Probabilistic Artificial Intelligence

Final Exam

Feb 2, 2016

Time limit:	120 minutes
Number of pages:	19
Total points:	100

You can use the back of the pages if you run out of space.
Collaboration on the exam is strictly forbidden.

(1 point) Please fill in your student ID and fullname (LASTNAME, FIRSTNAME) in all capital letters.

1. Propositional and First-order Logic (12 points)

A minesweeper world is a rectangular grid of m squares with n invisible mines scattered among them. Each square contains at most 1 mine. Any square may be probed by the minesweeper; if the probed square does not contain a mine, then a number from 0-8 will be revealed, indicating the number of mines that are *directly* or *diagonally* adjacent. The minesweeper has to find all the mines in a given minefield without detonating any of them.

(4 points) (i) We want to provide a first-order knowledge base that formalizes the knowledge of a player in a game state. We consider the following first-order vocabulary:

- Mine(x), a unary predicate, which denotes that the square x contains a mine
- Adj(x, y), a binary predicate, which means that the square x is adjacent to the square y
- Contains(x, n), a binary predicate, which denotes that the square x contains the number n

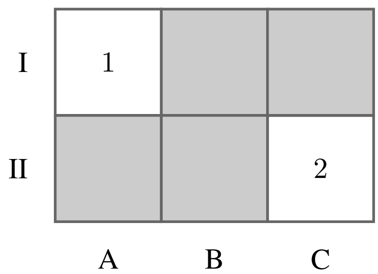
Using first-order logic, formalize the following knowledge:

- If a square contains the number 1, then there *cannot* be a mine in that square, and there is exactly one mine in the adjacent squares.

(8 points) (ii) Consider the game state provided in the figure below. The shaded squares represent locations that have not yet been probed. We use the propositional symbols

$$A_{\mathrm{I}}, A_{\mathrm{II}}, B_{\mathrm{I}}, B_{\mathrm{II}}, C_{\mathrm{I}}, C_{\mathrm{II}}$$

to denote that the corresponding square contains a mine. For example, A_{I} denotes that square (A, I) contains a mine, and $\neg A_{\mathrm{I}}$ denotes that there is no mine in square (A, I).



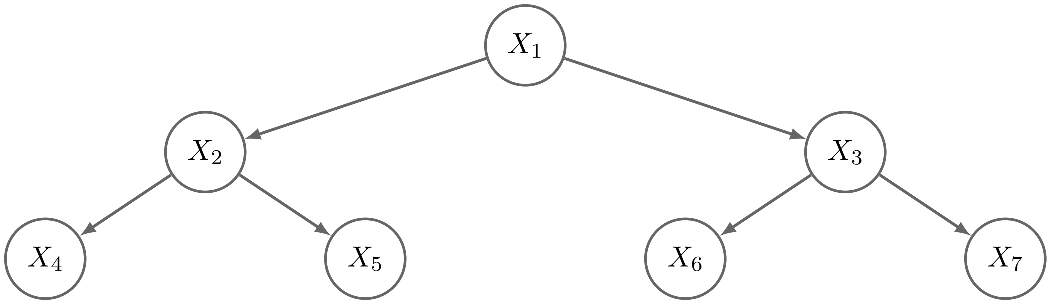
We have already established the following facts in propositional logic:

- $B_{\mathrm{I}} \wedge C_{\mathrm{I}} \Leftrightarrow \neg B_{\mathrm{II}}$
- $B_{\mathrm{I}} \Leftrightarrow \neg A_{\mathrm{II}} \wedge \neg B_{\mathrm{II}}$
- $B_{\mathrm{II}} \Leftrightarrow \neg A_{\mathrm{II}} \wedge \neg B_{\mathrm{I}}$

Using propositional resolution, and the three facts provided above, prove that there is no mine in square $(\mathrm{A}, \mathrm{II})$.

2. Information Cascade (20 points)

As part of developing an early warning system against earthquakes, the government of a city has installed a grid of sound alarms throughout the city to notify the people in case an earthquake is imminent. The nodes in the grid are arranged in the form of a binary tree. The root of the tree receives the noise-free signal about the earthquake, and the information is transmitted by a node to its left and right child nodes, thus creating a downward information cascade starting from the root to the leaves. Let there be a total of N nodes in the grid. The scenario is modeled by the following Bayesian network shown below for the case of $N = 7$ total nodes:



Bayesian network for $N = 7$ nodes

Each node X_i represents a random variable corresponding to the binary signal $\{0, 1\}$ observed at that node. For the root node X_1 where the information cascade begins, the probability of observing a signal of 1 is $2/3$. As the cascade unfolds, each node transmits the signal it received to its left and right children nodes. However, there is a transmission error of $1/4$ at the *receiving-end*, i.e. the probability that a node X_i observes the same signal as that observed by its parent $\pi(X_i)$ is $3/4$. The conditional probability tables of this information cascade process are shown in the tables below.

