CSE 291d Midterm Exam **Student ID #**: ____________________

Thursday February 15, 2007 *Note: do not write your actual name.*

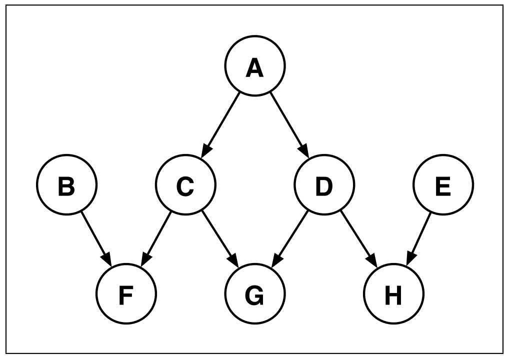
INSTRUCTIONS

There are six problems, for a total of one hundred 100 points.
If necessary, the blank side of each page can be used to show your work.
Where appropriate, show work for partial credit.

Question	**Score**	**Max**
1. Conditional independence		20
2. Markov blankets		20
3. Inference with sigmoid CPT		15
4. Polytrees		15
5. Naive Bayes model		15
6. Maximum likelihood parameter estimation		15
Total		100

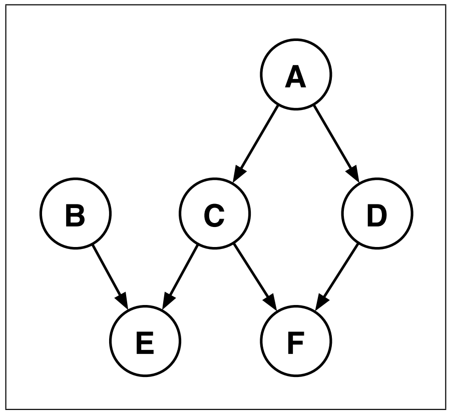
1. Conditional independence

For the belief network shown below, indicate whether the following statements of conditional independence are **true (T)** or **false (F)**.



________	$P(D \mid H)$	=	$P(D \mid E, H)$
________	$P(A, F \mid C)$	=	$P(A \mid C)\, P(F \mid C)$
________	$P(C, D \mid A)$	=	$P(C \mid A)\, P(D \mid A)$
________	$P(C, D \mid G)$	=	$P(C \mid G)\, P(D \mid G)$
________	$P(B \mid G)$	=	$P(B)$
________	$P(A)$	=	$P(A \mid E, B)$
________	$P(G)$	=	$P(G \mid H)$
________	$P(F, H \mid A)$	=	$P(F \mid A)\, P(H \mid A)$
________	$P(F, G, H \mid A)$	=	$P(F \mid A)\, P(G \mid A)\, P(H \mid A)$
________	$P(F, H \mid G)$	=	$P(F \mid G)\, P(H \mid G)$

2. Markov blankets



For the above belief network, consider the following statements of conditional independence. Indicate the largest subset of nodes $\mathcal{S} \subset \{A, B, C, D, E, F\}$ for which each statement is true. Note that one possible answer is the empty set $\mathcal{S} = \emptyset$ or $\mathcal{S} = \{\}$ (whichever notation you prefer). The first two have been done as examples.

$$P(A) = P(A|\mathcal{S}) \qquad \mathcal{S} = \{B\}$$

$$P(A|C) = P(A|\mathcal{S}) \qquad \mathcal{S} = \{B, C, E\}$$

$$P(A|C, D) = P(A|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(D) = P(D|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(D|A) = P(D|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(D|A, F) = P(D|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(D|A, C, F) = P(D|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(B) = P(B|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

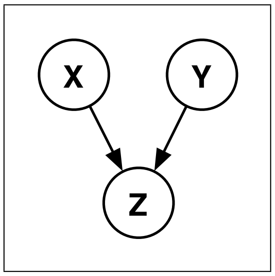
$$P(B|E) = P(B|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(B|C, E) = P(B|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(E) = P(E|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

$$P(A, B) = P(A, B|\mathcal{S}) \qquad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

3. Inference with sigmoid CPT



Variables: $X \in \{0,1\}$, $Y \in \{0,1\}$, $Z \in \{0,1\}$

Sigmoid CPT: $P(Z\!=\!1|X\!=\!x, Y\!=\!y) \;=\; [1 + e^{-w_x x - w_y y}]^{-1}$

Parameters: $w_x \in \Re$, $w_y \in \Re$

Consider the above belief network with binary random variables and sigmoid CPT. Suppose that the prior probabilities satisfy:

$$0 < P(X\!=\!1) < 1,$$
$$0 < P(Y\!=\!1) < 1.$$

In other words, both X and Y have some probability of being either zero or one. Likewise, in the sigmoid CPT, suppose that the weights satisfy:

