) = .5 * 6 + .5 * 9 = 7.5$ is 2.

Q5. [12 pts] Sampling

Consider the following Bayes net. The joint distribution is not given, but it may be helpful to fill in the table before answering the following questions.

$P(A)$	
$+a$	$2/3$
$-a$	$1/3$

A

B C

$P(B\|A)$		
$+a$	$+b$	$1/2$
$+a$	$-b$	$1/2$
$-a$	$+b$	$1/4$
$-a$	$-b$	$3/4$

$P(C\|A)$		
$+a$	$+c$	$1/2$
$+a$	$-c$	$1/2$
$-a$	$+c$	$2/3$
$-a$	$-c$	$1/3$

$P(A, B, C)$			
$+a$	$+b$	$+c$	$1/6$
$+a$	$+b$	$-c$	$1/6$
$+a$	$-b$	$+c$	$1/6$
$+a$	$-b$	$-c$	$1/6$
$-a$	$+b$	$+c$	$1/18$
$-a$	$+b$	$-c$	$1/36$
$-a$	$-b$	$+c$	$1/6$
$-a$	$-b$	$-c$	$1/12$

We are going to use sampling to approximate the query $P(C| + b)$. Consider the following samples:

Sample 1	Sample 2	Sample 3
$(+a, +b, +c)$	$(+a, -b, -c)$	$(-a, +b, +c)$

(a) [6 pts] Fill in the following table with the probabilities of drawing each respective sample given that we are using each of the following sampling techniques.

$$P(+b) = \frac{2}{6} + \frac{1}{12} = \frac{5}{12}$$

$P(\text{sample} \mid \text{method})$	Sample 1	Sample 2
Prior Sampling	$1/6$	$1/6$
Rejection Sampling	$\frac{1/6}{5/12} = 2/5$	0
Likelihood Weighting	$2/3 \cdot 1/2 = 1/3$	0

Lastly, we want to figure out the probability of getting Sample 3 by Gibbs sampling. We'll initialize the sample to $(+a, +b, +c)$, and resample A then C.

(b) [1 pt] What is the probability the sample equals $(-a, +b, +c)$ after resampling A?

$$P(-a| + b, +c) = \frac{P(-a, +b, +c)}{P(-a, +b, +c) + P(+a, +b, +c)} = \frac{1/18}{1/18 + 1/6} = \frac{1/18}{4/18} = \frac{1}{4}$$

(c) [1 pt] What is the probability the sample equals $(-a, +b, +c)$ after resampling C, given that the sample equals $(-a, +b, +c)$ after resampling A?

$$P(+c| - a, +b) = P(+c| - a) = \frac{2}{3}$$

(d) [1 pt] What is the probability of drawing Sample 3, $(-a, +b, +c)$, using Gibbs sampling in this way?

$$P(-a| + b, +c) \cdot P(+c| - a, +b) = \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}$$

(e) [2 pts] Suppose that through some sort of accident, we lost the probability tables associated with this Bayes net. We recognize that the Bayes net has the same form as a naïve Bayes problem. Given our three samples:

$$(+a, +b, +c), \ (+a, -b, -c), \ (-a, +b, +c)$$

Use naïve Bayes maximum likelihood estimation to approximate the parameters in all three probability tables.

$P(A)$		
	$+a$	$2/3$
	$-a$	$1/3$

$P(B\|A)$			
	$+a$	$+b$	$1/2$
	$+a$	$-b$	$1/2$
	$-a$	$+b$	1
	$-a$	$-b$	0

$P(C\|A)$			
	$+a$	$+c$	$1/2$
	$+a$	$-c$	$1/2$
	$-a$	$+c$	1
	$-a$	$-c$	0

(f) [1 pt] What problem would Laplace smoothing fix with the maximum likelihood estimation parameters above?

Laplace smoothing would help prevent overfitting to our very few number of samples. It would avoid the zero probabilities found in the parameters above. It would bring the estimated parameters closer to uniform, which in this case is closer to the original parameters than the maximum likelihood estimated parameters.