X_1	$P(X_1)$
1	$2/3$
0	$1/3$

X	$\pi(X)$	$P(X\|\pi(X))$
1	1	$3/4$
0	1	$1/4$
1	0	$1/4$
0	0	$3/4$

(4 points) (i) You are given evidence E that node X_2 observed signal 0 and node X_3 observed signal 1. Given this evidence, compute the conditional probability that the signal observed by the root X_1 is 1.

(4 points) (ii) You are given evidence E that node X_2 observed signal 0, and all other nodes $X_3, X_4, \ldots X_7$ observed signal 1. Given this evidence, compute the conditional probability that the signal observed by root X_1 is 1.

(2 points) (iii) You are given evidence E that node X_1 observed signal 1. Given this evidence, compute the conditional probability that the signal observed by X_3 is 1.

(4 points) (iv) You are given evidence E that node X_1 observed signal 1. Given this evidence, compute the conditional probability that the signal observed by X_7 is 1.

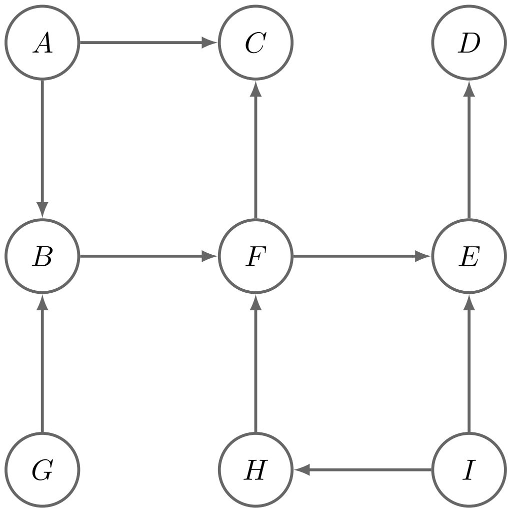
(6 points) (v) Given evidence E that node X_1 observed signal 1, let us define $\gamma(N)$ to be the expected fraction of nodes in the network that observed signal 1 (i.e. $X_i = 1$) . Formally, we can state $\gamma(N)$ by:

$$\gamma(N) = \frac{1}{N} \cdot \sum_{i=1}^{N} P(X_i = 1 | X_1 = 1)$$

Compute $\gamma(N)$ for the Bayesian network shown above, i.e. for $N = 7$.

3. Variable Elimination (15 points)

Considering the Bayesian network shown below, solve the following questions.



(4 points) (i) Find a minimal set $\mathcal{S}_d$ of variables that must be observed so that d-sep$(G; E \mid \mathcal{S}_d)$ holds. Minimal means that removing any element from $\mathcal{S}_d$ will make the statement false. Note that $\mathcal{S}_d$ can be $\emptyset$ if the property holds without any additional observations. Briefly explain your answer. We use d-sep$(X; Y \mid \mathcal{Z})$ to denote that X is d-separated from Y, given a set of observed variables $\mathcal{Z}$.

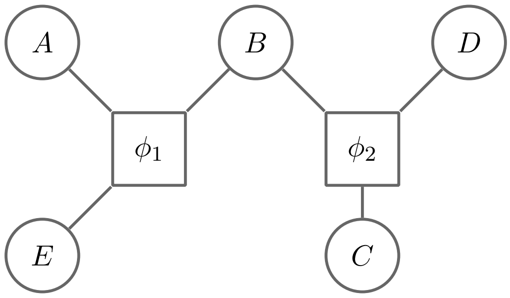
(2 points) (ii) Determine a minimal set of edges that should be removed from the network to convert it into a polytree. Briefly explain why the resulting network is a polytree.

(4 points) (iii) Consider the resulting polytree after removing the edges from the previous answer. Suggest an ordering for variable elimination which results in factors with the minimum size possible. Briefly explain how you selected that ordering.

(5 points) (iv) Consider the ordering from (iii). For the first four iterations of the variable elimination algorithm, determine which factors are removed and introduced.

4. Belief propagation (14 points)

Consider the factor graph shown below and its associated tables. Answer the following questions.



A	B	E	$\phi_1(A,B,E)$
0	0	0	0.25
0	0	1	0.00
0	1	0	0.75
0	1	1	1.00
1	0	0	0.00
1	0	1	0.75
1	1	0	1.00
1	1	1	0.25

B	C	D	$\phi_2(B,C,D)$
0	0	0	0.9
0	0	1	0.7
0	1	0	0.1
0	1	1	0.3
1	0	0	1.0
1	0	1	0.5
1	1	0	0.0
1	1	1	0.5

(6 points) (i) Compute the first message $\mu^{(1)}_{\phi_1 \to B}$ from factor node ϕ_1 to variable node B. Remember that the messages from variable nodes v to factor nodes u are initialized as:

$$\mu^{(0)}_{v \to u} = 1$$

(8 points) (ii) Compute the first message $\mu^{(1)}_{\phi_2 \to B}$ from factor node ϕ_2 to variable node B, and use the result to compute the estimated marginal distribution $\hat{P}^{(2)}(B)$.