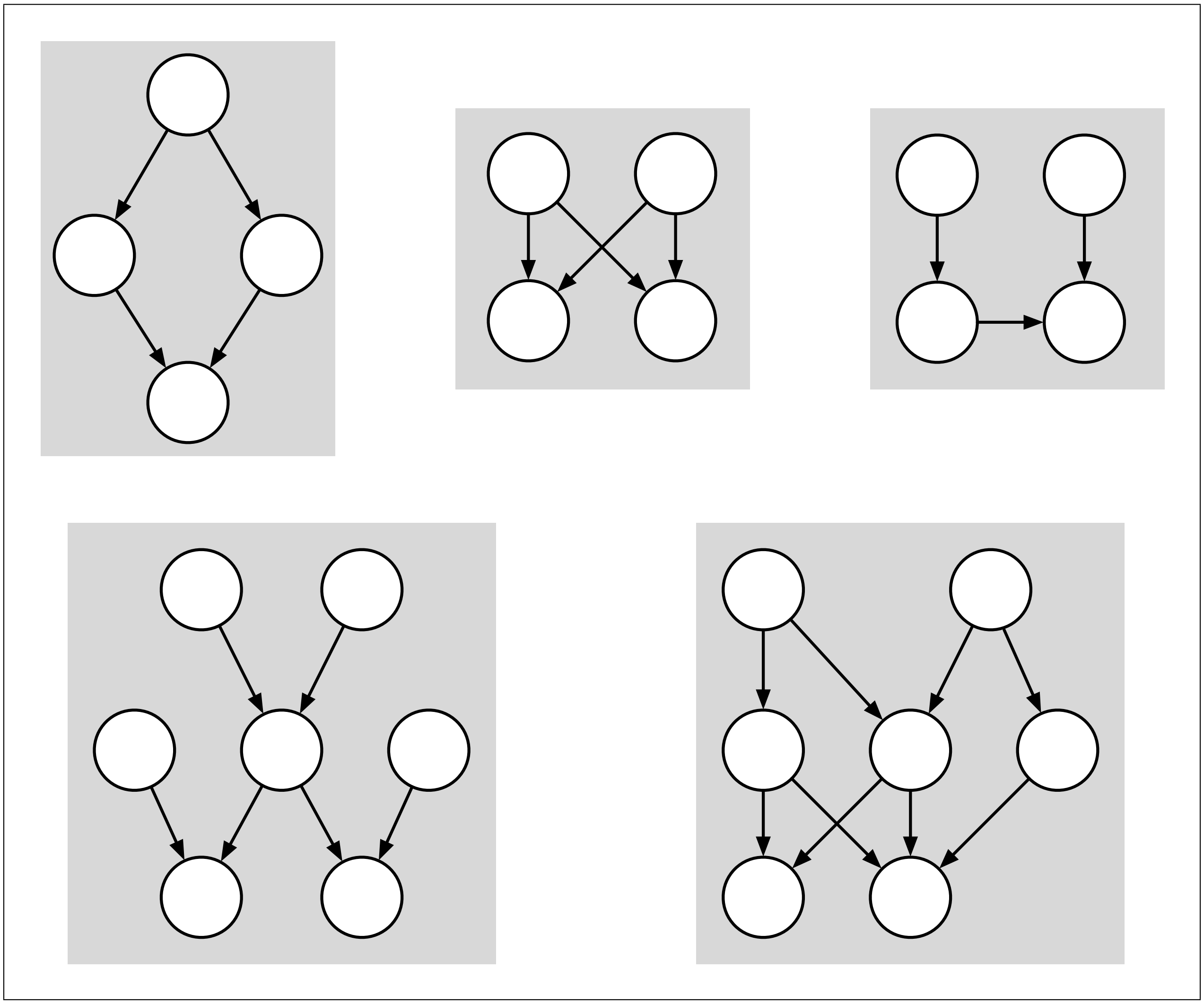
$$\begin{aligned} w_x &> 0, \\ w_y &< 0, \\ w_x + w_y &> 0. \end{aligned}$$

Note the opposite signs of the weights. For each of the following pairs, indicate whether the probability on the left is equal ($=$), greater than ($>$), or less than ($<$) the probability on the right.

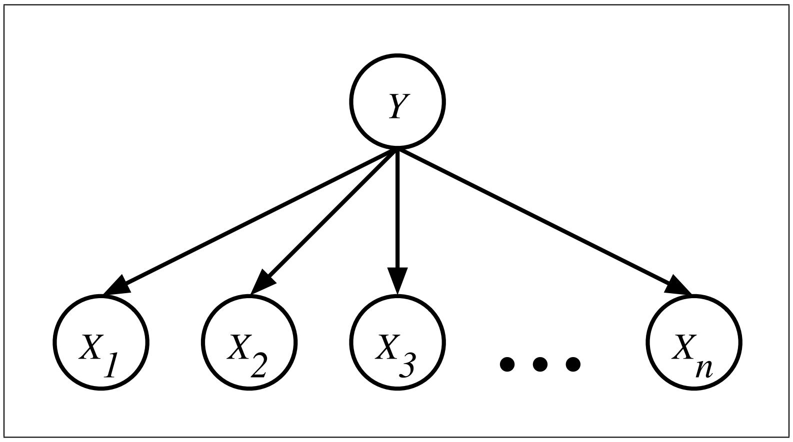
$P(X\!=\!1)$	☐	$P(X\!=\!1\vert Y\!=\!1)$
$P(X\!=\!1)$	☐	$P(X\!=\!1\vert Z\!=\!1)$
$P(Y\!=\!1)$	☐	$P(Y\!=\!1\vert Z\!=\!1)$
$P(Z\!=\!1\vert X\!=\!0, Y\!=\!0)$	☐	$P(Z\!=\!1\vert X\!=\!0, Y\!=\!1)$
$P(Z\!=\!1\vert X\!=\!1, Y\!=\!0)$	☐	$P(Z\!=\!1\vert X\!=\!0, Y\!=\!1)$
$P(Z\!=\!1\vert X\!=\!0, Y\!=\!0)$	☐	$P(Z\!=\!1\vert X\!=\!1, Y\!=\!1)$
$P(Y\!=\!0\vert Z\!=\!1)$	☐	$P(Y\!=\!0\vert Z\!=\!1, X\!=\!1)$
$P(X\!=\!1)\, P(Y\!=\!0)\, P(Z\!=\!1)$	☐	$P(X\!=\!1, Y\!=\!0, Z\!=\!1)$

