Midterm 2: CS 6375
Fall 2019

The exam is closed book (1 page cheat sheet allowed). Answer the questions in the spaces provided on the question sheets. If you run out of room for an answer, use an additional sheet (available from the instructor), write your name on the sheet and staple it to your exam.

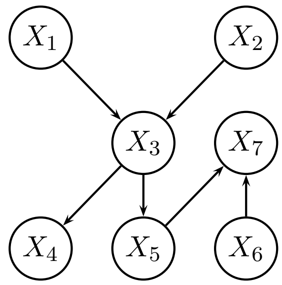
- **NAME** ______________________________

- **UTD-ID if known** ______________________________

Question	Points	Score
Bayesian Networks: Inference	10	
Markov Models and HMMs	10	
EM Algorithm	10	
Miscellaneous	10	
Total:	40	

Question 1: Bayesian Networks: Inference (10 points)

Consider the Bayesian network given below:



(a) (5 points) Let X_7 be the evidence variable. Schematically trace the operation of variable elimination (namely show the new functions created and how they are created) for computing $P(X_7 = x_7)$ where x_7 is a value in the domain of X_7. Use the elimination order $(X_3, X_1, X_2, X_4, X_5, X_6)$.

(b) (2 points) What is the time and space complexity of variable elimination *on the network and order* given on the previous page. Assume that each variable has d values in its domain.

(c) (3 points) Suggest an elimination order along which variable elimination will have smaller time and space complexity than the order $(X_3, X_1, X_2, X_4, X_5, X_6)$ on the network given on the previous page.

Question 2: Markov Models and HMMs (10 points)

Consider a Hidden Markov Model with a binary state X, namely X takes values from the domain $\{0,1\}$. We will denote the State Variable at time slice t by X_t. The prior belief distribution over the initial state X_1 is given by $P(X_1 = 0) = 0.7$, $P(X_1 = 1) = 0.3$.

The transition probabilities $P(X_{t+1}|X_t)$ are:

X_t	X_{t+1}	$P(X_{t+1} \mid X_t)$
0	0	0.9
0	1	0.1
1	0	0.6
1	1	0.4

The sensor model $P(E_t|X_t)$ is given by:

X_t	E_t	$P(E_t \mid X_t)$
0	0	β
0	1	$(1-\beta)$
1	0	$(1-\beta)$
1	1	β

where $\beta \in [0, 1]$.

(a) (5 points) At $t = 1$, we get the first sensor reading, $E_1 = 0$. What is the new belief distribution $P(X_1|E_1 = 0)$? Leave your answer in terms of β.

(b) (5 points) At $t = 2$, we get the second sensor reading, $E_2 = 1$. Use your answer from part (a) to compute $P(X_2 \mid E_1 = 0, E_2 = 1)$. Again, leave your answer in terms of β.

Question 3: EM Algorithm (10 points)

Consider the dataset given below. All variables (A, B, C and D) are Boolean (namely, they take values from the domain $\{0, 1\}$). "?" denotes a missing value.

A	B	C	D
0	1	1	?
1	0	0	?
0	1	1	?
1	1	0	?
0	0	1	?
1	1	1	?

(a) (10 points) Assume that D is a class variable and you are learning a Naive Bayes model. Starting with probabilities that are initialized uniformly (i.e., all probabilities are initialized to 0.5), calculate the parameters of this Naive Bayes model using the EM algorithm. Stop at convergence or after 3 iterations, whichever is earlier. Does the EM algorithm converge and after how many iterations?

Question 4: Miscellaneous (10 points)

(a) (5 points) Given a hypothesis space H defined over an instance space X, prove that if there exists a subset of X of size d (for some integer d) that is shattered by H, then for any $1 \leq k < d$ there also exists a subset of size k of X that is shattered by H.

(b) (5 points) **True/False**. K-means algorithm always converges to a global minima. Namely, it will find the global minima of the following function:

$$\sum_i \text{dist}(x_i, c_{a_i})$$

where i indexes the points, dist is the Euclidean distance measure, a_i is the index of the cluster to which the point x_i is assigned, and c_{a_i} is the cluster center of the cluster indexed by a_i. **Explain your answer. No credit if the explanation is incorrect.**