5. Inside an RL robot (18 points)

Imagine that you are a robot. Your task is to press one of the two big red buttons, labeled *Left* and *Right*. Whenever you press a button, a big capital letter (A, B, or C) flashes up in front of you and, depending on which letter is shown, you get a certain amount of reward.

After pressing the buttons at random for a while, you realize that there is an underlying pattern that determines the amount of reward that you obtain at each time step. You want to figure out which buttons you should press in order to maximize the future rewards.

(9 points) (i) As a first step you want to model the system based on your past experience. Assume you have recorded the past data as tuples, (s, a, s', r), containing the state s before transition, action a taken (button pressed), state s' after the transition, and the reward r received. Given the following observations, estimate the transition probabilities and state-dependent rewards and draw the corresponding MDP.

A, *Left*, 5, A
A, *Right*, 10, B
B, *Right*, 0, C
C, *Right*, 5, A
A, *Right*, 5, A
A, *Right*, 5, A
A, *Right*, 5, A
A, *Right*, 10, B
B, *Left*, 10, B
B, *Left*, 5, A
A, *Right*, 10, B
B, *Right*, 0, C
C, *Left*, 5, A

(9 points) (ii) Given the model of the system it is now time to find the optimal policy. You decide to discount future rewards with a factor of $\gamma = 0.5$ and want to use policy iteration to find the optimal button to press.

Recall that policy iteration starts with an arbitrary initial policy π. Until convergence, it iteratively computes the value function $V_\pi(x)$ for the current policy and then updates the current policy to be the greedy policy π_g w.r.t the computed $V_\pi(x)$. The greedy policy for a value function is given by

$$\pi_g(x) = \arg\max_a \sum_{x'} P(x'|x,a)\Big(r(x,a,x') + \gamma V_\pi(x')\Big),$$

and given a fixed policy π, the value function $V_\pi(x)$ satisfies the condition

$$V_\pi(x) = \sum_{x'} P(x'|x,\pi(x))[r(x,\pi(x),x') + \gamma V_\pi(x')].$$

Compute the optimal policy and its value function for the above MDP. *[Hint: To save time, start with an initial guess for the optimal policy and prove that it is greedy w.r.t. the corresponding value function, i.e. policy iteration terminates]*

6. Qmax **(8 points)**

Consider a grid world, in which we want to learn a policy that maximizes future rewards. The world is modeled as a deterministic MDP that is initially unknown to the learner.

At each state, there are two possible actions: *Left* and *Right*, which deterministically move the learner to the corresponding neighboring state (or have the learner stay in place, if the action tries to move out of the grid). At each state, an action receives the state-dependent reward shown in the grid-world below, irrespective of the new state after taking the action.

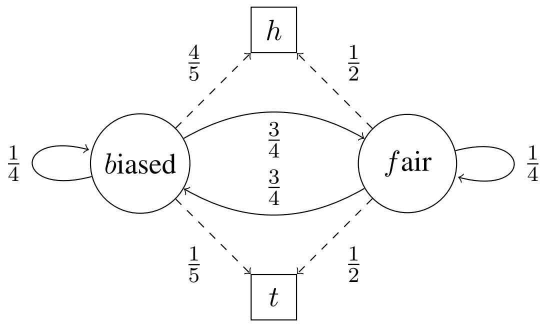
20	30
1	2

(8 points) (i) The learner uses optimistic Q-learning with a discount factor $\gamma = 0.5$ and learning-rate $\alpha = 0.5$ in order to learn the optimal policy. Recall that optimistic Q-Learning initializes the reward estimates for each state-action pair to a high value ($Q_{\max} = 100$). At each iteration, it picks the action with the highest estimated Q-value from all the available actions in the current state. Ties are broken by picking actions in the order (*Right*, *Left*).

Write down the first 5 state-action pairs that are taken by the learner when exploring the environment as well as the corresponding Q-value updates. The learner starts in field 1.

7. Hidden Markov Models (12 points)

Consider the following Hidden Markov Model. In each time step, a coin is flipped, resulting in heads (h) or tails (t). There are two coins, one is biased (b) and the other one is fair (f). After each coin flip (indexed by $i \in \mathbb{N}$), the coin that is used in the next time step changes with probability $\frac{3}{4}$. The prior for the two coins at $i = 1$ is $P(X_1 = b) = \frac{3}{5}$ and $P(X_1 = f) = \frac{2}{5}$. The fair coin results in heads and tails with $\frac{1}{2}$ probability each, whereas the biased coin results in heads with probability $\frac{4}{5}$ and in tails with probability $\frac{1}{5}$. This process is illustrated below:



(6 points) (i) Derive the probability $\lim_{i \to \infty} P(X_i = b)$, i.e. the probability of flipping the biased coin as $i \to \infty$.

(6 points) (ii) Derive the probability $\lim_{i \to \infty} P(X_i = b \mid Y_1 = t)$, i.e. the probability of flipping the biased coin as $i \to \infty$, given that you observe tails at $i = 1$.