Name: ______________________ SUID: ______________________

Stanford University

AA228/CS238: Decision Making under Uncertainty

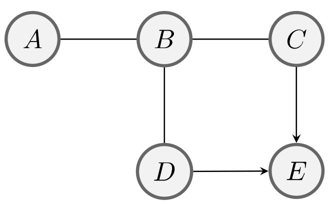
Autumn 2020

Prof. Mykel J. Kochenderfer • Remote • email: *mykel@stanford.edu*

MIDTERM 1 **Due date: October 2, 2020 (5pm)**

You have **90 minutes** to complete this exam. This exam is electronically timed; you do not have to keep track of your own time. To accommodate those in other timezones and complex working situations, you may choose any 90 minute window between 5pm PDT October 1, 2020 and 5pm PDT October 2, 2020 to take the exam. Answer all questions. You may consult any material (e.g., books, calculators, computer programs, and online resources), but you may not consult other people inside or outside of the class. If you need clarification on a question, please make a private post on Piazza. **Only what is submitted prior to the deadline will be graded.**

Question 1. (2pts) The following graph encodes a Markov equivalence class:



a) (1pt) How many members are in this class?

b) (1pt) Give the Markov blanket of node C.

Solution:

a) 4

b) B, D, E

Question 2. (4pts) Anna and Emma are shooting basketball free throws. In their game, they take four shots each, and the one who makes the most baskets wins. Before we see them play, we start with independent uniform priors over each player successfully making a basket with their shot. After taking three shots each, Anna made two baskets and Emma made three. What is the probability that their game results in a tie?

Solution: We use θ_A and θ_E to represent the probabilities that Anna and Emma makes a basket with their shot. With uniform priors, our posteriors become

$$\theta_A \sim \beta(1+2, 1+1) = \beta(3,2) \qquad \theta_E \sim \beta(1+3, 1+0) = \beta(4,1)$$

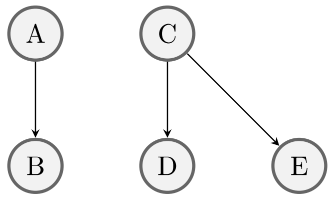
The expected value is then

$$\hat{\theta}_A = P(a^1) = 0.6 \qquad \hat{\theta}_E = P(e^1) = 0.8 \tag{1}$$

For the game to end in a tie, Anna must make a basket and Emma must miss:

$$P(\text{tie}) = \hat{\theta}_A(1-\hat{\theta}_E) = 0.12$$

Question 3. (7pts) We have the following Bayesian network with binary variables and dataset D consisting of five samples, where b_2, b_5, c_2, d_3 represent missing entries:



$$D = \begin{array}{c|ccccc} \text{A} & 1 & 1 & 0 & 1 & 0 \\ \text{B} & 1 & b_2 & 0 & 0 & b_5 \\ \text{C} & 1 & c_2 & 1 & 0 & 1 \\ \text{D} & 0 & 0 & d_3 & 1 & 1 \\ \text{E} & 0 & 1 & 1 & 1 & 1 \end{array}$$

a) (5pts) Perform a single iteration of expectation maximization given the priors $P(b^0 \mid a^1) = 0.7, P(b^0 \mid a^0) = 0.5$, $P(c^1) = 0.2$, $P(d^1 \mid c^1) = P(e^1 \mid c^1) = 0.6, P(d^1 \mid c^0) = P(e^1 \mid c^0) = 0.25$. You only need to solve for θ_b and θ_d. [For partial credit, please show your work by listing your sample weights.]

b) (2pts) Use the posterior mode to infer missing data entries b_5 and d_3 using the result from (a).

Solution:

a) By adding the possible permutations for each sample with missing entries, we find ten sample instantiations with the following weights:

s_1	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	$s_{2,4}$	$s_{3,1}$	$s_{3,2}$	s_4	$s_{5,1}$	$s_{5,2}$
1	1	1	1	1	0	0	1	0	0
1	1	1	0	0	0	0	0	1	0
1	1	0	0	1	1	1	0	1	1
0	0	0	0	0	1	0	1	1	1
0	1	1	1	1	1	1	1	1	1
1	$w_{2,1}$	$w_{2,2}$	$w_{2,3}$	$w_{2,4}$	0.6	0.4	1	0.5	0.5

We use $s_{i,j}$ to represent the jth completion of the ith sample. With the weights being proportional to $P(b \mid a)P(c)P(d^0 \mid c)P(e^1 \mid c)$, we find

$$\begin{aligned} w_{2,1} &= \alpha \times 0.3 \times 0.2 \times 0.4 \times 0.6 = 0.07273, \\ w_{2,2} &= \alpha \times 0.3 \times 0.8 \times 0.75 \times 0.25 = 0.2273, \\ w_{2,3} &= \alpha \times 0.7 \times 0.8 \times 0.75 \times 0.25 = 0.530, \\ w_{2,4} &= \alpha \times 0.7 \times 0.2 \times 0.4 \times 0.6 = 0.17 \end{aligned}$$

with the normalizing constant $\alpha = 5.0505$. The weighted counts for B are

$$\begin{bmatrix} 1.5 & 0.5 \\ w_{2,3} + w_{2,4} + 1 & 1 + w_{2,1} + w_{2,2} \end{bmatrix}$$