Q6. [6 pts] I Heard You Like Markov Chains

In California, whether it rains or not from each day to the next forms a Markov chain (note: this is a terrible model for real weather). However, sometimes California is in a drought and sometimes it is not. Whether California is in a drought from each day to the next itself forms a Markov chain, and the state of this Markov chain affects the transition probabilities in the rain-or-shine Markov chain. This is the state diagram for droughts:
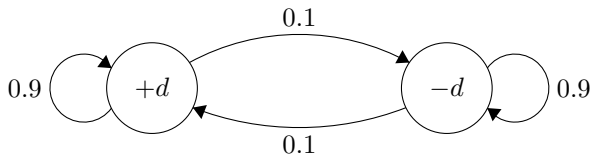
These are the state diagrams for rain given that California is and is not in a drought, respectively:
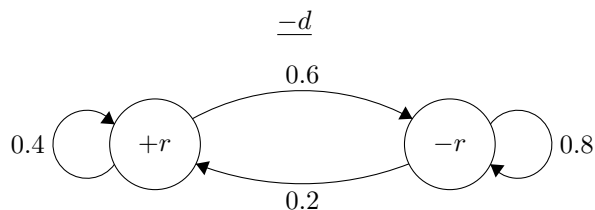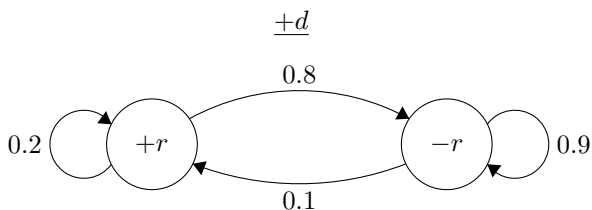
(a) [1 pt] Draw a dynamic Bayes net which encodes this behavior. Use variables D_{t-1}, D_t, D_{t+1}, R_{t-1}, R_t, and R_{t+1}. Assume that on a given day, it is determined whether or not there is a drought before it is determined whether or not it rains that day.
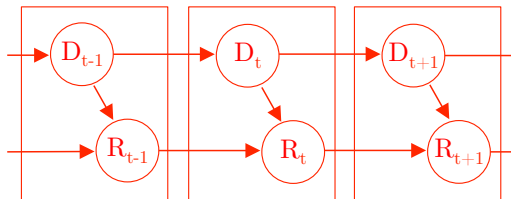
(b) [1 pt] Draw the CPT for D_t in the above DBN. Fill in the actual numerical probabilities.

$P(D_t\|D_{t-1})$		
$+d_{t-1}$	$+d_t$	0.9
$+d_{t-1}$	$-d_t$	0.1
$-d_{t-1}$	$+d_t$	0.1
$-d_{t-1}$	$-d_t$	0.9

(c) [1 pt] Draw the CPT for R_t in the above DBN. Fill in the actual numerical probabilities.

$P(R_t\|R_{t-1}, D_t)$			
$+d_t$	$+r_{t-1}$	$+r_t$	0.2
$+d_t$	$+r_{t-1}$	$-r_t$	0.8
$+d_t$	$-r_{t-1}$	$+r_t$	0.1
$+d_t$	$-r_{t-1}$	$-r_t$	0.9
$-d_t$	$+r_{t-1}$	$+r_t$	0.4
$-d_t$	$+r_{t-1}$	$-r_t$	0.6
$-d_t$	$-r_{t-1}$	$+r_t$	0.2
$-d_t$	$-r_{t-1}$	$-r_t$	0.8

Suppose we are observing the weather on a day-to-day basis, but we cannot directly observe whether California is in a drought or not. We want to predict whether or not it will rain on day $t+1$ given observations of whether or not it rained on days 1 through t.

(d) [1 pt] First, we need to determine whether California will be in a drought on day $t+1$. Derive a formula for $P(D_{t+1}|r_{1:t})$ in terms of the given probabilities (the transition probabilities on the above state diagrams) and $P(D_t|r_{1:t})$ (that is, you can assume we've already computed the probability there is a drought today given the weather over time).

$$P(D_{t+1}|r_{1:t}) = \sum_{d_t} P(D_{t+1}|d_t)P(d_t|r_{1:t})$$

(e) [2 pts] Now derive a formula for $P(R_{t+1}|r_{1:t})$ in terms of $P(D_{t+1}|r_{1:t})$ and the given probabilities.

$$P(R_{t+1}|r_{1:t}) = \sum_{d_{t+1}} P(D_{t+1}|r_{1:t})P(R_{t+1}|r_t, d_{t+1})$$

THIS PAGE IS INTENTIONALLY LEFT BLANK