4. Polytrees

In the figure below, circle the DAGs that are polytrees. In the other DAGs, shade in **one** node that could be instantiated to induce a polytree by the method of cutset conditioning.



5. Naive Bayes model



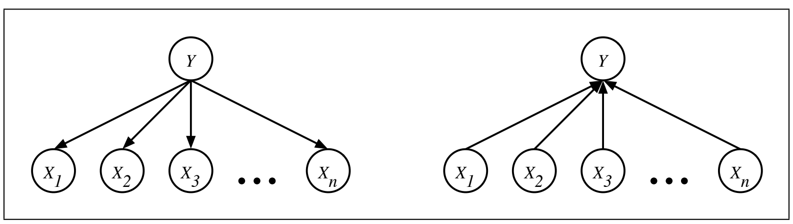
(a) For the belief network shown above, express the joint distribution $P(Y, X_1, X_2, \ldots, X_n)$ in terms of the conditional probability tables $P(Y)$ and $P(X_i|Y)$.

(b) Compute the posterior probability $P(Y = y|X_1, X_2, \ldots, X_n)$ in terms of these same CPTs.

(c) Consider a "complete" data set of N examples, assumed to consist of i.i.d. samples drawn from the joint distribution of this belief network. Let $N(y)$ count the number of examples with $Y\!=\!y$, and let $N_i(x,y)$ count the number of examples with $X_i\!=\!x$ and $Y\!=\!y$.

Express the maximum likelihood (ML) estimates for $P(Y=y)$ and $P(X_i=x|Y=y)$ in terms of these counts. It is only necessary to state your answers, not to derive them.

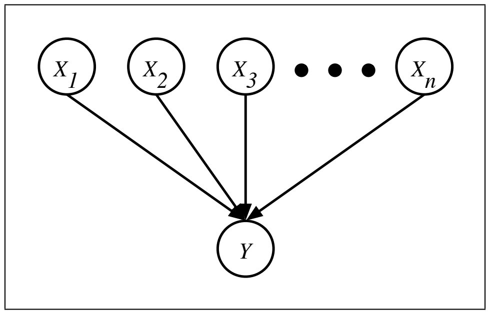
(d) Consider the belief network (shown on the right) with the same connectivity but opposite directionality in the edges of its DAG.



(i) Indicate one statement of conditional independence that is satisfied by the model on the left, but not the model on the right.

(ii) Indicate one statement of conditional independence that is satisfied by the model on the right, but not the model on the left.

6. Maximum likelihood parameter estimation



Consider the above belief network with **nonnegative** random variables $X_i \in [0, +\infty]$ and binary random variable $Y \in \{0, 1\}$. Also, consider the following conditional probability table:

$$P(Y=1|X_1=x_1, \ldots, X_n=x_n) \;=\; \exp\left(-\sum_{i=1}^{n} w_i x_i\right),$$

which is parameterized in terms of the **nonnegative** weights $w_i \geq 0$. A useful shorthand for the above CPT is simply:

$$P\left(Y=1|X=\vec{x}\right) = e^{-\vec{w}\cdot\vec{x}}.$$

(a) Consider a data set of i.i.d. examples $\{\vec{x}_t, y_t\}_{t=1}^{T}$. Compute the conditional log-likelihood of the data set,

$$\mathcal{L} \;=\; \sum_{t=1}^{T} \log P\left(Y=y_t|X=\vec{x}_t\right),$$

in terms of the weight vector $\vec{w}$. Note: you will want to simplify your expression as much as possible for the next part of this question.

(b) As shorthand in this problem, let $p_t = P(Y\!=\!1|X\!=\!\vec{x}_t)$. Show that the gradient of the conditional log-likelihood from part (a) is given by:

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \sum_{t=1}^{T} \left(\frac{p_t - y_t}{1 - p_t} \right) \vec{x}_t.$$

Note: show all of your intermediate steps **very clearly** to receive full credit.