When we normalize the rows, we obtain a maximum likelihood estimate

$$\hat{\theta}_b = \begin{bmatrix} 0.75 & 0.25 \\ 0.566 & 0.433 \end{bmatrix}$$

The weighted counts for D are

$$\begin{bmatrix} w_{2,2} + w_{2,3} & 1 \\ 1.4 + w_{2,1} + w_{2,4} & 1.6 \end{bmatrix}$$

When we normalize the rows, we obtain a maximum likelihood estimate

$$\hat{\theta}_d = \begin{bmatrix} 0.43 & 0.57 \\ 0.507 & 0.493 \end{bmatrix}$$

b) We infer $b_5 = 0$ and $d_3 = 0$ given these parameters. Due to rounding errors, you may end up with a uniform distribution over B when c^1. In that case, you could infer $d_3 = 1$ as well. Partial credit is awarded if we can see the inference from question 3.a.

Solution: Since A and B are independent, we can split in the samples and evaluate them separately.

s_1	$s_{2,1}$	$s_{2,2}$	$s_{3,1}$	$s_{3,2}$	s_4	s_5
1	1	0	1	1	0	1
0	0	0	1	0	1	1
0	1	1	1	1	1	1
1	$w_{2,1}$	$w_{2,2}$	0.6	0.4	1	1

with $w_{2,i} = \alpha P(c)P(d^0 \mid c)P(e^1 \mid c)$

$$w_{2,1} = \alpha \times 0.2 \times 0.4 \times 0.6 = 0.24$$
$$w_{2,2} = \alpha \times 0.8 \times 0.75 \times 0.25 = 0.76$$

and

s_1	$s_{2,1}$	$s_{2,2}$	s_3	s_4	$s_{5,1}$	$s_{5,2}$
1	1	1	0	1	0	0
1	1	0	0	0	1	0
1	0.3	0.7	1	1	0.5	0.5

The counts associated with B are

$$\begin{bmatrix} 1.5 & 0.5 \\ 1.7 & 1.3 \end{bmatrix}$$

The maximum likelihood estimate is

$$\hat{\theta}_b = \begin{bmatrix} 0.75 & 0.25 \\ 0.567 & 0.433 \end{bmatrix}$$

The counts associated with D are:

$$\begin{bmatrix} w_{2,2} & 1 \\ 1.4 + w_{2,1} & 1.6 \end{bmatrix}$$

The maximum likelihood estimate is

$$\hat{\theta}_d = \begin{bmatrix} 0.43 & 0.57 \\ 0.51 & 0.49 \end{bmatrix}$$

Question 4. (7pts) We would like to predict whether there will be a storm tomorrow based on historical data. We have three variables we have observed in the past: a binary variable S that indicates whether there was a storm coming, a binary variable W that indicates whether there was high wind, and a binary variable H that indicates for whether there was high humidity. Counts from our observations are shown in the table below:

S	W	H	count
0	0	0	30
0	0	1	15
0	1	0	25
0	1	1	5
1	0	0	0
1	0	1	5
1	1	0	5
1	1	1	15

We decide to use a naive Bayes model, where S is the class, and W and H are the observations.

a) Taking a maximum likelihood estimation approach and using the counts in the table, compute the parameters of the class prior distribution and the class conditional distributions.

b) Find the probability that there will be a storm tomorrow given that we observe high winds and high humidity.

Solution:

a) We need to find parameters for the following distributions: $P(S)$, $P(W \mid S)$, and $P(H \mid S)$. We give each of these in tables below:

S	$P(S)$
0	75/100
1	25/100

W	S	$P(W \mid S)$
0	0	45/75
0	1	1/5
1	0	30/75
1	1	4/5

H	S	$P(H \mid S)$
0	0	55/75
0	1	1/5
1	0	20/75
1	1	4/5

b) Now we would like to compute $P(s^1 \mid w^1, h^1)$. From the definition of conditional probability, we obtain:

$$P(s^1 \mid w^1, h^1) = \frac{P(w^1, h^1, s^1)}{P(w^1, h^1)}$$

With marginalization, we have

$$P(s^1 \mid w^1, h^1) = \frac{P(w^1, h^1, s^1)}{\sum_s P(w^1, h^1, s)}$$

Applying the chain rule, we have

$$P(s^1 \mid w^1, h^1) = \frac{P(w^1 \mid s^1)P(h^1 \mid s^1)P(s^1)}{\sum_s P(w^1 \mid s)P(h^1 \mid s)P(s)}$$

Expanding the denominator, we have

$$\frac{P(w^1 \mid s^1)P(h^1 \mid s^1)P(s^1)}{P(w^1 \mid s^0)P(h^1 \mid s^0)P(s^0) + P(w^1 \mid s^1)P(h^1 \mid s^1)P(s^1)}$$

Using the terms from our table, we have

$$\frac{\frac{4}{5}\frac{4}{5}\frac{1}{4}}{\frac{30}{75}\frac{20}{75}\frac{3}{4} + \frac{4}{5}\frac{4}{5}\frac{1}{4}} = \frac{\frac{16}{100}}{\frac{8}{100} + \frac{16}{100}} = \frac{2}{3}$$